

Sandra Mara Torres Müller

*Adaptação dos Modelos de Markov para um
Sistema de Segmentação e Classificação de
Sinais de Eletrocardiograma*

Vitória - ES

2006

Sandra Mara Torres Müller

*Adaptação dos Modelos de Markov para
Sistema um de Segmentação e
Classificação de Sinais de
Eletrocardiograma*

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para a obtenção do Grau de Mestre em Engenharia Elétrica, na área de concentração em Automação.

Orientadores: Prof. Dr. Teodiano Freire Bastos Filho e Prof. Dr. Mário Sarcinelli Filho.

Co-orientador: Dr. Rodrigo Varejão Andreão.

Vitória - ES

2006

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

M958a Müller, Sandra Mara Torres, 1978-
Adaptação dos modelos de Markov para um sistema de segmentação
e classificação de sinais de eletrocardiograma / Sandra Mara Torres
Müller. – 2006.
98 f. : il.

Orientadores: Teodiano Freire Bastos Filho e Mário Sarcinelli Filho.
Co-Orientador: Rodrigo Varejão Andreão.
Dissertação (mestrado) – Universidade Federal do Espírito Santo,
Centro Tecnológico.

1. Markov, Processos de. 2. Wavelets (Matemática). 3. Otimização
matemática. 4. Reconhecimento de padrões. I. Bastos Filho, Teodiano
Freire. II. Sarcinelli Filho, Mário. III. Andreão, Rodrigo Varejão.
IV. Universidade Federal do Espírito Santo. Centro Tecnológico. V. Título.

CDU: 621.3

Sandra Mara Torres Müller

Adaptação dos Modelos de Markov para um Sistema de Segmentação e Classificação de Sinais de Eletrocardiograma

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para a obtenção do Grau de Mestre em Engenharia Elétrica - Automação.

Aprovada em 03 de abril de 2006.

COMISSÃO EXAMINADORA

Prof. Dr. Teodiano Freire Bastos Filho
Universidade Federal do Espírito Santo
Orientador

Prof. Dr. Mário Sarcinelli Filho
Universidade Federal do Espírito Santo
Orientador

Dr. Rodrigo Varejão Andreão
Universidade Federal do Espírito Santo
Co-Orientador

Prof. Dr. Evandro Ottoni Teatini Salles
Departamento de Engenharia Elétrica - UFES

Prof. Dr. Fernando Gil Vianna Resende Junior
Departamento de Eletrônica - UFRJ

*Dedico este trabalho a minha família, a meus amigos
e a todos que me ajudaram e sempre me ajudam.*

Agradecimentos

Agradecer, aqui, só às pessoas que contribuíram de forma relevante para a elaboração do trabalho, segundo sugestões, me parece incompleto, porque são os momentos em que não se está trabalhando que nos fortalecem e nos incentivam para seguir os objetivos trilhados.

Assim, gostaria de agradecer a Deus, a meus pais Manoel e Cacilda, aos meus irmãos Marcos e André e a minha irmã Cláudia Janaina, pelo apoio de sempre, às minhas amigas e aos meus amigos, que não cito para não incorrer no erro de esquecer algum nome. Agradecer também, mesmo que eles não tenham a oportunidade de ler esses agradecimentos, aos amigos que fiz na França, onde tive a oportunidade de viver uma experiência muito especial.

Quero enviar um agradecimento especial aos colegas do LAI e também ao Anderson. Gostaria de agradecer aos meus orientadores Teodiano e Sarcinelli pela confiança em mim depositada, e também aos professores que sempre me ajudaram durante toda a graduação e agora no mestrado. Estendo meus agradecimentos ao professor Evandro pela indicação ao estágio na França, ao professor Jérôme, pela força e paciência com o francês, e, finalmente ao Rodrigo Varejão, pela ajuda e pela disposição de tirar as dúvidas que surgiram ao longo do trabalho, e que não foram poucas.

Agradeço também ao CNPq pela bolsa de mestrado no Brasil e ao projeto TELCARD-GET (projet incitatif Fondation Leprince-Ringuet) pela bolsa de pesquisa no Institute National des Télécommunications, França.

Je voudrais remercier, en français, à Bernadette Dorizzi pour l'accueil, à mes collègues de bureau et à mon encadrant Jérôme Boudy pour toute son aide, professionnel et personnel pendant mon séjour à l'INT.

“No fim tudo dá certo. Se não deu certo é porque ainda não chegou ao fim.”

Fernando Sabino

Resumo

Neste trabalho foram estudadas e implementadas três técnicas incrementais de adaptação de modelos ocultos de Markov (HMM - Hidden Markov Model) baseadas nos algoritmos de treinamento, que são a esperança da maximização (expectation maximization - EM), a k-means segmental (segmental k-means) e a máximo a posteriori (Maximum a Posteriori -MAP). Essas técnicas, muito utilizadas em reconhecimento de voz, são aqui usadas para sinais biomédicos, mais precisamente para sinal de eletrocardiograma (ECG). Para tal objetivo, utilizou-se uma plataforma, já desenvolvida, de segmentação e classificação de ECG, além de detecções de anomalias cardíacas como extra-sístole ventricular (ESV) e isquemia do miocárdio. Nessa plataforma, os modelos de Markov são empregados na etapa de segmentação do sinal de ECG, tendo em vista a identificação das formas de onda elementares que compõem um ciclo cardíaco. O desenvolvimento dessas técnicas permite, uma vez que a plataforma esteja funcionando como sistema real, um ajuste autônomo dos modelos às variações do sinal de ECG ao longo do tempo, assim como a outras variações presentes em um sistema real. As técnicas foram avaliadas a partir de experimentos usando duas bases de sinais de ECG: *QT database* e *European ST-T database*. Os resultados confirmam o ganho de desempenho obtido com a adaptação, permitindo uma modelagem do sinal ao longo do tempo mais apropriada. As técnicas desenvolvidas são indicadas também para outros tipos de sinais biomédicos, como o sinal de eletroencefalograma (EEG), por exemplo.

Abstract

In this work three incremental adaptation methods for the hidden Markov models (HMM) are studied and implemented, which are based on the Expectation-Maximization (EM), Segmental k-Means and Maximum a Posteriori (MAP) algorithms. These methods, already used in the speech recognition field, are applied here in the electrocardiogram (ECG) segmentation problem. For that, it was used an ECG analysis system able to segment and classify cardiac diseases, like premature ventricular contraction (PVC) and ischemia. The use of these methods allow us to adjust the models to the signal fluctuations commonly met during ambulatory recording. The methods can also be implemented for other kinds of biomedical signals, like electroencephalogram (EEG).

Lista de Figuras

1	Estrutura do coração e curso do fluxo sangüíneo pelas câmaras cardíacas.	p. 21
2	Eventos do ciclo cardíaco na função ventricular esquerda, mostrando as alterações da pressão atrial esquerda, da pressão ventricular esquerda, da pressão aórtica, do volume ventricular, do eletrocardiograma e do fonocardiograma.	p. 22
3	Eletrocardiograma normal.	p. 24
4	Registro da onda de despolarização (A e B) e da onda de repolarização (C e D) de uma fibra do músculo cardíaco.	p. 25
5	<i>Acima</i> , potencial de ação monofásico de fibra muscular ventricular durante o funcionamento normal do coração, mostrando a despolarização rápida e em seguida a repolarização ocorrendo lentamente durante a fase do platô, mas rapidamente próxima do fim. <i>Abaixo</i> , eletrocardiograma registrado simultaneamente.	p. 26
6	Disposição convencional dos eletrodos para o registro do eletrocardiograma pelas derivações padrão. O triângulo de Einthoven está disposto sobre o tórax.	p. 28
7	Eletrocardiogramas normais registrados pelas três derivações padrão.	p. 29
8	Conexões do corpo com o eletrocardiógrafo para registro das derivações torácicas.	p. 30
9	Eletrocardiogramas normais registrados pelas seis derivações torácicas.	p. 31
10	Eletrocardiogramas normais registrados pelas três derivações unipolares periféricas aumentadas.	p. 32
11	Ponto "J" como referencial zero de referência do eletrocardiograma.	p. 33
12	Corrente de lesão em infarto agudo da parede anterior.	p. 35
13	Corrente de lesão em infarto agudo apical de parede posterior.	p. 35

14	Interface GUI-Matlab desenvolvida em (OLIVEIRA; ANDREA, 2003-2004).	p. 36
15	Sistema de detecção de isquemia.	p. 37
16	Sistema de tele-vigilância TelePat.	p. 38
17	Arquitetura básica do sistema de detecção.	p. 38
18	Batimento cardíaco.	p. 39
19	Modelo do batimento cardíaco, utilizado por (COAST et al., 1990).	p. 40
20	Modelo do batimento cardíaco composto pela concatenação de HMMs.	p. 40
21	Decodificação da seqüência de observações com o uso do algoritmo <i>one-pass</i> .	p. 41
22	Sistema de segmentação automática do sinal de ECG.	p. 41
23	Arquitetura básica do algoritmo de (ANDREA, 2004).	p. 42
24	Sistema desenvolvido por (ANDREA, 2004).	p. 43
25	Tela de avaliação.	p. 44
26	Comparação com duas funções wavelets: a) usando a wavelet Chapéu Mexicano; b) usando a wavelet Morlet.	p. 45
27	Transformada wavelet para um sinal de ECG.	p. 45
28	Histograma das observações da onda P. A escala j é correspondente à transformada wavelet efetuada.	p. 46
29	Aprendizagem do modelo λ^k pelo método de Baum-Welch (RABINER, 1989).	p. 47
30	Adaptação direta e indireta de HMM (LEE; HUO, 2000).	p. 56
31	Sistema para o algoritmo EM incremental implementado.	p. 64
32	Arquitetura básica do algoritmo segmental K-means incremental.	p. 68
33	Sistema para o algoritmo segmental k-Means incremental implementado.	p. 69
34	Aproximação da verdadeira distribuição posterior por uma simplificada, sob o critério de mesma moda.	p. 74
35	Arquitetura básica do algoritmo MAP incremental.	p. 79
36	Sistema desenvolvido para abordagem MAP.	p. 80

37	Comportamento de detecção de onda P com a variação de ρ	p. 86
38	Transformada wavelet.	p. 95
39	Transformada wavelet com mudança de escala.	p. 95
40	Avaliação do valor da escala na transformada wavelet.	p. 96
41	Função Chapéu Mexicano.	p. 98

Lista de Tabelas

1	Especificações dos modelos.	p. 47
2	Valores das constantes para o algoritmo implementado.	p. 64
3	Valores das constantes para o algoritmo implementado.	p. 69
4	Número de etiquetas feitas para os 105 registros da base QT.	p. 82
5	Precisão e detecção de ondas.	p. 82
6	Detecção de complexos QRS.	p. 84
7	Detecção de Extrasístoles.	p. 85
8	Número de episódios etiquetados em 48 registros.	p. 86
9	Número de episódios isquêmicos em 48 registros.	p. 87
10	Detecção de ondas para determinados valores de ρ	p. 88
11	Detecção de complexo QRS para determinados valores de ρ para o Canal 1.	p. 88

Sumário

1	Introdução	p. 17
1.1	Estado da Arte	p. 18
1.2	Estrutura do Trabalho	p. 20
2	O Coração	p. 21
2.1	O Ciclo Cardíaco	p. 22
2.2	Excitação Rítmica do Coração	p. 23
2.3	O Eletrocardiograma Normal	p. 24
2.3.1	Valores Normais do Eletrocardiograma	p. 27
2.4	Derivações Eletrocardiográficas	p. 27
2.4.1	Derivações Bipolares Periféricas	p. 27
2.4.2	Derivações Torácicas (Derivações Pré-Cordiais)	p. 30
2.4.3	Derivações Unipolares Periféricas Aumentadas	p. 31
2.5	Interpretação Eletrocardiográfica	p. 31
3	Sistema Atual e Metodologia	p. 36
3.1	Introdução	p. 36
3.2	Sistema Atual	p. 39
3.3	Metodologia	p. 43
3.3.1	Extração de Parâmetros	p. 44
3.3.2	Segmentação Automática do Sinal de ECG	p. 46
3.3.3	Aprendizagem e Adaptação	p. 46

3.3.4	Especificação dos Modelos	p. 47
4	Técnicas de Adaptação	p. 48
4.1	Teoria de HMM	p. 48
4.1.1	Elementos de um HMM	p. 48
4.1.2	Os três problemas básicos do HMM	p. 49
4.1.2.1	Solução para o problema 1: cálculo da Máxima Verosimilhança (ML)	p. 49
4.1.2.2	Solução para o problema 2: algoritmo de Viterbi	p. 50
4.1.2.3	Solução para o problema 3	p. 51
4.2	Adaptação	p. 55
4.2.1	Algoritmos em Lote (Batch Algorithms)	p. 55
4.2.2	Algoritmos <i>On-line</i>	p. 57
4.2.3	Treinamento Incremental	p. 57
4.3	Aprendizagem EM	p. 59
4.3.1	Adaptação EM Incremental Implementada	p. 62
4.4	Aprendizagem Segmental k-Means	p. 64
4.4.1	Adaptação Segmental K-means Incremental Implementada	p. 67
4.5	Aprendizagem Bayesiana	p. 69
4.5.1	Aprendizagem MAP quasi-Bayes	p. 71
4.5.2	Aprendizagem MAP quasi-Bayes em Lote	p. 74
4.5.3	Aprendizagem MAP quasi-Bayes Incremental	p. 77
4.5.4	Adaptação MAP Incremental Implementada	p. 79
5	Resultados e Conclusões	p. 81
5.1	Resultados	p. 81
5.2	Discussão	p. 87
5.3	Conclusões e Perspectivas	p. 89

Referências	p. 91
Anexos	p. 94
Anexo A - Wavelets	p. 94
Transformada Wavelet	p. 94
Escala	p. 96
Funções Wavelets	p. 96

Lista de Símbolos

λ	Modelo Oculto de Markov
a_{ij}	Probabilidade de transição do estado i para o estado j
$b_j\{k\}$	Distribuição das probabilidades de emissão de símbolos
π	Probabilidade inicial dos estados
O	Observações, dados ou amostras
q_j	Estado j
$Q = q_1 q_2 \dots q_T$	Seqüência de estados
$p(O \lambda)$	Probabilidade da observação dado o modelo ou verossimilhança
$\delta_t(i)$	Maior probabilidade ao longo do caminho, no tempo t
$\psi_t(j)$	Vetor para a recuperação da melhor seqüência de estados
$L(\lambda)$	Logaritmo da verossimilhança
$X(q)$	Distribuição de probabilidade dos estados sobre as variáveis ocultas
$F(X, \lambda)$	Função custo do algoritmo EM
$E[.]$	Operador Esperança
$H(X(q))$	Entropia da distribuição $X(q)$
$Q(\lambda, \lambda)$	Função auxiliar no algoritmo EM
$\gamma_t(j, m)$	Probabilidade de emissão para caso contínuo
c_{jm}	Coefficiente de mistura de gaussianas
μ_{jm}	Vetor média
U_{jm}	Matriz de covariância
\bar{t}_k^n	Estatísticas do bloco n
\hat{q}	Caminho mais provável
$p(\lambda O)$	Probabilidade a posteriori
$p_0(\lambda)$	Distribuição a priori de λ
$p(O)$	Probabilidade das observações
$g(\lambda)$	Aproximação de $p(\lambda)$
ϕ	Função de hiper-parâmetros
$g(\lambda \varphi^n)$	Aproximação da probabilidade a posteriori
$R(\lambda \lambda)$	Função auxiliar do algoritmo MAP

1 *Introdução*

O eletrocardiograma (ECG) é um gráfico obtido quando os potenciais de um campo elétrico com origem no coração são registrados à superfície do organismo. Os sinais são detectados por eletrodos metálicos ligados aos membros e à parede torácica e são depois amplificados e registrados pelo eletrocardiógrafo. Deve notar-se que no ECG apenas são registradas diferenças de potencial instantâneas entre os eletrodos. Apesar das suas limitações, o ECG é o exame auxiliar mais usado no diagnóstico de doenças cardíacas. Isto resulta do fato de ser um exame não invasivo, barato e extremamente versátil. Criado em 1903 pelo fisiologista holandês Willem Einthoven, o eletrocardiograma registra a atividade elétrica do coração, a qual se apresenta alterada no aparecimento ou na evolução de mais de 90% das doenças cardíacas, principalmente as arritmias e os infartos agudos.

A cada ano, cerca de 140 mil pessoas morrem de doenças do coração no Brasil, segundo dados da Organização Mundial da Saúde. Cerca de 90% dessas mortes, inclusive as decorrentes de mal súbito, poderiam ser evitadas com o diagnóstico básico de um simples eletrocardiograma, seguido de tratamento e acompanhamento médicos adequados. “Uma boa consulta cardiológica e um eletrocardiograma bem interpretado podem diagnosticar mais de 90% dos problemas cardíacos, poupando milhares de vidas”, segundo o Dr. Carlos Alberto Pastore, diretor do Serviço de Eletrocardiologia do Instituto do Coração do Hospital das Clínicas (Incor-HCFMUSP) (USP, 2005).

O exame é eficaz, ainda, para acompanhar pacientes sob uso de medicamentos, como antibióticos, antialérgicos, antidepressivos e até antiarrítmicos, que, em alguns casos, agem negativamente sobre o coração, quando este é acometido de uma doença silenciosa, ainda não diagnosticada. Muitas vezes, a doença é congênita e não apresenta qualquer sintoma, mas, ao interagir com medicação inadequada, pode ser fatal.

“A avaliação cardiológica anual, acompanhada de eletrocardiograma, de pacientes acima de 40 anos e com história familiar de doença do coração ou fatores de risco importantes como colesterol alto, pressão alta, obesidade, tabagismo, etc, e pode prevenir os

eventos fatais”, alerta o médico. Ao longo da evolução do método de diagnóstico, impulsionada pelas necessidades específicas dos médicos, foram surgindo diversas variantes do exame que, em sua forma original, é feito com o paciente deitado e em repouso. O sistema *holter* - eletrocardiograma de 24 horas - para avaliação de arritmias, é um dos resultados desse desenvolvimento.

As alterações eletrocardiográficas no infarto do miocárdio foram identificadas pela primeira vez por Pardee, em 1920 (NICOLAU et al., 2003), como correspondendo a ondas de isquemia, lesão e necrose celular, representadas por aspectos morfológicos específicos no ECG. Vale lembrar que isquemia é a supressão da circulação sanguínea em determinada parte do organismo, o que pode levar a alterações do funcionamento deste organismo. Atualmente, o eletrocardiograma é componente fundamental e indispensável na avaliação precoce e na estratificação de risco em pacientes com infarto agudo do miocárdio, face às valiosas informações nele contidas.

O eletrocardiograma inicial faz o diagnóstico de infarto agudo do miocárdio somente em 60% dos pacientes com tal doença, sendo anormal, porém não definitivo, em 25% dos mesmos, e, nos restantes 15%, situa-se dentro da normalidade. Essa sensibilidade relativamente baixa do eletrocardiograma inicial melhora de forma significativa quando se analisam eletrocardiogramas seriados, ou seja, uma análise ao longo do tempo, chegando próximo a 95% de diagnósticos corretos de infarto do miocárdio.

Mesmo com outras tecnologias para a análise do sinal cardíaco, como o ecocardiograma ou tomografia computadorizada do tórax, o eletrocardiograma convencional continua sendo fundamental no diagnóstico e prognóstico de pacientes com infarto agudo do miocárdio, fornecendo informações únicas e, freqüentemente, mais importantes que aquelas obtidas com tecnologias mais sofisticadas e dispendiosas (NICOLAU et al., 2003).

1.1 Estado da Arte

Devido à importância da análise do eletrocardiograma, torna-se necessária a criação de um sistema de análise autônomo que auxilie o médico no diagnóstico do paciente e na avaliação do seu histórico cardíaco que, por vezes, resulta na análise de horas de eletrocardiograma, o que é um trabalho repetitivo e passível de erros. Assim, a segmentação do sinal eletrocardiográfico ainda constitui um desafio atual, sobretudo para os registros de longa duração. Vale lembrar que o ato de segmentar o sinal cardíaco corresponde à localização das suas formas elementares de onda e segmentos que serão mostrados nesta

dissertação.

Há, na literatura, muitos trabalhos na segmentação autônoma do sinal de ECG, como (CHAZAL; REILLY, 2003), que detecta contração ventricular normal e prematura e utiliza características baseadas no formato das formas de onda do ECG e nos seus intervalos como entrada para os classificadores. Estes são baseados em dois modelos: um modelo discriminante linear e um modelo de rede neural *feed forward*. Pode-se citar também (SABRY-RISK et al., 1999), onde foi desenvolvido um classificador neural que distingue as características normais e anormais do sinal de ECG.

Outra abordagem que tem se mostrado eficiente na segmentação do sinal de ECG, são os modelos ocultos de Markov - HMM (do inglês Hidden Markov Models). Esta técnica, muito utilizada para reconhecimento de voz, é aplicada também em processamento de sinais biomédicos, como sugerido em (COHEN, 1998), e tem como principal vantagem sua dinâmica temporal, que permite uma melhor modelagem desses sinais. Pode-se citar (COAST et al., 1990) como um dos primeiros trabalhos na área de segmentação e classificação do sinal de ECG utilizando HMM. Ali é desenvolvida uma abordagem markoviana dependente do paciente para a análise de arritmias cardíacas. O mesmo sinal de ECG é utilizado para aprendizagem e teste dos modelos. O sistema construído necessita de uma pré-etiquetagem manual cada vez que um novo sinal de ECG do indivíduo é analisado, constituindo um ponto fraco. Assim, os modelos são treinados de forma supervisionada, o que aumenta o ganho do sistema, mas que diminui sua autonomicidade.

Pode-se citar também o trabalho de Bardonova et al. (2003), que utiliza sinais eletrocardiográficos de animais, de forma invasiva para a detecção de isquemia do miocárdio. Nesse trabalho, assim como Coast, ele se utiliza um único modelo de Markov para a classificação do sinal, e não há uma etapa de segmentação para a identificação de outras anormalidades cardíacas. Ressalta-se que nenhum desses trabalhos desenvolve a adaptação de modelos de Markov para o ajuste às modificações do sinal que ocorrem ao longo do tempo. A adaptação da modelagem de um sistema é necessária para que o mesmo possa se ajustar às flutuações no sinal que podem ocorrer ao longo do tempo.

Existem trabalhos sobre adaptação para métodos estatísticos, como as redes neurais (WATROUS; TOWELL, 1995). Nesse caso, o autor utiliza a extração de características do paciente para o treinamento da rede em um sistema de monitoração de eletrocardiograma. O treinamento dos modelos dura cerca de 6 dias e a adaptação é feita pelo método do gradiente descendente. Há processos de adaptação que se utilizam de filtros adaptativos, como em (SORIA et al., 1996), mas o sistema desenvolvido não envolve a segmentação do

sinal de ECG, mas sim a identificação do começo e final de certas partes do batimento.

O trabalho desenvolvido em (ANDREAO, 2004) compõem-se de um sistema de segmentação e classificação do sinal de ECG. Esta classificação é composta basicamente pela identificação de extrasístoles ventricular e de episódios isquêmicos. Para uma boa classificação é necessária a adaptação, ao longo do tempo, dos modelos markovianos a cada paciente. Essa adaptação já é feita em (ANDREAO, 2004). O presente trabalho visa uma melhoria no sistema de adaptação já implementado. Assim, o objetivo desta dissertação é o estudo e a implementação de novas técnicas que serão apresentadas nesta dissertação. Para isso, fez-se um estudo comparativo de algoritmos de treinamento de HMM usados para a adaptação dos mesmos. Os algoritmos estudados foram: *expectation-maximization* - EM, *segmental k-means* e *maximum a posteriori* - MAP. Existem grupos distintos que trabalham com algoritmos baseados no critério de *maximum likelihood* - ML (EM e segmental k-means), como (DIGALAKIS, 1999), ou que trabalham com o critério MAP, como (HUO; LEE, 1997). Assim, este trabalho apresenta sua originalidade não só no fato da comparação de algoritmos baseados em diferentes critérios de otimização, mas também pelo fato de aplicá-los à análise de sinais eletrocardiográficos. Os resultados, que serão apresentados aqui, mostram o desempenho das técnicas implementadas e comprovam que elas melhoraram a adaptação do sistema.

1.2 Estrutura do Trabalho

Na Introdução foi vista a motivação para o desenvolvimento deste trabalho, seguido de um estado da arte atual, situando o presente trabalho no mesmo contexto de um outro já desenvolvido.

No Capítulo 2 é apresentada a fisiologia do coração e seu funcionamento para que se possa analisar corretamente o sinal de ECG.

No Capítulo 3 é apresentado o sistema proposto em (ANDREAO, 2004), sobre o qual o presente trabalho foi desenvolvido, reforçando algumas de suas características.

As técnicas de adaptação implementadas são apresentadas no Capítulo 4 e os resultados e as conclusões são explicitados no Capítulo 5.

No Anexo A é apresentada a teoria de wavelets, as quais são utilizadas na extração de parâmetros do sinal.

2 O Coração

Neste capítulo inicia-se a discussão sobre o coração e suas características, de acordo com Guyton e Hall (2002). O coração, mostrado na Figura 1, constitui-se, na realidade, de duas bombas distintas: o coração direito, que bombeia sangue para os pulmões, e o coração esquerdo, que bombeia sangue para os órgãos periféricos. Cada um desses corações é uma bomba pulsátil com duas câmaras, sendo elas o átrio e o ventrículo. O átrio funciona, principalmente, como bomba de escoamento, ajudando a movimentar o sangue para o ventrículo. O ventrículo, por sua vez, fornece a força principal que propõe o sangue para a circulação pulmonar, pelo ventrículo direito, ou para a circulação periférica, no caso do ventrículo esquerdo.

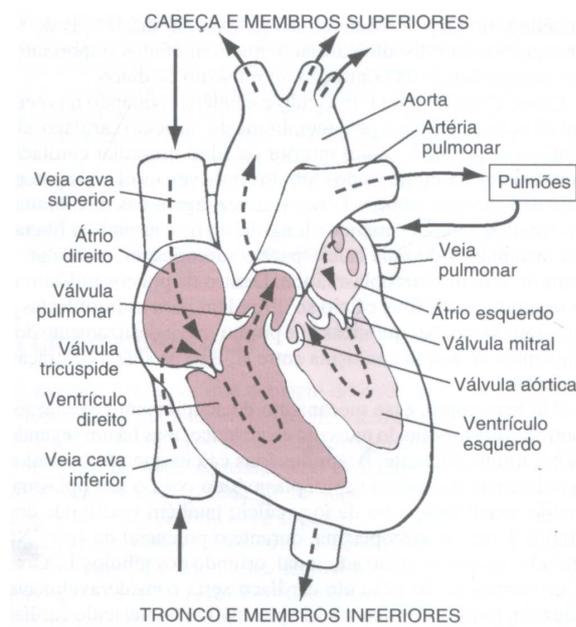


Figura 1: Estrutura do coração e curso do fluxo sanguíneo pelas câmaras cardíacas.

2.1 O Ciclo Cardíaco

Os eventos cardíacos que ocorrem desde o início de cada batimento cardíaco até o início do próximo são chamados de ciclo cardíaco. O ciclo cardíaco consiste em um período de relaxamento do músculo, chamado de diástole, durante o qual o coração se enche com sangue, seguido de um período de contração, chamado de sístole.

A Figura 2 mostra os diferentes eventos durante o ciclo cardíaco, para o lado esquerdo do coração. Os três traçados superiores mostram as variações da pressão na aorta, no ventrículo e átrio esquerdo, respectivamente. O quarto traçado mostra as variações do volume ventricular, o quinto, o eletrocardiograma, e o sexto, um fonocardiograma, que é o registro dos sons produzidos pelo coração durante o bombeamento. A válvula A-V corresponde às válvulas tricúspide e mitral. O eletrocardiograma apresentado na Figura 2 mostra as ondas P, Q, R, S e T que serão discutidas adiante. Elas representam as tensões elétricas geradas pelo coração e registradas pelo eletrocardiógrafo na superfície do corpo.

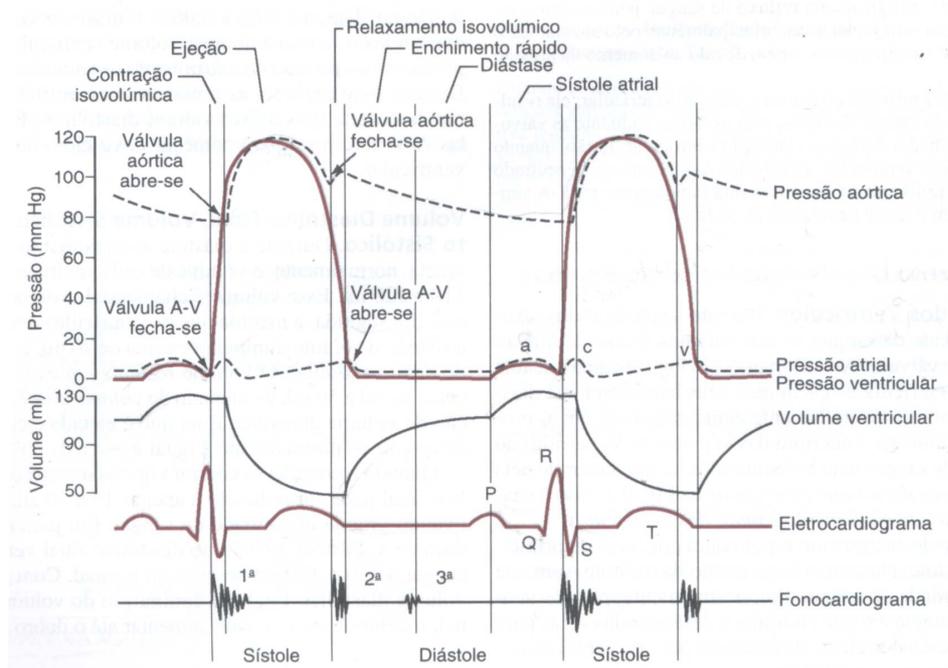


Figura 2: Eventos do ciclo cardíaco na função ventricular esquerda, mostrando as alterações da pressão atrial esquerda, da pressão ventricular esquerda, da pressão aórtica, do volume ventricular, do eletrocardiograma e do fonocardiograma.

A onda P é causada pela despolarização atrial, e isso é seguido pela contração atrial, que provoca pequena elevação da curva da pressão atrial, imediatamente após a onda P. Cerca de 0,16 s após o início da onda P, o complexo QRS aparece, como resultado da despolarização dos ventrículos, iniciando a contração dos ventrículos e provocando o início

da elevação da pressão ventricular, também mostrada na Figura 2. Portanto, o complexo QRS começa pouco antes do início da sístole ventricular. Finalmente, nota-se a onda T ventricular no eletrocardiograma. Ela representa a fase de repolarização dos ventrículos quando as fibras musculares ventriculares começam a relaxar. Por conseguinte, a onda T ocorre pouco antes do término da contração ventricular.

No traçado da pressão atrial da Figura 2 são notadas três principais elevações de pressão, chamadas de ondas de pressão atrial *a*, *c* e *v*. A onda *a* é causada pela contração atrial. Já a onda *c* ocorre quando os ventrículos começam a se contrair. Isso é causado, em parte, pelo pequeno refluxo de sangue para os átrios no início da contração ventricular, mas, principalmente, pelo abaulamento das válvulas A-V em direção aos átrios, devido ao aumento da pressão nos ventrículos. Finalmente, a onda *v* ocorre próximo ao fim da contração ventricular; ela resulta do fluxo lento de sangue das veias para os átrios, enquanto as válvulas A-V estão fechadas durante a contração ventricular. Então, quando termina a contração ventricular, as válvulas A-V abrem-se, permitindo que esse sangue armazenado nos átrios flua rapidamente para os ventrículos, causando o desaparecimento da onda *v*.

2.2 Excitação Rítmica do Coração

O coração é dotado de um sistema eletrogênico, ou seja, possui um processo de auto-excitação elétrica, especializado para gerar impulsos ritmados, que produzem a contração rítmica do músculo cardíaco, e conduz esses impulsos rapidamente através do coração. Quando esse sistema funciona normalmente, os átrios contraem-se cerca de 0,17 s antes da contração dos ventrículos, o que permite o seu enchimento antes que eles bombeiem o sangue para os pulmões e para a circulação periférica. Outra importância especial desse sistema é permitir que todas as porções dos ventrículos se contraíam quase que simultaneamente, o que é essencial para a geração efetiva de pressão nas câmaras ventriculares.

Esse sistema rítmico e condutor do coração é suscetível à lesão por doença cardíaca, em especial pela isquemia dos tecidos cardíacos resultante do fluxo sanguíneo coronário insuficiente. A conseqüência é, geralmente, ritmo cardíaco bizarro ou seqüência anormal de contração das câmaras cardíacas, sendo que a eficiência do bombeamento do coração fica, em geral, gravemente afetada, podendo resultar em morte.

2.3 O Eletrocardiograma Normal

Quando o impulso cardíaco passa pelo coração, a corrente elétrica também se propaga para os tecidos adjacentes ao coração. Uma pequena parte da corrente se dissemina por toda a superfície do corpo. Se forem colocados eletrodos sobre a pele em pontos opostos do coração, os potenciais elétricos gerados por essa corrente podem ser registrados, registro este que é conhecido como eletrocardiograma (ECG). Um eletrocardiograma normal, para dois batimentos cardíacos, é mostrado na Figura 3. Note-se que a menor divisão na vertical corresponde a 0,1 mV e na horizontal a 0,04 s.

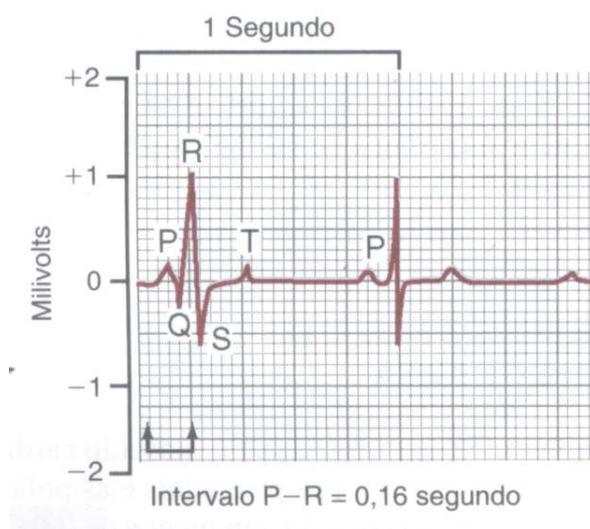


Figura 3: Eletrocardiograma normal.

O eletrocardiograma normal é formado pela onda P, pelo complexo QRS e pela onda T. O complexo QRS é freqüentemente, mas nem sempre, formado por três ondas distintas, a onda Q, a onda R e a onda S.

A onda P é causada pelos potenciais elétricos gerados quando os átrios se despolarizam, antes da contração atrial. O complexo QRS é causado pelos potenciais gerados quando os ventrículos se despolarizam antes da sua contração, isto é, à medida que a onda de despolarização se propaga pelos ventrículos. Por conseguinte, tanto a onda P como os componentes do complexo QRS são ondas de despolarização. A onda T é causada pelos potenciais gerados quando os ventrículos se recuperam do estado de despolarização. Esse processo ocorre normalmente no músculo ventricular de 0,25 s a 0,35 s após a despolarização, e a onda T é conhecida como onda de repolarização.

Dessa forma, o eletrocardiograma é composto por ondas de despolarização e repola-

rização. A Figura 4 mostra uma fibra muscular cardíaca nos quatro estágios de despolarização e repolarização, com o vermelho indicando a despolarização.

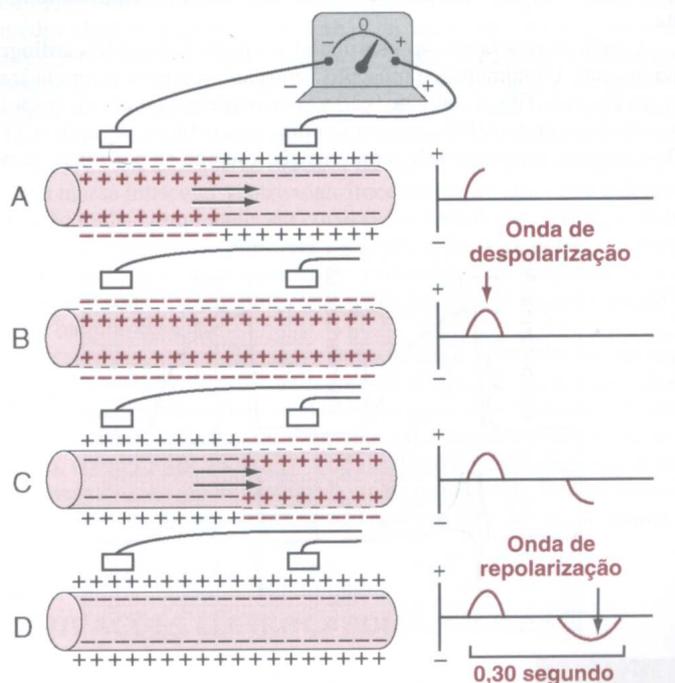


Figura 4: Registro da onda de despolarização (A e B) e da onda de repolarização (C e D) de uma fibra do músculo cardíaco.

Durante a despolarização, o potencial negativo normal no interior da fibra desaparece e o potencial da membrana se inverte, isto é, torna-se ligeiramente positivo no interior e negativo no exterior. Na Figura 4A, a despolarização demonstrada pelas cargas positivas no interior e pelas cargas negativas no exterior, ambas em vermelho, está se propagando da esquerda para a direita. A primeira metade da fibra já está despolarizada enquanto que a outra metade ainda está polarizada. Portanto, o eletrodo esquerdo no exterior da fibra está em área de negatividade, e o eletrodo direito está em área de positividade, o que faz o aparelho registrar um valor positivo. À direita da fibra muscular é mostrado o potencial entre os dois eletrodos registrado pelo aparelho de medida. Observe-se que, quando a despolarização atingiu o ponto médio, na Figura 4A, o registro atingiu o valor máximo.

Na Figura 4B, a despolarização se estende por toda a fibra muscular e o registro, à direita, retornou à linha de base zero, visto que os eletrodos estão agora em área de igual negatividade. A onda completa é uma onda de despolarização, visto que resulta da propagação da despolarização ao longo da fibra muscular. A Figura 4C mostra a repolarização no ponto médio da fibra muscular, com a positividade retornando ao exterior

da fibra. Nesse ponto, o eletrodo esquerdo está em área de positividade e o eletrodo direito está em área de negatividade. Conseqüentemente, o registro torna-se negativo, como mostrado à direita. Na Figura 4D, a fibra muscular está completamente repolarizada e ambos os eletrodos estão agora em área de positividade de modo que nenhum potencial é registrado entre eles. Dessa forma, no registro à direita, o potencial retorna mais uma vez ao valor zero. Essa onda negativa completa é uma onda de repolarização ao longo da fibra muscular.

O potencial de ação monofásico do músculo ventricular, como já dito, dura normalmente de 0,25 s a 0,35 s. A parte superior da Figura 5 mostra o potencial de ação monofásico registrado por um microeletrodo inserido no interior da fibra muscular ventricular. A deflexão desse potencial de ação é causada por despolarização e o retorno do potencial à linha de base é causado por repolarização.

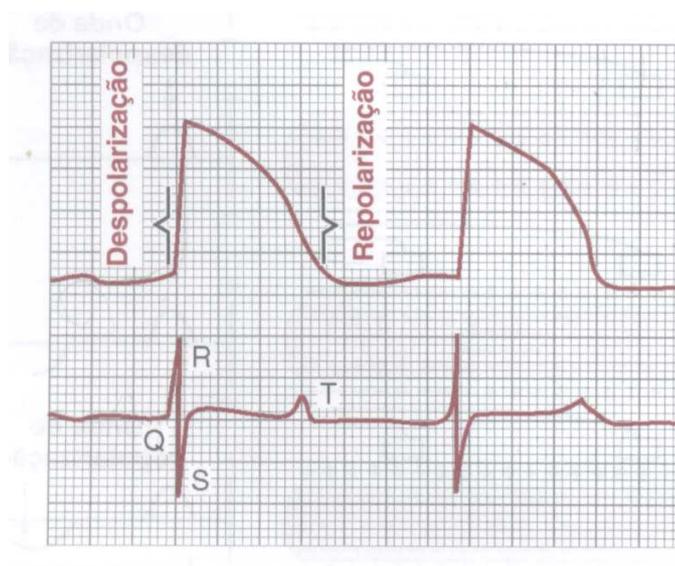


Figura 5: *Acima*, potencial de ação monofásico de fibra muscular ventricular durante o funcionamento normal do coração, mostrando a despolarização rápida e em seguida a repolarização ocorrendo lentamente durante a fase do platô, mas rapidamente próxima do fim. *Abaixo*, eletrocardiograma registrado simultaneamente.

Observa-se na metade inferior da Figura 5 o registro simultâneo do eletrocardiograma desse mesmo ventrículo, o qual mostra as ondas QRS aparecendo no início do potencial de ação monofásico e a onda T aparecendo no final. Ressalta-se que nenhum potencial é registrado no eletrocardiograma quando o músculo ventricular está completamente polarizado ou completamente despolarizado. Somente quando o músculo está parcialmente polarizado ou parcialmente despolarizado é que a corrente flui de uma parte dos ventrículos para outra e, portanto, alguma corrente flui para a superfície do corpo para produzir o eletrocardiograma.

2.3.1 Valores Normais do Eletrocardiograma

As amplitudes normais das ondas no eletrocardiograma dependem da maneira que os eletrodos são colocados na superfície do corpo e de sua distância ao coração. Quando um eletrodo é colocado diretamente sobre os ventrículos e um segundo eletrodo é colocado em outro ponto do corpo, distante do coração, a amplitude do complexo QRS pode ser de até 3 a 4 mV. Essa tensão é ainda pequena quando comparada com o potencial de ação monofásico de 110 mV registrado diretamente na membrana do músculo cardíaco. Quando os eletrocardiogramas são registrados com eletrodos nos dois braços ou em um braço e uma perna, a amplitude do complexo QRS, em geral, é de cerca de 1 mV do pico da onda R à base da onda S; a amplitude da onda P fica entre 0,1 e 0,3 mV; e a da onda T fica entre 0,2 e 0,3 mV.

Para o intervalo PQ ou PR, ou seja, o tempo entre o início da onda P e o início do complexo QRS, que corresponde ao intervalo entre o início da excitação elétrica dos átrios e o início da excitação dos ventrículos, o valor normal é cerca de 0,16 s. O intervalo PQ é também chamado de intervalo PR, devido à ausência freqüente da onda Q. Para o intervalo QT, que corresponde à contração do ventrículo, que dura quase do começo do início da onda Q (ou onda R, se a onda Q está ausente) ao fim da onda T, o valor normal é de cerca de 0,35 s. Já a freqüência cardíaca normal é de cerca de 72 batimentos por minuto.

2.4 Derivações Eletrocardiográficas

2.4.1 Derivações Bipolares Periféricas

A Figura 6 mostra as conexões elétricas entre os membros do paciente e o eletrocardiógrafo, para o registro dos eletrocardiogramas pelas chamadas derivações bipolares periféricas padrão. O termo bipolar significa que o eletrocardiograma é registrado por meio de dois eletrodos localizados nos diferentes lados do coração, neste caso nos membros. Assim, uma derivação não é um fio único ligado ao corpo, mas a combinação de dois fios e seus eletrodos para formar um circuito completo com o eletrocardiógrafo, representado por um registrador elétrico. O potencial médio do corpo, que serve como referencial para o eletrocardiógrafo, apesar de não mostrado na Figura 6, é medido na perna direita do paciente.

No registro da derivação periférica I, o terminal negativo do eletrocardiógrafo está

conectado ao braço direito e o terminal positivo ao braço esquerdo. Portanto, quando o ponto sobre o tórax, onde o braço direito se conecta ao tórax, está eletronegativo em relação ao ponto torácico, onde o braço esquerdo se prende, o eletrocardiógrafo registra positivamente. Quando o oposto é verdadeiro, o registro é negativo.

No registro da derivação periférica II, o terminal negativo do eletrocardiógrafo está conectado ao braço direito e o terminal oposto à perna esquerda. Portanto, quando o braço está negativo em relação à perna, o eletrocardiógrafo registra positivamente.

No registro da derivação periférica III, o terminal negativo do eletrocardiógrafo está conectado ao braço esquerdo e o terminal oposto à perna esquerda. Isso significa que o eletrocardiógrafo registra positivamente quando o braço esquerdo está negativo em relação à perna esquerda.

Na Figura 6 foi traçado um triângulo, chamado de Triângulo de Einthoven, em torno da área cardíaca. Esse é um meio esquemático de mostrar que os dois braços e a perna esquerda formam os vértices de um triângulo ao redor do coração. Os dois vértices na parte superior do triângulo representam os pontos onde os dois braços se conectam eletricamente com os líquidos do coração e o vértice inferior é o ponto onde a perna esquerda se conecta com esses líquidos.

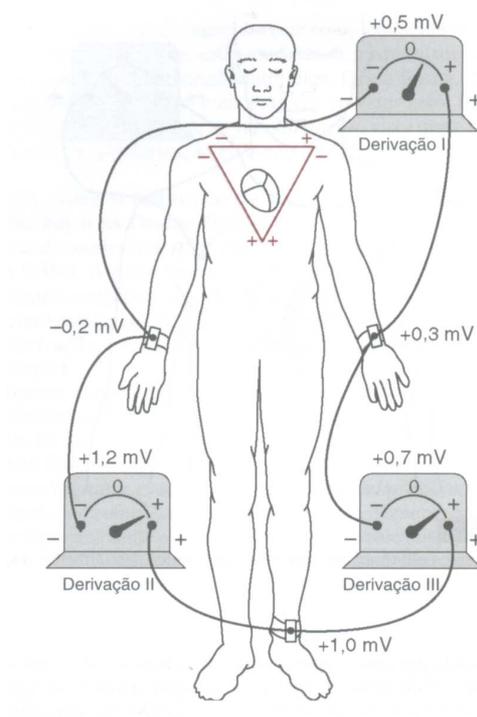


Figura 6: Disposição convencional dos eletrodos para o registro do eletrocardiograma pelas derivações padrão. O triângulo de Einthoven está disposto sobre o tórax.

A lei de Einthoven, semelhante à lei de Thèvenin, propõe que caso os potenciais elétricos de duas das três derivações bipolares periféricas forem conhecidos em determinado instante, a terceira pode ser determinada matematicamente a partir das duas primeiras, simplesmente pela sua soma com a consideração da polaridade. Um exemplo é apresentado na Figura 6, onde a soma das voltagens das derivações I e III é igual a voltagem da derivação II.

A Figura 7 mostra os registros eletrocardiográficos das derivações I, II e III. Note-se que, os eletrocardiogramas dessas três derivações são similares entre si, pois todos registram ondas P e T positivas e a parte principal do complexo QRS também é positiva. Ao se analisar os três eletrocardiogramas, pode-se mostrar, com medidas cuidadosas e observação das polaridades, que, em qualquer instante, a soma dos potenciais das derivações I e III é igual ao potencial da derivação II, validando assim a lei de Einthoven.



Figura 7: Eletrocardiogramas normais registrados pelas três derivações padrão.

Como o registro de todas as derivações bipolares periféricas são similares entre si, não importa muito qual derivação está sendo registrada quando se deseja diagnosticar as diferentes arritmias cardíacas, visto que o diagnóstico das arritmias depende principalmente das relações temporais entre as diferentes ondas do ciclo cardíaco. Por outro lado, quando se deseja diagnosticar lesão ventricular ou atrial, as derivações registradas são muito importantes, visto que as anomalias da contração cardíaca ou da condução do impulso cardíaco mudam acentuadamente o padrão eletrocardiográfico em algumas derivações, sem afetar as outras.

2.4.2 Derivações Torácicas (Derivações Pré-Cordiais)

Muitas vezes os eletrocardiogramas são registrados com um eletrodo colocado diretamente na superfície anterior do tórax, por sobre o coração, em um entre seis pontos distintos, de 1 a 6, apresentados na Figura 8. Esse eletrodo está conectado ao terminal positivo do eletrocardiógrafo e o eletrodo negativo, chamado de eletrodo indiferente, está conectado, por meio de resistências elétricas, ao braço direito, braço esquerdo e à perna esquerda, ao mesmo tempo. Normalmente, seis derivações torácicas padrão são registradas na parede torácica anterior, e o eletrodo torácico é colocado, respectivamente, nos seis pontos indicados na Figura 8.

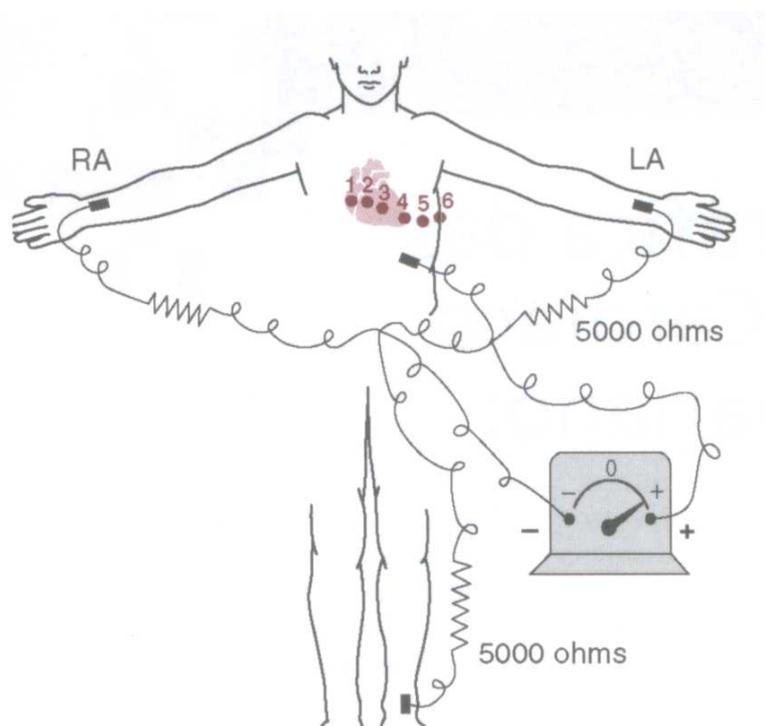


Figura 8: Conexões do corpo com o eletrocardiógrafo para registro das derivações torácicas.

Os diversos registros obtidos por esse método, mostrados na Figura 9, são conhecidos como derivações V_1 , V_2 , V_3 , V_4 , V_5 e V_6 . Dado que a superfície cardíaca fica próxima à parede torácica, cada derivação torácica registra, principalmente, o potencial elétrico da musculatura cardíaca, imediatamente abaixo do eletrodo. Portanto, anormalidades relativamente diminutas nos ventrículos podem causar alterações acentuadas nos eletrocardiogramas registrados pelas derivações torácicas.

Nas derivações V_1 e V_2 , os registros de QRS do coração normal são principalmente negativos, visto que, como mostrado na Figura 8, o eletrodo torácico nessas derivações está

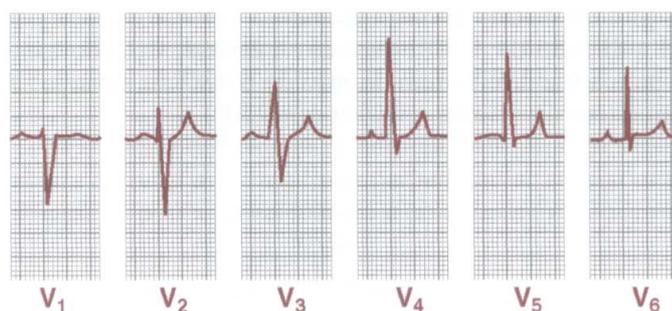


Figura 9: Eletrocardiogramas normais registrados pelas seis derivações torácicas.

muito mais próximo da base do coração que do ápice, e a base do coração é a direção da eletronegatividade durante a maior parte do processo de despolarização ventricular. Por outro lado, os complexos QRS nas derivações V_4 , V_5 e V_6 são em grande parte positivos, visto que o eletrodo torácico nessas derivações está próximo do ápice do coração, que é a direção da eletropositividade durante a maior parte da despolarização.

2.4.3 Derivações Unipolares Periféricas Aumentadas

Outro sistema de derivações de grande uso é o da derivação unipolar periférica aumentada. Nesse tipo de registro, dois dos membros são conectados por meio de resistências elétricas ao terminal negativo do eletrocardiógrafo e o terceiro membro é conectado ao terminal positivo. Quando o terminal positivo está no braço direito a derivação é chamada de aVR; quando no braço esquerdo, de aVL, e, quando na perna esquerda, de aVF.

Registros normais das derivações unipolares periféricas aumentadas são mostradas na Figura 10. Todos são similares aos registros das derivações periféricas padrão, exceto que o registro da derivação aVR está invertido.

2.5 Interpretação Eletrocardiográfica

Qualquer alteração no padrão da transmissão do impulso cardíaco pode causar potenciais elétricos anormais em torno do coração e, conseqüentemente, alterar as formas das ondas no eletrocardiograma. Por essa razão, quase todas as anormalidades graves do músculo cardíaco podem ser detectadas pela análise dos contornos das diferentes ondas nas diversas derivações eletrocardiográficas.

Muitas e distintas anormalidades cardíacas, especialmente as que lesam o músculo



Figura 10: Eletrocardiogramas normais registrados pelas três derivações unipolares periféricas aumentadas.

cardíaco, fazem com que parte do coração permaneça parcial ou totalmente despolarizada todo o tempo. Quando isso ocorre, a corrente flui entre as áreas patologicamente despolarizadas e as áreas normalmente polarizadas, mesmo entre os batimentos cardíacos. Isso é chamado de corrente de lesão. Ressalta-se que a parte lesada do coração é negativa, porque essa parte está despolarizada e emite cargas negativas para os líquidos circundantes, enquanto o restante do coração é positivo.

Algumas das anormalidades que causam corrente de lesão são: i) trauma mecânico, que torna as membranas tão permeáveis que a repolarização total não pode ocorrer; ii) processos infecciosos, que lesam as membranas musculares e; iii) isquemia de áreas localizadas do músculo, provocada por oclusão coronária local, que é a causa mais comum de corrente de lesão no coração. Durante a isquemia, nutrientes suficientes do suprimento sanguíneo coronário não estão disponíveis para manter a despolarização da membrana do músculo cardíaco normal.

Poder-se-ia pensar que os aparelhos cardiográficos poderiam determinar os períodos em que nenhuma corrente está fluindo ao redor do coração. Entretanto, existem no corpo muitas correntes fugidias, tais como as correntes que resultam dos potenciais cutâneos e de diferentes concentrações iônicas nos diversos líquidos do corpo. Portanto, quando dois eletrodos são conectados entre os braços ou entre um braço e uma perna, essas correntes fugidias impossibilitam a pré-determinação do ponto exato de referência zero no eletrocardiograma. Por essas razões, o seguinte procedimento é adotado para se determinar o ponto zero de referência: primeiro, deve-se detectar o ponto exato no qual a onda de despolarização completa sua passagem pelo coração, o que ocorre no fim do complexo QRS. Exatamente nesse ponto, todas as regiões dos ventrículos estão despolarizadas, in-

cluindo as regiões lesadas e as normais, de modo que nenhuma corrente flui ao redor do coração. Até mesmo a corrente de lesão desaparece nesse ponto. Portanto, o potencial do eletrocardiograma, nesse instante, tem voltagem zero. Esse ponto é chamado de ponto “J” no eletrocardiograma, como mostrado na Figura 11.

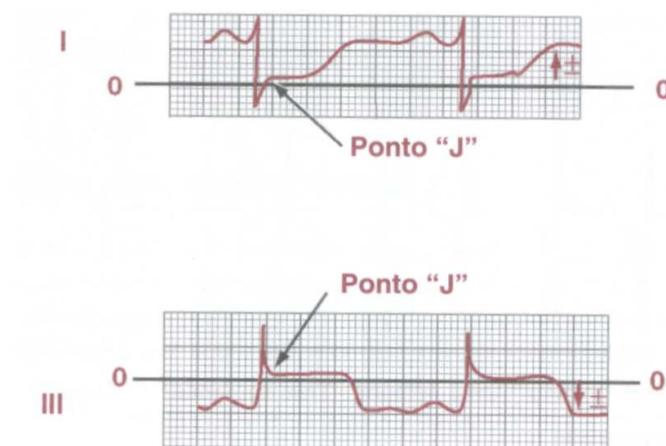


Figura 11: Ponto "J" como referencial zero de referência do eletrocardiograma.

Em seguida, uma linha horizontal é traçada através do eletrocardiograma, no nível do ponto J, e essa linha horizontal é o nível do potencial zero nesse eletrocardiograma, a partir do qual todos os potenciais causados pelas correntes de lesão devem ser medidos.

A Figura 11 mostra eletrocardiogramas de coração lesionado, registrado pelas derivações I e III, que mostram correntes de lesão. Em outras palavras, o ponto J de cada um desses dois eletrocardiogramas não está sobre a mesma linha que o segmento TP (final da onda T e início da onda P). A linha horizontal foi traçada pelo ponto J para representar o nível de voltagem zero em cada um dos dois registros. A voltagem da corrente de lesão, em cada derivação, é a diferença entre o nível do segmento TP do eletrocardiograma (que fica entre os batimentos cardíacos quando há corrente de lesão) e o nível do potencial de voltagem zero, como mostrado pelas duas setas nas respectivas derivações. Na derivação I, a voltagem registrada causada pela corrente de lesão está acima do nível do potencial zero e, portanto, é positiva. Inversamente, na derivação III, o segmento TP está abaixo do nível de voltagem zero; portanto, a voltagem de corrente de lesão na derivação III é negativa.

A parte do eletrocardiograma que ocorre entre a extremidade do complexo QRS e o início da onda T é chamada de segmento ST. O ponto J fica no início desse segmento.

Portanto, cada vez que uma corrente de lesão ocorre em uma das derivações eletrocardiográficas, verifica-se que os segmentos ST e TP do eletrocardiograma não estão no mesmo nível do registro. Na verdade, é o segmento TP, e não o segmento ST, que é deslocado do eixo zero. A maioria das pessoas, no entanto, estão condicionadas a considerar o segmento TP do eletrocardiograma com o nível de referência zero, em vez do ponto J. Portanto, quando uma corrente de lesão está evidente no eletrocardiograma, parece que o segmento ST é deslocado do nível normal desse eletrocardiograma, o que é chamado de desvio do segmento ST, que é utilizado na identificação de isquemia. Quando se vê um desvio do segmento ST, sabe-se, imediatamente, que esse eletrocardiograma apresenta as características de uma corrente de lesão. De fato, a maioria dos eletrocardiógrafos não se refere à corrente de lesão mas, simplesmente, faz menção ao desvio do segmento ST, o que significa a mesma coisa.

O fluxo sanguíneo insuficiente para o músculo cardíaco deprime o metabolismo do músculo, por três razões: falta de oxigênio, acúmulo excessivo de dióxido de carbono e falta de nutrientes suficientes. Conseqüentemente, a repolarização das membranas não ocorre nas áreas de isquemia miocárdia grave. Frequentemente, o músculo cardíaco não morre porque o fluxo sanguíneo é suficiente para manter a vida do músculo, embora não seja suficiente para provocar a repolarização das membranas. Enquanto esse estado persiste, a corrente de lesão continua a fluir durante o período sistólico (período TP) de cada batimento cardíaco.

A isquemia extrema do músculo cardíaco ocorre após oclusão coronária, e uma forte corrente de lesão flui da área infartada dos ventrículos, durante o intervalo TP, entre os batimentos, como mostrado nas Figuras 12 e 13. Por conseguinte, uma das características diagnósticas mais importantes dos eletrocardiogramas após trombose coronária aguda é a corrente de lesão.

A isquemia moderada é a causa mais comum de aumento na duração da despolarização no músculo cardíaco. Quando ela ocorre em apenas uma área do coração, o período de despolarização dessa área aumenta desproporcionalmente em relação ao de outras partes. Como conseqüência, pode acontecer mudanças definitivas na onda T. A isquemia pode resultar de oclusão coronária progressiva e crônica, oclusão coronária aguda ou insuficiência coronária relativa que ocorre durante o exercício.

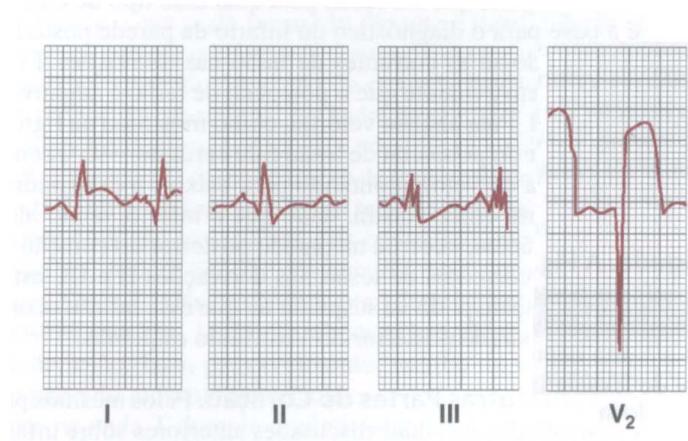


Figura 12: Corrente de lesão em infarto agudo da parede anterior.

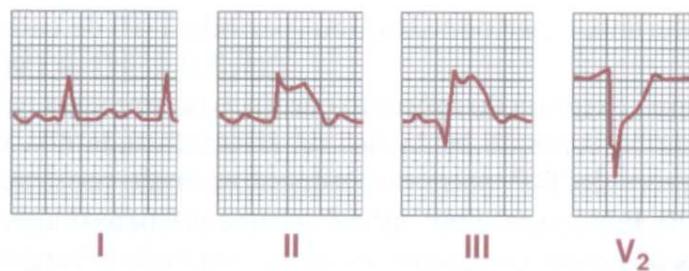


Figura 13: Corrente de lesão em infarto agudo apical de parede posterior.

3 Sistema Atual e Metodologia

3.1 Introdução

A implementação deste trabalho se deu com o uso de uma interface gráfica (GUI-Graphical User Interface) em ambiente MatLab desenvolvida em (OLIVEIRA; ANDREAO, 2003-2004), a qual é vista na Figura 14. O sistema realiza a classificação do sinal de ECG em quatro passos: *Extração de Características, Aprendizagem, Segmentação e Classificação*.

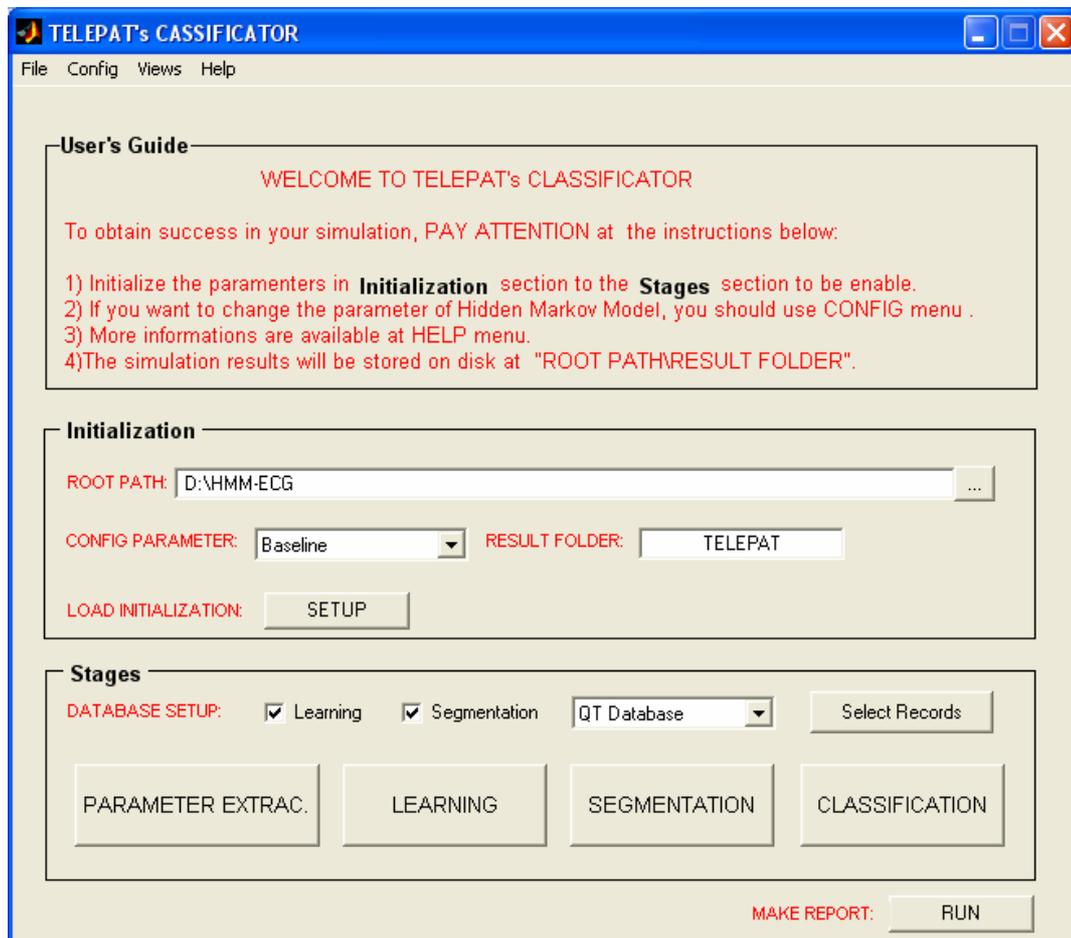


Figura 14: Interface GUI-Matlab desenvolvida em (OLIVEIRA; ANDREAO, 2003-2004).

O núcleo do sistema é composto pelos algoritmos de processamento de sinais eletrocardiográficos, que estão estruturados de forma hierárquica, como mostrado na Figura 15.

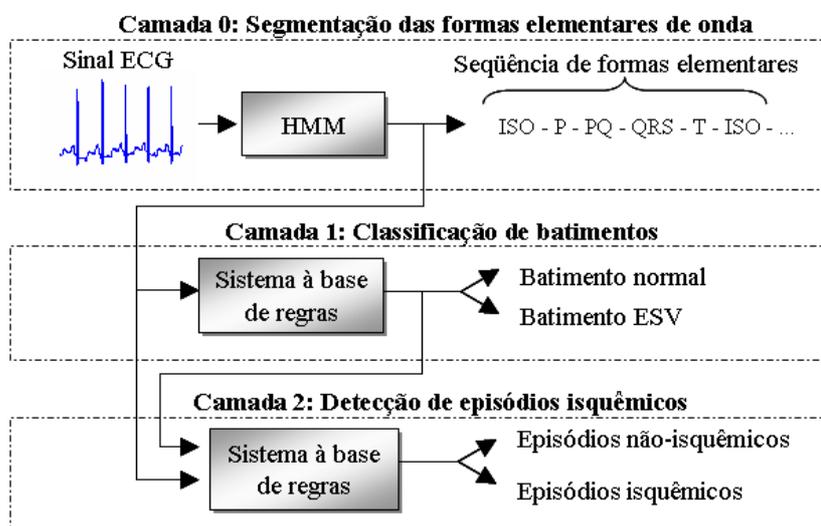


Figura 15: Sistema de detecção de isquemia.

A *Camada 0* tem por função identificar uma seqüência de formas elementares de onda a partir de uma seqüência de observações do sinal de ECG. Já a *Camada 1* faz a classificação de batimentos ESV (extrasystole ventriculaire), que são batimentos que apresentam extrasístole, ou normais, que abrangem todas as outras categorias de batimentos. Na *Camada 2* é feita a detecção de episódios isquêmicos a partir de uma análise ao longo do tempo da evolução do deslocamento do segmento ST. É importante frisar que o presente trabalho se restringe à implementação de novas técnicas de adaptação dos modelos, ou seja, os algoritmos que serão apresentados aqui são implantados na *Camada 0*.

O sistema apresentado na Figura 15 foi projetado para antecipar a ocorrência de uma doença cardíaca, a saber, a isquemia. Esse trabalho, melhor detalhado em (ANDREAO, 2004), faz parte de um projeto francês de tele-vigilância, chamado *Tele-Pat*, que tem por objetivo a realização de um sistema de alarme automático para o acompanhamento à distância de pessoas idosas ou com doenças cardíacas, (LAMKADDEM, 2005). Esse sistema registra eletrocardiogramas, por modo portátil, e o transmite a uma base de recepção doméstica de tele-vigilância. Uma conexão do conjunto é garantida com um servidor remoto localizado no centro de tele-vigilância médico, o qual pode reagir aos alarmes vindos do dispositivo de vigilância na casa do paciente ou mesmo explorar continuamente as informações biomédicas do paciente. O sistema é mostrado na Figura 16.

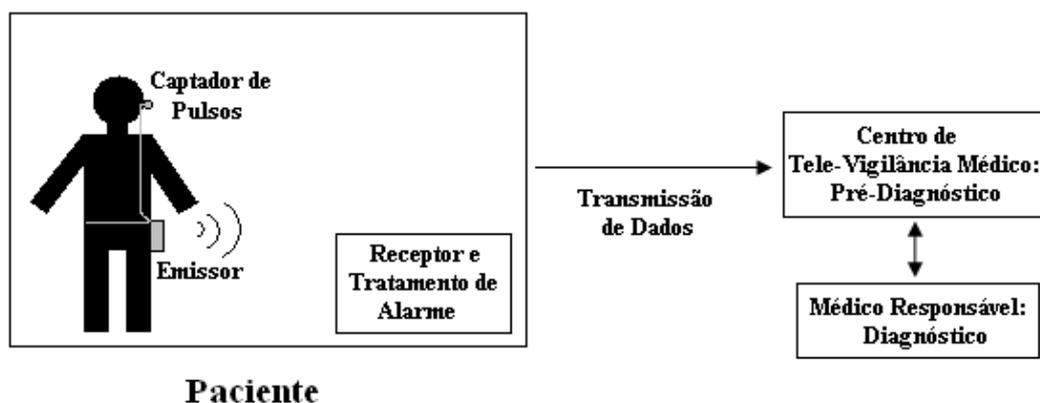


Figura 16: Sistema de tele-vigilância TelePat.

O tratamento de alarme apresentado na Figura 16 corresponde à análise de movimentação do corpo, com a medição dos pulsos cardíacos, e pode se estender à identificação de doenças cardíacas utilizando a plataforma desenvolvida em (ANDREAO, 2004). Assim, o sistema, já treinado com ECG normal e com aqueles que apresentam anomalias, é capaz de gerar um alarme para prevenção de infarto do miocárdio.

O presente trabalho, que opera na *Camada 0*, Figura 15, tem por objetivo identificar as formas de onda do sinal de ECG, conforme mostra a Figura 17. O sistema deve ser capaz de processar o sinal de diversos pacientes, com a finalidade de trabalhar com as características de cada pessoa. Para tal fim, é realizado um processo de aprendizagem onde o sistema aprende as formas elementares de onda do sinal de ECG, a partir da marcação feita por cardiologistas em uma base de dados de treinamento. Essa marcação, também chamada de etiquetas ou rótulos, é feita em batimentos normais, uma vez que os batimentos com alguma anomalia são marcados com a respectiva anomalia e não com a marcação das formas elementares. A adaptação ao paciente é entendida como a aprendizagem das formas de onda observadas do sinal de ECG de cada indivíduo.

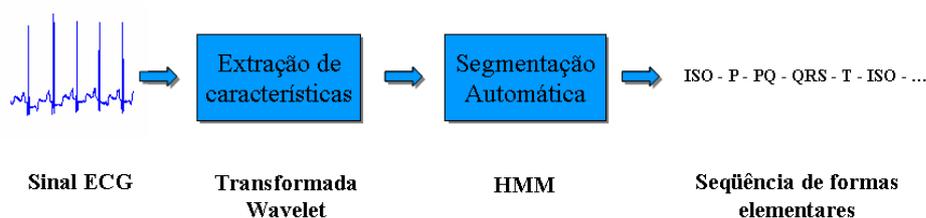


Figura 17: Arquitetura básica do sistema de detecção.

A vantagem do sistema adaptativo está na compensação das variações no sinal de ECG produzidas no decorrer da análise, que podem ser causadas por:

- i) deslocamento dos eletrodos, movimentos do paciente, e outros problemas com a captura do sinal;
- ii) variações no sinal de ECG de pessoa para pessoa, e variações na mesma pessoa durante o dia.

Uma vez que se utilize o processo de adaptação, o banco de dados necessário para o treinamento será menor, pois o sistema sempre estará em treinamento, o que melhora o desempenho do sistema automático de análise do sinal de ECG apresentado na Figura 17. Deve-se ressaltar que o estudo sobre adaptação apresentado no presente trabalho foi aplicado sobre o sistema desenvolvido em (ANDREAO, 2004), e por isso é necessário conhecer as características do atual sistema para a proposição de implementações de técnicas de adaptação.

3.2 Sistema Atual

O sistema desenvolvido em (ANDREAO, 2004) utiliza de uma abordagem markoviana para a segmentação do batimento cardíaco, rerepresentado na Figura 18, com aplicação à detecção de isquemia. Um dos primeiros trabalhos nessa linha foi o desenvolvido por (COAST et al., 1990). Ali, os autores fazem cada estado do modelo de Markov, com a topologia esquerda-direita, corresponder a uma forma elementar do sinal de ECG, conforme visto na Figura 19. Isso faz com que o modelo aprenda o batimento dentro de sua globalidade, com a relação direta entre os estados e as ondas. Por exemplo, se é desejado que o modelo aprenda a morfologia de uma nova onda P ou um novo complexo QRS, todo o modelo do batimento deve ser treinado, o que aumenta a complexidade do problema.

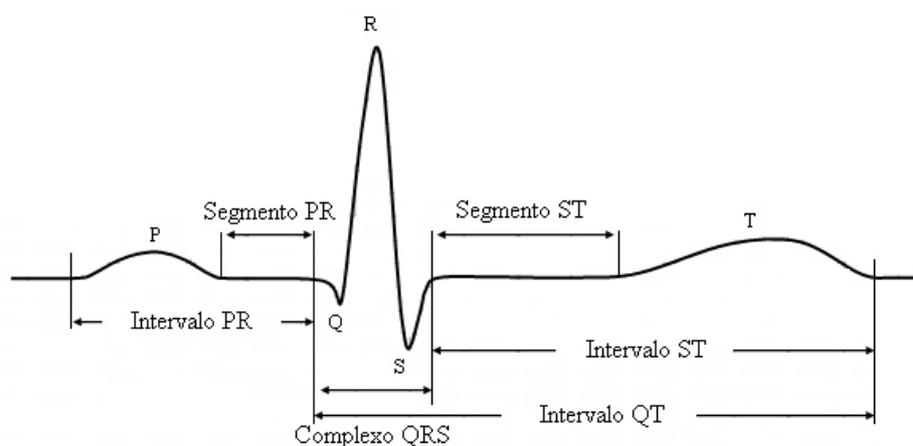


Figura 18: Batimento cardíaco.

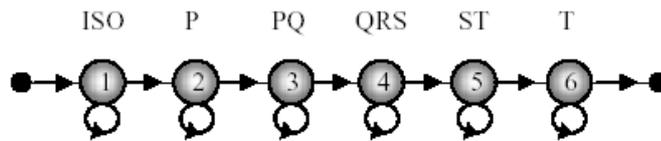


Figura 19: Modelo do batimento cardíaco, utilizado por (COAST et al., 1990).

Diferentemente, em (ANDREAO, 2004) cada forma elementar de onda e segmentos do sinal de ECG são modelados por um HMM, que, concatenadas, facilitam a aprendizagem de novas morfologias. A Figura 20 apresenta o modelo do batimento composto de HMMs concatenados. Vale lembrar que a onda ISO, que vem de linha isoeletrica, corresponde ao intervalo entre o final da onda T até o início da onda P (ou do complexo QRS, caso não haja onda P). É possível notar a introdução de transições entre os modelos, como: i) a transição da onda P para a onda ISO, para modelar as ondas P não seguidas de uma atividade ventricular; ii) transição de ISO para PQ sem passar por P, para modelar as arritmias ventriculares e supra-ventriculares sem onda P visível; iii) transição entre onda T e ISO, para o modelamento contínuo de uma seqüência de batimentos.

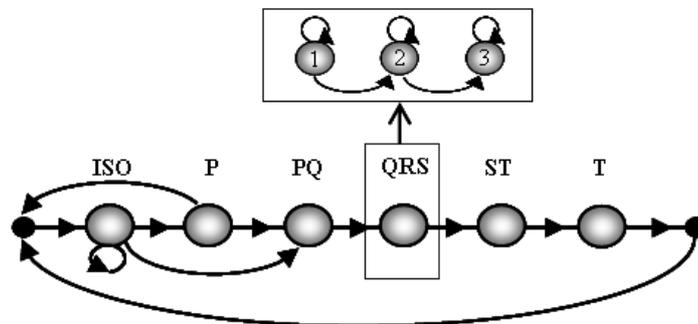


Figura 20: Modelo do batimento cardíaco composto pela concatenação de HMMs.

A concatenação dos modelos é feita pelo uso do algoritmo *one-pass*, onde os níveis correspondem às ondas do batimento. Esse método foi adaptado para efetuar o alinhamento temporal dos modelos à seqüência de observações O_i , $i = 1, \dots, T$, segundo o modelo do batimento proposto. A Figura 21 é um exemplo da decodificação de uma seqüência de observações, cujo resultado é a posição, no tempo, de cada forma elementar do batimento, o que, neste exemplo, significa a seqüência dos modelos ISO-P-PQ-QRS-ST-T-ISO-P.

O processo de adaptação é composto por duas fases, mostradas na Figura 22. Na Fase 1, os modelos genéricos são adaptados pela primeira vez às morfologias do sinal de ECG do indivíduo, sendo feito com os primeiros 20 segundos do traçado de ECG. Já a Fase 2 visa a adaptar os parâmetros do modelo às flutuações que podem aparecer, ao longo

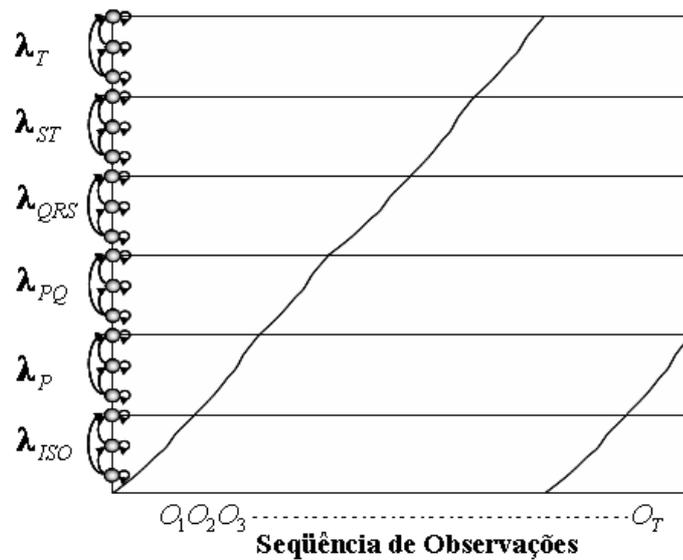


Figura 21: Decodificação da sequência de observações com o uso do algoritmo *one-pass*.

do tempo, sobre o sinal da pessoa. Nessa fase, a matriz de covariância das distribuições gaussianas dos estados não é reestimada, mas sim o vetor média. A matriz é fixada nos valores obtidos na Fase 1. A outra reestimação que é feita na Fase 2 é a da matriz de probabilidade de transição de estados.

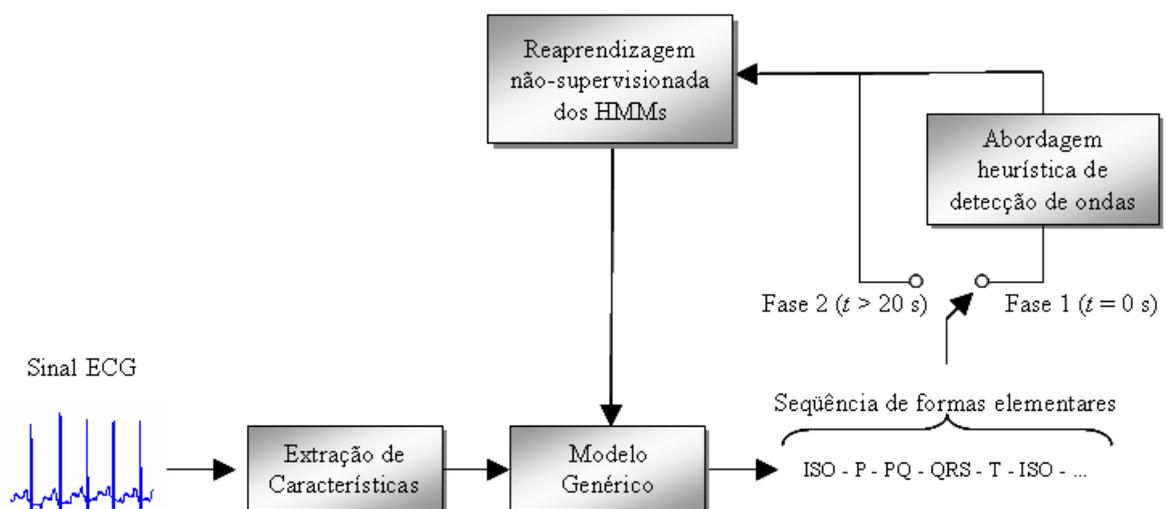


Figura 22: Sistema de segmentação automática do sinal de ECG.

A estrutura do algoritmo de adaptação Baum-Welch (RABINER, 1989), da família EM, é ilustrado na Figura 23. Note-se que cada registro de traçado de ECG contido na base usada, a base QT (LAGUNA et al., 1997), em geral possui 225000 pontos (15 min), com um período de amostragem de 4 ms. Esses pontos são analisados (segmentados e adaptados) em janelas de 20 s, ou seja, de 5000 pontos, dentro da rotina *hmm_segmentation_gui.m*.

Observa-se então, que ao contrário de uma abordagem em lote, que utiliza todos os pontos de uma só vez, o sistema atual não usa todo o registro de uma única vez para se efetuar os cálculos, mas o faz em blocos menores.

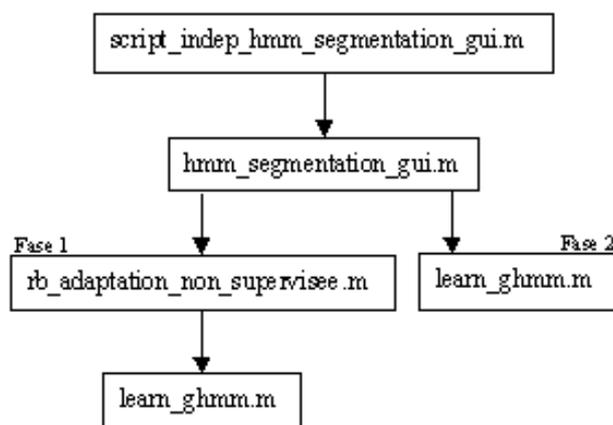


Figura 23: Arquitetura básica do algoritmo de (ANDREAO, 2004).

Essa rotina chama uma outra chamada, *rb_adaptation_non_supervisee.m*, que realiza a Fase 1 do processo de adaptação. Essa fase consiste em adaptar, pela primeira vez, os modelos genéricos às morfologias do sinal de ECG do indivíduo nos primeiros 20 segundos do seu traçado, ou seja, no primeiro bloco de 5000 pontos. Este processo é feito com o chamado da função *learn_ghmm.m*.

Após a primeira fase, é, então, chamada outra vez a rotina *learn_ghmm.m*, dentro do corpo principal de *hmm_segmentation_gui.m*, onde é implementada a Fase 2, que visa a adaptar os parâmetros do modelo às flutuações que podem aparecer ao longo do tempo sobre o sinal de ECG. Essa rotina será utilizada a cada bloco ou conjunto de 5000 pontos até o final do traçado. Dentro de cada um desses blocos, a rotina é chamada para cada modelo das formas elementares de onda, como o modelo de onda P, de onda T, de complexo QRS, de linha ISO, entre outros. E então, os exemplos de ocorrência da forma elementar de onda em análise, dentro desse bloco, são guardados e depois passados para *learn_ghmm.m*, para os cálculos de adaptação.

Sobre o processo de adaptação, é importante citar que, dentro de cada bloco e depois para cada forma elementar de onda, os passos E (Expectation) e M (Maximization) do algoritmo Baum-Welch são executados até no máximo 10 vezes, ou até a convergência, a qual é medida pela diferença da verossimilhança entre duas iterações. O número máximo de iterações foi determinado de forma empírica, assim como o limiar aceito para a convergência (ANDREAO, 2004). Após *learn_ghmm.m*, os modelos são reestimados.

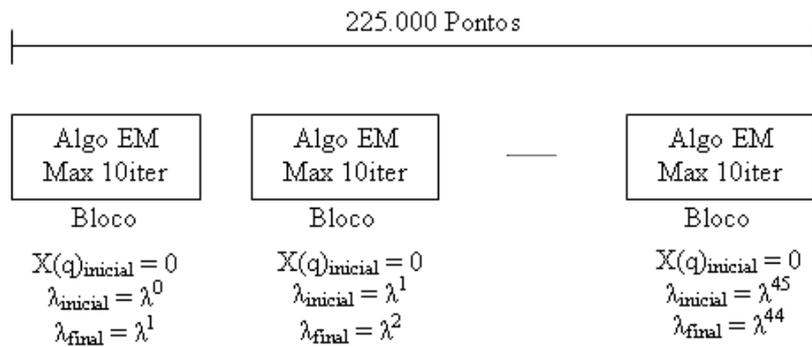


Figura 24: Sistema desenvolvido por (ANDREA, 2004).

É importante ressaltar que na passagem de um bloco a outro nenhuma informação sobre o bloco anterior é guardada, a não ser os próprios modelos λ reestimados. Essas informações correspondem às variáveis estatísticas que compõem a distribuição de probabilidade dos estados sobre o conjunto de variáveis ocultas, $X(q)$. Assim, cada bloco é inicializado com os modelos encontrados no grupo anterior, mas o valor inicial da distribuição $X(q)$, que armazena as informações do modelo, é igual a zero. Isto pode ser observado na Figura 24.

Observa-se que a abordagem adotada em (ANDREA, 2004) não é uma abordagem em lote propriamente dita, uma vez que não é utilizado todo o conjunto de dados, de uma única vez, para as estimações feitas no algoritmo Baum-Welch (RABINER, 1989), mas também não se torna uma abordagem incremental, pois executa 10 iterações até a convergência por verossimilhança.

3.3 Metodologia

Para a realização deste trabalho, primeiro, estudou-se a extração de características com o uso da interface já apresentada neste capítulo. Isso permitiu o estudo sobre a influência da transformada wavelet na extração de parâmetros do sinal de ECG. Depois dessa etapa, o passo seguinte foi o estudo da segmentação do sinal de ECG, o que engloba o estudo da teoria de HMM (que será apresentada nos próximos capítulos) e sua implementação na plataforma utilizada.

3.3.1 Extração de Parâmetros

A partir da teoria apresentada no Anexo A, estudaram-se os resultados da transformada wavelet sobre o sinal de ECG em relação à escolha da função wavelet e da escala utilizada. Para isso, com o caminho *View*→*Parameter Extraction*, na interface, é visualizada uma janela (Figura 25), que permite escolher a função wavelet, o canal e o arquivo de gravação. Note-se que a transformada wavelet é utilizada em outros trabalhos envolvendo processamento de eletrocardiograma, como em (PROVAZNIK et al., 2000).

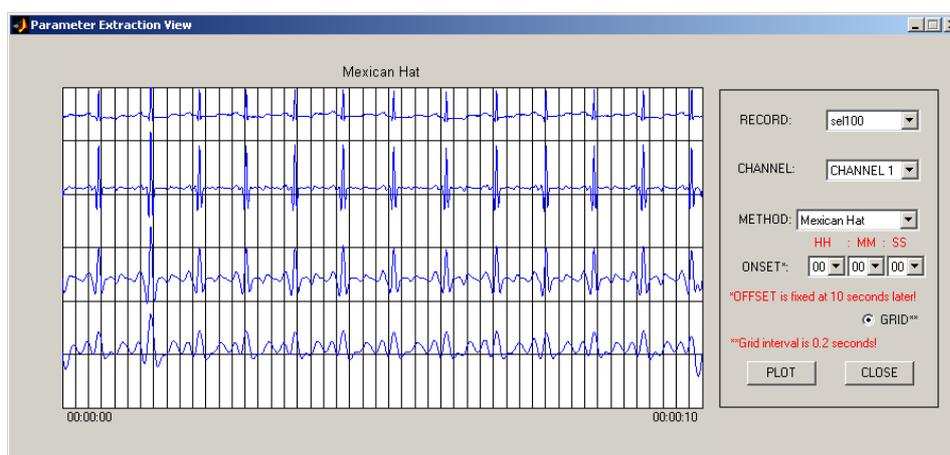


Figura 25: Tela de avaliação.

Um estudo comparativo é apresentado na Figura 26, utilizando as funções Chapéu Mexicano e Morlet. Nota-se que, com a função Chapéu Mexicano, é possível ressaltar características do sinal de ECG como o complexo QRS, o mesmo não acontecendo para a wavelet Morlet. Após um estudo desenvolvido em (ANDREAO, 2004), foi eleita a wavelet Chapéu Mexicano como aquela que melhor caracteriza o sinal de ECG.

A Figura 27 mostra a representação de um sinal de ECG, no primeiro quadro, e a transformada wavelet com a função Chapéu Mexicano, em três escalas distintas, nos quadros seguintes. A extração de características do sinal de ECG é feita com a Transformada Wavelet Contínua (CWT) (ver anexo A), que permite uma representação robusta do sinal e um modelamento preciso das ondas e das rupturas do sinal. A função wavelet escolhida foi a Chapéu Mexicano com escala de dois, $s = 2^j$, onde $j = 2, 3$ e 4 .

O resultado da transformada corresponde às observações do sistema, e as três escalas combinadas se tornam uma representação que engloba as informações úteis do batimento cardíaco.

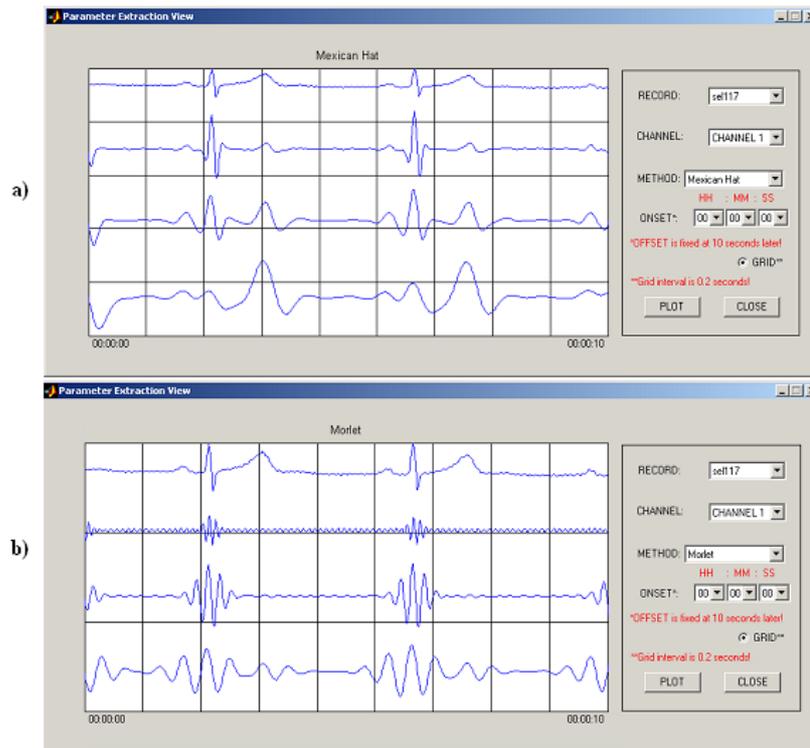


Figura 26: Comparação com duas funções wavelets: a) usando a wavelet Chapéu Mexicano; b) usando a wavelet Morlet.

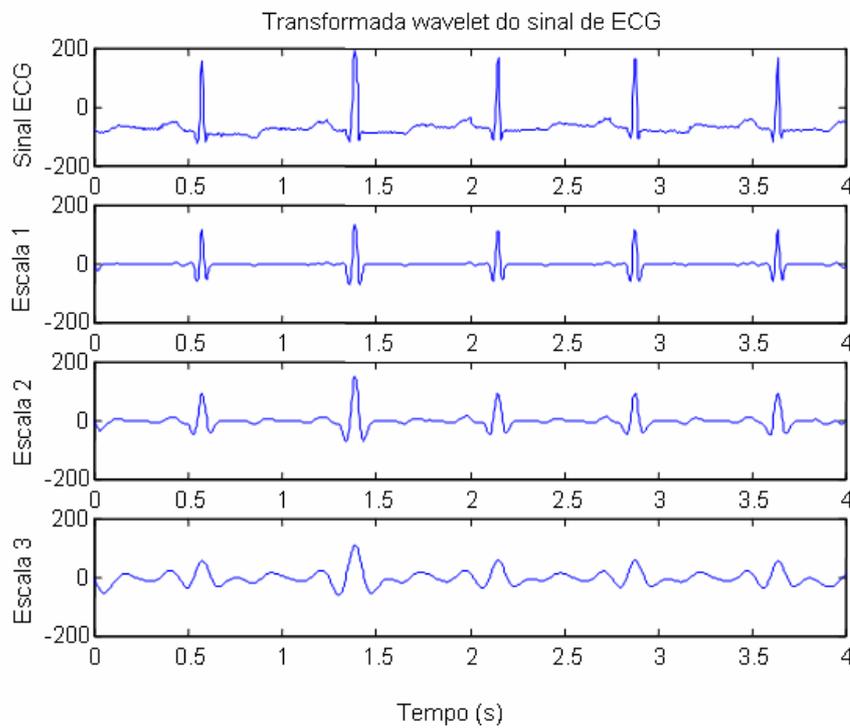


Figura 27: Transformada wavelet para um sinal de ECG.

3.3.2 Segmentação Automática do Sinal de ECG

A segmentação do sinal de ECG pode ser vista como a decodificação de uma seqüência de observações em função das formas elementares do batimento. O processo de segmentação utiliza o algoritmo de Viterbi, e é controlado por um conjunto de regras, das quais se pode citar: limiar de amplitude, correção da segmentação da onda T e supressão de falsos complexos QRS (ANDREAO, 2004).

A probabilidade de emissão $b_j(o_t)$, um dos parâmetros dos modelos de Markov, geralmente é dada por uma mistura de gaussianas. Contudo, a partir dos histogramas das observações da base de aprendizagem, exemplificados pela Figura 28, determinou-se que as observações podem ser modeladas por uma única gaussiana por estado de cada HMM. Apesar que uma mistura de gaussianas consegue modelar melhor um conjunto de observações, ela demanda um grande custo computacional, o que não é justificável pelos testes realizados em (ANDREAO, 2004).

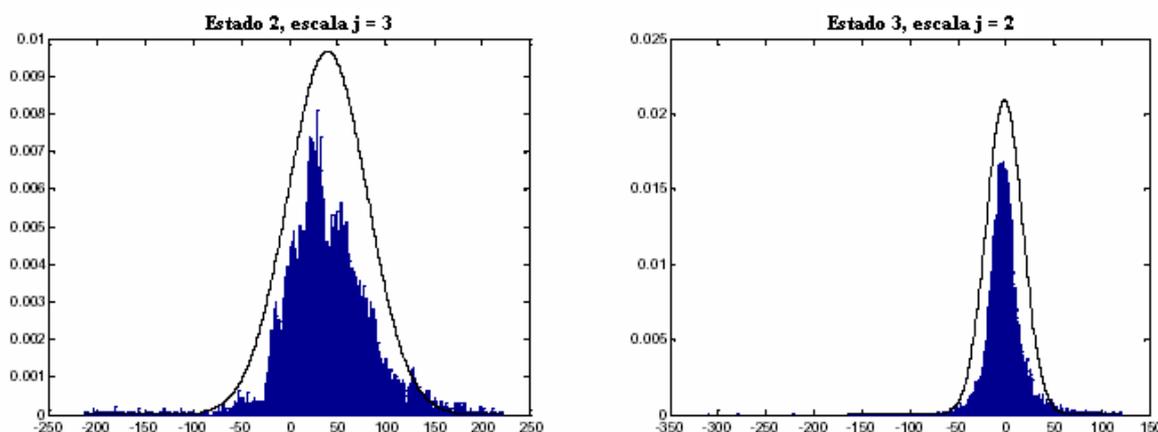


Figura 28: Histograma das observações da onda P. A escala j é correspondente à transformada wavelet efetuada.

3.3.3 Aprendizagem e Adaptação

Uma etapa relacionada à segmentação do sinal de ECG é a aprendizagem dos modelos de Markov. Essa aprendizagem, ou seja, a estimação dos parâmetros dos modelos markovianos, é feita pelo método de Baum-Welch (BW) (RABINER, 1989), explicado mais adiante, que maximiza a verossimilhança $P(O|\lambda)$, dadas as observações e dado o modelo. Os modelos de cada forma elementar do batimento são treinados por um conjunto de aprendizagem próprio. A partir de um modelo inicial, no qual uma boa estimativa se faz necessária, o método BW executa a aprendizagem iterativa dos modelos, conforme pode

ser visto na Figura 29, e o critério de convergência é a comparação da verossimilhança dos dados de aprendizagem entre duas iterações consecutivas.

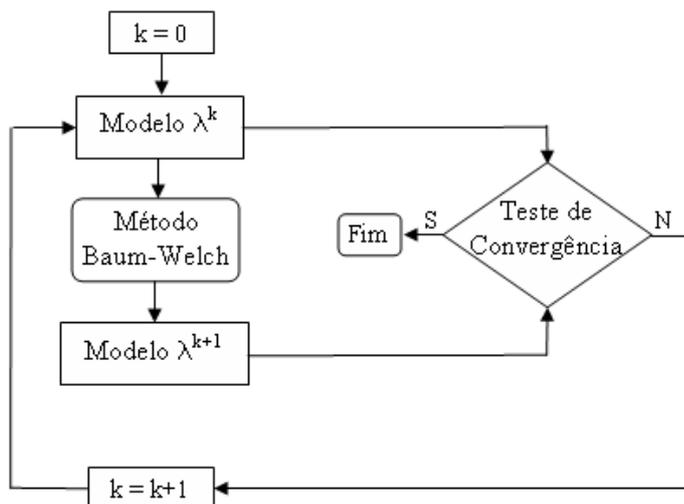


Figura 29: Aprendizagem do modelo λ^k pelo método de Baum-Welch (RABINER, 1989).

3.3.4 Especificação dos Modelos

As características dos modelos utilizados para cada forma elementar de onda são mostradas na Tabela 1. É importante dizer que todos os HMMs possuem a configuração *esquerda-direita* e é utilizada uma única gaussiana para cada estado de cada modelo.

Tabela 1: Especificações dos modelos.

Modelo	P	PQ	QRS	ST	T	ISO
Nº de estados	3	2	3	2	6	3
Nº de exemplos de treinamento	660	719	770	339	760	697

4 *Técnicas de Adaptação*

Neste capítulo serão estudadas as técnicas de adaptação implementadas no sistema desenvolvido em (ANDREAIO, 2004). Para a total compreensão dessas técnicas é necessário abordar, primeiramente, a teoria dos modelos ocultos de Markov (HMM).

4.1 Teoria de HMM

Estudados a partir dos anos 60, os métodos estatísticos de Markov se tornaram usuais por dois grandes motivos. Primeiro, porque os modelos de Markov possuem uma boa estrutura matemática, que serve de base teórica em um grande número de aplicações. A outra razão é que, quando os modelos são aplicados apropriadamente, eles funcionam relativamente bem para aplicações importantes como reconhecimento de voz e tratamento de sinais biomédicos.

4.1.1 Elementos de um HMM

Pode-se definir formalmente os elementos de um HMM, segundo Rabiner (1989), pelas variáveis a seguir:

- N é o número de estados do modelo. Mesmo que os estados sejam escondidos, que é o caso deste trabalho, eles estão relacionados a fenômenos físicos do sinal a ser tratado. Os estados são chamados de $S = \{S_1, S_2, \dots, S_N\}$ e o estado no tempo t é chamado de q_t .
- M é o número de símbolos distintos existentes. Os símbolos propriamente ditos são elementos do conjunto $V = \{v_1, v_2, \dots, v_M\}$.
- \mathbf{A} é a matriz de probabilidade de transição de estados. Os elementos da matriz \mathbf{A} podem ser definidos como $a_{ij} = P[q_{t+1} = S_j | q_t = S_i]$, ou seja, a probabilidade de se estar no estado S_j no tempo $t + 1$, dado que o estado anterior foi S_i .

- B é a distribuição das probabilidades de emissão de símbolos, e é definida por:
 $B = \{b_j(k)\}$, onde $b_j(k) = P[v_k \text{ em } t | q_t = S_j]$
- π é a probabilidade inicial dos estados.

Pode-se, então, representar um modelo de Markov λ como $\lambda = (A, B, \pi)$ e o sistema irá processar as observações $O = (O_1 O_2 \dots O_T)$, que correspondem ao sinal medido. T é o número de observações.

4.1.2 Os três problemas básicos do HMM

Existem três problemas a serem resolvidos por HMM e que são utilizados em muitas aplicações, a saber:

- problema 1: dada uma seqüência de observações e um modelo λ , qual a probabilidade dessas observações acontecerem, dado o modelo? Isto é, calcular $P(O|\lambda)$.
- problema 2: dada a seqüência de observações e o modelo λ , como escolher uma seqüência de estados $Q = q_1 q_2 \dots q_T$ que melhor representa a seqüência de observações?
- problema 3: como ajustar os parâmetros do modelo $\lambda = (A, B, \pi)$ para maximizar $P(O|\lambda)$?

O primeiro problema trata de dizer qual o modelo, dentre os implementados, que melhor representa uma seqüência de observações, ou seja, dizer a que classe pertencem as observações. Já o segundo problema segmenta, ou seja, tenta encontrar a melhor seqüência de estados para uma dada observação. É importante dizer que há vários critérios de otimização a se utilizar, e mais adiante será proposta uma opção. Por fim, o último problema visa otimizar os parâmetros, a fim de criar um modelo o mais próximo possível do real, usando, para isso, uma seqüência de observações em um processo de treinamento, que é essencial para a adaptação dos parâmetros do modelo.

4.1.2.1 Solução para o problema 1: cálculo da Máxima Verossimilhança (ML)

Uma das maneiras para se calcular a probabilidade da seqüência de observação $O = O_1 O_2 \dots O_T$ dado o modelo λ , ou seja, para se calcular $P(O|\lambda)$, é contar todas as seqüências de estados de mesmo comprimento T , isto é, $Q = q_1 q_2 \dots q_T$, onde q_1 é o estado inicial.

A probabilidade da seqüência de observação para a seqüência de estados dada acima pode ser escrita como

$$P(O|Q, \lambda) = \prod_{t=1}^T P(O_t|Q_t, \lambda),$$

onde se assume a independência estatística das observações, ou seja, $P(O_1O_2\dots O_T) = \prod_{i=1}^T P(O_i)$. Assim, tem-se que:

$$P(O|Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \dots b_{q_T}(O_T).$$

Depois de equacionar $P(O|Q, \lambda)$, é possível definir $P(Q|\lambda)$ como

$$P(Q|\lambda) = \pi_{q_1} a_{q_1q_2} a_{q_2q_3} \dots a_{q_{T-1}q_T}.$$

Então, a probabilidade das observações e da seqüência de estados acontecerem ao mesmo tempo é o produto dos dois termos acima, isto é,

$$P(O, Q|\lambda) = P(O|Q, \lambda)P(Q|\lambda).$$

A expressão acima é chamada de expressão de Máxima Verossimilhança, e, assim, a probabilidade desejada, $P(O|\lambda)$, é obtida pelo somatório das probabilidades sobre todas as possíveis seqüências de estados q , obtendo-se

$$\begin{aligned} P(O|\lambda) &= \sum_Q P(O|Q, \lambda)P(Q|\lambda) \\ &= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1q_2} b_{q_2}(O_2) \dots a_{q_{T-1}q_T} b_{q_T}(O_T). \end{aligned}$$

Devido ao custo computacional para o cálculo de $P(O|\lambda)$, de acordo com a expressão acima, existe um algoritmo chamado *Forward-Backward*, que diminui o número de operações para o cálculo da probabilidade, que pode ser visto em (RABINER, 1989).

4.1.2.2 Solução para o problema 2: algoritmo de Viterbi

Diferentemente do problema 1, que possui uma solução exata, o problema 2 pode ser resolvido de várias maneiras, dependendo do critério de otimização utilizado para encontrar a seqüência ótima de estados associada à seqüência de observação dada. Por exemplo, um critério possível seria escolher os estados q_t os quais tem, individualmente, a maior probabilidade de ocorrência. Este critério de otimização maximiza o número esperado de estados individualmente corretos. Este critério apresenta alguns problemas com a seqüência resultante, pois esta solução determina o estado mais provável a cada instante sem levar em consideração a probabilidade de ocorrência de seqüência de estados.

Pode-se, então, modificar o critério de otimização. O mais utilizado é encontrar a única melhor seqüência de estados, ou seja, maximizar $P(Q|O, \lambda)$, que é equivalente a maximizar $P(Q, O|\lambda)$. A solução é, então, chamada de Algoritmo de Viterbi, que é baseado em métodos de programação dinâmica. Para encontrar a melhor seqüência de estados, dada a observação, define-se $\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1 q_2 \dots q_t = S_i, O_1 O_2 \dots O_t | \lambda]$ como a maior probabilidade ao longo do caminho, no tempo t , que considera as t observações que terminam no estado S_i . Por indução, tem-se que

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) a_{ij} \right] b_j(O_{t+1}).$$

Para recuperar a seqüência de estados, é necessário manter os dados que maximizam a expressão acima para cada t e j . Isto é feito utilizando o vetor $\psi_t(j)$. O algoritmo de Viterbi pode ser visto como o procedimento a seguir

- Inicialização:

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N$$

$$\psi_1(i) = 0$$

- Recursão:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T \text{ e } 1 \leq j \leq N$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T \text{ e } 1 \leq j \leq N$$

- Terminação:

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)]$$

$$q_t^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)]$$

- Volta ao caminho da seqüência de estados:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1$$

4.1.2.3 Solução para o problema 3

Este é o mais difícil dos três problemas, já que não existe um meio analítico conhecido para se achar os parâmetros do modelo que maximizem a probabilidade da seqüência de observações. O que se faz é escolher o modelo $\lambda = (A, B, \pi)$, tal que $p(O|\lambda)$ é localmente maximizado utilizando o algoritmo *Expectation-Maximization* ou o método Baum-Welch,

(RABINER, 1989) que é o algoritmo EM adaptado para HMM.

- **Algoritmo Expectation-Maximization (EM).** Este algoritmo pode ser utilizado para obter a máxima verossimilhança, estimada em problemas com variáveis ocultas ou não-observáveis. Assim, ele utiliza a probabilidade $p(O, q|\lambda)$ como ferramenta para o cálculo de $p(O|\lambda)$, com o objetivo de estimar os parâmetros do modelo λ . Para isso, busca-se a maximização do logaritmo de $p(O|\lambda)$, ou seja, de

$$L(\lambda) = \log p(O|\lambda) = \log \sum_q p(O, q|\lambda).$$

Para a maximização de $L(\lambda)$, utiliza-se uma abordagem simplificada, considerando $X(q)$ como a distribuição de probabilidade dos estados sobre o conjunto de variáveis escondidas.

$$\begin{aligned} \log \sum_q p(O, q|\lambda) &= \log \sum_q X(q) \frac{p(O, q|\lambda)}{X(q)} \\ \log \sum_q p(O, q|\lambda) &\geq \sum_q X(q) \log \frac{p(O, q|\lambda)}{X(q)} \\ \log \sum_q p(O, q|\lambda) &\geq \sum_q X(q) \log p(O, q|\lambda) - \sum_q X(q) \log X(q) \end{aligned}$$

Pode-se então definir uma função $F(X, \lambda)$ como a parte direita da desigualdade acima, ou

$$F(X, \lambda) = \sum_q X(q) \log p(O, q|\lambda) - \sum_q X(q) \log X(q). \quad (4.1)$$

É possível, então, fazer uma relação entre a função $F(X, \lambda)$ e a função $L(\lambda)$, como

$$\log \sum_q p(O, q|\lambda) \geq F(X, \lambda). \quad (4.2)$$

A equação 4.1 pode ser escrita também de outra forma, a saber,

$$F(X, \lambda) = E[\log p(O, q|\lambda)] + H(X(q)), \quad (4.3)$$

onde $H(X(q))$ é a entropia da distribuição $X(q)$.

O algoritmo EM busca a maximização de $F(X, \lambda)$, que é feita em duas etapas, a Esperança (E) e a Maximização (M). Na primeira etapa, é feita a reestimação da distribuição $X(q)$, mantendo-se fixo os parâmetros do modelo, e busca-se a maximização de $F(X, \lambda)$, ou seja,

$$X_{k+1} \leftarrow \arg \max_X F(X, \lambda^k), \quad (4.4)$$

onde λ^k representa os valores atuais dos parâmetros do modelo na iteração k . Assim, a maximização de $F(X, \lambda)$ na equação 4.1 acontece quando

$$X_{k+1}(q) = p(q|O, \lambda^k). \quad (4.5)$$

e então a desigualdade 4.2 se torna a igualdade $F(X_{k+1}, \lambda^k) = L(\lambda^k)$

Na etapa de Maximização, a função $F(X, \lambda)$ é maximizada modificando os parâmetros do modelo e mantendo fixa a distribuição $X(q)$, ou seja,

$$\lambda^{k+1} \leftarrow \arg \max_{\lambda} F(X_{k+1}, \lambda). \quad (4.6)$$

Isto é feito maximizando a primeira parte da equação 4.1, já que a segunda parte não depende de λ , ou seja,

$$\lambda^{k+1} = \arg \max_{\lambda} \sum_q p(q|O, \lambda^k) \log p(q, O|\lambda). \quad (4.7)$$

Assim, define-se uma função auxiliar Q como sendo o argumento de maximização

$$Q(\lambda^k, \lambda) = \sum_q p(q|O, \lambda^k) \log p(q, O|\lambda), \quad (4.8)$$

ou seja,

$$Q(\lambda^k, \lambda) = E [\log p(q, O|\lambda)]. \quad (4.9)$$

Vê-se, então, que Q é a esperança matemática do logaritmo da densidade conjunta das variáveis escondidas q e das observáveis O . Assim, pode-se afirmar que maximizar a função Q é equivalente a maximizar $L(\lambda)$, já que quando uma mudança dos parâmetros do modelo fazem aumentar $Q(\lambda^k, \lambda)$ isso também aumentará $L(\lambda)$.

Pode-se, então, maximizar a esperança matemática Q para cada valor dos parâmetros do modelo e, em seguida, reestimá-los o que leva a maximização de $p(O|\lambda)$ para cada iteração k , a não ser que já se esteja em um máximo local de Q . Para o caso de HMM, como já foi dito, o algoritmo Baum-Welch (RABINER, 1989), detalhado a seguir, foi especialmente desenvolvido para o método de aprendizagem baseado no algoritmo EM.

- **Algoritmo Baum-Welch (BW)**. O treinamento de um HMM é um caso particular do algoritmo EM. Na etapa E, a distribuição a posteriori dos estados do HMM, que é uma variável escondida, é estimada pela equação 4.5. Esta estimacão faz uso dos algoritmos *forward* e *backward*, vistos em (RABINER, 1989), que diminuem os recursos computacionais.

Primeiro, a distribuição a posteriori pode ser escrita em função da probabilidade de junção das variáveis escondidas e observáveis, a saber

$$p(q|O, \lambda^k) = \frac{p(q, O|\lambda^k)}{\sum_{q'} p(q', O|\lambda^k)}.$$

O numerador da equação acima também pode ser escrito em função dos parâmetros do modelo, ou seja

$$\begin{aligned} p(q, O|\lambda^k) &= \prod_{t=1}^T p(q_t|o_t, \lambda^k) p(q_0|\lambda^k) p(q_t|q_{t-1}, \lambda^k) \\ p(q, O|\lambda^k) &= \pi_{q_0} \prod_{t=1}^T b_{q_t}(o_t) a_{q_{t-1}q_t} \\ \log p(q, O|\lambda^k) &= \log \pi_{q_0} + \sum_{t=1}^T \log b_{q_t}(o_t) + \sum_{t=1}^T \log a_{q_{t-1}q_t}. \end{aligned}$$

Assim, depois de calcular $p(q, O|\lambda^k)$, a etapa M é feita da seguinte forma:

- Buscam-se os valores dos parâmetros do modelo que anulam a derivada da função $Q(\lambda^k, \lambda)$ em relação a cada parâmetro do HMM, utilizando multiplicadores de Lagrange;
- Respeitam-se as restrições que garantem a validade das probabilidades de emissão $b_j(k)$, a de transição de estados a_{ij} e a probabilidade inicial π .

Como resultado do algoritmo, tem-se as equações de reestimação dos parâmetros do modelo λ :

- Probabilidade inicial:

$$\pi_j^{k+1} = \frac{p(q_0 = j, O|\lambda^k)}{p(O|\lambda^k)}$$

- Probabilidade de transição de estados:

$$a_{ij}^{k+1} = \frac{\sum_{t=1}^T p(q_{t-1} = i, q_t = j, O|\lambda^k)}{\sum_{t=1}^T p(q_{t-1} = i, O|\lambda^k)}$$

- Probabilidade de emissão (caso contínuo):

$$\gamma_t(j, m) = \left[\frac{p(q_t = j, O|\lambda^k)}{p(O|\lambda^k)} \right] \left[\frac{c_{jm} N(o_t, \mu_{jm}, U_{jm})}{\sum_{m=1}^M c_{jm} N(o_t, \mu_{jm}, U_{jm})} \right]$$

$$c_{jm}^{k+1} = \frac{\sum_{t=1}^T \gamma_t(j, m)}{\sum_{t=1}^T \sum_{m=1}^M \gamma_t(j, m)}$$

$$\mu_{jm}^{k+1} = \frac{\sum_{t=1}^T \gamma_t(j, m) \cdot o_t}{\sum_{t=1}^T \gamma_t(j, m)}$$

$$U_{jm}^{k+1} = \frac{\sum_{t=1}^T \gamma_t(j, m) \cdot (o_t - \mu_{jm})(o_t - \mu_{jm})^{Tr}}{\sum_{t=1}^T \gamma_t(j, m)}$$

onde $(X)^{Tr}$ representa a matriz transposta de X , c_{jm} representa o coeficiente da mistura de gaussianas, e μ_{jm} e U_{jm} são o vetor média e a covariância da distribuição gaussiana, respectivamente.

4.2 Adaptação

O processo de adaptação do sistema desenvolvido em (ANDREAO, 2004) é fundamentado na reaprendizagem dos modelos de Markov. Devido à importância da etapa de adaptação, serão estudados os principais algoritmos, suas características e possibilidades de implementação.

Existem duas formas de adaptação, listadas a seguir (LEE; HUO, 2000) e mostradas na Figura 30:

- indireta, onde os modelos são adaptados utilizando uma matriz de transformação;
- direta, onde os modelos são adaptados de acordo com um método de adaptação.

É importante ressaltar que os algoritmos estudados aqui adotam a técnica direta, pois assim cada modelo pode ser adaptado separadamente quando desejado.

4.2.1 Algoritmos em Lote (Batch Algorithms)

Uma das formas de se adaptar HMMs é utilizar algoritmo em lote. Seu conceito, de acordo com Digalakis (1999), consiste em que todos os dados de adaptação precisam ser armazenados e os parâmetros do modelo necessitam de múltiplas iterações para serem estimados, ou seja, passa-se várias vezes pelos dados armazenados, chamadas de épocas.

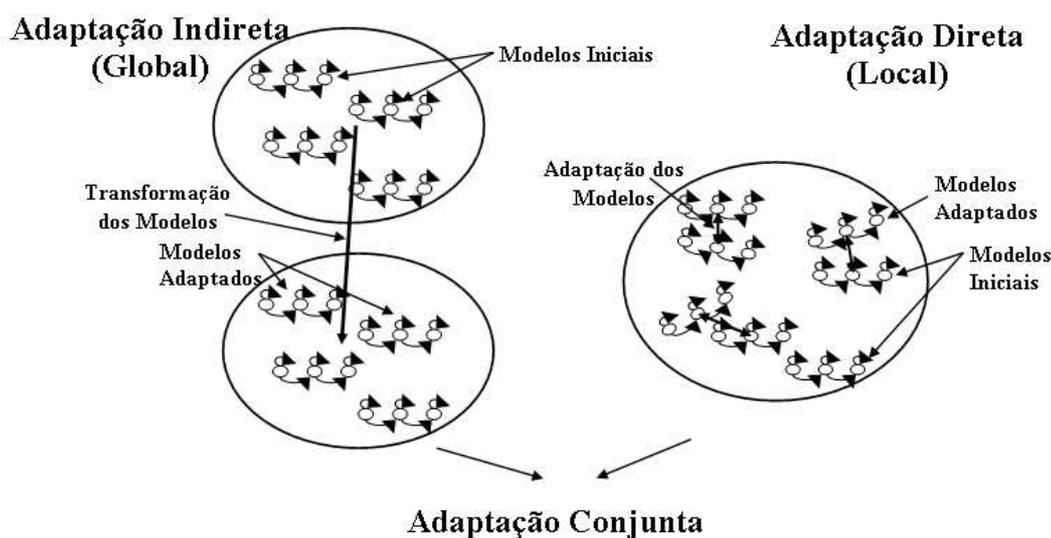


Figura 30: Adaptação direta e indireta de HMM (LEE; HUO, 2000).

Devido a essa característica, esses algoritmos apresentam a limitação de que o usuário deve ter uma grande quantidade de dados armazenados para servir de base para o processo de adaptação. Então, os parâmetros do sistema serão adaptados *off-line*, ou em lote, sem que o usuário possa ver os benefícios da adaptação aplicados a estes dados.

A maioria das aplicações práticas necessitam de algoritmos de adaptação *on-line* e incremental, ou seja, os parâmetros devem ser atualizados depois de cada observação ou pequenos subconjuntos dentro do conjunto total de dados. Isso faz com que o desempenho da aprendizagem melhore progressivamente, à medida que o sistema for sendo usado. Assim, os subconjuntos se beneficiam da adaptação dos modelos adaptados durante a passagem pelos subconjuntos anteriores. Então é possível sentir o efeito da adaptação dentro do conjunto grande de dados.

Como em (HUO; LEE, 1997), pode-se fazer uma relação entre os algoritmos em lote, ou *off-line*, e os algoritmos *on-line*, ou seqüenciais, considerando que a vantagem dos algoritmos seqüenciais em relação aos algoritmos *off-line* não está necessariamente no resultado final, mas sim na eficiência computacional, na condição de armazenagem reduzida e no fato de que o resultado pode ser obtido sem ter que esperar que todos os dados sejam processados.

Ressalte-se que, para qualquer tipo de algoritmo de adaptação, deseja-se que ele seja não-supervisionado, para que possa se adaptar automaticamente às novas observações.

4.2.2 Algoritmos *On-line*

Pode-se definir um algoritmo *on-line* como aquele que é desenvolvido para executar uma adaptação contínua, usando um sistema de memória limitado. Para que essa definição seja colocada em prática, a implementação de um algoritmo *on-line* deve fazer com que ele se torne (HUO; MA, 2001):

- incremental: os parâmetros do modelo devem se adaptar continuamente aos novos dados de adaptação, sem precisar de um armazenamento de um grande número prévio de dados de treinamento ou adaptação. Isso aumenta a eficiência computacional e diminui as condições de armazenamento;
- adaptativo: como as variações dos parâmetros do modelo correspondem às variações dos dados de observação, o sistema de adaptação deve ter alguns mecanismos de “esquecimento” para reduzir o efeito das antigas observações em relação aos novos dados de entrada;
- eficiente: o algoritmo de adaptação deve manter ou melhorar o desempenho com uma pequena quantidade de dados de adaptação, e se aproximar assintoticamente do desempenho máximo da adaptação *off-line*, à medida que o número de dados de adaptação for aumentando.

4.2.3 Treinamento Incremental

De acordo com as características dos algoritmos *on-line* já citadas, pode-se ver que não é possível utilizar o tradicional algoritmo EM para reestimar os parâmetros do sistema. Assim, foi criada uma versão incremental do método em um sistema de reconhecimento de voz desenvolvido em (GOTOH; SILVERMAN, 1996), com a máxima verossimilhança (Maximum Likelihood - ML) como critério de otimização. Esse algoritmo (GOTOH; SILVERMAN, 1996) foi implementado para um processo de treinamento de HMMs sem perda de desempenho. Treinamentos convencionais de HMMs são feitos usando algoritmos EM que utilizam como critério de convergência a máxima verossimilhança (ML), que é o caso do sistema atual. De acordo com Neal e Hinton (1993) e Gauvain e Lee (1994), usando uma variante incremental, melhoras substanciais na velocidade de treinamento podem ser obtidas com uma estimação máximo a posteriori (MAP). Esse algoritmo seleciona sequencialmente os dados do subconjunto de treinamento, atualiza os parâmetros do subconjunto e então faz as iterações até a convergência. E assim, no trabalho desenvolvido

por Gotoh e Silverman (1996) não é necessária estimar uma distribuição a priori, como no treinamento bayesiano que será visto adiante.

A justificativa teórica para uma variante incremental do algoritmo EM é dada em (NEAL; HINTON, 1993), onde se estabelece que se as estatísticas, ou seja, a distribuição $X(q)$, para o passo de estimação E, são incrementalmente coletadas, os parâmetros do modelo são freqüentemente estimados e a convergência acelera, já que a informação de dados novos contribui para a estimação dos novos parâmetros mais rapidamente que no algoritmo padrão.

A teoria usada em (GOTOH; SILVERMAN, 1996) é aplicada a dados que pertencem à *Família Exponencial*, o que se aplica também ao sinal de ECG, uma vez que ele é modelado por gaussianas, que nada mais são do que curvas exponenciais. É também necessário que os dados sejam independentes, ou seja, $p(O_1, O_2, \dots, O_T) = \prod_{i=1}^T p(O_i)$.

O algoritmo proposto por Gotoh e Silverman (1996) pode ser descrito pelas seguintes etapas:

- Inicialização
 - Inicializar os parâmetros λ^0 .
 - Separar o conjunto de observações em N subconjuntos $O = O^n$, $n = 1, \dots, N$.
 - Definir distribuições iniciais $X_0^n(q)$ adequadas para cada subconjunto O^n .
 - Inicializar o contador da iteração k em zero.
- Passo de estimação (E)
 - Escolher um subconjunto O^n
 - Dado λ^k , calcular os termos forward/backward e calcular $X_{k+1}^n(q)$ como $X_{k+1}^n(q) = E[t(O^n, q^n | O^n, \lambda^k)]$.
 - Atualizar $X_{k+1}^l(q) \leftarrow X_k^l(q)$ para os blocos $l \neq n$. Isso não requer praticamente cálculo algum.
 - Estabelecer a distribuição para todo o conjunto de dados como: $X_{k+1}(q) = X_k(q) - X_k^n(q) + X_{k+1}^n(q)$.
- Passo de maximização (M)
 - Ajustar os parâmetros do modelo $\lambda^{(k+1)}$ para ser a solução da igualdade $E[t(O, q) | \lambda] = X_{k+1}$.

- Se não houver convergência, fazer $k = k + 1$ e repetir a partir do passo de estimação.

Observe-se que os subconjuntos O^n podem ser selecionados sequencialmente ou aleatoriamente, e que, para cada iteração, o subconjunto escolhido pode ser diferente da iteração anterior. Um problema que pode aparecer nessa abordagem incremental é o de memória para armazenar as estatísticas de cada subconjunto $X^n(q)$ separadamente. Isso pode ocorrer quando o número de subconjuntos é muito grande.

Isso rendeu ao trabalho desenvolvido por (GOTOH; SILVERMAN, 1996) uma redução de tempo em horas de CPU de 9,1 para 2,7, sem prejudicar o desempenho de reconhecimento. Note-se que em (GOTOH; SILVERMAN, 1996) trabalha-se com reconhecimento da fala. Foi observado, neste trabalho, que há uma melhora significativa no tempo de processamento quando há um número grande de subconjuntos, pois assim os parâmetros dos modelos serão freqüentemente estimados. Ressalta-se que as melhorias foram obtidas sem o uso da estimação MAP, que requer uma estimação a priori, o que é muito difícil de se conseguir na maioria dos casos. Devido a algumas dificuldades de se guardar informações (distribuições) para grandes HMMs, é então introduzida a idéia de um "fator de esquecimento" que consideraria com uma importância menor as informações de subconjuntos mais antigos.

4.3 Aprendizagem EM

Note-se que, para a adaptação *off-line*, os dados precisam ser primeiramente armazenados para depois se executar os passos EM (Expectation-Maximization) várias vezes até a convergência, ou seja, a cada iteração do algoritmo é necessário fazer uma passagem pelos dados, resultando ao final em várias passagens pelos dados ou observações. Assim, os parâmetros do sistema devem ser adaptados *off-line* antes que o usuário possa usufruir dos benefícios dessa adaptação.

Para a adaptação *on-line*, é necessário, então, que seja feita uma única passagem pelos dados, para que os parâmetros possam ser atualizados incrementalmente. Digalakis (1999) propõe que, para facilitar a integração de uma adaptação incremental com um sistema de reconhecimento de fala, é preferível que o treinamento dos modelos utilize um procedimento segmental k-means em vez do treinamento Baum-Welch. Para este trabalho, continuar-se-á com o treinamento BW já utilizado em (ANDREAO, 2004) e a adaptação incremental também será feita usando BW. Para se chegar ao algoritmo de adaptação incremental implementado na próxima seção, será feita uma recapitulação da

teoria do algoritmo EM.

O algoritmo EM pode ser visto como aquele que maximiza a função conjunta

$$F(X, \lambda) = E[\log p(O, q|\lambda)] + H(X(q)).$$

Considerando as observações como $O = O^n$, $n = 1, \dots, N$, onde O^n pode ser uma observação ou um grupo de observações, q^n como as variáveis ocultas correspondentes e que as observações são independentes, então a função $F(X, \lambda)$ pode ser escrita como a soma das funções custo de cada bloco:

$$F(X, \lambda) = \sum_{i=1}^N F^i(X^i, \lambda)$$

$$F^i(X^i, \lambda) = E[\log p(O^i, q^i|\lambda)] + H(X^i(q)), \quad i = 1, \dots, N. \quad (4.10)$$

Tanto a versão em lote quanto a versão incremental do algoritmo EM (no qual o BW é baseado), conforme já visto, é produzida por uma alternância entre a maximização da função $F(X, \lambda)$ sobre a distribuição $X(q)$ (passo E) e sobre os parâmetros λ (passo M). Assim, a $(k + 1)$ -ésima iteração do algoritmo EM em lote simplesmente maximiza, no passo E, a distribuição conjunta $X(q)$ sobre todas as variáveis não-observáveis q , como a seguir:

Algoritmo EM em lote

- Passo E

$$X_{k+1}(q) = \prod_{i=1}^N X_{k+1}^i(q^i), \text{ onde:}$$

$$X_{k+1}^i(q^i) = p(q^i|O^i, \lambda^k)$$

para maximizar as funções

$$F_i(X_i, \lambda^k).$$

e assim obter a função $F(X, \lambda^k)$.

- Passo M

Encontrar os parâmetros de λ^{k+1} que maximizem $F(X_{k+1}, \lambda)$.

Para a elaboração do algoritmo EM incremental, o passo E irá maximizar a distribuição conjunta $X(q)$ definida sobre somente um subconjunto das variáveis ocultas. Por exemplo, para a $(k + 1)$ -ésima iteração, onde está sendo processada somente a n -ésima observação, a versão EM incremental será:

Algoritmo EM incremental

- Passo E

Selecionar O^n e fazer:

$$X_{k+1}(q) = \prod_{i=1}^N X_{k+1}^i(q^i), \text{ onde:}$$

$$\begin{cases} X_{k+1}^n(q^n) = p(q^n | O^n, \lambda^k) & \text{para } i = n \\ X_{k+1}^i(q^i) = X_k^i(q^i) & \text{para } i \neq n \end{cases}$$

para maximizar somente a função:

$$F_n(X_n, \lambda^k)$$

mantendo $F_i(X_i, \lambda^k)$ com o valor prévio para as outras observações.

e assim obter a função $F(X, \lambda^k)$.

- Passo M

Encontrar os parâmetros de λ^{k+1} que maximizem $F(X_{k+1}, \lambda)$.

Note que o algoritmo EM incremental efetua o passo E para cada O^n , que, como já dito, pode ser uma observação ou um grupo de observações, e vários passos M são feitos até a convergência. Ressalta-se que, assim como no trabalho de Gotoh e Silverman (1996), a cada iteração é escolhido no passo E um novo subconjunto O^n , que poderá ser repetido se a seleção de O^n for aleatória.

Quando a distribuição das variáveis observáveis e ocultas é um membro da família exponencial, como no caso da gaussiana, o passo E do algoritmo EM se reduz ao cálculo das estatísticas, e o passo M ao cálculo da máxima verossimilhança e estimação dos parâmetros do modelo, dadas as estatísticas. Observe-se que no passo E já não se faz mais necessária a maximização da função $F(X, \lambda^k)$.

Na versão incremental, para família exponencial, as estatísticas são mantidas incrementalmente. Na $(k+1)$ -ésima iteração, as estatísticas da n -ésima observação O^n processada pelo algoritmo são calculadas usando as k estimações do modelo. Essas estatísticas, \bar{t}_{k+1}^n , substituem o seu valor prévio, \bar{t}_k^n , o qual foi calculado usando os parâmetros do modelo estimado quando a n -ésima observação foi previamente processada. Esse processamento prévio pode ter ocorrido durante o cálculo das estatísticas iniciais e a escolha de O^n pode ser aleatória. Assim, as estatísticas acumuladas no algoritmo EM incremental para família exponencial são armazenadas usando uma atualização contínua dos parâmetros estimados, o que se opõe à versão EM em lote, onde depois de cada passo M as estatísticas de todo o conjunto de treinamento são rearmazenadas usando os últimos parâmetros estimados. Usando $t(O^n, q^n)$ para denotar as estatísticas da n -ésima observação e suas variáveis ocultas associadas, então a k -ésima iteração do algoritmo EM incremental, para membros da família exponencial, pode ser descrita como:

Algoritmo EM incremental para família exponencial

• Passo E

Selecionar a n -ésima observação e calcular a esperança para as estatísticas

$$\bar{t}_{k+1}^n = E[t(O^n, q^n) | O^n, \lambda^k] \quad \text{para } i = n, \text{ onde:}$$

$$X_{k+1}^n(q^n) = p(q^n | O^n, \lambda^k)$$

$$\bar{t}_{k+1}^i = \bar{t}_k^i \quad \text{para } i \neq n$$

$$\bar{t}_{k+1} = \sum_{i=1}^N \bar{t}_{k+1}^i = \bar{t}_k + \bar{t}_{k+1}^n - \bar{t}_k^n$$

• Passo M

Encontrar os parâmetros de λ^{k+1} que maximizem a verossimilhança das observações dado \bar{t}_{k+1} , ou seja, encontrar a solução de $E[t(O, q) | \lambda^{k+1}] = \bar{t}_{k+1}$.

Note-se que, por simplicidade e para associar o formalismo acima com o usado por Neal e Hinton (1993) e Digalakis (1999), $E_{X_{k+1}^n}[\cdot]$ é escrito apenas como $E[\cdot]$ que denota a esperança da distribuição sobre as variáveis ocultas, q^n , dado por X_{k+1}^n .

Observe-se que no algoritmo EM incremental, seja para família exponencial ou não, se se considerar que as observações são processadas incrementalmente, tem-se ao final de todas as observações, uma única passagem sobre os dados durante o passo E, e que, no entanto, o passo M é executado várias vezes durante a passagem total.

Pode se ver, também, que é necessário manter uma cópia separada do valor atual das estatísticas \bar{t}_{k+1}^i para cada bloco de observações. Isso torna o algoritmo computacionalmente custoso, a menos que o conjunto de dados seja dividido em poucos blocos. Pode-se notar a semelhança entre o algoritmo EM incremental para família exponencial apresentado por Digalakis (1999), usado para adaptação, e o apresentado por Gotoh e Silverman (1996) usado para estimação, assim como o problema de armazenagem das estatísticas de cada observação ou conjunto de observações O^n , o que pode tornar o algoritmo até mesmo impraticável.

Os conceitos apresentados pelas abordagens explicadas foram utilizados para a concepção do algoritmo EM incremental implementado, que será abordado a seguir.

4.3.1 Adaptação EM Incremental Implementada

No conceito de adaptação *on-line*, onde se faz uma única passagem pelos dados, a necessidade de cópias separadas das estatísticas pode ser eliminada se os valores iniciais para as estatísticas de cada bloco \bar{t}_0^n forem ajustados para zero, e se as observações forem

selecionadas seqüencialmente. De acordo com Neal e Hinton (1993), a iteração do EM incremental, que a partir daqui é denotado assim mesmo seja para a família exponencial ou não, é implementada usando estatísticas que são mantidas incrementalmente, começando com uma estimativa inicial, a qual pode ser ou não consistente com a estimativa inicial dos parâmetros e, portanto, podem ser igualadas a zero, ou seja, $t_0^i = 0$, $i = 1, \dots, N$.

A escolha de valores iniciais iguais a zero para as estatísticas parciais permite acumular as estatísticas que forem surgindo no processo de adaptação. Como será feita uma única passagem pelos dados, não há necessidade de se armazenar as estatísticas parciais de cada bloco, mas somente a global, ou seja, $\bar{t}_{k+1} = \bar{t}_k + \bar{t}_{k+1}^n$.

Para implementar o algoritmo EM incremental na plataforma desenvolvida por Andreao (2004) tentou-se primeiramente dividir cada bloco de 5000 pontos em partes menores ainda, ou seja, a cada observação. Isso significa que os passos EM seriam calculados para cada amostra, o que resultou em uma situação instável, com matrizes de covariância singulares. Essa instabilidade é facilmente explicável, pois não há dados suficientes para se estimar a média e a covariância das gaussianas.

Após uma releitura do trabalho de Digalakis (1999), chegou-se à conclusão que a única modificação a ser feita era a mudança do número de iterações do algoritmo BW do valor 10, anteriormente usado, para 1, como ilustrado na Figura 31. Então, uma vez que os valores iniciais das estatísticas já são iguais a zero, com o número de iterações igual a um tem-se uma única passagem pelos dados. A única coisa que falta é atualizar as estatísticas incrementalmente, o que será feito em uma próxima etapa. É importante observar que essa modificação tem por objetivo implementar o algoritmo EM incremental para família exponencial, o que é justificável, uma vez que a gaussiana pertence a essa família.

Ressalta-se mais uma vez, como já dito no Capítulo 3 e observado na Figura 31, que cada bloco é inicializado com os modelos do bloco anterior, mas o valor inicial da distribuição $X(q)$, que armazena as informações dos modelos, é sempre zero.

A Tabela 2 mostra os valores das constantes utilizadas. Uma vez implementada a abordagem EM incremental, buscou-se outros métodos, que também utilizem a estimação ML, para serem implementados de forma incremental. A abordagem escolhida, a partir da literatura, foi a segmental k-means e o próximo passo foi, então, o estudo e a implementação da sua versão incremental.

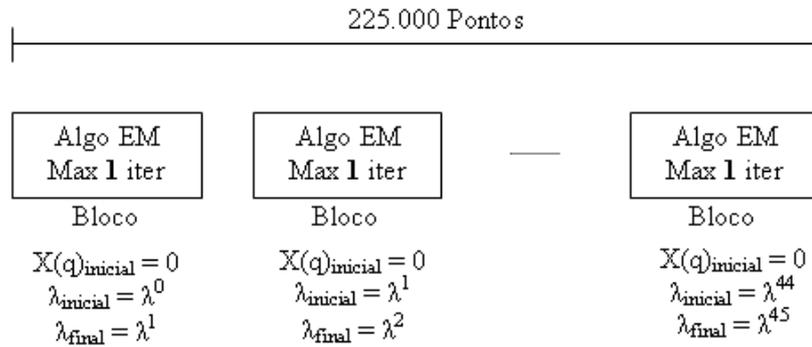


Figura 31: Sistema para o algoritmo EM incremental implementado.

Tabela 2: Valores das constantes para o algoritmo implementado.

Constantes	
N	45
k_{max}	1

4.4 Aprendizagem Segmental k-Means

O treinamento segmental k-Means é baseado na seqüência de estados mais provável, que é justamente a saída do algoritmo de Viterbi, facilitando a implementação do algoritmo. Para implementar a adaptação *online* usando o algoritmo de treinamento segmental k-means, utiliza-se sua versão incremental, assim como foi feito com o algoritmo EM. Deve-se, porém, estudar a versão em lote desse algoritmo de adaptação.

- **Segmental K-means em Lote**

O algoritmo de treinamento segmental k-means maximiza a verossimilhança conjunta das observações e das seqüências de estado escondidas associadas às observações, $p(O, q|\lambda)$, em vez do critério de máxima verossimilhança (ML) $p(O|\lambda)$ usado no treinamento EM. O algoritmo segmental k-means desenvolve esta otimização iterativamente, alternando entre duas etapas: a primeira, que visa encontrar a seqüência de estados mais provável, dados os parâmetros atuais do modelo, e a segunda, que visa obter uma nova estimacão para os parâmetros do modelo otimizando a verossimilhança conjunta das observações e da seqüência de estados mais provável encontrada na etapa anterior.

A relação entre o treinamento segmental k-means e o EM pode ser vista escrevendo-se a função auxiliar do algoritmo EM. Esta última equação é uma fusão das equações

4.8 e 4.9, e é maximizada a cada iteração para gerar novas estimações dos parâmetros do modelo:

$$Q(\lambda^k, \lambda^{k+1}) = E \left[\log p(q, O | \lambda^{k+1}) \right] = \sum_q p(q|O, \lambda^k) \log p(q, O | \lambda^{k+1}). \quad (4.11)$$

Em contrapartida, a função auxiliar do algoritmo segmental k-means é

$$\log p(\hat{q}, O | \lambda^{k+1}) = \sum_q \delta(q - \hat{q}) \log p(q, O | \lambda^{k+1}), \quad (4.12)$$

onde

$$\delta(q - \hat{q}) = \begin{cases} 1, & \text{se } q = \hat{q} \\ 0, & \text{caso contrário} \end{cases}$$

Portanto, a seqüência de estados oculta \hat{q} é justamente a seqüência de estados mais provável, ou seja, a equação 4.12 é a probabilidade dessa seqüência. Então, \hat{q} pode ser estimado como

$$\hat{q} = q(\lambda^k) = \arg \max_q p(q, O | \lambda^k).$$

Como já visto, a distribuição $X(q)$ no algoritmo segmental k-means é

$$X(q) = \delta(q - \hat{q}),$$

e a equação 4.3, que descreve a função $F(X, \lambda)$ para o algoritmo EM, ou seja,

$$F(X, \lambda) = E[\log p(O, q | \lambda)] + H(X(q)),$$

é reescrita, para o algoritmo segmental k-means, como:

$$F(X, \lambda) = \sum_q \delta(q - \hat{q}) \log p(O, q | \lambda) - \sum_q \delta(q - \hat{q}) \log \delta(q - \hat{q}). \quad (4.13)$$

Nesse caso, $H(X(q))$ é zero e, considerando que a esperança da função $\log p(O, q | \lambda)$ é calculada usando a distribuição $X(q) = \delta(q - \hat{q})$, a função $F(X, \lambda)$ se torna

$$F(X, \lambda) \Big|_{X(q)=\delta(q-\hat{q})} = \log p(O, \hat{q} | \lambda). \quad (4.14)$$

Então, como já foi dito, o algoritmo segmental k-means maximiza $F(X, \lambda)$, alternando entre as maximizações sobre as diferentes seqüências de estados ocultas \hat{q} (etapa de segmentação) e sobre os parâmetros do modelo λ .

Pode-se resumir o algoritmo segmental k-means em lote como a seguir:

Algoritmo Segmental K-Means em lote

– Segmentação

Fazer:

$$X_{k+1}(q) = \prod_{i=1}^N X_{k+1}^i(q^i), \text{ onde:}$$

$$X_{k+1}^i(q^i) = \delta(q^i - \hat{q}_{k+1}^i)$$

onde a seqüência mais provável de estados (MLSS - Most Likely State Sequence) é calculada por

$$\hat{q}_{k+1}^i = \arg \max_{q^i} p(q^i, O^i | \lambda^k).$$

– Reestimação

Encontrar os parâmetros de λ^{k+1} que maximizam $F(X_{k+1}, \lambda)$.

• Segmental k-Means Incremental

Assim como foi feito com o algoritmo EM incremental (DIGALAKIS, 1999; NEAL; HINTON, 1993), pode-se criar uma versão incremental do algoritmo segmental k-means inserindo uma etapa de reestimação depois do cálculo da verossimilhança da seqüência mais provável \hat{q}_n^{k+1} de cada observação O_n , ou seja,

Estimação

$$X_{k+1}(q) = \delta(q - \hat{q}^{k+1}),$$

onde a seqüência de estados com maior verossimilhança é calculada somente para a observação O_n :

$$\hat{q}_j^{k+1} = \begin{cases} \arg \max_{q_j} p(q_j, O_j | \lambda^k), & \text{se } j = n \\ \hat{q}^k, & \text{se } j \neq n \end{cases}$$

Maximização

É o mesmo processo do algoritmo EM, ou seja, a equação 4.7, que é

$$\lambda^{k+1} = \arg \max_{\lambda} \sum_q p(q|O, \lambda^k) \log p(q, O|\lambda).$$

Resumindo, tem-se:

Algoritmo Segmental K-Means Incremental

– Segmentação

Selecionar O^n e fazer:

$$X_{k+1}(q) = \prod_{i=1}^N X_{k+1}^i(q^i), \text{ onde:}$$

$$X_{k+1}^i(q^i) = \delta(q^i - \hat{q}_{k+1}^i)$$

onde a seqüência mais provável de estados será:

$$\begin{cases} \hat{q}_{k+1}^n = \arg \max_{q^n} p(q^n, O^n | \lambda^k) & \text{para } i = n \\ \hat{q}_{k+1}^i = \hat{q}_k^i & \text{para } i \neq n \end{cases}$$

– Reestimação

Encontrar os parâmetros de λ^{k+1} que maximizam $F(X_{k+1}, \lambda)$.

Para a família exponencial, o passo de segmentação do algoritmo k-means se reduz a acumular as estatísticas sobre os estados da seqüência mais provável. Dessa forma, as estatísticas são mantidas incrementalmente, como é visto a seguir:

Algoritmo Segmental K-Means incremental para família exponencial

– Passo E

Selecionar a n -ésima observação e encontrar o caminho mais provável correspondente:

$$\hat{q}_{k+1}^n = \arg \max_{q^n} p(q^n, O^n | \lambda^k)$$

e atualizar as estatísticas:

$$\bar{t}_{k+1}^n = t(O^n, q_{k+1}^n) \quad \text{para } i = n$$

$$\bar{t}_{k+1}^i = \bar{t}_k^i \quad \text{para } i \neq n$$

$$\bar{t}_{k+1} = \sum_{i=1}^N \bar{t}_{k+1}^i = \bar{t}_k + \bar{t}_{k+1}^n - \bar{t}_k^n$$

– Passo M

Encontrar os parâmetros de λ^{k+1} que maximizam a verossimilhança das observações, dado \bar{t}_{k+1} .

4.4.1 Adaptação Segmental K-means Incremental Implementada

Depois de se estudar o algoritmo, foi necessário estudar a implementação do mesmo sobre a plataforma desenvolvida por Andreao (2004). Para isso, usou-se como referência o trabalho de Salicetti (1996), onde é possível encontrar as equações que foram utilizadas.

As modificações feitas no sistema apresentado na Figura 23 podem ser explicadas considerando que a função *learn_ghmm.m*, na Fase 2, foi chamada para analisar um bloco de 5000 pontos. Já se têm, dentro dessa janela de 20 s, os exemplos da forma elementar de onda a ser analisada. Esses exemplos se encontram armazenados dentro de uma estrutura Matlab chamada *data*. Considerando que foram achados N_e exemplos, têm-se também as N_e seqüências ótimas armazenadas na estrutura *path*, criada dentro de *hmm_segmentation_gui.m*. As estruturas *path* e *data* são passadas para a função que agora se chama *learn_ghmm_km.m*, que é a própria *learn_ghmm.m* modificada para a adaptação segmental k-means incremental. As modificações podem ser vistas na Figura 32.

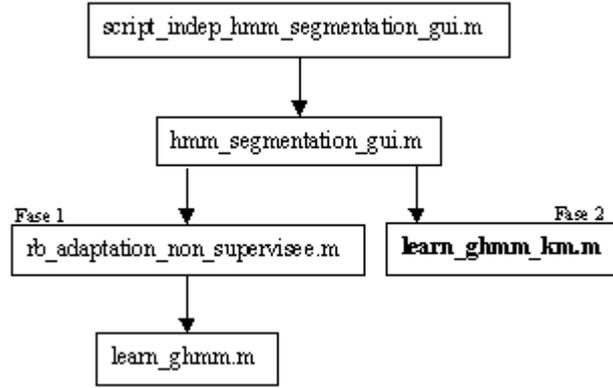


Figura 32: Arquitetura básica do algoritmo segmental K-means incremental.

A adaptação via algoritmo segmental k-means incremental requer que se calcule o valor $n_{ij}(q^*)$, que é o número de transições do estado i ao estado j , onde a variável q^* é o caminho ótimo de cada exemplo de cada forma elementar de onda. O valor $n_{ij}(q^*)$ será obtido para os N_e exemplos encontrados. É válido lembrar que N_e varia de acordo com o número de exemplos encontrados em cada bloco.

Na etapa de reestimação, os parâmetros são reestimados por:

- Probabilidade inicial:

$$\pi_j = \frac{\sum_{n_e=1}^{N_e} \delta(q_1^*(n_e) - j)}{N_e}$$

- Probabilidade de transição de estados:

$$a_{ij} = \frac{n_{ij}(q_t^*)}{\sum_{j=1}^N n_{ij}(q_t^*)}$$

- Probabilidade de emissão (caso contínuo):

$$\mu_j = \frac{\sum_{t=1}^T \delta(q_t^* - j) O_t}{\sum_{t=1}^T \delta(q_t^* - j)}$$

$$\text{cov}_j = \frac{\sum_{t=1}^T \delta(q_t^* - j) (O_t - \mu_j) (O_t - \mu_j)'}{\sum_{t=1}^T \delta(q_t^* - j)},$$

onde T é o número de amostras dentro de cada exemplo.

A adaptação incremental segmental k-means não apenas é fácil de integrar ao deco-

dificador Viterbi, como também é adequada para a adaptação não-supervisionada, pois nesse tipo de adaptação é feita uma única passagem pelos dados e informações suficientes são coletadas ao mesmo tempo. Por fim, os modelos do reconhecedor são reestimados ao final da passagem de todas as sentenças de adaptação. O desempenho de reconhecimento do sistema pode ser avaliado com um conjunto de teste separado.

É importante dizer que para o cálculo da média e da covariância são usados todos os N exemplos, apesar de isto não estar implícito nas equações, ou seja, são utilizados os N caminhos ótimos para se buscar uma média. Para o cálculo da verossimilhança, é calculada a probabilidade ótima, dados os parâmetros do modelo, através da função *viterbi_path.m*, a qual é chamada dentro da função *learn_ghmm_km.m*.

Assim, na Figura 33 é apresentado o sistema utilizando o algoritmo segmental k-means implementado, e na Tabela 3 são mostrados os valores das constantes utilizadas.

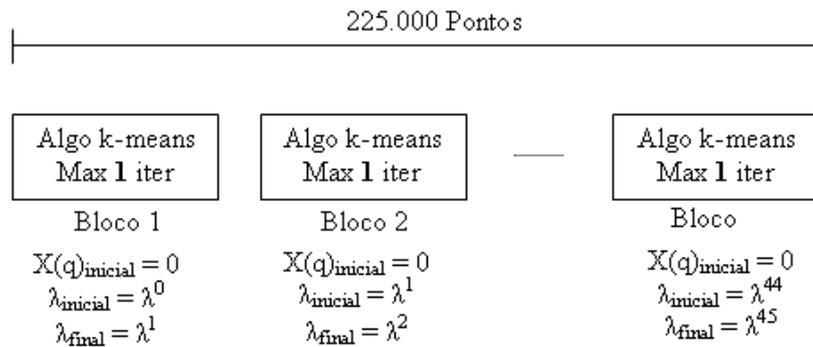


Figura 33: Sistema para o algoritmo segmental k-Means incremental implementado.

Tabela 3: Valores das constantes para o algoritmo implementado.

Constantes	
N	45
k_{max}	1
N_e	variante

4.5 Aprendizagem Bayesiana

Uma outra abordagem para a adaptação *on-line* é utilizando o treinamento ou aprendizagem Bayesiana. Essa abordagem deve ser capaz de atualizar os parâmetros da distribuição *a priori* e *a posteriori*, e também de atualizar os parâmetros do modelo simultaneamente, utilizando o último dado de adaptação. Na abordagem Bayesiana, usa-se

o algoritmo MAP (Maximum a Posteriori) como método de estimação (GAUVAIN; LEE, 1994). Tem-se também um suporte teórico de aprendizagem, proposto em (LEE; LIN; JUANG, 1991) para estimar os parâmetros vetor média e a matriz de covariância de HMM de densidade contínua (CDHMM) com múltiplas gaussianas.

Assim, estudou-se e efetuou-se a implementação do algoritmo MAP. É importante ressaltar que uma vez que esse algoritmo é baseado na teoria da inferência Bayesiana, não era evidente a possibilidade do seu uso na arquitetura já desenvolvida por Andreato (2004), a qual foi estruturada para uma abordagem de máxima verossimilhança.

Pode-se ver em (RABINER; JUANG, 1993) que, para a abordagem Bayesiana, se os parâmetros do modelo são considerados fixos mas desconhecidos, a máxima verossimilhança (ML) estimada para λ , dada a seqüência de treinamento O , é encontrada resolvendo a equação de máxima verossimilhança, ou seja,

$$\frac{\partial}{\partial \lambda} p(O|\lambda) = 0$$

Assumindo que λ é uma variável aleatória, com uma distribuição *a priori* $p_0(\lambda)$, então o máximo *a posteriori* (MAP) estimado para o modelo é obtido resolvendo

$$\frac{\partial}{\partial \lambda} p(\lambda|O) = 0,$$

dada a seqüência de treinamento O . Usando o teorema de Bayes, pode-se reescrever $p(\lambda|O)$ como

$$p(\lambda|O) = \frac{p(O|\lambda)p_0(\lambda)}{p(O)}.$$

A influência do parâmetro prévio $p_0(\lambda)$ no processo de solução torna-se explícita, e uma boa inicialização é essencial para a convergência do algoritmo.

Como já conhecido, a abordagem de inferência Bayesiana se torna um método conveniente para combinar as amostras do sinal e a informação anterior. Assumindo que os parâmetros HMM são variáveis aleatórias, essa informação anterior é expressa na forma de uma distribuição anterior que chamamos de distribuição *a priori*, a qual é combinada com a função de verossimilhança, via teorema de Bayes, para formar uma distribuição *a posteriori* na qual as inferências são baseadas. Conseqüentemente, a flexibilidade de se incorporar uma quantidade variável de informação anterior faz com que a inferência Bayesiana seja indicada para lidar com o problema de uma quantidade limitada de dados de amostragem relevantes.

A idéia desse tipo de aprendizagem Bayesiana adaptativa para HMM não é uma novi-

dade. Rabiner e Juang (1993) propõem o uso da estimação bayesiana para a adaptação do parâmetro média dos modelos, quando esta é modelada por uma gaussiana; Lee, Lin e Juang (1991) investigam vários métodos de treinamento bayesiano para adaptação ao locutor em tarefas de reconhecimento de palavras isoladas onde os parâmetros da densidade de observação de estados de múltiplas gaussianas com matriz de covariância diagonal foram adaptados, mas não adaptava todos os parâmetros do modelo. Posteriormente, Gauvain e Lee (1994) conseguiram estender a adaptação bayesiana para lidar com parâmetros de densidades de múltiplas gaussianas com matriz de covariância diagonal. Atualmente, conforme apresentado por Gauvain e Lee (1992), a aprendizagem MAP foi estendida para todos os parâmetros de um HMM com densidade de observação de estados de múltiplas gaussianas.

O algoritmo proposto em (HUO; LEE, 1997) é adaptativo em sua essência e faz com que ele possa ser usado para desempenhar o papel integral de aprendizagem *on-line* adaptativa dos parâmetros do modelo utilizando somente os dados atuais para acompanhar continuamente as variações das condições. A seguir, será apresentada a teoria do algoritmo MAP desenvolvido por Huo e Lee (1997), a qual permite encontrar a solução MAP clássica de λ usando o algoritmo EM, o que já foi comprovado por Dempster, Laird e Rubin (1977).

4.5.1 Aprendizagem MAP quasi-Bayes

Considere um HMM de densidade contínua (CDHMM) onde a *pdf* (do inglês probability density function - função densidade de probabilidade) de observação de estados é considerada como uma mistura de *pdfs* de múltiplas gaussianas, ou seja,

$$b_j(o_t) = \sum_{m=1}^M c_{jm} N(o_t, \mu_{jm}, U_{jm}),$$

onde

$$b_j(o_t) = N(o_t, \mu_j, U_j) = \frac{1}{(2\pi)^{1/d} |U_j|^{1/2}} \exp\left(-\frac{1}{2} (o_t - \mu_j)^T U_j^{-1} (o_t - \mu_j)\right),$$

e

$$\sum_{m=1}^M c_{jm} = 1, 1 \leq j \leq N,$$

onde c_{jm} é o coeficiente de mistura das gaussianas.

A diferença entre as estimações ML e MAP se encontra na suposição de uma distri-

buição anterior apropriada dos parâmetros a ser estimada. Assim, assumindo o modelo λ como um vetor de parâmetros a ser estimado a partir das amostras \mathbf{O} com *pdf* $p(O|\lambda)$, aqui aproximada por $F(X, \lambda)$, e se $p(\lambda)$ é a *pdf* anterior de λ , aqui aproximada por $g(\lambda)$, então a estimação MAP dos parâmetros do modelo, segundo a inferência bayesiana, é definida como a moda da *pdf* posterior de λ , denotada como $p(\lambda|O)$, ou seja,

$$\lambda^{k+1} \leftarrow \arg \max_{\lambda} p(\lambda|O), \quad (4.15)$$

$$\lambda^{k+1} \leftarrow \arg \max_{\lambda} p(O|\lambda)p(\lambda), \quad (4.16)$$

$$\lambda^{k+1} \leftarrow \arg \max_{\lambda} F(X, \lambda)g(\lambda). \quad (4.17)$$

Se λ é considerado como fixo mas desconhecido, então não existe conhecimento sobre λ , o que é equivalente a uma *pdf* anterior não-informativa, ou seja, $g(\lambda) = \text{constante}$ e então a equação 4.16 se reduz à formulação ML já apresentada.

Assim, dentro de uma análise teórica, considere o conjunto $\check{O}^n = \{O^1, O^2, \dots, O^n\}$ como o de n amostras de observações independentes que são usadas para estimar os parâmetros de λ . Sabe-se que o conhecimento inicial sobre λ está relacionado à densidade a priori $p(\lambda)$ e que a inferência bayesiana formal de λ é baseada na densidade a posteriori a seguir, ou seja,

$$p(\lambda|\check{O}^n) = \frac{p(\check{O}^n|\lambda)p(\lambda)}{\int_{\Omega} p(\check{O}^n|\lambda)p(\lambda)d\lambda}, \quad (4.18)$$

onde Ω denota uma região admissível do espaço de parâmetros. A equação 4.18 nada mais é que a aplicação da lei de Bayes:

$$p(\lambda|\check{O}^n) = \frac{p(\check{O}^n|\lambda)p(\lambda)}{p(\check{O}^n)}.$$

Assumindo-se que os dados ou amostras de treinamento/adaptação O^i são apresentados sucessivamente, é possível obter uma expressão recursiva para a *pdf* a posteriori de λ dado \check{O}^n , a qual é

$$p(\lambda|\check{O}^n) = \frac{p(O^n|\check{O}^{n-1}, \lambda)p(\lambda|\check{O}^{n-1})}{p(O^n|\check{O}^{n-1})} = \frac{p(O^n|\lambda)p(\lambda|\check{O}^{n-1})}{\int_{\Omega} p(O^n|\lambda)p(\lambda|\check{O}^{n-1})d\lambda}. \quad (4.19)$$

Assim, considerando a *pdf* inicial como $p(\lambda|\check{O}^0) = p(\lambda)$, o uso repetido da equação 4.19 produz uma seqüência de densidades $p(\lambda|\check{O}^1)$, $p(\lambda|\check{O}^2)$, e assim por diante. Isso serve de base para se fazer uma inferência bayesiana recursiva dos parâmetros λ , como pode

ser visto em (SPRAGINS, 1965).

Infelizmente, a implementação desta técnica de aprendizagem para o treinamento incremental de CDHMM traz algumas dificuldades computacionais. Isso faz com que algumas aproximações sejam necessárias para aliviar os custos computacionais. O procedimento proposto é aplicar a recursão Bayes da equação 4.19 incrementalmente, com uma ou mais observações ou amostras consideradas a cada instante.

O algoritmo bayesiano de aprendizagem de λ proposto implica na especificação de uma densidade a priori para λ , e no subsequente cálculo recursivo da densidade a posteriori aproximada. Para o caso de CDHMM, onde a média e a covariância são assumidas como aleatórias, a *pdf* anterior inicial de λ , $p(\lambda)$, é aproximada por $g(\lambda)$, que é assumida como um produto das densidades Dirichlet e normal-Wishart (HUO; LEE, 1997), ou seja,

$$g(\lambda) = g(\lambda') \cdot \prod_{j=1}^N \prod_{m=1}^M g(\mu_{jm}, U_{jm}),$$

onde

$$g(\lambda') \propto \prod_{j=1}^N \left\{ \pi_i^{\eta_i - 1} \cdot \left(\prod_{j=1}^N a_{ij}^{\eta_{ij} - 1} \right) \cdot \left(\prod_{m=1}^M c_{jm}^{\nu_{jm} - 1} \right) \right\}$$

e

$$g(\mu_{jm}, U_{jm}) \propto \prod_{d=1}^D \frac{1}{|U|_{jm}^{(\alpha_{jm} - 1/2)}} \times \exp \left[-\frac{\tau_{jm}}{2U_{jm}} (\mu_{jm} - \kappa_{jm}) (\mu_{jm} - \kappa_{jm})^t \right] \times \exp \left[-\frac{\beta_{jm}}{U_{jm}} \right].$$

Pode-se observar que a função $g(\lambda)$ é parametrizada pelo conjunto de hiper-parâmetros positivos: η_i , η_{ij} , ν_{jm} ; τ_{jm} , κ_{jm} , α_{jm} e β_{jm} , usados para a estimação dos parâmetros do modelo $\lambda = (\pi, A, B)$ e que D é a dimensão do vetor média. Ressalta-se que esses hiper-parâmetros são definidos para o caso de matriz de covariância diagonal, que é o caso deste trabalho.

Então, a solução para tornar o algoritmo de aprendizagem Bayes recursivo realizável é aproximar a verdadeira distribuição a posteriori $p(\lambda|\check{O}^n)$ na equação 4.19 pela distribuição tratável aproximada $g(\lambda|\varphi^n)$, correspondente ao produto $F(X, \lambda)g(\lambda)$ apresentado na equação 4.17, onde φ^n denota os hiper-parâmetros atualizados depois de uma observação \mathbf{O}^n . Assim, a estimação MAP “aproximada” dos parâmetros de um CDHMM é obtido por

$$\lambda^n = \arg \max_{\lambda} g(\lambda|\varphi^n).$$

O termo “aproximada” depende do critério adotado durante a aproximação. Uma

solução para φ^n é

$$\varphi^n = \arg \min_{\varphi} \int p(\lambda | \check{O}^n) \log \frac{p(\lambda | \check{O}^n)}{g(\lambda | \varphi)} d\lambda.$$

Infelizmente, não existe uma solução explícita de forma fechada para essa aproximação e uma otimização geral é necessária para estimar os hiper-parâmetros. Portanto, em vez de usar essa aproximação, é proposto neste trabalho um método chamado de aprendizagem quasi-Bayes, que é, ao mesmo tempo, computacionalmente simples e eficaz.

4.5.2 Aprendizagem MAP quasi-Bayes em Lote

O procedimento quasi-Bayes é uma solução aproximada que mantém a essência do procedimento Bayes formal. No sentido em que a distribuição posterior aproximada tenha uma média idêntica àquela da distribuição posterior verdadeira, as propriedades da convergência foram estabelecidas.

A cada passo da aprendizagem Bayes recursiva, o procedimento quasi-Bayes proposto aproxima a distribuição posterior resultante $p(\lambda | \check{O}^n)$ pela distribuição “aproximada” tratável $g(\lambda | \varphi^n)$, sob a condição de que as duas distribuições terão a mesma moda. A idéia é ilustrada na Figura 34. Esse procedimento também será chamado aqui de algoritmo MAP em lote.

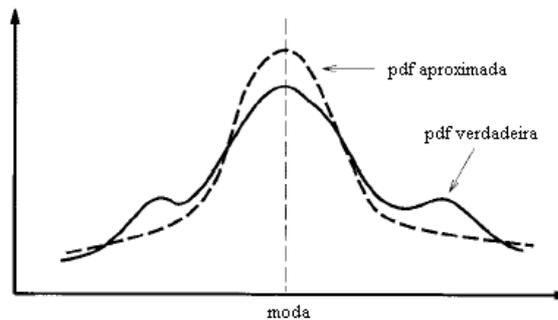


Figura 34: Aproximação da verdadeira distribuição posterior por uma simplificada, sob o critério de mesma moda.

Especificamente, considerando-se que, no instante n , tem-se a seqüência de treinamento $O^n = (o_1^n, o_2^n, \dots, o_{T_n}^n)$, a seqüência não-observada do estado $q^n = (q_1^n, q_2^n, \dots, q_{T_n}^n)$, e a seqüência associada às etiquetas das componentes não-observadas da mistura de gaussianas, e que o conhecimento a priori sobre λ é aproximado por $g(\lambda | \varphi^{n-1})$, então, usando o formalismo apresentado em (LEE; HUO, 2000), tem-se a estimação MAP aproximada λ^n de λ , repetindo-se os seguintes passos:

- Passo E

Calcular

$$R(\lambda|\lambda^{n-1,k-1}) = \rho \cdot \log g(\lambda|\varphi^{n-1}) + E[\log p(O^n, q^n|\lambda)|O^n, \lambda^{n-1,k-1}], \quad (4.20)$$

onde $0 < \rho \leq 1$ é o fator de esquecimento, com $\rho = 1$ significando que não há esquecimento.

- Passo M

Escolher

$$\lambda^{n-1,k} = \arg \max_{\lambda} R(\lambda|\lambda^{n-1,k-1}) \quad (4.21)$$

onde $k = 1, 2, \dots, K$ é o índice da iteração e K é o número total de iterações.

O procedimento EM acima apresentado vem do fato que $F(X|\lambda) = Q(\lambda^k, \lambda) + H(X(q))$ e que quando $Q(\lambda^k, \lambda) > Q(\lambda^k, \lambda^k)$ então $F(X|\lambda) > F(X|\lambda^k)$. Assim, pode-se usar o mesmo procedimento para estimar a moda da densidade posterior maximizando agora a função auxiliar $R(\lambda^k, \lambda) = Q(\lambda^k, \lambda) + \log g(\lambda)$ a cada iteração, em vez de $Q(\lambda^k, \lambda)$, como na abordagem EM. Note que a função custo da aprendizagem quasi-Bayes, representada pela equação 4.20, é composta de duas partes, uma que representa a estimação EM (segundo termo à direita da igualdade) somada à outra parte que representa a distribuição aproximada a priori com peso ρ (primeiro termo à direita da igualdade).

Torna-se importante, nesse momento, fazer a análise da influência do conhecimento a priori sobre o processo de adaptação. Se o conhecimento é muito forte, ou um certo lote de dados de adaptação já foi incrementalmente processado, os novos dados de adaptação, normalmente, terão um pequeno efeito sobre a atualização dos parâmetros no treinamento incremental. Para se acompanhar continuamente as variações dos parâmetros do modelo correspondentes às informações apresentadas pelos novos dados, é necessário introduzir algum mecanismo de esquecimento para reduzir o efeito das observações passadas em relação aos novos dados de entrada. O mecanismo proposto é um esquecimento exponencial implementado pelo uso do coeficiente de esquecimento ρ na equação 4.20.

Os hiper-parâmetros da função $g(\lambda|\hat{\varphi})$ são reestimados como a seguir:

$$\hat{\eta}_i = \rho \cdot (\eta_i^{n-1} - 1) + 1 + \gamma_1(i) \quad (4.22)$$

$$\hat{\eta}_{ij} = \rho.(\eta_{ij}^{n-1} - 1) + 1 + \sum_{t=1}^T \gamma_t(i, j) \quad (4.23)$$

$$\hat{\nu}_{jm} = \rho.(\nu_{jm}^{n-1} - 1) + 1 + c_{jm} \quad (4.24)$$

$$\hat{\tau}_{jm} = \rho.\tau_{jm}^{n-1} + c_{jm} \quad (4.25)$$

$$\hat{\kappa}_{jm} = \frac{\rho\tau_{jm}^{n-1}\kappa_{jm}^{n-1} + c_{jm}\bar{o}_{jm}}{\rho\tau_{jm}^{n-1} + c_{jm}} \quad (4.26)$$

$$\hat{\alpha}_{jm} = \rho.(\alpha_{jm}^{n-1} - 0.5) + 0.5 + 0.5c_{jm} \quad (4.27)$$

$$\hat{\beta}_{jm} = \rho\beta_{jm}^{n-1} + \frac{1}{2}S_{jm} + \frac{\rho\tau_{jm}^{n-1}c_{jm}}{2(\rho\tau_{jm}^{n-1} + c_{jm})} \times (\bar{o}_{jm} - \kappa_{jm}^{n-1})(\bar{o}_{jm} - \kappa_{jm}^{n-1})^t, \quad (4.28)$$

onde

$$\gamma_t(i) = p(q_t = i | O, \lambda) \quad 1 \leq t \leq T - 1 \quad (4.29)$$

$$\gamma_t(i, j) = p(q_t = i, q_{t+1} = j | O, \lambda) \quad 1 \leq t \leq T - 1 \quad (4.30)$$

$$\zeta_t(j, m) = p(q_t = j | O, \lambda) \quad 1 \leq t \leq T - 1 \quad (4.31)$$

$$c_{jm} = \sum_{t=1}^T \zeta_t(j, m) \quad (4.32)$$

$$\bar{o}_{jm} = \sum_{t=1}^T \zeta_t(j, m) o_t / c_{jm} \quad (4.33)$$

$$S_{jm} = \sum_{t=1}^T \zeta_t(j, m) (o_t - \bar{o}_{jm})(o_t - \bar{o}_{jm})^t. \quad (4.34)$$

Os termos acima podem ser calculados pelo algoritmo de *forward-backward*, realizando, assim, a adaptação MAP em lote. O termo n representa o índice da iteração, j o índice do estado e m o índice da gaussiana. As equações de reestimação EM dos parâmetros do modelo podem ser derivadas pela moda de $g(\lambda|\hat{\varphi})$, e são dadas por

$$\pi_i = \frac{\hat{\eta}_i - 1}{\sum_{i=1}^N (\hat{\eta}_i - 1)} \quad (4.35)$$

$$a_{ij} = \frac{\hat{\eta}_{ij} - 1}{\sum_{j=1}^N (\hat{\eta}_{ij} - 1)} \quad (4.36)$$

$$c_{jm} = \frac{\hat{\nu}_{jm} - 1}{\sum_{m=1}^M (\hat{\nu}_{jm} - 1)} \quad (4.37)$$

$$\mu_{jm} = \hat{\kappa}_{jm} \quad (4.38)$$

$$U_{jm} = \frac{2\rho\beta_{jm}^{n-1} + \rho\tau_{jm}^{n-1}(\mu_{jm} - \kappa_{jm}^{n-1})(\mu_{jm} - \kappa_{jm}^{n-1})^t + \sum_{t=1}^T \zeta_t(j, m)(o_t - \mu_{jm})(o_t - \mu_{jm})^t}{\rho(2\alpha_{jm}^{n-1} - 1) + \sum_{t=1}^T \zeta_t(j, m)}. \quad (4.39)$$

Utilizando as equações acima, obtém-se uma série de *pdfs* aproximadas $g(\lambda|\hat{\varphi})$, cuja

moda está se aproximando da moda da *pdf* posterior verdadeira, apresentada na equação 4.19. Assim, a *pdf* aproximada será calculada por

$$p(\lambda|O^n) = \frac{p(O^n|\lambda)g(\lambda|\varphi^{n-1})}{\int_{\Omega} p(O^n|\lambda)g(\lambda|\varphi^{n-1})d\lambda}. \quad (4.40)$$

Os hiper-parâmetros φ^n são atualizados pela última iteração EM, usando as equações 4.22 a 4.28 para satisfazer a condição

$$g(\lambda|\varphi^n) \propto \exp\{R(\lambda|\lambda^{n-1+\frac{k-1}{K}})\}, \quad (4.41)$$

e então os parâmetros do CDHMM são atualizados.

4.5.3 Aprendizagem MAP quasi-Bayes Incremental

O procedimento MAP em lote, visto na seção anterior, pode ser estendido para o procedimento segmental ou Viterbi, substituindo as equações 4.29 a 4.31 pelas equações

$$\gamma_t(i) = \delta(q_t - i), \quad (4.42)$$

$$\gamma_t(i, j) = \delta(q_t - i)\delta(q_{t+1} - j), \quad (4.43)$$

$$\zeta_t(j, m) = \gamma_t(i) \cdot \frac{c_{jm}N(o_t, \mu_{jm}, U_{jm})}{\sum_{m=1}^M c_{jm}N(o_t, \mu_{jm}, U_{jm})}, \quad (4.44)$$

onde $q = (q_1, q_2, \dots, q_T)$ é a seqüência de estados mais provável e $\delta(\cdot)$ é a função delta Kronecker. Então, a proposta feita neste trabalho é continuar a usar o procedimento que utiliza a equação 4.20 e 4.21, mas com $\gamma_t(i)$, $\gamma_t(i, j)$ e $\zeta_t(j, m)$ como apresentados nas equações 4.42 a 4.44, ou seja, para a seqüência de estados mais provável. Assim, tem-se a influência das observações já processadas, dada por $\log g(\lambda|\varphi)$ com peso ρ , e o uso do caminho ótimo, o que reduz os custos computacionais. Vale notar que a verdadeira influência de ρ se dá no momento do cálculo dos hiper-parâmetros de acordo com as equações 4.22 a 4.28 e com a reestimação dos parâmetros do modelo que se efetua a partir dos hiper-parâmetros. Os resultados obtidos por essa proposta são apresentados no Capítulo 5, e o algoritmo é dado por

- Passo E
Calcular

$$L^n(\lambda) = \rho \cdot \log g(\lambda|\varphi^{n-1}) + E \left[\log p(O^n, \hat{q}^n|\lambda)|O^n, \lambda^{n-1} \right] + L^{n-1}(\lambda) \quad (4.45)$$

onde $L(\lambda)$ é a função custo utilizada.

- Passo M

Escolher $\lambda = \lambda^n$ para maximizar $L^n(\lambda)$ e também atualizar os hiper-parâmetros para obter $\hat{\varphi}^n$.

Observe que a função $L(\lambda)$ deve ser atualizada incrementalmente, e sua estimação inicial se dá por meio de

$$L^0(\lambda) = \log p(\lambda) = \log g(\lambda|\varphi^0).$$

Uma questão que ainda resta resolver é sobre a estimação inicial dos hiper-parâmetros. A densidade anterior inicial $g(\lambda|\varphi^0)$ é assumida como um membro de uma família pré-determinada de distribuição anterior. Devido à dificuldade de se possuir um conhecimento completo da distribuição anterior, é adotada a abordagem empírica, apresentada em (HUO; LEE, 1997), para estimar os valores iniciais dos hiper-parâmetros φ^0 . Considerando que a abordagem MAP é usada para a adaptação dos modelos, e que estes já possuem uma versão inicial advinda de um treinamento, ou seja, os modelos genéricos (mg), os hiper-parâmetros iniciais são dados empiricamente por

$$\eta_i^0 = 1 + \varepsilon_1 \gamma_1^{(mg)}(i) \quad (4.46)$$

$$\eta_{ij}^0 = 1 + \varepsilon_1 \sum_t \gamma_t^{(mg)}(i, j) \quad (4.47)$$

$$\nu_{jm}^0 = 1 + \varepsilon_1 \sum_t \zeta_t^{(mg)}(j, m) \quad (4.48)$$

$$\tau_{jm}^0 = \varepsilon_1 \sum_t \zeta_t^{(mg)}(j, m) \quad (4.49)$$

$$\kappa_{jm}^0 = \mu_{jm}^{(mg)} \quad (4.50)$$

$$\alpha_{jm}^0 = 0.5 + 0.5\varepsilon_1 \sum_t \zeta_t^{(mg)}(j, m) \quad (4.51)$$

$$\beta_{jm}^0 = 0.5\varepsilon_1 U_{jm}^{(mg)} \sum_t \zeta_t^{(mg)}(j, m), \quad (4.52)$$

onde $0 < \varepsilon_1 \leq 1$ é o coeficiente que controla a importância do conhecimento anterior ou para equilibrar a contribuição entre os dados de adaptação, dados pelo modelo genérico (mg), e os dados de adaptação.

4.5.4 Adaptação MAP Incremental Implementada

Depois de se estudar o algoritmo incremental apresentado na seção anterior, é necessário estudar a implementação do mesmo sobre a plataforma desenvolvida em (ANDREAO, 2004). É importante notar que a estrutura de aprendizagem quasi-Bayes apresentada nas seções anteriores é flexível o suficiente para processar uma observação a cada instante ou um bloco de observações pequeno o suficiente para garantir os requisitos computacionais da atualização bayesiana.

Assim como na implementação do algoritmo segmental k-means incremental, as modificações feitas no sistema desenvolvido em (ANDREAO, 2004) correspondem à Fase 2, com a implementação da função *learn_ghmm_map.m* apresentada na Figura 35.

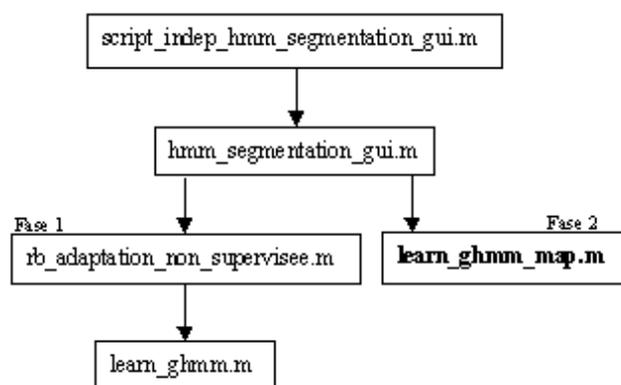


Figura 35: Arquitetura básica do algoritmo MAP incremental.

É importante dizer o sistema desenvolvido em (ANDREAO, 2004) foi agora modificado para se guardar as informações dos blocos anteriores, representadas por $X(q)$, conforme mostrado na Figura 36. Guardar as informações, também chamadas de estatísticas de blocos anteriores, nos leva a uma real implementação de um algoritmo incremental, de acordo com Digalakis (1999) e Gotoh e Silverman (1996).

A variável ε_1 foi definida como $1/\text{num_exem}$, onde *num_exem* corresponde ao número de exemplos analisados a cada instante. Este é o valor sugerido em (HUO; LEE, 1997), apesar de ela poder ser especificada pelo usuário ou poder ser determinada a posteriori pela medida da similaridade entre os dados de adaptação e os modelos existentes.

Uma questão a se responder é qual deve ser o valor ótimo de ρ , uma vez que ele significa a influência das observações passadas na adaptação dos modelos. Assim, para se buscar o valor ótimo de ρ , foram feitos experimentos para verificar o efeito dos hiperparâmetros na adaptação dos modelos. Os testes realizados para a escolha de ρ são

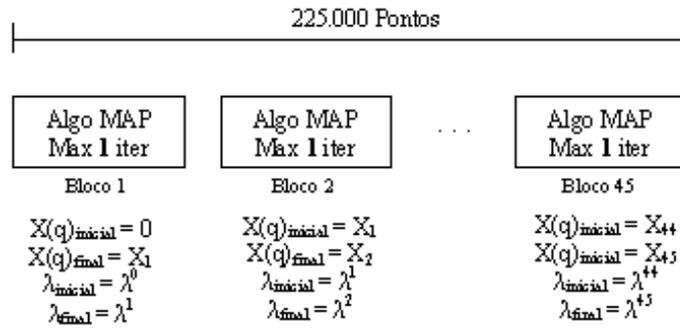


Figura 36: Sistema desenvolvido para abordagem MAP.

apresentados no Capítulo 5, só interessando, até o presente momento, que o seu valor ótimo encontrado que foi $\rho = 0,4$. Este valor ótimo pode ser explicado de forma que é necessário acelerar o esquecimento das estatísticas dos blocos anteriores para que o modelo acompanhe da melhor maneira a evolução do sinal de ECG no tempo. No entanto, o completo esquecimento das mesmas parece ser maléfico, pois se perde, aí, dados para bem restimar os parâmetros do modelo. Assim, o valor ótimo seria um valor intermediário, que foi dado como 0,4.

Uma vez encontrados os valores de ρ e ϵ_1 , foi executada toda a série de testes, cujos resultados são apresentados no Capítulo 5. Uma outra etapa feita foi a checagem da robustez da estimação MAP, criando uma abordagem que será chamada mais adiante de **MAP Incremental Robusto**. Nela, os exemplos adquiridos durante a fase de segmentação do sinal de ECG, em um dado bloco de 20 s, são utilizados para a adaptação, mesmo que não sejam representativos. Ressalte-se que, para as outras abordagens, foi determinado empiricamente por (ANDREAO, 2004) um número mínimo de exemplos que garantissem pelo menos 30 amostras para cada gaussiana do modelo, de forma a permitir uma boa estimação dos seus parâmetros. Nesse sentido, a adaptação MAP será feita mesmo que para poucos dados, o que é o ponto forte de se usar essa abordagem de adaptação. Os resultados para todas as abordagens são apresentados no próximo capítulo.

5 *Resultados e Conclusões*

Neste capítulo são mostrados os resultados encontrados, assim como as conclusões obtidas a partir das simulações. Será feita uma avaliação e uma comparação entre o algoritmo desenvolvido em (ANDREAO, 2004), o algoritmo EM incremental, o algoritmo segmental k-means incremental e o algoritmo MAP incremental, estes desenvolvidos neste trabalho.

5.1 Resultados

Foram feitos quatro protocolos de simulação: anotações das ondas, anotações de QRS, anotações de PVC e anotações de isquemia, que se dividem em duas partes.

- **Primeira parte**

Para esta parte é utilizada a base QT (LAGUNA et al., 1997). Detalhando um pouco mais sobre tal base, ela é composta por 105 arquivos de 15 minutos de duração. Cada registro contém o sinal de duas derivações, com taxa de amostragem de 250 Hz. Os registros são acompanhados pelas anotações feitas por cardiologistas. Pelo menos 30 batimentos por registro são etiquetados com a indicação das ondas, ou formas elementares de onda (P, QRS e T) do batimento. A Tabela 4 mostra o número de formas elementares de onda da base QT etiquetadas pelo cardiologista. Entre os 105 registros, 82 contêm as etiquetas da classe do batimento entre mais de 13 anomalias diagnosticadas por um cardiologista.

- 1ª Etapa: anotações das ondas. Nesta etapa utilizam-se os 105 registros da base para se calcular os parâmetros que avaliam o desempenho de segmentação do sistema, aqui indicada pela detecção, onde se calcula a porcentagem de ondas P, QRS e T corretamente detectadas de acordo com as etiquetadas pelo cardiologista.

As abordagens avaliadas foram: ***Sem Adaptação***, para o caso onde os mode-

los não são adaptados, *Andreão*, para a abordagem já feita em (ANDREAO, 2004), *EM Incremental*, *Segmental K-means Incremental*, *MAP Incremental* e *MAP Incremental Robusto*, estas estudadas no Capítulo 4. Para esta primeira etapa, os resultados são apresentados na Tabela 5, que mostra os valores de detecção obtidos para um canal.

Tabela 4: Número de etiquetas feitas para os 105 registros da base QT.

Número de etiquetas para os 105 registros		
Onda P	Complexo QRS	Onda T
3194	3623	3543

Tabela 5: Precisão e detecção de ondas.

Detecção de ondas

	Onda P	Complexo QRS	Onda T
<i>Sem Adaptação</i>			
Detecção (%)	87,26	99,31	99,27
<i>Andreão</i>			
Detecção (%)	92,89	99,86	99,83
<i>EM Incremental</i>			
Detecção (%)	91,45	100,00	99,97
<i>Seg. K-means Inc.</i>			
Detecção (%)	92,36	100,00	99,97
<i>MAP Incremental</i>			
Detecção (%)	93,11	100,00	99,97
<i>MAP Inc. Robusto</i>			
Detecção (%)	91,30	100,00	99,97

- 2ª Etapa: anotações de QRS. Destina-se à detecção de complexos QRS, e é feita com 78 registros da base QT. Nesta etapa são adotados dois critérios de desempenho para avaliação de detecção das abordagens, que são utilizados pelos cardiologistas. O primeiro critério é a sensibilidade (Se), que nada mais é que a fração de batimentos corretamente detectados, ou seja,

$$Se = \frac{VP}{VP + FN},$$

onde VP (verdadeiro positivo) é o número de batimentos corretamente detectados e FN (falso negativo) é o número de batimentos que não foram detectados pelo sistema. O denominador da última equação corresponde ao número total de batimentos etiquetados da base.

O segundo critério de avaliação é o valor preditivo positivo (PP) que corresponde à capacidade do sistema de detectar batimentos verdadeiros, ou seja,

$$PP = \frac{VP}{VP + FP},$$

onde FP (falso positivo) é o número de batimentos detectados pela abordagem que não estão de acordo com as etiquetas da base. Um batimento detectado está de acordo com a base quando a distância entre a posição das duas etiquetas é inferior a 150 ms.

Deve ser citado que as abordagens possuem a capacidade de trabalhar em um contexto multi-canal (duas derivações), e isto é feito com a “fusão” de canais. A fusão de canais, dentro desta primeira parte, visa a confirmar a detecção de um batimento, uma vez que ele é detectado dentro de todas as derivações. É considerado que um batimento detectado em um canal está de acordo com o detectado pelo outro canal quando a distância entre a posição das duas etiquetas é inferior a 200 ms. Assim, a lógica booleana ‘E’ é aplicada na fusão. Os resultados obtidos podem ser vistos na Tabela 6.

- 3ª Etapa: anotações de PVC (Premature Ventricular Contraction) ou ESV (Extra Systole Ventriculaire). Detecta a ocorrência de extrasístole ventricular, que é uma arritmia ventricular caracterizada pela presença de um complexo QRS pré-maturo, de morfologia diferente (complexo QRS mais largo) e com a ausência de uma onda P precedente. Nesta etapa, utiliza-se os mesmos registros e critérios de desempenho usados na segunda etapa, e é importante notar que os batimentos que não são etiquetados como normal nem ESV dentro da base são considerados normais dentro deste trabalho. Os resultados são apresentados na Tabela 7.

- **Segunda parte**

Aqui é analisada a aparição de episódios isquêmicos. Essa análise consiste em medir o deslocamento da amplitude do segmento ST e da onda T, segundo as regras estabelecidas por um grupo de cardiologistas no momento da construção da base européia ST-T (TADDEI et al., 1991). Assim, fez-se necessário o uso dessa base,

Tabela 6: Detecção de complexos QRS.

Anotações de QRS					
	VP	FP	FN	Se (%)	PP (%)
<i>Sem Adaptação</i>					
Canal1	82514	1862	159	99,81	97,79
Canal2	82437	2090	236	99,71	97,53
Fusão	82294	20	379	99,54	99,98
<i>Andreão</i>					
Canal1	82576	1021	97	99,88	98,78
Canal2	82527	1718	146	99,82	97,96
Fusão	82484	67	189	99,77	99,92
<i>EM Incremental</i>					
Canal1	82575	948	98	99,88	98,86
Canal2	82539	1699	134	99,84	97,98
Fusão	82487	66	186	99,78	99,92
<i>Seg. K-means Inc.</i>					
Canal1	82576	853	97	99,88	98,98
Canal2	82504	2144	169	99,80	97,47
Fusão	82462	33	211	99,74	99,96
<i>MAP Incremental</i>					
Canal1	82576	845	97	99,88	98,99
Canal2	82498	2523	175	99,79	97,03
Fusão	82458	26	215	99,74	99,97
<i>MAP Inc. Robusto</i>					
Canal1	82579	1030	94	99,89	98,77
Canal2	82531	2161	142	99,83	97,45
Fusão	82510	18	163	99,80	99,98

a qual é composta de 90 registros de duas horas de 79 indivíduos. Cada registro contém o sinal de duas derivações amostradas à taxa de 250 Hz.

Ressalte-se que tal base é utilizada somente para as fases de segmentação e classificação dos episódios, uma vez que ela não possui etiquetas de onda P, QRS, T. Assim, os modelos iniciais são treinados com os parâmetros extraídos da base QT.

- 1ª Etapa: anotações de isquemia. Nessa etapa foram utilizados 48 registros da base ST-T, os mesmos usados em (ANDREAO, 2004). As etiquetas da base se referem aos episódios isquêmicos associados a mudanças do segmento ST

Tabela 7: Detecção de Extrasístoles.

	Detecção de Extrasístoles					Detecção de Batimentos Normais				
	VP	FP	FN	Se (%)	PP (%)	VP	FP	FN	Se (%)	PP (%)
<i>Sem Adaptação</i>										
Canal 1	471	6	1225	27,77	98,74	80845	3054	132	99,84	99,36
Canal 2	391	15	1305	23,05	96,31	80756	3365	221	99,73	96,00
Fusão	1156	214	540	68,16	84,38	80437	507	540	99,33	99,37
<i>Andreão</i>										
Canal 1	1050	519	646	61,91	66,92	80866	1154	111	99,86	98,59
Canal 2	956	682	740	56,37	58,36	80864	1425	113	99,86	98,27
Fusão	1529	454	167	90,15	77,11	80466	108	511	99,37	99,87
<i>EM Incremental</i>										
Canal 1	1043	367	653	61,50	73,97	80884	1161	93	99,89	98,58
Canal 2	957	1046	739	56,43	47,78	80865	1369	112	99,86	98,34
Fusão	1523	432	173	89,80	77,90	80513	113	464	99,43	99,86
<i>Seg. K-means Inc.</i>										
Canal 1	1047	187	649	61,73	84,85	80887	1107	90	99,89	98,65
Canal 2	951	1461	745	56,07	39,43	80846	1462	131	99,84	98,22
Fusão	1508	360	188	88,92	80,73	80536	110	441	99,46	99,86
<i>MAP Incremental</i>										
Canal 1	1001	423	695	59,02	70,29	80893	1137	84	99,90	98,61
Canal 2	946	1823	750	55,78	34,16	80836	1444	141	99,83	98,25
Fusão	1515	371	181	89,33	80,33	80497	109	480	99,41	99,86
<i>MAP Inc. Robusto</i>										
Canal 1	995	425	701	58,67	70,07	80888	1320	89	99,89	99,39
Canal 2	927	1379	769	54,66	40,20	80870	1534	107	99,87	98,14
Fusão	1488	380	208	87,74	79,66	80513	153	464	99,43	99,81

e da onda T. Todavia, para uma análise multicanal, os episódios etiquetados em cada derivação são combinados por uma lógica booleana ‘OU’. O número de episódios etiquetados em cada derivação e na combinação entre elas são apresentados na Tabela 8.

Os critérios de avaliação continuam sendo a sensibilidade (Se) na detecção de episódios e o valor preditivo positivo (PP), que representa a capacidade de detectar episódios verdadeiros. Os resultados obtidos podem ser vistos na

Tabela 8: Número de episódios etiquetados em 48 registros.

Número de Episódios Etiquetados		
Derivação	Episódios ST	Episódios T
1	89	115
2	84	58
1 'OU' 2	120	117

Tabela 9. Os resultados, aqui, são apresentados pela taxa bruta, onde os índices são calculados quando todos os registros são considerados como um único registro total e a taxa média, onde os valores são calculados para cada registro separadamente.

- **Determinação de ρ .**

A detecção de QRS não foi relevante para a escolha de ρ . Assim, sua escolha foi feita a partir dos resultados da detecção de ondas, mais precisamente da detecção da onda P. Para se buscar o valor ótimo de ρ foram feitos experimentos para verificar o efeito dos hiper-parâmetros na adaptação dos modelos. Os testes realizados foram a detecção de ondas e de complexos QRS (para o canal 1, base QT) com ρ variando de 0 a 1 e os resultados obtidos são apresentados nas Tabelas 10 e 11. Pode-se observar que o uso das estatísticas dos blocos anteriores melhora o desempenho da detecção, sobretudo no que diz respeito à detecção de onda P. Para isso, é necessário ponderar a contribuição dos hiper-parâmetros. O comportamento da detecção de onda P com a variação de ρ pode ser visto na Figura 37. Observa-se também que o melhor resultado encontrado foi para $\rho = 0,4$, que foi escolhido como valor ótimo.

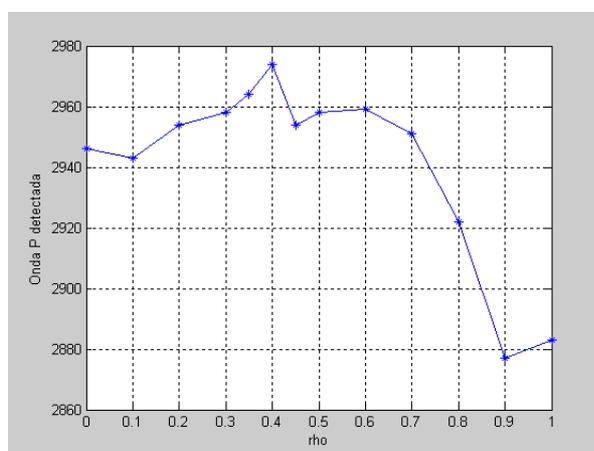
Figura 37: Comportamento de detecção de onda P com a variação de ρ .

Tabela 9: Número de episódios isquêmicos em 48 registros.

	Detecção de Episódios Isquêmicos			
	Taxa Bruta		Taxa Média	
	Se (%)	PP (%)	Se (%)	PP (%)
<i>Sem Adaptação</i>				
Canal 1	86,52	51,45	90,56	71,62
Canal 2	76,19	51,24	83,97	78,33
Fusão	74,17	73,33	77,97	81,96
<i>Andreão</i>				
Canal 1	77,53	55,93	83,71	82,44
Canal 2	76,19	58,33	83,97	79,92
Fusão	75,83	78,57	82,94	83,97
<i>EM Incremental</i>				
Canal 1	87,64	61,11	90,76	81,51
Canal 2	78,57	57,27	85,20	75,96
Fusão	81,67	73,60	86,38	80,84
<i>Seg. K-means Inc.</i>				
Canal 1	86,52	65,18	89,80	82,46
Canal 2	77,38	53,04	83,73	72,73
Fusão	79,17	79,82	82,86	85,47
<i>MAP Incremental</i>				
Canal 1	80,90	62,28	84,46	85,22
Canal 2	76,19	57,14	82,25	78,66
Fusão	81,67	78,33	86,27	84,67
<i>MAP Inc. Robusto</i>				
Canal 1	80,90	56,67	84,28	75,72
Canal 2	76,19	57,66	83,14	82,72
Fusão	76,67	77,97	82,18	84,57

5.2 Discussão

As tabelas apresentadas na seção anterior podem levar a diversas conclusões. Elas certamente comprovam a necessidade de se adaptar os modelos, uma vez que sem adaptação os resultados se apresentam inferiores. Em relação às abordagens desenvolvidas, é importante lembrar que todas elas são do tipo incremental, o que facilita a comparação. Pode-se ver também que a abordagem segmental k-Means incremental apresenta resultados muito

Tabela 10: Detecção de ondas para determinados valores de ρ .

Detecção de Onda			
ρ	P	QRS	T
0,00	2946	3623	3542
0,10	2943	3623	3542
0,20	2954	3623	3542
0,30	2958	3623	3542
0,35	2964	3623	3542
0,40	2974	3623	3542
0,45	2954	3623	3542
0,50	2958	3623	3542
0,60	2959	3623	3542
0,70	2951	3623	3542
0,80	2922	3623	3542
0,90	2877	3623	3542
1,00	2883	3623	3542

Tabela 11: Detecção de complexo QRS para determinados valores de ρ para o Canal 1.

Detecção de QRS					
ρ	VP	FP	FN	Se(%)	PP (%)
0,00	82576	853	97	99,88	98,98
0,10	82575	856	98	99,88	98,97
0,20	82577	798	96	99,88	99,04
0,30	82575	907	98	99,88	98,91
0,35	82575	835	98	99,88	99,00
0,40	82576	845	97	99,88	98,99
0,45	82577	862	96	99,88	98,97
0,50	82577	853	96	99,88	98,98
0,60	82579	831	94	99,89	99,00
0,70	82581	849	92	99,89	98,98
0,80	82581	905	92	99,89	98,92
0,90	82583	1192	90	99,89	98,58
1,00	82589	1149	84	99,90	98,63

bons, mas a abordagem MAP incremental apresenta resultados ainda melhores.

Tem-se como principal vantagem das abordagens segmental k-means e MAP a possibilidade de uma verdadeira adaptação *on-line*. A abordagem EM incremental também

possui essa vantagem, pois as versões incrementais dos algoritmos de adaptação podem melhorar a convergência e reduzir os custos computacionais, o que os conduz a uma aplicação *on-line*. Entretanto, uma maneira de verificar qual abordagem incremental é mais indicada à adaptação *on-line* seria checar o tempo de cálculo ou o número de operações computacionais executadas. Já que os algoritmos segmental k-means incremental e MAP incremental foram implementados em Matlab e o algoritmo EM em C++, não é possível comparar os tempos, mas utiliza-se a comparação do número de laços ‘for’ utilizados pelos mesmos no algoritmo de atualização dos modelos. O número de laços ‘for’ dos dois primeiros se resume a 2 com dimensões ‘1:Q’ (número de estados) e 1 com dimensões ‘1:T’ (número de amostras), já o terceiro, possui esse mesmo número de laços ‘for’ apenas para efetuar o procedimento forward, além dos outros existentes. Assim, conclui-se que a abordagem MAP, ou quasi-Bayes, é a mais apropriada para uma adaptação *on-line* do sinal de ECG, devido aos resultados de detecção e à redução dos custos computacionais.

Em se tratando de abordagem MAP, é possível verificar que mesmo quando os dados não são suficientes ela se comporta de maneira razoável (abordagem *MAP Inc. Robusto*), apresentando bons resultados de detecção de onda P e detecção de isquemia. O que leva a refletir, em se tratando de detecção de isquemia, sobre a importância do número de batimentos corretamente detectados (Se) e a capacidade de detecção de verdadeiros batimentos (PP), pois é a principal diferença entre as abordagens MAP implementadas.

Pode-se observar, também, que assim como o algoritmo EM incremental o segmental k-means incremental apresenta uma pequena perda na detecção de onda P, o que a princípio não é uma prioridade para a detecção de isquemia. Vale citar que a detecção de complexos QRS, de acordo com o número de etiquetas feitas pelo cardiologista, apresentado na Tabela 4, foi de 100%.

5.3 Conclusões e Perspectivas

Esses resultados são muito importantes para a criação de um sistema de adaptação *on-line* aplicado a um sistema portátil de captura de sinais de ECG, o qual está sendo desenvolvido no *Institut National des Télécommunications - França*, devido à redução dos custos computacionais dos algoritmos de adaptação. Viu-se que essa redução não afetou, mas sim, melhorou o desempenho do sistema.

O objetivo desse trabalho foi o estudo de técnicas de adaptação de modelos ocultos de

Markov para a análise do sinal de ECG. A comprovação de seus benefícios foi verificada em um sistema de segmentação e classificação de sinais de ECG já desenvolvido (ANDREAO, 2004). As técnicas implementadas foram baseadas em critérios de otimização de máxima verossimilhança (algoritmo EM e algoritmo segmental k-means) e de máximo a posteriori (algoritmo MAP). Foi comprovado que a adaptação é realmente necessária e que, entre as abordagens implementadas, a que apresentou melhores resultados foi a abordagem bayesiana, com a aprendizagem quasi-Bayes (HUO; LEE, 1997), executada pelo algoritmo MAP. Esta se apresenta melhor, não só pelos resultados mas também, pela requisição de poucos dados para o seu processamento.

Assim, será feito um estudo da possível aplicação das técnicas apresentadas a outros sinais biomédicos como o EEG (eletroencefalograma). Já existem alguns trabalhos nessa área como (OBERMAIER; GUGER; PFURTSCHELLER, 2001), onde se pode fazer a classificação de tarefas mentais usando HMM e (CHIAPPA; BENGIO, 2004) que utiliza HMM para modelar os ritmos de EEG. Os resultados obtidos também podem ser usados para o estudo da adaptação baseada em *Multiple-Stream* encontrada em (HUO; MA, 2001), a qual utiliza uma arquitetura onde é possível acoplar mais que uma técnica de adaptação, e em que cada técnica exerce uma função distinta no processo de adaptação, ou mesmo para o reconhecimento de assinatura (SALICETTI, 1996).

Referências

- ANDREAIO, R. V. *Segmentation de battements ECG par approche markovienne : application à la détection de ischémies*. Tese (Doutorado) — Ecole doctorale INT-UTT, decembre 2004.
- ANDREAIO, R. V. et al. ST-Segment analysis using hidden Markov model beat segmentation: application to ischemia detection. *Computers in Cardiology*, p. 384–384, 2004.
- ANDREAIO, R. V. et al. Efficient ECG multi-level wavelet classification through neural network dimensionality reduction. *In proceedings of the IEEE Workshop on Neural Network for Signal Processing*, p. 395–404, 2002.
- BARDONOVA, J. et al. Detection of myocardial ischemia using hidden Markov models. *Proc. of the 25th Annual International Conference of the IEEE EMBS*, p. 2869–2872, 2003.
- BROOKS, D. H. et al. Best basis segmentation of ECG signal using novel optimality criteria. *In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, p. 2750–2753, 1996.
- CHAZAL, P. de; REILLY, R. B. Automatic classification of ECG beats using waveform shape and heart beat interval features. *In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, p. II–269 – II–272, 2003.
- CHIAPPA, S.; BENGIO, S. HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems. *ESANN proceedings - European Symposium on Artificial Neural Networks*, p. 199–204, 2004.
- COAST, D. A. Segmentation of high-resolution ECGs using hidden Markov models. *In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 1, p. 67–70, 1993.
- COAST, D. A. et al. An approach to cardiac arrhythmia analysis using hidden Markov models. *IEEE Transactions on Biomedical Engineering*, v. 37, n. 9, p. 826–836, September 1990.
- COCO, K. F. Notas de aula da disciplina sinais e sistemas. *Universidade Federal do Espírito Santo - Brasil*, 2003.
- COHEN, A. Hidden Markov models in biomedical signal processing. *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, v. 20, n. 3, p. 1145–1150, 1998.

- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, v. 39, n. 1, p. 1–38, 1977.
- DIGALAKIS, V. V. Online adaptation of hidden Markov models using incremental estimation algorithms. *IEEE Transactions on Speech and Audio Processing*, v. 7, n. 3, p. 253–261, May 1999.
- GAUVAIN, J.-L.; LEE, C.-H. MAP estimation of continuous density HMM: Theory and applications. *Proc. DARPA Speech and Natural Language Workshop*, p. 185–190, February 1992.
- GAUVAIN, J.-L.; LEE, C.-H. Maximum a posteriori estimation for multivariate gaussian mixture observations of Markov chains. *IEEE Transactions on Speech and Audio Processing*, v. 2, n. 2, p. 291–298, April 1994.
- GOTOH, Y.; SILVERMAN, H. F. Incremental ML estimation of HMM parameters for efficient training. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, p. 585–588, 1996.
- GUYTON, A. C.; HALL, J. E. *Tratado de Fisiologia Médica*. [S.l.]: Guanabara Koogan, 2002.
- HUO, Q.; LEE, C.-H. On-line adaptive learning of the continuous density hidden Markov model based on approximate recursive bayes estimate. *IEEE Transactions on Speech and Audio Processing*, v. 5, n. 2, p. 161–172, March 1997.
- HUO, Q.; MA, B. On-line adaptive learning of the continuous density hidden Markov model based on multiple-stream prior evolution and posterior pooling. *IEEE Transactions on Speech and Audio Processing*, v. 9, n. 4, p. 388–398, May 2001.
- LAGUNA, P. et al. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ecg. *Computers in Cardiology*, p. 673–676, 1997.
- LAMKADDEM, S. *Rapport de stage 3ème année du cycle ingénieur*. [S.l.], 2005.
- LEE, C.-H.; HUO, Q. On adaptive decision rules and decision parameter adaptation for automatic speech recognition. *Proc. IEEE*, v. 88, n. 8, p. 1241–1269, August 2000.
- LEE, C.-H.; LIN, C.-H.; JUANG, B.-H. A study on speaker adaptation of the parameters of continuous density hidden Markov models. *IEEE Transactions on Signal Processing*, v. 39, n. 4, p. 806–814, April 1991.
- LI, C.; ZHENG, C.; TAI, C. Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, v. 42, n. 1, p. 21–28, January 1995.
- MALLAT, S. *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic Press, 1998.
- NEAL, R. M.; HINTON, G. E. A view of the EM algorithm that justifies incremental and other variants. *submitted to Biometrika*, February 1993.

- NICOLAU, J. C. et al. Aplicações clínicas do eletrocardiograma no infarto agudo do miocárdio. *Revista da Sociedade de Cardiologia de São Paulo*, v. 9, n. 3, 2003.
- OBERMAIER, B.; GUGER, C.; PFURTSCHHELLER, G. Information transfer rate in a five-classes brain-computer interface. *IEEE Trans. on Rehabilitation Engineering*, v. 9, p. 283–288, September 2001.
- OLIVEIRA, F. I. de; ANDREA, R. V. *TELEPAT Classifier, MatLab-GUI Interface*. Evry-France, 2003–2004.
- PROVAZNIK, I. et al. Wavelet transform in ECG signal processing. *In Analysis of Biomedical Signals and Images. Proceedings of the 15th International Conference BIOSIGNAL 2000*, p. 21–25, 2000.
- RABINER, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, v. 77, n. 2, p. 257–286, February 1989.
- RABINER, L. R.; JUANG, B. H. *Fundamentals of Speech Recognition*. 1993. Prentice Hall.
- SABRY-RISK, M. et al. Highly accurate higher order statistics based neural network classifier of specific abnormality in eletrocardiogram signals. *In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, p. 1033–1036, 1999.
- SALICETTI, S. G. *Une approche neuronale predictive pour la reconnaissance en-ligne de l'écriture cursive*. Tese (Doutorado) — Université Paris VI, decembre 1996.
- SORIA, E. et al. Adaptive system for the location of characteristic points in an ECG. *18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1996.
- SPRAGINS, J. A note on the iterative application of Bayes' rule. *IEEE Transactions on Information Theory*, IT-11, n. 4, p. 544–549, October 1965.
- TADDEI, A. et al. The european ST-T database: development, distribution and use. *Computers in Cardiology*, p. 170–180, 1991.
- TORRENCE, C.; COMPO, G. P. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, v. 79, n. 1, p. 61–78, January 1998.
- TUMER, M. B.; BELFORE, L. A.; ROPELLA, K. M. Applying hierarchical fuzzy automata to automatic diagnosis. *Conf. North American Fuzzy Inf. Proc. Soc.*, p. 315–319, 1998.
- USP, C. de Comunicação Social da Universidade de S. P. *Prevenção e diagnóstico precoce, receita para 90% dos problemas cardíacos*. 2005.
- WATROUS, R.; TOWELL, G. A patient-adaptive neural network ECG patient monitoring algorithm. *Computers in Cardiology*, p. 10–13, 1995.

Anexos

Anexo A - Wavelets

As wavelets têm três principais aplicações:

- i) Banco de Filtros: dependendo da base wavelet escolhida é possível decompor e reconstruir um sinal utilizando suas diferentes bandas de frequência.
- ii) Filtragem Adaptativa: no que se refere a uma análise de correlação. A saída será ótima quando o sinal de entrada se parecer mais com a função wavelet escolhida.
- iii) Localização no tempo e na frequência: a transformada wavelet analisa as componentes do sinal a partir de wavelets com escalas de tempo distintas e sua localização em frequência está relacionada com o grau de regularidade da função, que por sua vez está relacionada ao número de derivadas contínuas existentes.

Transformada Wavelet

Fazendo uma breve revisão da matemática das wavelets, sabe-se (COCO, 2003) que elas são funções caracterizadas por terem a média igual a zero, ou seja, $\int_{-\infty}^{+\infty} \psi(t) dt = 0$. Se forem normalizadas, elas possuem energia unitária, ou seja $\|\psi\| = 1$, e, geralmente, são centradas no eixo $t = 0$.

A transformada wavelet pode ser vista como uma convolução entre o sinal f e a função wavelet escolhida, ou seja

$$Wf(u, s) = f * \bar{\psi}_s(u),$$

onde

$$\bar{\psi}_s(t) = \frac{1}{\sqrt{s}} \psi^* \left(\frac{-t}{s} \right)$$

é chamada de wavelet escalada.

Escrevendo a wavelet escalada no domínio da frequência, tem-se que

$$\hat{\psi}_s(\omega) = \sqrt{s}\hat{\psi}^*(s\omega).$$

A transformada wavelet pode ser ilustrada como na seqüência a seguir:

- i) Efetua-se o produto interno entre a wavelet escalada e cada trecho do sinal, obtendo-se o coeficiente de correlação C que representa o grau de similaridade entre a wavelet escalada e a parte do sinal comparada (Figura 38(a));
- ii) Desloca-se a wavelet para a direita e repete-se o passo anterior. Essa etapa continua enquanto existirem amostras do sinal (Figura 38(b)), e obtém-se como resultado a Figura 38(c);
- iii) Escala-se a wavelet e repetem-se os passos anteriores (Figura 39).

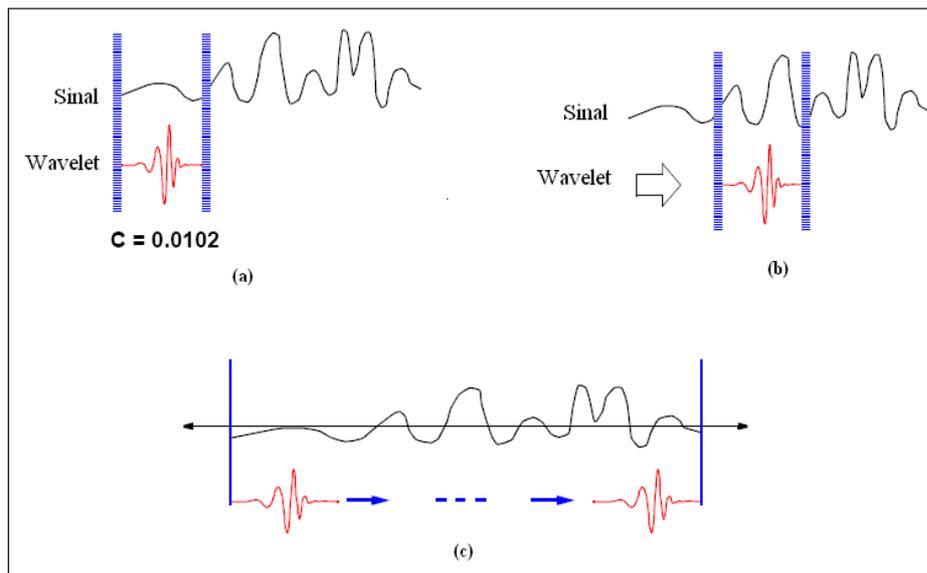


Figura 38: Transformada wavelet.

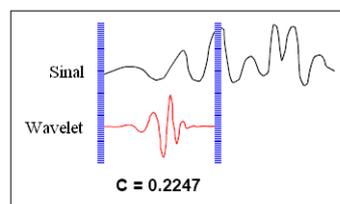


Figura 39: Transformada wavelet com mudança de escala.

Escalas

Para este trabalho foi escolhida como escala a Escala Dyádica ou Escala de Dois, onde $s = 2^j$. Isto se justifica pela simplificação dos cálculos e por não haver perda de informações, já que essa é uma escala redundante e completa, segundo (MALLAT, 1998).

A Figura 40 mostra a transformada wavelet de um sinal de ECG, para diferentes escalas.

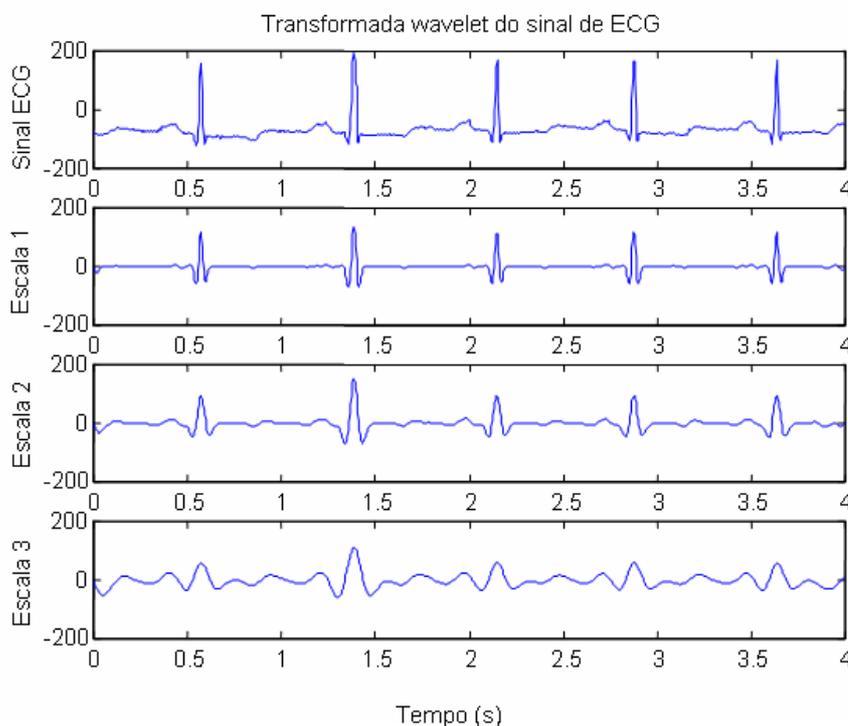


Figura 40: Avaliação do valor da escala na transformada wavelet.

Assim, é possível observar que pequenas escalas estão associadas às componentes de alta frequência do sinal transformado, assim como as escalas maiores refletem as suas componentes de baixa frequência. Para o sinal de ECG, é importante mencionar que quando a escala é muito grande ($j > 5$), a energia do complexo QRS diminui e a energia do artefato de movimento e do ruído aumentam, limitando, então, o tamanho da escala a ser utilizada.

Funções Wavelets

É necessário, então, ter algum critério para a escolha da wavelet a ser usada. Alguns fatores podem ser considerados:

- *Base Ortogonal ou Não-ortogonal:* a base ortogonal é utilizada para processamento de sinais, pois realiza uma representação compacta do sinal. Mas para a análise de série de tempo, um deslocamento não periódico na série de tempo produz um espectro diferente da wavelet. Já a transformada não-ortogonal é altamente redundante nas escalas grandes, onde o espectro da wavelet em momentos adjacentes é altamente correlacionado. É usada, então, para análise de série no tempo, onde variações contínuas e suaves na amplitude da wavelet são esperadas.
- *Complexa ou Real:* a wavelet complexa retorna a informação de amplitude e de fase, aplicando-se à captura do comportamento oscilatório do sinal. Já a função real retorna uma única componente, destinando-se a isolar picos ou descontinuidades.
- *Largura:* uma função wavelet estreita no tempo terá uma boa resolução temporal mas uma pequena resolução na frequência, enquanto que uma função estreita se apresenta com uma resolução pobre temporal e boa resolução na frequência.
- *Forma:* o formato da função wavelet escolhida deve refletir o tipo de características que se deseja extrair do sinal.

Assim é que, depois de estudos e experimentações, a wavelet escolhida para melhor ressaltar as características a serem extraídas do sinal de ECG foi a wavelet Chapéu Mexicano (Mexican Hat), a qual corresponde à segunda derivada da função gaussiana, conforme visto na Figura 41, tendo como expressão matemática

$$\psi(t) = \frac{2}{\sqrt{3}\sigma\pi^{1/4}} \left(1 - \frac{t^2}{\sigma^2}\right) e^{\left(-\frac{t^2}{2\sigma^2}\right)}.$$

No domínio da frequência tem-se

$$\hat{\psi}_s(\omega) = \sqrt{s} \frac{\sqrt{8}\sigma^{5/2}\pi^{1/4}}{\sqrt{3}} (s\omega^2) e^{\left(-\frac{(\sigma s\omega)^2}{2}\right)}.$$

Pode-se observar que as pequenas escalas são associadas a filtros passa-altas, enquanto as grandes escalas correspondem a filtros passa-baixas.

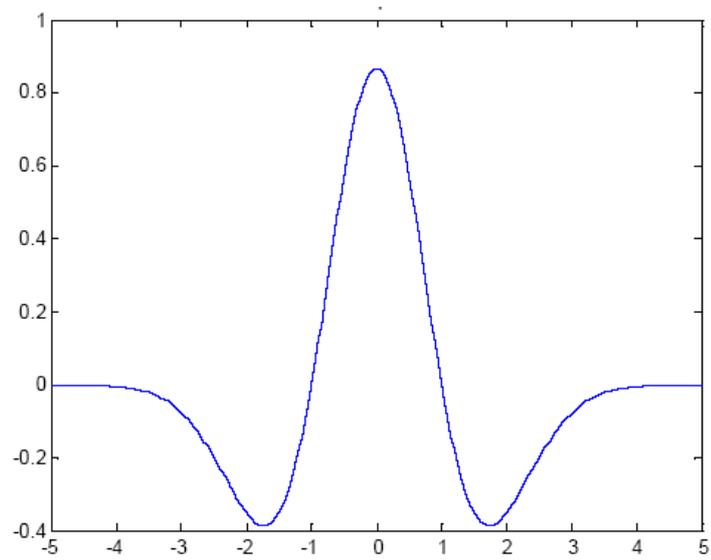


Figura 41: Função Chapéu Mexicano.