

Ana Maria Perfeito Tomé

SISTEMA PARA MICROANÁLISE E CLASSIFICAÇÃO
AUTOMÁTICA DO ELECTROENCEFALOGRAMA DE SONO

Dissertação de Doutoramento apresentada à Universidade de Aveiro

A impressão desta dissertação foi subsidiada pelo INIC

Universidade de Aveiro

1990

AGRADECIMENTO

A todos que de algum modo contribuíram para que pudesse realizar este trabalho o meu agradecimento e gratidão e em especial:

- ao Prof. José Carlos Príncipe que propôs e orientou o trabalho, leu e criticou esta tese, por todo o seu apoio
- ao Dr. António Martins da Silva, do Hospital Geral de Santo António, pela sua contribuição nos aspectos clínicos, sem a qual este trabalho teria sido impossível
- ao Eng^o Ernesto Martins pela sua contribuição para o desenvolvimento do software de sistema e constante disponibilidade para a resolução dos problemas de hardware do HIDRA.
- à Prof^a Beatriz Sousa Santos que leu esta tese, pelas suas sugestões e conselhos
- aos Profs. Francisco Vaz e Guedes de Oliveira pelo apoio na criação das infra-estruturas necessárias à realização deste trabalho.

Ana Maria Perfeito Tomé

INDICE

Agradecimento	i
Índice	iii
Abstract	v
Resumo	vi

Cap 1: Análise Bibliográfica

1	Análise Visual	1
2	Sistemas Automáticos para Classificação do Sono	3
2.1	Análise de Amplitude	4
2.2	Sistemas Híbridos	4
2.3	Análise Espectral	8
2.4	Fuzzy Sets na Classificação de Sono	11
2.5	Reconhecimento de Padrões na Classificação de Sono	12
2.6	Expert System na Classificação de Sono	14
2.7	Autómato Baseado na Teoria da Evidência	18
2.8	Classificação de Sono Baseada em Modelos	19
3	Hipóteses e Objectivos	21
3.1	Organização	23
4	Características Técnicas do Sistema	23

Cap 2: Detecção Automática de Transientes no EEG

1	Introdução	25
2	Definição das actividades básicas	26
3	Princípios básicos do método de detecção	27
4	Filtros na Análise do EEG	29
4.1	Condições de fase linear nos filtros FIR	30
5	Conversão da frequência de amostragem	35
5.1	Sistema linear variante no tempo	36
5.1.1	Decimação	39
5.1.2	Características de frequência dos filtros para decimação	40
6	Aproximação da função de transferência dum FIR	46
6.1	Eliminação de bandas	46
7	Implementação dos filtros FIR	49
7.1	Efeitos do comprimento de palavra finita nos filtros	50
7.1.1	Erros de quantificação no sinal de entrada	51
7.1.2	Erros de quantificação nos coeficientes	53
7.1.3	Erros aritméticos	55
7.1.3.1	Erros de arredondamento	56
7.1.3.2	Scaling	59

Cap 3: Detecção de Transientes no HIDRA

1	Introdução	63
2	Modelo de Processamento	65
2.1	Encadeamento temporal das tarefas de processamento	69
2.2	Organização das tarefas de processamento em cada preprocessor	70
3	Descrição das tarefas de processamento	72
3.1	Detectores	75
3.2	Estatísticas	79
3.3	Comunicação	80
3.4	Estrutura dos fichiros de dados	81
4	Especificação e teste dos filtros	83
4.1	Erros na implementação dos filtros	89

Cap 4: Classificação

1	Introdução	93
2	Redes Neurais	94
2.1	Aprendizagem em Redes Neurais	96
2.1.1	Perceptrão de camada simples	97
2.1.2	Perceptrão multicamada	99
2.2	Redes Neurais em Classificação	104
3	Aplicação das Redes Neurais à Classificação de Sono	107

Cap 5: Resultados

1	Introdução	109
1.1	Recolha de dados	109
2	Detectores	110
2.1	Avaliação global dos detectores	112
2.2	Estudo estatístico das actividades detectadas	115
2.2.1	Actividade Sigma	116
2.2.2	Actividade Delta	118
2.2.3	Conclusões	121
3	Validação do método de classificação	121
3.1	Entradas Binárias	122
3.1.1	Aprendizagem	123
3.1.2	Resultados	125
3.2	Entradas analógicas	128
3.3	Conclusões	134
4	Classificação dos dados obtidos pelo HIDRA	137
4.1	Evolução das actividades ao longo da noite	140

Cap 6: Conclusões 143

Referências	157
-------------	-----

ABSTRACT

Automatic sleep analysis has been a subject of interest since the sixties, mainly because of the growing importance of various sleep disorders. The visual analysis of sleep recordings is a tedious, laborious, time-consuming and expensive task. During a single night recording, 400-500 meters of paper are produced and the visually analysis usually takes several hours of an specialist time.

This work presents an environment where it is possible to do EEG microanalysis as well as sleep staging. On a multiprocessor system a real-time and multirate implementation is done for the detection of EEG and EOG activities. At the end of real-time analysis on EEG/EOG signals we have all the activities described by temporal parameters (such as duration, period and amplitude). A study of off-line parameter statistics is also presented, which proved that drug and age effects can be easily achieved in this environment. These statistics can also be used to validate the detected events to improve the system performance on sleep scoring.

The sleep scoring is done by a feedforward neural network trained with the back-propagation algorithm. The input of the neural network is a feature vector that describes the amount of the detected EEG activities within a minute and where the sleep stage is the output. The performance of the neural network is the degree of agreement with the human scorer (88%). This work also compares the performance of a single and multilayer perceptrons under different training strategies. For this application single and multilayer perceptron performed at some level (specially if both can learn the training set), and the best strategy to train the network is the use of human *apriori* knowledge. The neural network performs at the same level as other, much more difficult to implement, pattern recognition algorithms.

RESUMO

A investigação automática de sono teve início na década de 60, principalmente pelo interesse nas perturbações de sono manifestadas por cada vez maior número de pessoas. A classificação visual de um registo de sono é uma tarefa monótona, morosa e cara. Um registo poligráfico de uma noite de sono tem cerca de 500m de comprimento e, um especialista pode levar algumas horas a analisá-lo.

Este trabalho apresenta um sistema onde é possível a classificação automática em estádios de sono e a microanálise dos eventos detectados. No sistema de multiprocessamento (HIDRA), as actividades do EEG/EOG são detectadas em tempo real utilizando uma filosofia de multifrequências de amostragem. No final da noite, os parâmetros temporais das actividades (duração, amplitude, período) estão armazenados em disco num ficheiro binário. A partir da informação armazenada é também apresentado um estudo estatístico do parâmetros, o que mostra que neste sistema os estudos clínicos podem ser facilmente concretizados. Além disso, as estatísticas foram também utilizadas para melhorar a concordância na classificação.

A classificação em estádios de sono é feita por uma rede neuronal treinada com o algoritmo "back propagation". A entrada da rede é constituída por vectores que descrevem a percentagem das actividades detectadas no minuto e, a saída é o estádio de sono. O grau de concordância com a classificação visual é a medida de "performance" da rede. A concordância é estudada para o perceptrão e redes multicamada considerando estratégias de treino distintas. Nesta aplicação, com o perceptrão e redes multicamada obtêm-se concordâncias idênticas, especialmente se ambas as topologias aprenderem o conjunto de treino. A melhor estratégia para treinar uma rede é utilizar o conhecimento humano. A concordância obtida pelas redes tem valores idênticos a outro métodos de reconhecimento de padrões bastante mais elaborados.

CAP 1: ANÁLISE BIBLIOGRÁFICA

1: Análise Visual

A utilização do EEG (electroencefalograma) nos estudos de sono começou em 1937 [Loomis et al] com a primeira classificação da actividade eléctrica do cérebro humano durante o sono. Esta classificação distinguia pelo menos cinco padrões no traçado poligráfico do EEG, associando a cada um deles um nível de consciência. Em 1953 Aserinsky e Kleitman, com a descoberta dos movimentos oculares (REM) propuseram a divisão do sono em duas fases: Sono REM (associando a esta fase o período de sonho) e Sono NREM. Como consequência foi proposta uma nova classificação baseada na visualização simultânea do Electroencefalograma (EEG) e do Electrooculograma (EOG). Assim, foram propostos 4 estádios de sono durante o período NREM, cuja atribuição era unicamente dependente do sinal EEG, e o estágio REM, determinado pelo aparecimento de grupos de movimentos oculares no canal EOG. Os critérios para esta classificação revelaram-se pouco precisos, conseqüentemente reuniu-se um comité internacional cujas conclusões [Rechtschaffen and Kales-1968] constituem hoje a classificação de sono mais divulgada. Para esta classificação, e com o polígrafo à velocidade de 15mm/s ou 10mm/s, devem registar-se simultaneamente:

- 1 canal de EEG (C3- A2), podem utilizar-se outros canais desde que aquele seja incluído.

- 2 canais oculares, para que seja possível a separação dos movimentos oculares dos movimentos lentos de grande amplitude nos canais frontais do EEG.

- 1 canal de EMG (Electromiograma) é fortemente recomendado, por ajudar a corroborar o estágio REM.

A classificação é feita por épocas de 30s ou 20s (dependendo da velocidade do papel), nos estádios seguintes:

- Estádio 0 (acordado): actividade alfa e/ou EEG de amplitude baixa e frequência variada. Neste estágio, algumas vezes, o EMG é de grande amplitude e existem movimentos oculares.

- Estádio 1: EEG de amplitude relativamente baixa e frequência variada mas com predominância na gama 2-7 Hz. Este estágio ocorre normalmente na transição do estágio 0 para outros estádios, ou a seguir a movimentos do corpo durante o sono. Neste caso ocorre por períodos curtos: de 1 a 7 minutos. Pode ainda acrescentar-se que este estágio é também caracterizado pela ausência de complexos K e fusos de sono (actividade sigma).

- Estádio 2: Aparecimento dos fusos de sono e complexos K e a ausência de actividade lenta que determine os estádios 3/4. Sendo os fusos de sono e os complexos K actividades isoladas, pode haver períodos sem o aparecimento destes eventos que não justifiquem uma mudança de estágio. Se esses períodos forem inferiores a 3 minutos e o EEG tiver características de estágio 1, estas épocas devem ser classificadas como estágio 2 se não houver aumento de amplitude no EMG nem movimentos "arousal". Se alguma destas modificações ocorrer as épocas devem ser classificadas como estágio 2 até esse instante e como estágio 1 até voltar a surgir um complexo K ou fuso de sono.

- Estádio 3: Existência de pelo menos 20%, mas inferior a 50%, de actividade lenta (2 Hz ou inferior) de elevada amplitude no canal EEG. Fusos de sono podem, ou não, existir neste estágio.

- Estádio 4: Mais de 50% de actividade lenta (actividade delta) de grande amplitude no canal EEG. Fusos de sono podem, ou não, aparecer neste estágio.

- Estádio 5 (REM): este estágio é caracterizado por EEG de baixa amplitude e frequência variada, e movimentos oculares no canal EOG. A amplitude do EMG é baixa, isto é, pelo menos mais baixa que a do estágio anterior. Como os REM aparecem em grupos pode haver épocas em que o sinal EEG tenha as características referidas, mas não existam movimentos oculares rápidos (REM). Por isso existem recomendações adicionais para a classificação deste estágio. Em geral, pode dizer-se que todas as épocas, com EEG de amplitude baixa e frequência variada, contíguas a épocas em que ocorrem grupos de REM devem ser classificadas como estágio REM. No entanto, o nível de tónus muscular (EMG) pode ajudar a corroborar este estágio, pois a amplitude daquele sinal é mais baixa durante este estágio que nos outros. Assim, a verificação de transição da amplitude deste sinal pode ser utilizada para marcar o início/fim do estágio 5.

2: Sistemas Automáticos para Classificação de Sono

Durante uma noite de sono é gerada uma grande quantidade de dados (um registo com cerca de 500m de papel). Por isso, a análise visual dum registo é uma tarefa morosa e cansativa para os especialistas, além de que é impossível para o especialista manter os critérios de análise ao longo de todo o registo. Deste modo, e praticamente desde o início da investigação em sono, apareceram sistemas que propõem a análise automática do sono quantificando os sinais EEG, EOG e EMG, isto é, sistemas cujos critérios de decisão se baseiam em medidas efectuadas no próprio sinal e que portanto não têm o carácter subjectivo da análise visual. Por

outro lado, o especialista pode ter acesso a um conjunto de informação que complementa o seu conhecimento sobre os sinais em estudo. E neste campo apareceram uma gama variada de estudos, embora com objectivos diversos: descrever as formas de onda que caracterizam os diversos estádios de sono, estabelecer a natureza cíclica de algumas actividades, imitar o especialista na classificação dos registos em estádios de sono e ainda determinar o efeito de drogas, da idade e doenças nas características do sono.

2.1: Análise de Amplitude

A análise de amplitude do EEG foi o primeiro método proposto para a classificação do sono [Agnew-1967]. Neste trabalho os autores salientaram a relação entre o integral do valor absoluto da amplitude do EEG e o estágio de sono. Estes autores sugeriram que esta medida de amplitude no EEG conjuntamente com a detecção dos REM e da actividade muscular poderiam constituir um método objectivo de classificação de sono. Em trabalho posterior, o mesmo autor [1973] concluiu que o valor integrado da amplitude (designado por nível de electrogénese) não é suficiente para diferenciar os estádios NREM de sono. Este estudo apresenta os valores integrados de amplitude do EEG de sono de adultos jovens. O índice de amplitude calculado para cada minuto do EEG é consistente de noite para noite em cada indivíduo, mas não o é entre indivíduos.

2.2: Sistemas Híbridos

Os sistemas híbridos foram os primeiros a ser utilizados com êxito em estudos de classificação de sono. Estes sistemas são constituídos por uma parte analógica e outra digital. Na parte analógica é vulgarmente implementada a detecção

dos eventos (actividades) considerados relevantes para a classificação do sono, conseguindo-se deste modo uma grande redução no volume de dados; na parte digital, e a partir dos dados armazenados, é feita a classificação em estádios de sono.

A parte analógica do sistema desenvolvido por Gaillard [1972/1973] é constituída por um conjunto de 12 filtros passa banda, tipo Butterworth, em cada uma das bandas do EEG (na gama 0 a 87,5 Hz), cuja saída rectificada e integrada é a entrada de um discriminador. O sistema inclui ainda um discriminador de declive do EEG, um detector de EOG e um detector de amplitude do EMG. Com este sistema são analisados 2 canais de EEG (frontal e central), 1 canal ocular (EOG) e 1 canal de EMG. A parte digital amostra as saídas dos discriminadores à frequência de 6.25 Hz, ficando um bit duma palavra de 16 com o valor 0 ou 1, conforme aquele sinal de saída atinge um valor superior ou inferior a um nível predeterminado. Resumindo, a presença ou ausência de determinada actividade considerada relevante nos sinais de entrada durante aquele período fica codificada numa palavra de 16 bits. Esta informação é analisada em conjunto antes da classificação em estádios de sono, para eliminar alguns falsos positivos. Por exemplo, os movimentos oculares lentos são detectados no canal frontal do EEG como ondas delta, por isso só é considerada actividade delta se esta aparecer conjuntamente nos canais frontal e central do EEG. Outro problema surge com a detecção dos fusos de sono mas, neste caso, por os filtros serem sensíveis a actividades de frequência elevada (como é o caso dos artefactos musculares e das ondas beta), assim a informação de fusos de sono é ignorada se fôr simultânea com aquelas actividades. Após estas correcções são determinadas as actividades básicas para cada época: alfa, fusos, complexos K, ondas triangulares, ondas delta. A partir das actividades básicas para o minuto é feita classificação em estádios de sono implementando um

conjunto de regras inspiradas no manual de Rechtschaffen e Kales. Começa por testar-se o estágio 0, seguidamente os estádios 3/4 e, finalmente, os estádios 2, 1/REM. O estágio 0 é determinado pela quantidade de alfa e nível de tónus muscular; os estádios 3 e 4 dependem da quantidade de delta; a ausência de fusos de sono, complexos K, delta e tónus muscular compatível com sono REM determina o estágio REM (note-se que a informação do canal EOG não é utilizada); fusos e complexos K caracterizam o estágio 2; finalmente, com actividade muscular com amplitude acima do nível de estágio REM, é classificado o estágio 1. Este sistema foi utilizado em estudos de sono num grupo de 10 indivíduos com idades compreendidas entre 23 e 47 anos onde os autores obtiveram 83.7% de concordância com a análise visual. As maiores discrepâncias surgiram na classificação dos estádios 3 e 4. Este sistema foi utilizado pelos autores em estudos de influências de drogas na estrutura do sono [1973,1975].

Outro sistema híbrido (SAHC) foi desenvolvido por Smith e seus colaboradores [Karacan-1971, Su -1974, Smith-1975]. O sistema, na sua primeira versão, era constituído por vários detectores analógicos ligados a um PDP-8. Com este sistema são analisados 3 canais de EEG e 2 de EOG, sendo detectadas as actividades delta, beta, alfa, fusos de sono, artefactos musculares e REM. Cada detector é composto por um bloco de filtragem passa-banda (na zona de frequência da actividade a detectar), discriminadores de período e amplitude e um bloco de reconhecimento de padrões. Esta metodologia é idêntica à proposta neste estudo por isso falaremos nela com mais detalhe no próximo capítulo. Mas em resumo, pode dizer-se que o período e amplitude duma onda são medidos com o fim de verificar se correspondem aos valores pretendidos para determinado tipo de actividade; no entanto a referida actividade só é considerada relevante se tiver uma duração mínima, pelo que se contam o número de ondas. O detector de REM utiliza

2 canais EOG para eliminar a influência da actividade lenta dos canais frontais de EEG no canal ocular. As regras de classificação do sono em estádios são ligeiramente diferentes das contidas no manual de Rechtschaffen e Kales. Por exemplo, o estágio 0 é determinado não só pela duração da actividade alfa como pela duração da actividade muscular. Além disso, a actividade beta é utilizada para separar o estágio 2 dos estádios 1/5 e acordado. Outra alteração proposta consiste em apenas considerar o estágio REM após 45 minutos de registo. Este sistema foi aplicado em estudos de sono nos 5 grupos etários, 5 indivíduos (2 noites de sono) de cada grupo. A percentagem média de concordância foi da ordem dos 80%; com o valor mínimo de 49.9% no grupo etário dos 3-5 anos e máximo de 83,4% no grupo dos 24-35 anos. Concluindo os autores que os valores de percentagem de concordância eram aceitáveis em todos os grupos etários excepto no 1º grupo, o que sugere que as regras de classificação devem ser alteradas nesta faixa etária. Este sistema foi também utilizado em estudos sobre o efeito de drogas no sono [Smith-1972/1976].

Um sistema com a mesma metodologia foi implementado por Saariki [1980] e utilizado em estudos de sono normal e anormal (indivíduos alcoólicos) por Hasan [1983]. As percentagens de concordância obtidas são idênticas às obtidas pelo SAHC. Mas este sistema utiliza mais fontes de informação: a análise do canal EMG, e inclui a actividade teta no algoritmo de classificação do sono em estádios (distinção do estágio 0 do estágio 1). Estes autores consideram a actividade teta essencial para a separação do estágio acordado dos outros estádios. Outra alteração proposta é considerarem a classificação dum época dependente da classificação da época anterior.

Recentemente surgiu uma implementação digital do SAHC [Príncipe e Smith-1986]. Os autores utilizaram um Texas TI-9900, em que desenvolveram algoritmos de processamento digital, amostrando os sinais com frequências diferentes. Além disso, inspirados na metodologia seguida pelo SAHC, implementaram os detectores desenvolvendo técnicas de "multitasking" e subrotinas reentrantes. Naturalmente que os filtros implementados são digitais e escolhem as frequências de amostragem de acordo com os erros de medida considerados razoáveis na determinação dos períodos das onda.

2.3: Análise espectral

Em 1969 Johnson [1977] fez um dos primeiros estudos utilizando análise espectral em segmentos de 1 minuto dum canal de EEG. O estudo foi efectuado num grupo de 13 jovens adultos divididos em 2 grupos: um com actividade alfa elevada e outro baixa. Os autores concluíram que a frequência da actividade delta está sempre presente em todos os estádios, embora apresente um pico mais acentuado nos estádios 3 e 4. Os fusos de sono aparecem unicamente nos estádios 2, 3 e 4. Nos indivíduos de actividade alfa elevada aparece um pico na zona de frequência daquela actividade em sono REM embora seja a uma frequência inferior (0.5 a 2 Hz) ao pico existente no estádio 0. Nos indivíduos de baixa actividade alfa não se nota nenhuma diferença significativa no espectro dos estádios 1 e 0.

Com estes dados Lubin [1969] fez uma tentativa de classificação automática do sono em estádios utilizando análise de regressão múltipla. Este estudo foi retomado por Larsen [1970] efectuando a classificação dos mesmos espectros mas com discriminadores não lineares. A percentagem de concordância foi de 85%. Concluindo estes autores que, com a análise espectral, se pode obter um conjunto

de parâmetros que permitem a classificação do sono em estádios com a possível excepção da separação dos estádios 1/5. Este resultado é bastante contestado por Smith [1978] por se tratar duma população seleccionada e com segmentos de EEG também previamente escolhidos.

No entanto a análise espectral tem sido utilizada para apresentar duma forma gráfica e comprimida a evolução duma noite de sono -CSA ("compressed spectral array") [Bickford et al-1973]. Calcula-se a FFT de 8s de EEG e os espectros são sucessivamente representados graficamente suprimindo a "hidden line". Ao lado e em colunas sucessivas são também representados o integral de amplitude do EEG, EMG e EOG. Esta técnica pode dar rapidamente a ideia da evolução do sono ao longo da noite. No entanto não se podem fazer estudos estatísticos sobre as actividades pois não resulta nenhuma informação quantitativa.

Recentemente [1988] Solinsky et al desenvolveram uma técnica idêntica que designaram por CDSA ("color density spectral array") pois neste caso a potência espectral é codificada com um código de cores. Estes autores fizeram um estudo comparativo entre a classificação visual a partir dos registos poligráficos e dos registos do CDSA. A classificação dos CDSA era feita por um especialista treinado na sua leitura e que não tinha qualquer conhecimento dos registos poligráficos. A percentagem de concordância variou entre 85-92%, mas para a classificação dos CDSA os autores introduziram algumas modificações nas regras de Rechtschaffen para a distinção entre os estádios 2, 3 e 4. Isto porque a distinção entre estes estádios é feita à custa da percentagem de actividade delta presente na época e esta informação não está disponível numa FFT. Por estas razões os autores estudaram a potência total na zona de frequência do delta em épocas de 30 s (note-se que a FFT

é calculada em segmentos de 5s) num conjunto de registos previamente estudados calculando limiares que utilizaram para este estudo.

Paralelamente, Stanus [1987] e seus colaboradores apresentaram 2 técnicas de classificação do sono, baseadas em parâmetros espectrais que designaram por determinística e estocástica. Estas técnicas aplicadas em 30 indivíduos têm 75% de percentagem de concordância com a análise visual. São analisados simultaneamente 3 canais de EEG, 2 de EOG, 1 de EMG. Os sinais de EEG e EOG são filtrados analógicamente e digitalizados à frequência de 50Hz. O método determinístico é baseado na detecção das actividades do EEG, EOG e EMG. Neste método e para o sinal EEG, é feita uma segmentação do sinal em janelas de 2.5s com sobreposição de 1.25, sendo o sinal modelado por um filtro autoregressivo de 4º ordem. No caso do EEG a localização dos polos do filtro no plano z são relacionadas com a actividade de EEG presente na janela, isto é, a existência duma actividade pressupõe a existência de um par de polos com módulo superior a 0.8 e situados na zona de frequência da referida actividade. No caso do sinal EOG é calculado um coeficiente de correlação entre os dois canais, sempre que seja inferior a um limiar existem movimentos oculares. O EMG é amostrado a muito baixa frequência (10Hz), e o desvio padrão da sequência aleatória obtida é considerado proporcional à envolvente do sinal. Estes resultados são associados para cada época de 20s num vector. O sistema de classificação dos vectores é uma árvore lógica construída a partir das regras de Rechtschaffen and Kales. No método estocástico, a partir do cálculo da FFT de janelas de 20s do sinal é calculada a densidade espectral para cada uma das bandas das actividades do EEG; em seguida é calculado um valor relativo: aquela densidade a dividir pela densidade total. Os resultados associados num vector são classificados pela fórmula de Bayes: a probabilidade *a posteriori* é calculada a partir da probabilidade *a priori* e condicional dos estádios. É utilizado

um modelo Gaussiano paramétrico para o cálculo das funções densidade de probabilidade condicionais cuja estimativa é baseada em épocas classificadas visualmente. Em seguida, é feita outra classificação adicionando a informação do EOG e EMG. Neste caso a classificação é também baseada numa árvore lógica onde são analisados 5 minutos de informação para a classificação de 20s (época situada a meio da janela de 5min.).

2.4: Fuzzy Sets na Classificação de Sono

Considerando a classificação de sono como um problema de clustering em cinco zonas determinadas pelas características de amplitude e frequência do EEG, Kumar [1977] desenvolveu um sistema analógico. O autor considera que os limites entre "clusters" não são bem definidos, sendo possível sobreposição. Por isso, considera a utilização da teoria dos "fuzzy sets" vantajosa porque além de poder tratar o problema da sobreposição entre clusters, também dá a possibilidade de, na fase de classificação, aplicar termos usados na análise visual (por exemplo: frequência misturada, não elevada, muito, etc.). No sistema analógico é feita a separação da amplitude do EEG em 4 níveis com comparadores com histerese, a fim de permitir alguma sobreposição. Um aspecto interessante neste codificador de amplitude é que o número de impulsos e tempo entre impulsos representa a frequência. O estágio de sono é determinado pela integração com uma constante de tempo de 30s daqueles impulsos. A histerese e limiares de comparação nos comparadores são ajustados de modo a conseguir a sobreposição desejada nos "clusters" no subespaço de frequências. Este sistema não permite a separação estágio 1 do estágio 5, uma vez que só analisa um canal de EEG. Para um conjunto de 2 noites de dados foi obtido 90% de percentagem de concordância com a análise visual de 2 especialistas.

2.5: Reconhecimento de Padrões na Classificação de Sono.

Martin [1972] fez um estudo de classificação de sono utilizando conjuntamente técnicas de análise espectral (FFT) e de reconhecimento de padrões. Os sinais EEG (canal central) e EOG depois de uma pré-filtragem eram digitalizados à frequência de 68.5Hz. No canal EOG são detectados os REM; no sinal EEG é calculada a FFT para segmentos de 30s, com a resolução de 1Hz. Além disso é calculada a percentagem de actividade delta em cada época (30s) com um algoritmo que detecta a amplitude entre picos e períodos das ondas. Esta medida temporal da actividade delta foi incluída porque os autores não conseguiram distinguir os estádios 3 /4 dos outros estádios com a FFT dos segmentos. Assim, começa-se por classificar os estádios 3 e 4 com a informação dos detectores delta; em seguida com a informação das 34 riscas espectrais, e utilizando o conjunto de programas DAID [Discriminant Analysis Iterative Design-Viglione 1970], classificam as épocas em estádios 1/5, 2 ou acordado; a informação do canal ocular é utilizada para separar os estádios 1 e 5. Como resultados os autores salientam 80% de concordância com a análise visual, em 5 indivíduos com idades compreendidas entre 17-21 anos, além disso analisaram as percentagens de concordância para cada um dos estádios, chegando à conclusão que obtinham os piores valores para os estádios 1 e 3. Para minorar estes erros os autores propõem a análise da actividade alfa no canal occipital, por ser mais visível neste canal e a detecção dos complexos K uma vez que neste sistema esta actividade é caracterizada como delta, e pelo que alguns estádios 2 são muitas vezes classificados como 3 (sobretudo em épocas de transição 2-3).

função discriminante linear. Variando a ordem do modelo autoregressivo, entre 5 e 11 e com segmentos com duração de 5s, obtêm percentagens de concordância da ordem dos 85% na classificação de 323 segmentos diferentes dos contidos no conjunto de aprendizagem. Nesta fase de classificação os autores salientaram os erros obtidos na classificação dos complexos K, que são muitas vezes classificados como delta se no segmento existirem 2 complexos K, e ainda na classificação dos fusos de sono que, muitas vezes são classificados como actividade de fundo. Nos canais de EMG e EOG efectua medidas temporais; no sinal EMG calculam o somatório do quadrado das amplitudes, e salientam que este parâmetro tem valores característicos dependentes do estágio de sono: muito baixo nos estádios 0 e REM e elevado nos outros estádios; para o sinal EOG é calculada a diferença de primeira ordem na vizinhança dum pico, apresentando este parâmetro valores característicos conforme o estágio de sono: valor muito elevado nos estádios 1, 2, 3, 4, pequeno no estágio 0 e intermédio (entre os outros valores) no estágio REM. Finalmente, a classificação em estádios de sono proposta pelos autores consiste em reunir a informação obtida na análise dos sinais EEG, EOG e EMG do seguinte modo: calcular a percentagem de cada padrão de EEG presente em 20s, em seguida, e conjuntamente com as medidas propostas para o EOG e EMG aplicar um conjunto de regras idênticas às contidas no Rechtschaffen e Kales. A base de dados experimental foi bastante restrita: 202 minutos. Com estes dados obtiveram percentagens de concordância que variam entre 71.3% e 75.4% não separando o estágio 3 do 4.

2.6: Expert System na Classificação de Sono

Para cada 30s do sinal é definido um vector com 13 parâmetros característicos relacionados com: detecção das actividades básicas do EEG, uma

medida da actividade de fundo, número de picos por segundo do EMG, e as detecções REM nos canais oculares. A partir destes vectores Ray et al [1986] implementaram um "expert system" para a classificação do sono. As actividades teta, alfa, sigma e actividade de fundo (característica do estágio 1) são detectadas calculando uma medida de correlação para cada intervalo de 0.6s. Este valor é o integral de sinal EEG multiplicado pela exponencial complexa com frequência igual aos limites das gamas de detecção daquelas actividades (5-7, 9-11, 13-15). Estes valores são normalizados e os limiares de cada uma das gamas são somados e comparados com um nível predefinido, se algum deles for superior, o intervalo fica atribuído à respectiva actividade. Caso não tenha sido atribuído a nenhuma das actividades, fica caracterizado como actividade de fundo. A actividade REM é detectada, a partir de uma medida adaptativa (hn) que calcula o valor do 90% dos picos do EOG, assim um pico será um REM se tiver 2 vezes o valor hn. A actividade delta é detectada com um algoritmo de picos e passagens por zero. Um conjunto de vectores característicos de cada estágio de sono é utilizado para gerar um conjunto de regras de lógica multivariável, as regras que os autores designam por "context-free". Além destas regras os autores propuseram um outro grupo de regras designadas por "context", escritas pelo especialista, e que contêm directivas do tipo: se entre épocas de estágio 5, houver 2 épocas de estágio 1, estas deverão ser classificadas como estágio 5. Assim, a cada vector de dados são aplicadas o primeiro grupo de regras, ficando deste modo atribuído à época um estágio de sono ou então assinalando diagnóstico impossível. Após uma ou mais épocas sem classificação é invocado o segundo tipo de regras. Para este estudo foram gravados 2 canais de EEG (central e occipital), 1 EOG e 1 EMG de 7 indivíduos com idades compreendidas entre os 17-21 anos. Todos os canais foram digitalizados à frequência de 100Hz. Os autores utilizaram 2 métodos diferentes para gerar as regras "context-free": 7 conjuntos de regras foram gerados com 50% dos dados,

escolhidos aleatoriamente, de cada estágio para cada um dos registros; outro grupo de regras foi gerado por escolha aleatória de dados de todos os registros. Em seguida foram classificados todos os registros com todos os conjuntos de regras. O grau de concordância é maior quando o registro é classificado com as regras respectivas (geradas a partir desse mesmo registro), variando entre 90-95%. A percentagem baixa consideravelmente, para 80-87%, com o último conjunto de regras. Os autores consideram que a causa desta descida no grau de concordância se justifica porque existe sobreposição entre regras e além disso devido à variabilidade de dados entre os sujeitos que o especialista compensa, isto é, há indivíduos com pequena quantidade de actividade alfa e outros com grande quantidade; outros podem ter actividade alfa durante o sono delta; etc. Por estas razões os autores propõem o agrupamento dos registros por tipos de modo a que as regras geradas para cada uma delas sejam discriminativas.

Chang et al [1988] desenvolveram um "expert system" cuja estrutura hierárquica de manipulação da informação reflecte o conhecimento humano dos sinais em escalas diferentes (a forma de onda, o minuto, e o contexto). Assim um "front-end" (baseado no SAMICOS) transfere informação das ondas detectadas nos sinais EEG e EOG (8 bytes/s) ao "host", que armazena e atribui coeficientes de confiança aos dados recebidos. Na segunda camada, e para cada minuto de informação, é feita a classificação em estágios de sono. Finalmente, na terceira camada, esta classificação é revista atendendo à informação disponível em todas as camadas e à informação contextual criada por uma janela deslizante de 5 minutos. Os autores referem que as regras de classificação são, em qualquer das camadas, baseadas nos critérios tradicionais de classificação do sono em estágios (regras de Rechtschaffen e Kales), embora, o sistema possa ser facilmente modificado revendo ou acrescentando regras. O grau de concordância médio é de 83.6% , ou

87.6% se juntar os estádios lentos de sono, para uma gama de casos com idades compreendidas entre 5 e 70 anos.

Jansen [1989] propõe-se implementar um sistema baseado em programação por objectos. Assim, todos os eventos básicos e situações (os objectos) devem ser definidos como entidades de determinado tipo caracterizadas por um conjunto de atributos, onde cada atributo está relacionado com a morfologia e relações espaço-temporais do objecto em causa. Para gerir a informação, o sistema tem 5 componentes básicos: 1) a estrutura de dados central que controla o fluxo de informação; 2) o conjunto de descrição dos objectos; 3) o especialista que opera sobre os objectos detectando eventos ou características dos mesmos; 4) um gestor e ordenador de tarefas; 4) o módulo de detecção que a partir da descrição do evento acciona o especialista apropriado. Esta filosofia foi unicamente testada na detecção de fusos de sono e de complexos K. Por exemplo, para a detecção dos fusos criam-se 2 objectos e 5 especialistas. Um dos objectos tem como atributos: "burst" de frequência (um especialista que detecta um evento com essa característica); ocorrência dominante (canais centrais); frequência - amplitude (o especialista que verifica os critérios de duração e amplitude que caracterizam o fuso). Se qualquer dos especialistas não rejeitar ou aceitar claramente o evento, desencadeia outro processo de detecção noutra canal invocando outro objecto que basicamente tem os mesmos atributos do objecto descrito. Os autores referem que esta filosofia pode ser aplicada à classificação do sono em estádios, em que a partir da análise de amplitude e frequência da actividade de fundo, pode ser formulada uma hipótese dum estádio. Em seguida seriam desencadeados um conjunto de operações no sinal procurando eventos específicos que confirmem ou rejeitem a hipótese formulada. O processamento no sinal é por isso "goal-driven" pois só se faz um determinado

processamento nos sinais de entrada quando existe um objectivo em vista: a verificação ou confirmação duma hipótese.

2.7: Autómato Baseado na Teoria da Evidência

Em 1988 [Principe et al] apresentaram um método de classificação em estádios de sono baseado na teoria da evidência matemática. Este método pressupõe que cada época (minuto) esteja descrita num vector que contém a quantidade (em segundos ou número de ondas) de cada actividade na referida época. Assim, utilizam como "front-end" o SAMICOS que faz a detecção das actividades básicas do EEG/EOG. O sistema de classificação é considerado como uma máquina de estados finitos (em que o estado inicial é o estádio 0), dependendo a transição ou não-transição de estado da combinação das ondas detectadas. Além disso os autores consideram que há transições impossíveis, como por exemplo a passagem directa do estádio 0 aos estádios 3 e 4. Na implementação do sistema distinguem-se três fases. Na primeira um conjunto de regras heurísticas, designadas funções de fuzificação, transformam os dados de entrada em graus de confiança parciais numa determinada transição. Em seguida, são aplicadas um conjunto de regras para cada transição possível, calculando-se o grau de evidência final nos estádios possíveis com a teoria da evidência. O estádio escolhido será aquele que apresenta grau de evidência superior. Finalmente, para a terceira fase, criaram-se um conjunto de regras que relacionam a informação que transcende o minuto: por exemplo, períodos de estádio 1 inferiores a 3 minutos contíguos a estádio 5 devem ser classificados como estádio 5. Outro aspecto que estas regras contemplam é a reclassificação de estádios com valores de evidência baixos (<0.4), considerando que isso acontece por dois motivos: falta de informação nos vectores de dados ou informação contraditória. Para implementar as regras de fuzificação os autores

consideram, tal como está implícito no manual do Rechtschaffen e Kales, que a cada vector de dados está associado (ou existe a evidência) determinado estágio. Dividindo a presença das ondas em categorias (não existente, baixa, média, alta), a cada uma destas pode estar associado determinado estágio. Para quantificar esta informação, os autores utilizam fundamentalmente 2 tipos de funções: escada e adaptativa (em que os limiares dos intervalos são calculados para cada indivíduo). Deste modo, o número de segundos (ou número de ondas) de determinada actividade é transformando no grau de evidência parcial que determinada transição (estágio) ocorra. Verificando-se assim que a fuzificação do alfa e artefactos musculares determinam o estágio 0, os fusos de sono o estágio 2, as ondas delta os estádios 3/4, e o REM e as ondas beta os estádios 1/5. Este algoritmo foi testado com 5 registos, obtendo-se 84% de grau de concordância com a análise visual.

2.8: Classificação de Sono baseada em Modelos

Kemp et al [1986] utilizaram um modelo baseado em cadeias de Markov para gerar estádios de sono. As probabilidades de transição entre estádios foram calculadas para 95 hipnogramas de 23 indivíduos. Os autores referem que os parâmetros calculados (probabilidades de transição variáveis no tempo) caracterizam o mecanismo do sono, pois reflectem as periodicidades conhecidas no fenómeno (alternância REM-NREM); diminuição do sono lento ao longo da noite, assim como a variabilidade inter-indivíduo. Os autores referem que as simulações geram hipnogramas que não se distinguem dos reais. Kemp et al [1985,1986,1987] propõem também modelos estocásticos para monitorização de estádios de sono e respectivas actividades básicas. Novamente, a transição de estágio é descrita por uma cadeia de Markov, mas o modelo previamente descrito é assimétrico, isto é, há estádios que ocorrem mais frequentemente e tomam maior

consideram, tal como está implícito no manual do Rechtschaffen e Kales, que a cada vector de dados está associado (ou existe a evidência) determinado estágio. Dividindo a presença das ondas em categorias (não existente, baixa, média, alta), a cada uma destas pode estar associado determinado estágio. Para quantificar esta informação, os autores utilizam fundamentalmente 2 tipos de funções: escada e adaptativa (em que os limites dos intervalos são calculados para cada indivíduo). Deste modo, o número de segundos (ou número de ondas) de determinada actividade é transformando no grau de evidência parcial que determinada transição (estágio) ocorra. Verificando-se assim que a fuzificação do alfa e artefactos musculares determinam o estágio 0, os fusos de sono o estágio 2, as ondas delta o estádios 3/4, e o REM e as ondas beta os estádios 1/5. Este algoritmo foi testado com 5 registos, obtendo-se 84% de grau de concordância com a análise visual.

2.8: Classificação de Sono baseada em Modelos

Kemp et al [1986] utilizaram um modelo baseado em cadeias de Markov para gerar estádios de sono. As probabilidades de transição entre estádios foram calculadas para 95 hipnogramas de 23 indivíduos. Os autores referem que os parâmetros calculados (probabilidades de transição variáveis no tempo) caracterizam o mecanismo do sono, pois reflectem as periodicidades conhecidas no fenómeno (alternância REM-NREM); diminuição do sono lento ao longo da noite, assim como a variabilidade inter-indivíduo. Os autores referem que as simulações geram hipnogramas que não se distinguem dos reais. Kemp et al [1985,1986,1987] propõem também modelos estocásticos para monitorização de estádios de sono e respectivas actividades básicas. Novamente, a transição de estágio é descrita por uma cadeia de Markov, mas o modelo previamente descrito é assimétrico, isto é, há estádios que ocorrem mais frequentemente e tomam maior

percentagem de tempo, e como estas características são variáveis de indivíduo para indivíduo, os autores propõem uma simplificação no referido modelo. Portanto sugerem um modelo simétrico e homogéneo, isto é, com probabilidade de transição entre estádios (*a priori*) contante e igual para todos eles, formulando em seguida o problema de decisão como o cálculo duma função densidade de probabilidade *a posteriori* (fórmula de Bayes) a partir das probabilidades condicionais de cada estádio dados um conjunto de observações. O conjunto de observações é constituído pelas actividades alfa, sigma e delta do EEG, REM e SEM (movimentos oculares lentos) do EOG e actividade muscular do EMG. Sendo os sinais EEG, EOG e EMG modelados respectivamente por processos de Wiener, Poisson e Gaussiano. O grau de concordância médio em 6 indivíduos é de 70.6% que os autores referem como 4.5% inferior à concordância média entre 6 especialistas, referindo ainda que a discordância é maior na distinção do estádios 0 e 1, se juntarem estes dois o grau de concordância sobe para 81.8%.

Rosa [1989] sugere também um modelo estocástico do EEG de sono. O modelo é composto por 5 malhas de retroacção em paralelo, duas entradas (uma de ruído gaussiano e outra de impulsos aleatórios). Cada ramo de retroacção está relacionado com uma actividade básica do EEG (alfa, beta, teta, sigma e teta) e é constituído por um filtro passa-banda e um bloco de ganho variável, em série. Por razões computacionais o modelo é discretizado e simplificado (retirando, os ramos de teta, alfa e beta). A simulação, com ganhos variáveis na malha de retroacção e impulsos na entrada de amplitude e durações variáveis, gera sinais de EEG que se assemelham aos sinais reais. A partir deste modelo são dimensionados estimadores de actividade tónica e detectores de actividade fásica. Estudam em especial a detecção de complexos K e ondas Vertex. Na persepectiva do modelo de detecção as actividades fásicas correspondem a uma mudança brusca do processo de

excitação. No caso dos complexos K e ondas Vertex a perturbação consiste numa variação brusca do ruído na forma dum impulso.

3: Hipóteses e Objectivos

Como vimos os métodos de classificação automática do sono em estádios são baseados na hipótese de que a informação contida nos sinais EEG, EOG, EMG (útil na classificação visual) é parametrizável e classificável; concluindo desse modo que cada época pode ser descrita por um conjunto de parâmetros que poderemos considerar coordenadas dum espaço multidimensional. Pressupondo ainda que épocas semelhantes do ponto de vista da análise visual serão pontos muito próximos nesse espaço, podemos considerar cada estádio como uma classe em que todas as épocas do mesmo estádio se aglomeram. Neste caso de classificação do sono, temos portanto um conjunto de classes conhecido *a priori* - os estádios de sono. Como foi notório nesta revisão bibliográfica, e como é característica de qualquer problema de classificação, temos duas fases distintas no processo: a extracção dos parâmetros que representam a informação importante (do ponto de vista da análise visual) contida numa época e a classificação.

No Departamento de Electrónica e Telecomunicações foi implementado um sistema de multiprocessamento [Guedes de Oliveira et al-1988, Cruz-1990] que permite desenvolver esta análise de processamento hierárquico. Neste sistema, a cada um dos preprocessadores pode ficar atribuída uma tarefa (por exemplo, a detecção dos eventos básicos de EEG ou EOG), enviando toda a informação recolhida ao processador central que a armazena e classifica. Portanto, partindo da hipótese que os sinais são parametrizáveis, na primeira fase do processo propomos que sejam implementados algoritmos em tempo real de modo a conseguir uma

grande redução no volume de dados a armazenar. No entanto, sendo os preprocessadores constituídos por processadores comuns (sem qualquer unidade de cálculo aritmético auxiliar) os algoritmos devem implementar definições simples. Referimo-nos por isso a algoritmos ditos do domínio dos tempos, isto é, o método de extracção baseado em características temporais das actividades. Propomos ainda que os parâmetros característicos das actividades sejam também armazenados. Assim, ao contrário dos sistemas descritos que unicamente armazenam a informação necessária à classificação em estádios de sono (percentagem das actividades na época), nós propomos que as medidas temporais, amplitude e período, das actividades (impossíveis de obter por análise visual) sejam também armazenadas. Com esta informação pretendemos efectuar microanálise dos sinais, que em muitos estudos pode complementar a análise do especialista; e utilizar as estatísticas dos parâmetros de cada indivíduo para melhorar o processo de detecção e eventualmente o grau de concordância com a classificação visual.

Todos autores apresentam os seus resultados em termos do grau de concordância dos respectivos sistemas automáticos com a classificação visual. A concordância varia entre 70% e 90%, embora não se possam comparar directamente estes valores, pois além destes resultados não serem obtidos com os mesmos dados (ou sinais que correspondam ao mesmo grupo de pessoas), também não se referem ao mesmo conjunto de classes. Finalmente há autores que referem como sendo satisfatório um grau de concordância próximo dos 90% por ser este o grau concordância requerido entre dois especialistas do mesmo laboratório. As regras de R+K (Rechtschaffen e Kales) inspiram, directa ou indirectamente, os processos de decisão nos classificadores automáticos. Alguns autores consideram, face aos resultados obtidos, que o conjunto de regras é incompleto; sugerindo alguns, a introdução no processo de classificação de informação respeitante a actividades não claramente mencionadas nas referidas regras, por exemplo, Hasan

refere teta e Smith beta e artefactos musculares. Neste trabalho será utilizada uma rede neuronal para a classificação do sono em estádios. Tratando-se de um novo método, e uma vez que dispomos de dados já classificados por outros métodos, faremos a validação deste novo classificador com este grupo de dados. Além disso, sendo aplicação de redes neuronais em classificação um problema de formulação simples, propomos o estudo da influência relativa das actividades do EEG no grau de concordância do classificador.

3.1: Organização

No capítulo 2, trata-se o problema da detecção de transientes por métodos no domínio dos tempos. Referem-se por isso, os problemas teóricos de processamento de digital de sinal envolvidos neste processo.

No capítulo 3, apresenta-se a implementação no sistema de multiprocessamento hierárquico referido - o HIDRA.

No capítulo 4 referem-se os aspectos teóricos da utilização de redes neuronais em classificação: topologias, processo de aprendizagem e teste.

Finalmente no capítulo 5, trataremos os resultados. Em que se classificam dados obtidos por um sistema de detecção automática referido na bibliografia (o SAMICOS), bem como dados obtidos pelo método proposto (no capítulo 2) e implementado no HIDRA (tal como é descrito no capítulo 3).

4: Características Técnicas do Sistema

Este sistema de multiprocessamento hierárquico designado por HIDRA (Hierarchical Instrument for Distributed Real-Time Analysis) é basicamente constituído por 5 módulos [Cruz -1990] :

1- microprocessador central com CPU MOTOROLA 6809 e bus G-64 , é um produto da GESPAC que corre o sistema da MICROWARE OS-9, um sistema de operação "multitasking" em tempo real.

2- A memória central que é também um placa GESPAC de 64k de RAM estática.

3- Os processadores periféricos, também com CPU MOTOROLA 6809, tem acesso ao bus central G-64 e a um bus de dados multiplexado donde acedem ao sinal de entrada. Para o acesso ao bus central é implementada uma estrutura de "daisy-chain", de tal modo que a prioridade em casos de pedidos simultâneos fica dependente da posição do preprocessor na estrutura. Além disso se a um preprocessor for concedido acesso ao bus nunca pode ser interrompido, mesmo que o preprocessor de elevada prioridade queira estabelecer comunicação.

4- Um módulo de aquisição de dados com 16 canais, com ADC de 12 bits à taxa programável de 200, 250 e 1KHz por canal. Este módulo está ligado ao preprocessadores pelo bus de dados que além dos 12 bits de dados tem mais 4 para selecção do canal de aquisição.

5- Um módulo gerador de interrupts para comunicação do processador central com os preprocessadores. Esta característica do sistema é bastante útil, nesta aplicação, durante a configuração do sistema em que é enviado um NMI para os preprocessadores lerem o respectivo bloco memória de central onde foram previamente carregados os programas de detecção. Em seguida é enviado um FIRQ para que a execução dos programas seja iniciada [Martins et al- 1990].

CAP 2: DETECÇÃO AUTOMÁTICA DE TRANSIENTES NO EEG

1: Introdução

No EEG um evento (actividade básica ou ritmo básico) é por definição uma onda ou grupo de ondas que se destacam da actividade de fundo. Os especialistas de electroencefalografia reconhecem estas ondas pelas características temporais associando-lhes por vezes um padrão global e/ou relação inter-canal. Os métodos de detecção destas actividades no domínio dos tempos baseiam-se na aplicação das definições (amplitude, frequência, duração etc.) recolhidas da literatura clínica [Ktonas-1984, Príncipe-1985]. Algumas destas definições são de natureza qualitativa (por exemplo: ondas agudas, abruptas), mas que os métodos automáticos validaram e quantificaram. O interesse dos métodos no domínio dos tempos é a caracterização dos eventos por medidas desta natureza que podem ser conseguidas com algoritmos simples, e por isso implementáveis em tempo real; esta é a exigência mais importante na análise (monotorização) de EEG/EOG de longa duração como é o caso dos estudos de sono. Contudo, após detecção automática das actividades, cada evento pode ficar descrito pelo conjunto de parâmetros utilizados na detecção, que conjuntamente com o instante de ocorrência caracteriza o sinal de forma mais fina e à custa dum volume de dados perfeitamente comportável para um microcomputador. Além disso, com os parâmetros armazenados (cujos valores são impossíveis de obter por análise visual) é possível criar estatísticas que em muitos estudos podem complementar a análise do clínico. Referimo-nos, por exemplo, a estudos de influência drogas na estrutura do sono e nos parâmetros das actividades básicas do EEG [Johnson-1979, Smith-1972/1976,

Wauquier-1985, Wolf-1985] ou de variação dos mesmos parâmetros com a idade [Gondeck-1974, Príncipe-1982, Smith-1977].

2: Definição das actividades básicas

Como referimos, as actividades básicas do EEG durante o sono vêm descritas na literatura clínica [Spehlmann-1981] pela sua forma, amplitude e frequência. Neste contexto, frequência significa inverso do número de picos do mesmo sinal que ocorrem numa janela temporal limitada, isto é, o especialista conta o número de picos numa unidade de tempo (segundo) e este número traduz a frequência do sinal. Note-se que esta definição de frequência é diferente da normalmente utilizada em Engenharia Electrónica.

As actividades alfa, sigma e beta (figura 2.1) são descritas como grupos de ondas quase-sinusoidais com frequências respectivamente nas gamas : 8-12, 12-16 e 16-30 cps. As actividades alfa e sigma são consideradas ondas de amplitude média¹ (20-50 μv) enquanto as ondas beta têm baixa amplitude (<20 μv). As actividades alfa e sigma apresentam alguma modulação em amplitude sendo por isso designadas por "spindles". Embora sejam actividades generalizadas, é muitas vezes referido que a actividades alfa, beta e sigma são mais visíveis nos canais occipital, frontal e central, respectivamente .

As actividades teta e delta (figura 2.1) são ondas que podem aparecer em grupo ou isoladas e têm frequências com valores inferiores às actividades previamente descritas: 2-8 e 0.5-2 cps, respectivamente. As ondas teta são consideradas de amplitude média enquanto as delta são de amplitude elevada

¹Note que estas referências pressupõem saídas gráficas normalizadas: 50 $\mu\text{v}/\text{cm}$

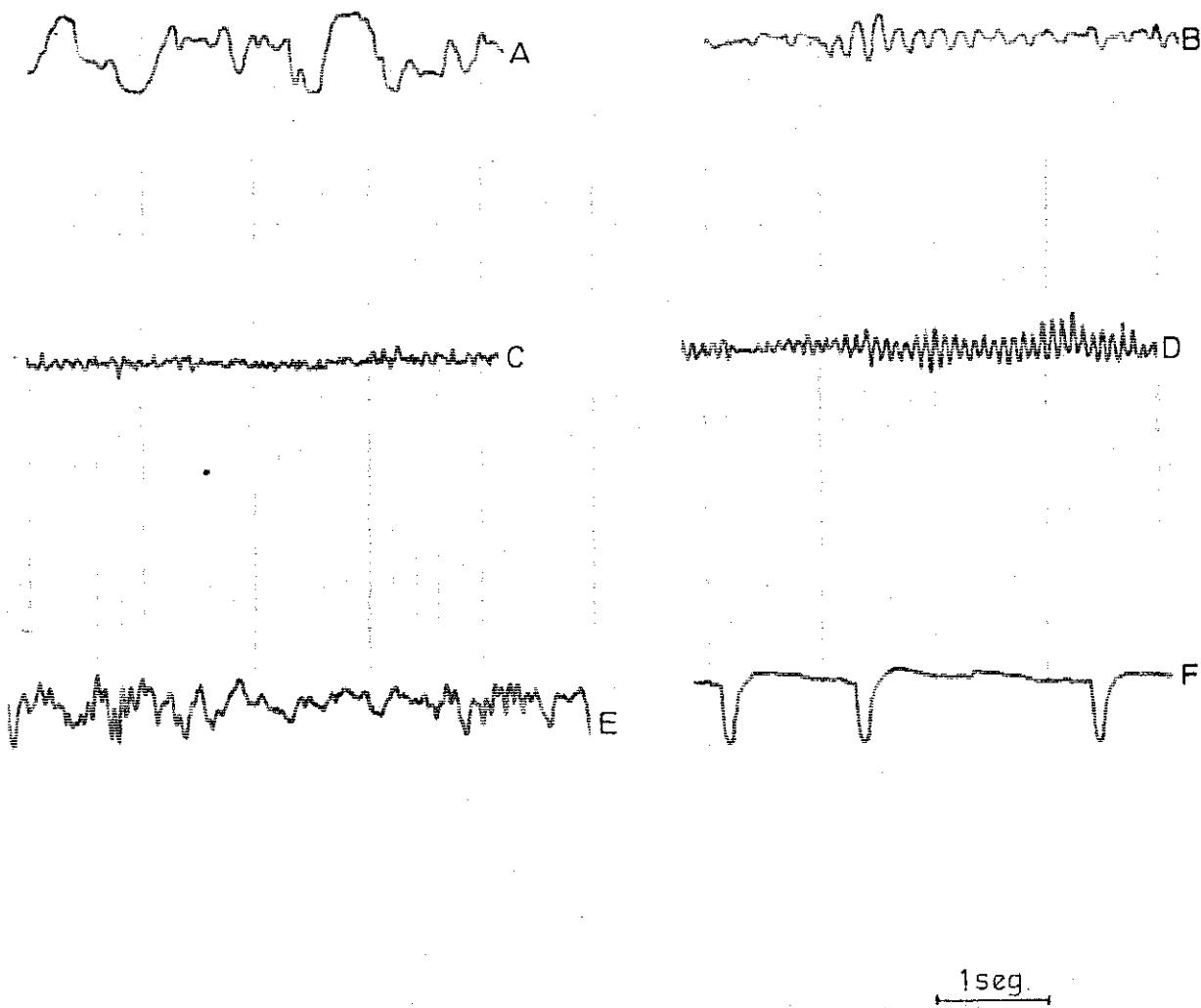


figura 2.1: Actividades do EEG e EOG: A actividade delta; B actividade teta; C actividade beta; D actividade alfa; E actividade sigma (fusos de sono); F movimentos oculares rápidos (REM).

(>50 μ v). São também actividades generalizadas, embora se refira que, por vezes, são mais visíveis no frontal.

Os complexos K são uma actividade bifásica que se destaca na actividade de fundo (de baixa amplitude), pela sua amplitude elevada e baixa frequência. Este evento aparece sincronizado em todos os canais e é muitas vezes seguido de actividade sigma. Tem sido uma actividade muito difícil de detectar por métodos automáticos por se confundir com actividade delta [Bremer-1970, Jansen-1989], sobretudo quando se tem informação referente a apenas um canal.

No canal ocular, são importantes os movimentos oculares rápidos [Minard-1971, Martinerie-1980] que geralmente aparecem em grupos e são ondas de grande amplitude e com declive acentuado num dos flancos (figura 2.1)

Resumindo, os parâmetros a detectar são basicamente o período e amplitude das ondas e, em alguns casos, parâmetros que são derivados daqueles: duração, declives e tempos de subida. Os parâmetros são comparados com limiares fixos. Assim, só são consideradas ondas válidas aquelas cujos parâmetros obedecem àqueles critérios, isto é, as que tiverem valores num intervalo definido previamente; além disso, é de todo o interesse armazenar os parâmetros destas ondas para análise posterior ("off-line").

3: Princípios básicos do método de detecção

Como foi referido os parâmetros a detectar são o período e a amplitude das ondas, ou outros cujos valores também dependem da detecção de picos ou passagens por zero. Mesmo considerando o sinal EEG como uma sobreposição de ondas quase-sinusoidais, não é fácil a aplicação ao sinal de um algoritmo de detecção de picos, já que altas ou baixas frequências sobrepostas ao ritmo básico em estudo dificultam e distorcem as medidas. As altas frequências constituem

mínimos locais muito difíceis de identificar e as baixas frequências produzem flutuações no sinal, resultando deste modo erros no período e na amplitude do evento. Para minimizar estes erros e, ao mesmo tempo facilitar a respectiva medição, utilizam-se filtros para atenuar as frequências fora da banda de interesse. Por exemplo, se o interesse residir nas ondas de baixa frequência (actividade lenta) e filtrar de modo a eliminar as frequências mais elevadas sobrepostas nessa actividade, a amplitude do sinal pode ser tomada como o pico (amostra de maior valor absoluto) entre dois zeros consecutivos do sinal filtrado. Consequentemente, a filtragem facilita a medição dos referidos parâmetros, bastando um simples algoritmo de detecção de picos e/ou passagens por zero. Mas, naturalmente que a filtragem pode introduzir erros e, como veremos em seguida, devemos ter especial cuidado com o comportamento de fase dos filtros, visto que neste caso estamos interessados na envolvente de sinais de curta duração (transientes).

Outra característica associada aos parâmetros temporais está relacionada com a precisão das medidas. O teorema da amostragem estabelece que se um sinal analógico for amostrado a duas vezes a frequência máxima do sinal, o sinal digital contém toda a informação do sinal original. Conseguindo-se por isso recuperar o sinal analógico da convolução de todas as amostras digitais ($-\infty$ a $+\infty$) com uma função interpoladora (que não é causal). Neste caso queremos efectuar uma medida sobre o conjunto finito de amostras do sinal digital; mas, quando as amostras são tomadas como o sinal, a medida vem afectada dum erro. Considerando um evento do EEG como uma sinusóide (ou grupo de sinusóides) de frequência f_0 sabemos que o erro relativo máximo Δf na determinação da frequência a partir do inverso do tempo que medeia dois máximos do mesmo sinal, quando a amostragem é efectuada à frequência f_s [Smith -1980] é:

$$\frac{\Delta f}{f_0} \cong \frac{f_0}{f_s} \quad (2.1)$$

Sendo assim, o erro relativo da medida também depende da frequência de amostragem. Para poder efectuar medidas com a precisão necessária numa detecção robusta de eventos, a frequência de amostragem para um evento pode ser superior à frequência de Nyquist na análise de EEG; a não ser que se efectuem as medidas numa versão interpolada do sinal digital, o que pressupõe bastantes mais cálculos (necessidade de filtragem).

4: Filtros na Análise do EEG

Alguns autores [Gailhard-1973, Smith-1988] têm apontado como motivo para falsas detecções, distorções introduzidas no sinal durante a fase de filtragem; mas estas conclusões têm aparecido de maneira qualitativa, não existindo nenhum critério para formular e desenhar os filtros com características óptimas para esta aplicação. Contudo os investigadores são unânimes ao considerar que os filtros devem apresentar uma boa resposta temporal, isto é, fase linear e resposta impulsional de curta duração e oscilações pouco acentuadas, de tal modo que as características temporais dos ritmos base sejam pouco distorcidas. Esta maneira de especificar os filtros é pouco comum; em geral, no projecto de filtros têm-se em conta as características de frequência: atenuação elevada na banda de atenuação, banda de transição pequena (com característica abrupta), e característica constante ("flat") na banda de passagem. A resposta impulsional dum filtro com estas especificações apresenta oscilações de elevada amplitude e longa duração. Consequentemente, estes filtros além de distorcerem as componentes de frequência dentro da banda de passagem, mascaram as actividades de frequência elevada de grande amplitude como é o caso dos artefactos musculares (aparecem como

características da banda). Por outro lado, artefactos lentos (deslocações dos eléctrodos) de grande amplitude produzem também oscilações à saída.

Recentemente, apareceu um estudo [Shirakawa, 1987] que apresenta para modelo dos ritmos básicos do EEG um grupo de sinusóides de duração limitada ("burst") e estuda a distorção introduzida nas sinusóides por vários tipos de filtros. Os resultados mostram que, quanto maior for a largura de banda dos filtros, menor é a distorção e, para a mesma largura de banda, a distorção aumenta com a ordem dos filtros. Com os filtros FIR (resposta impulsional finita) de fase linear a distorção é menor que a causada por filtros IIR (resposta impulsional infinita) de Butterworth, Chebyshev ou elípticos. Além da característica de fase linear os filtros FIR, têm em geral topologias estáveis e pouca sensibilidade a erros aritméticos embora seja necessária uma ordem elevada para poder concretizar características de atenuação aceitáveis. Outro conceito importante que também condiciona a escolha dos filtros é o de conversão da frequência de amostragem das sequências digitais. Nas secções seguintes analisaremos em detalhe alguns dos problemas relacionados com a implementação e aplicação de filtros digitais, especialmente dos filtro FIR, descrevendo também a topologia escolhida para esta aplicação.

4.1: Condições de fase linear nos filtros FIR

Um sistema linear causal com resposta impulsional finita, pode ser caracterizado por uma sequência $h(n)$, com $0 \leq n \leq N-1$, cuja transformada de z é calculada pela expressão seguinte:

$$H(z) = \sum_{n=0}^{N-1} h(n) z^{-n} \quad (2.2)$$

Sendo a transformada de Fourier de um sinal igual à transformada de z na circunferência unitária, isto é com $z = e^{j\omega}$, temos

$$H(e^{j\omega}) = \sum_{n=0}^{N-1} h(n) e^{-j\omega n} \quad (2.3)$$

Portanto a função de transferência de um sistema discreto é uma função periódica de período 2π . Além disso, se a resposta impulsional $h(n)$ tiver valores reais, podemos escrever $H(e^{j\omega})$ em representação de módulo e fase:

$$H(e^{j\omega}) = \pm |H(e^{j\omega})| e^{j\theta(\omega)} \quad (2.4)$$

consequentemente, o filtro tem fase linear num período da função de transferência se

$$\theta(\omega) = -\alpha\omega + \beta \quad (-\pi < \omega < \pi) \quad (2.5)$$

Para caracterizar a função de fase é usual definir o atraso de grupo e o atraso de fase. O atraso de grupo representa o atraso do filtro a um grupo de sinusóides e é descrito pela equação:

$$D(\omega) = -\frac{d\theta(\omega)}{d\omega} \quad (2.6)$$

Esta definição é uma medida de tempo em função da frequência, traduzindo a dispersão introduzida no sinal após a passagem por um filtro. Deste modo, os filtros com fase linear têm $D(\omega)$ com valor constante (α) para toda a gama de frequências. Isto é, à saída do filtro todas as componentes de frequência do sinal de

entrada têm o mesmo atraso. Assim, os sinais da banda de passagem do filtro mantêm a mesma envolvente (forma) após a passagem pelo filtro; portanto, fase linear (atraso de grupo constante) é uma das características mais importantes dos filtros usados na detecção de transientes.

E o atraso de fase representa o atraso do filtro a uma sinusóide é definido por

$$\tau(\omega) = -\frac{\theta(\omega)}{\omega} \quad (2.7)$$

Sendo assim, a partir da característica de fase apresentada distinguem-se dois grupos de filtros de fase linear: os verdadeiramente lineares e os "piecewise" lineares. Os primeiros têm $\beta=0$, conseqüentemente os atrasos de grupo e fase são iguais e de valor constante α . No outro grupo, com $\beta \neq 0$, só o valor de atraso de grupo tem valor constante α .

Assim, para filtros de fase linear podemos escrever a igualdade seguinte:

$$\sum_{n=0}^{N-1} h(n) e^{-j\omega n} = \pm |H(e^{j\omega})| e^{-j(\alpha\omega-\beta)} \quad (2.8)$$

que pode ser decomposta em duas partes: real e imaginária dadas respectivamente por:

$$\pm |H(e^{j\omega})| \cos(\alpha\omega-\beta) = \sum_{n=0}^{N-1} h(n) \cos(\omega n) \quad (2.9)$$

$$\pm |H(e^{j\omega})| \sin(\alpha\omega-\beta) = \sum_{n=0}^{N-1} h(n) \sin(\omega n) \quad (2.10)$$

Dividindo estas igualdades

$$\frac{\text{sen}(\alpha\omega-\beta)}{\text{cos}(\alpha\omega-\beta)} = \frac{\sum_{n=0}^{N-1} h(n) \text{sen}(\omega n)}{\sum_{n=0}^{N-1} h(n) \text{cos}(\omega n)} \quad (2.11)$$

Podemos reescrever esta expressão sob a forma:

$$\sum_{n=0}^{N-1} h(n) \text{cos}(\omega n) \text{sen}(\alpha\omega-\beta) - \sum_{n=0}^{N-1} h(n) \text{sen}(\omega n) \text{cos}(\alpha\omega-\beta) = 0 \quad (2.12)$$

Conclui-se que a condição suficiente para que os filtros tenham fase linear é:

$$\sum_{n=0}^{N-1} h(n) \text{sen}[(\alpha-n)\omega-\beta] = 0 \quad (2.13)$$

Com $\beta=0$, e com α e n inteiros, esta equação só é verdadeira quando os termos do somatório forem iguais dois a dois, isto é com:

$$\alpha = \frac{N-1}{2}$$

$$h(n) = h(N-n-1) \quad 0 \leq n \leq N-1 \quad (2.14)$$

Resumindo, um filtro é de fase linear com atraso de fase e de grupo constantes se a resposta impulsional tiver pares de valores simétricos relativamente ao meio da sequência. Assim, os filtros de fase linear têm resposta impulsional simétrica, que pode ser de dois tipos conforme o comprimento da respectiva

resposta impulsional (N) for par ou impar. O valor do atraso de grupo e de fase é igual a metade do comprimento da sequência que descreve a resposta impulsional do filtro (ordem do filtro).

Para $\beta \neq 0$ a igualdade (2.13) só se verifica com $\beta = \pi/2$ e, conseqüentemente, no somatório existem pares de valores iguais mas de sinal contrário, isto é

$$\alpha = \frac{N-1}{2} \quad \beta = \pm \frac{\pi}{2}$$

$$h(n) = -h(N-n-1) \quad 0 \leq n \leq N-1 \quad (2.15)$$

Concluindo, os filtros têm fase "piecewise" linear, se a resposta impulsional for anti-simétrica relativamente ao meio da sequência. Também neste grupo de filtros existem dois tipos conforme o comprimento da resposta impulsional for par ou impar.

As condições de simetria na resposta impulsional do filtro, isto é, nos coeficientes do polinómio $H(z)$ obtido do desenvolvimento do somatório da equação (2.2), implicam que os zeros desta função de transferência tenham uma localização específica no plano z . Os zeros aparecem em pares e são a imagem espelho um do outro, isto é, sendo z_0 um zero de $H(z)$, então $1/z_0$ também é zero de $H(z)$. Supondo ainda que a resposta impulsional tem valores reais, então os zeros da função de transferência são complexos conjugados. Assim, todos os zeros não situados na circunferência unitária aparecem em grupos de quatro; enquanto que os zeros situados na circunferência unitária aparecem em grupos de dois, visto que o recíproco coincide com o conjugado. Quando os zeros são reais e estão situados na circunferência unitária (isto é, $z = \pm 1$) o recíproco e o complexo conjugado coincidem, podendo ser identificados isoladamente. Concluiu-se deste

modo que a função de transferência do filtro, $H(z)$, pode ser naturalmente factorizada em produtos de polinómios de 4ª ordem, 2ª ordem e 1ª ordem, mantendo cada um dos factores as características de simetria de $H(z)$, e por isso fase linear.

5: Conversão da frequência de amostragem

Um dos conceitos fundamentais em processamento digital de sinal é, sem dúvida o de amostragem: representação dum sinal analógico numa sequência de números que contenham as características do sinal analógico. Ao longo do tempo foram discutidos muitos processos de amostragem, contudo o mais comum é a amostragem uniforme (ou periódica). Neste tipo de amostragem as amostras do sinal analógico são obtidas uniformemente espaçadas dum período T ; sendo a escolha da taxa de amostragem $F=1/T$ condicionada pelo Teorema de Amostragem (também conhecido por teorema de Nyquist) para que possa haver correspondência entre o sinal analógico e as respectivas amostras. Além disso, a taxa de amostragem condiciona o processamento do sinal, determinando o rigor da respectiva análise se utilizarmos simplesmente as amostras. Mas, muitas vezes, o sinal já foi amostrado com um determinado período T , não sendo este valor o mais conveniente para a aplicação em vista. Portanto, o objectivo da conversão da frequência de amostragem é converter este sinal amostrado num outro com período T' , de tal modo que, esta nova sequência continue a representar o sinal analógico. Este processo de conversão da taxa de amostragem pode ter dois aspectos distintos: aumentar (interpolação) ou diminuir (decimação) a taxa de amostragem. Em qualquer dos casos pretende-se efectuar essa conversão sem recorrer a uma conversão digital/analógica do sinal.

5.1: Sistema linear variante no tempo

Para facilitar a formulação deste problema, consideremos como primeira abordagem um passo intermédio: a reconstrução do sinal analógico, a partir das amostras $x(n)$. O sinal analógico é em seguida amostrado com nova taxa. Em primeiro lugar, as amostras $x(n)$ são filtradas por um filtro analógico, obtendo-se o sinal $x(t)$. Este sinal pode ser descrito como a convolução da resposta impulsional do filtro, $h(t)$, com as amostras $x(n)$. Como é óbvio, só as amostras de $h(t)$ espaçadas de T interferem na reconstrução do sinal analógico. Podemos então escrever o somatório seguinte:

$$\sum_{n=-\infty}^{+\infty} x(n) h(t - nT) \quad (2.16)$$

Sendo a resposta impulsional finita, aquele somatório tem limites finitos e iguais a N_1 e N_2 que são respectivamente os valores inferior e superior de n para os quais $x(n)$ coincide com a resposta impulsional. Em seguida $x(t)$ é amostrado com período T' pelo que podemos escrever

$$y(m) = x(t)|_{t=mT'} = \sum_{n=N_1}^{N_2} x(n)h(mT'-nT) \quad (2.17)$$

Desta expressão concluímos que sendo $T \neq T'$, para cada valor de m pode existir um conjunto diferente de amostras de $h(t)$ envolvidas no cálculo de $y(m)$, pois o produto do sinal $x(n)$ com a resposta impulsional atrasada não vai coincidir sempre nos mesmos pontos de $h(t)$. Assim, a operação é linear (a convolução) mas variante no tempo (não depende do mesmo grupo de amostras da resposta impulsional).

Mas, analisemos mais detalhadamente esta afirmação, para isso propomos a seguinte mudança de variável:

$$k = \left\lfloor \frac{mT'}{T} \right\rfloor - n \quad (2.18)$$

onde $\lfloor u \rfloor$ significa o menor inteiro contido em u . Calculando o valor de n na expressão 2.18 e substituindo em 2.17, vem

$$\begin{aligned} y(m) &= \sum_{k=K_1}^{K_2} x \left(\left\lfloor \frac{mT'}{T} \right\rfloor - k \right) h \left(mT' - \left\lfloor \frac{mT'}{T} \right\rfloor + kT \right) = \\ &= \sum_{k=K_1}^{K_2} x \left(\left\lfloor \frac{mT'}{T} \right\rfloor - k \right) h \left[(k + \delta_m)T \right] \end{aligned} \quad (2.19)$$

Nesta expressão torna-se perfeitamente claro que as amostras de $h(t)$ envolvidas no cálculo de $y(m)$ são sempre diferentes. As amostras são espaçadas de T mas o desfasamento δ_m é um valor positivo menor que a unidade; este valor é dado pela diferença entre um valor real e o inteiro imediatamente inferior, e marca o primeiro valor coincidente com $x(n)$ nos sucessivos atrasos de $h(t)$:

$$\delta_m = \frac{mT'}{T} - \left\lfloor \frac{mT'}{T} \right\rfloor \quad (2.20)$$

Resumindo, a conversão de um sinal $x(n)$ com período de amostragem T num sinal $y(m)$ com período T' é, como se queria demonstrar, uma operação linear mas variante no tempo. Suponhamos que os períodos de amostragem são submúltiplos de um período comum, isto é podemos escrever a relação

$$\frac{T'}{T} = \frac{M}{L} \quad (2.21)$$

sendo M e L valores inteiros. Podemos calcular o valor de δ_m substituindo (2.21) na expressão (2.20)

$$\delta_m = \frac{mM}{L} - \left\lfloor \frac{mM}{L} \right\rfloor = \frac{1}{L} \left[mM - \left\lfloor \frac{mM}{L} \right\rfloor \right] = \frac{1}{L} [(mM) \oplus L] \quad (2.22)$$

Onde $(k) \oplus L$ significa k módulo L. Deste modo δ_m só pode ter L valores distintos: 0, 1/L, 2/L, ..., (L-1)/L. Portanto no caso de se tratar unicamente de decimação (L=1) existe um único conjunto de valores de h(t) envolvidos no cálculo de y(m). Em alternativa, se for interpolação (M=1) existem L conjuntos distintos de amostras de h(t) para o cálculo de y(m). Assim, a conversão directa dum sinal digital x(n) num outro y(m) cujos períodos estejam relacionados por um número racional deve ser feita por um sistema linear variante no tempo. Onde, a resposta impulsional deste sistema é constituída por L conjuntos diferentes de amostras de h(t). No caso da relação de períodos de amostragem ser um número irracional já não é possível o cálculo da resposta impulsional do sistema linear digital, pois seria necessário um número infinito de conjuntos de amostras de h(t).

Conclui-se, deste modo, que a operação de conversão da frequência de amostragem de sequências só é realizável (sem recorrer a um sinal analógico) quando se pretendem conversões relacionadas por razões de valores inteiros. Trataremos em seguida com maior detalhe o problema da decimação. No entanto, a interpolação é o caso dual da decimação [Crochiere e Rabiner-1981/1983] por isso as soluções apresentadas para esta operação continuam a ser válidas para a interpolação.

5.1.1: Decimação

Podemos abordar este problema de conversão da taxa de amostragem do ponto de vista da análise de sinais em frequência. Sendo a sequência $x(n)$ a versão amostrada do sinal $x(t)$ à taxa de amostragem de T_s segundos ou $F_s=1/T_s$, sabemos que o espectro de $x(n)$ é periódico, com período 2π (ou F_s). Sendo cada período igual ao espectro de $x(t)$, se amostragem não tiver erro de "aliasing". Suponhamos que o espectro de $x(n)$ na banda base (1ª repetição do espectro), é sempre diferente de zero, i.e.

$$|X(e^{j\omega})| \neq 0 \quad |\omega| < \pi \quad (2.23)$$

Para se poder baixar a frequência de amostragem neste sinal sem haver erro de "aliasing", o sinal precisa de ser filtrado. Assim para poder baixar a frequência M vezes, o filtro deve ter as características seguintes:

$$H(e^{j\omega}) = \begin{cases} 1 & |\omega| < \frac{\pi}{M} \\ 0 & \frac{\pi}{M} < |\omega| < \pi \end{cases} \quad (2.24)$$

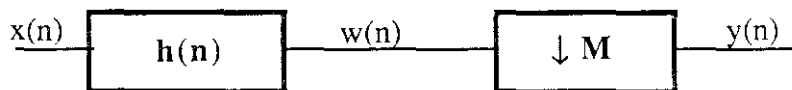


figura 2.2: Diagrama de blocos dum decimador de frequência.

O sinal de saída deste filtro, $y(n)$, pode agora ser amostrado à nova frequência de amostragem, isto é, basta guardar uma em cada M amostras e obtemos o sinal com

à frequência pretendida. Este processo pode ser representado pelo diagrama de blocos da figura anterior em que $h(n)$ é a resposta impulsional do filtro, portanto $w(n)$ é convolução linear e invariante no tempo de $x(n)$ e $h(n)$. A saída $y(m)$ é dada por

$$y(m) = w(mM) = \sum_{k=-\infty}^{+\infty} h(k) x(mM-k) \quad (2.25)$$

Ou, em alternativa, com $k=mM-n$ vem

$$y(m) = \sum_{n=-\infty}^{+\infty} h(mM-n)x(n) \quad (2.26)$$

Da primeira expressão verificamos que a resposta impulsional do filtro não é função de m . Existe portanto um único grupo de coeficientes para o cálculo de $y(m)$. No entanto, a operação não é invariante no tempo, isto é sendo $x(n)$ a entrada e $y(m)$ a resposta do sistema, se atrasarmos o sinal de entrada de um número inteiro de amostras (k) não podemos dizer que a saída é $y(m-k/M)$, a menos que o atraso seja um múltiplo inteiro de M .

5.1.2: Características de frequência dos filtros para decimação

A característica de frequência do filtro considerada na secção anterior não tem, como sabemos, realizabilidade. Para podermos escolher uma característica implementável vamos calcular o espectro da saída em função do espectro da entrada. Assim, veremos em que medida a filtragem não ideal pode introduzir erros neste processo.

Vamos supor que se obtém uma sequência $w'(n)$ a partir de $w(n)$ multiplicando por com um trem de impulsos de período M , de tal modo que $w'(n) = w(n)$ nos mesmos instantes que $y(m)$ e nos restantes tem valor zero.

$$w'(n) = w(n) \left[\frac{1}{M} \sum_{l=0}^{M-1} e^{j \frac{2\pi}{M} ln} \right] \quad (2.27)$$

O termo entre parêntesis é o desenvolvimento em série de Fourier do referido trem de impulsos. Sabemos que $y(m) = w'(mM) = w(mM)$, além disso podemos calcular a transformada de z de $y(m)$:

$$\begin{aligned} Y(z) &= \sum_{m=-\infty}^{+\infty} y(m) z^{-m} = \sum_{m=-\infty}^{+\infty} w'(m) z^{-m/M} = \\ &= \sum_{m=-\infty}^{+\infty} w(m) \left[\frac{1}{M} \sum_{l=0}^{M-1} e^{j \frac{2\pi}{M} lm} \right] z^{-m/M} \end{aligned} \quad (2.28)$$

Trocando a ordem dos somatórios obtém-se a relação seguinte entre as transformadas de z de $w(n)$ e $y(m)$.

$$Y(z) = \frac{1}{M} \sum_{l=0}^{M-1} W(e^{-j2\pi l/M} z^{1/M}) \quad (2.29)$$

Sendo a transformada de z do filtro $H(z)$ e $W(z) = H(z)X(z)$, pode obter-se a relação entre a transformada de z do sinal de saída e o de entrada. A transformada de Fourier desta relação calcula-se substituindo z por $e^{j\omega'}$

$$Y(e^{j\omega'}) = \sum_{l=0}^{M-1} H(e^{j(\omega' - 2\pi l)/M}) X(e^{j(\omega' - 2\pi l)/M}) \quad (2.30)$$

Esta equação traduz a influência da filtragem no espectro do sinal de saída, se não podermos desprezar os termos do somatório $l \neq 0$, há a interferência das componentes de frequência do sinal de entrada filtradas acima de ω/M . Na impossibilidade de realizar um filtro com característica ideal, a escolha de $H(e^{j\omega})$ deve ter em conta este erro de modo a que não tenha qualquer influência nos resultados pretendidos. O filtro passa-baixo que melhor se aproxima do ideal deve ter uma banda de transição abrupta e atenuação elevada na banda de atenuação. Um filtro FIR com estas características, tem necessariamente uma ordem muito elevada. Em alternativa, se quisermos evitar o "aliasing" no sinal de saída numa de comprimento inferior à banda base, isto é de 0 a ω_c com $\omega_c < \pi$, podemos introduzir algumas alterações na especificação dos filtros que podem conduzir a diminuição significativa do seu comprimento. O objectivo é controlar as zonas de "aliasing" no sinal de saída. Uma vez que sabemos a relação entre as frequências de amostragem (do sinal de entrada e de saída) podemos prever quais as zonas de frequência do sinal de entrada que vão introduzir "aliasing" na banda de interesse da saída, bem como as que não têm qualquer interesse. Assim, sendo a frequência de amostragem à saída $2\pi/M$, as zonas de frequência que interferem com a zona de interesse estão centradas nos múltiplos de $2\pi/M$ e têm largura de banda $2\omega_c/M$ (a largura de banda da zona de interesse). Por isso, o filtro deve ter as características seguintes, com $r = 1, 2, \dots, M-1$

$$H(e^{j\omega}) = \begin{cases} 1 & |\omega| < \frac{\omega_c}{M} \\ 0 & \frac{2\pi r - \omega_c}{M} < |\omega| < \frac{2\pi r + \omega_c}{M} \end{cases} \quad (2.31)$$

Estes filtros são designados por multibanda, e se $\omega_c/M < \pi/M$ existe uma diferença significativa entre o comprimento do filtro multibanda e o comprimento do filtro passa-baixo com características equivalentes [Crochiere e Rabiner-1983]. Por característica equivalente entende-se parâmetros idênticos na banda de passagem e bandas de atenuação em ambos os filtros. No entanto, deve ter-se em atenção o ganho do filtro multibanda nas zonas não especificadas pois não se pode amplificar o ruído, sob pena deste se tornar mais significativo que o sinal.

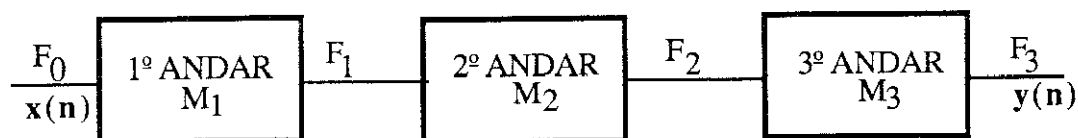


figura 2.3: Diagrama de blocos de um decimador com 3 andares.

Outro conceito importante, nesta técnica de decimação, é o de implementação em cascata. Neste contexto, cascata significa a implementação da decimação pretendida em vários andares, tal como está representado no diagrama de blocos da figura 2.3. Para isso factoriza-se o factor de conversão:

$$M = \prod M_i \quad (2.32)$$

Para cada um dos factores M_i implementa-se um andar de decimação. À primeira vista este método parece não oferecer qualquer vantagem, pois em vez de um filtro são necessários M_i filtros, sendo, por isso, levados a pensar num consequente aumento do tempo de cálculo. Mas na prática acontece exactamente o contrário, pois os filtros podem ter bandas de transição maiores o que além de diminuir os

problemas de implementação em aritmética de vírgula fixa diminui significativamente o comprimento do filtro.

Os filtros podem ser passa-baixo ou multibanda. Vamos especificar os filtros passa-baixo para a implementação em cascata, embora estes filtros sejam demasiado restritivos sobretudo nos primeiros andares de decimação. Neste caso, e para facilitar a exposição, trataremos as especificações em frequência analógica, isto é, não normalizada com a frequência de amostragem. Suponhamos que o sistema tem I andares, e que a largura de banda do sinal de saída é $F_I/2$. Se considerarmos a largura de banda do sinal de saída com duas componentes: a banda de interesse, com $0 \leq f \leq F_p$, e uma banda de transição, com $F_p \leq f \leq F_I/2$; podemos ter duas situações: uma em que preservamos toda a banda do sinal de saída e, outra em que permitimos "aliasing" na banda de transição. No último caso, a largura de banda (F_s) no sinal de saída pode ter como limite superior qualquer valor entre F_p e $F_I/2$; mas no primeiro caso a largura de banda é $F_s = F_I/2$ e, evitamos o erro de "aliasing" na largura de banda do sinal de saída. Para os outros andares a banda de interesse é, obviamente, a mesma do andar de saída. Agora, a banda de transição é que pode ser alargada permitindo "aliasing" para frequências superiores a F_s ; já que este erro vai ser eliminado pelos andares seguintes de decimação. Deste modo, e para o primeiro andar podemos definir a banda de transição como $F_s \leq f \leq (F_1 - F_s)$, em vez de $F_s \leq f \leq F_I/2$, sendo naturalmente a banda de atenuação de $(F_1 - F_s) \leq f \leq F_0/2$; no segundo andar a banda de transição é $F_s \leq f \leq (F_2 - F_s)$ e a banda de atenuação é $(F_2 - F_s) \leq f \leq F_1/2$; etc. O espectro do sinal ao longo dos três andares de filtragem de uma estrutura com 3 andares de decimação, está representado na figura 2.4. E, note-se que as bandas de transição vão sendo cada vez mais abruptas, por isso nas implementações com vários andares os factores de decimação devem ser mais elevados nos primeiros andares, isto é

$$M_1 > M_2 > \dots > M_1 \quad (2.33)$$

Isto porque, em geral, a ordem dos filtros depende de modo inverso da largura da banda de transição e esta pode ser alargada nos primeiros andares de decimação uma vez que se permite erro de "aliasing".

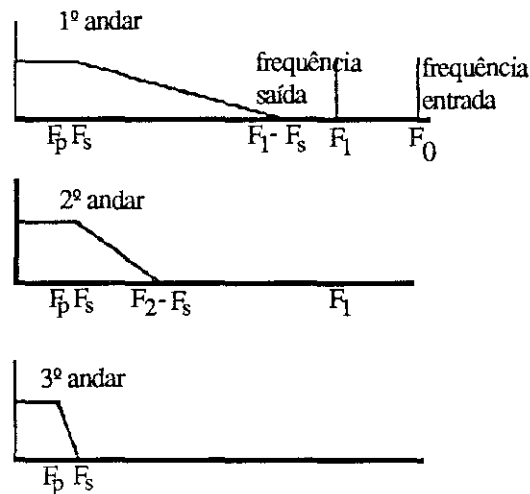


figura 2.4: Espectro do sinal ao longo dos andares de decimação

Para implementarmos um filtro com vários andares introduzimos mais variáveis no processo de especificação do filtro : o número de andares e em seguida a escolha dos factores de decimação para cada andar. Como em qualquer aproximação, há sempre um conjunto de factores que determinam uma determinada opção, isto é, não existe uma solução óptima. Crochiere e Rabiner tratam este assunto como um problema de optimização. Dadas as especificações do filtro, pretende-se é optimizar a solução quanto a tempo de cálculo e armazenamento, variando para isso o número total de andares e respectivos factores de decimação. Em conclusão, estes autores referem que a melhoria mais significativa daqueles parâmetros ocorre na passagem de um andar para dois. Além disso essas vantagens

são notórias quando o factor de decimação global é elevado e a banda de transição pretendida tem valor bastante baixo.

6: Aproximação da função de transferência dum FIR

A literatura sobre processamento digital de sinal refere vários métodos de aproximação da função de transferência de filtros FIR [Gold e Rabiner-1975]. Mas, neste caso, como estamos interessados numa implementação realizável num microprocessador comum (sem qualquer unidade de processamento aritmético auxiliar), a complexidade dos filtros é um factor muito importante. A estrutura deve ter o mínimo de multiplicações que devem ser substituídas, se possível, por um pequeno número de "shifts" e adições. Por esta razão, optámos por um método [Principe,1986] que, embora conduza a características de atenuação pouco invejáveis ("ripple" na banda de atenuação com valores superiores a 0.001 e banda de transição pouco abrupta) pode ser facilmente implementado, além de que o método de aproximação da respectiva função de transferência não requer cálculos complicados. Convém, também lembrar que o filtro na estrutura do detector pretende simplesmente atenuar os sinais fora da banda, sendo a saída sujeita a outros critérios de decisão que possibilitam uma discriminação mais rigorosa. Este método, como veremos em seguida, baseia-se na colocação de zeros sobre a circunferência de raio unitário eliminando as zonas de frequência sem interesse.

6.1: Eliminação de bandas

Este método consiste basicamente em criar uma série de bandas de passagem por colocação de zeros igualmente espaçados na circunferência de raio

unitário. Em seguida, eliminam-se as bandas sem interesse colocando novamente zeros nos picos dessas bandas.

O primeiro grupo de zeros é criado por funções do tipo seguinte:

$$G(z) = 1 + z^{-N} \quad (2.34)$$

Sendo a resposta em frequência destas funções:

$$G(e^{j\omega}) = 1 + e^{-j\omega N} = 2\cos\left(\frac{\omega N}{2}\right) e^{-j\omega N/2} \quad (2.35)$$

que é de fase linear com zeros na circunferência unitária em

$$z_i = \frac{\pi}{N} + \frac{2\pi}{N} k \quad k=0,1,\dots,N-1 \quad (2.36)$$

Isto é, o módulo da função de transferência é uma função sinusoidal com N zeros, portanto as bandas de passagem correspondem às arcadas do módulo da função sinusoidal. Assim, para eliminar bandas sem interesse colocam-se zeros nos locais dos respectivos picos, isto é, nos máximos da função sinusoidal. Os polinómios pretendidos têm zeros complexos conjugados na circunferência de raio unitário:

$$H_i(z) = z^{-2} - 2\cos\theta_i z^{-1} + 1 \quad (2.37)$$

sendo θ_i escolhido de modo a que os zeros do polinómio se situem nos picos do primeiro grupo funções, isto é

$$\theta_i = \frac{2\pi}{N} k \quad k = 1,2,\dots,\frac{N-1}{2} \quad (2.38)$$

A resposta em frequência destas funções é também linear mas o módulo vai alterar duma maneira não simétrica a banda de passagem do filtro pretendido.

$$H_1(e^{j\omega}) = 2 \left[\cos\omega - \cos\theta_1 \right] e^{-j\omega} \quad (2.39)$$

Outro grupo de funções com zeros equidistantes na circunferência unitária, e que é fundamental para a aproximação da função de transferência de filtros passa-banda e passa-alto uma vez que têm um zero em $z=1$, é constituído pelo seguinte polinómio:

$$Q(z) = 1 - z^{-N} \quad (2.40)$$

Sendo a resposta em frequência destas funções :

$$Q(e^{j\omega}) = 2 \operatorname{sen}\left(\frac{\omega N}{2}\right) e^{-j\left(\frac{N\omega}{2} + \frac{\pi}{2}\right)} \quad (2.41)$$

que é de fase "piecewise" linear, cujo módulo é também uma função sinusoidal em que os zeros estão também equidistantes na circunferência de raio unitário, mas neste caso o primeiro aparece em $z=1$.

A banda de passagem pode também ser definida pelos polinómios de 2ª ordem definidos em (2.37), eliminando a banda sem interesse com a colocação de zeros criados por funções do mesmo tipo ou pelas funções (2.34) ou (2.40). Em alguns casos esta solução pode trazer vantagens, pois a largura das arcadas de sinusóide destas funções é ω_s/N (distância entre dois zeros consecutivos), se ω_s fôr fixo a escolha de N pode não conduzir a uma largura de banda conveniente.

Estes filtros têm "ripple" elevado na banda de atenuação para ordens baixas; no entanto são facilmente implementáveis uma vez que têm poucas multiplicações

(visto que a maioria dos coeficientes são unitários) e, além disso podemos ajustar os coeficientes $2\cos\theta_i$ de modo a terem valores resultantes da soma de potências negativas de 2. Desta forma, as multiplicações existentes são reduzidas a somas e "shifts".

7: Implementação dos filtros FIR

Quando a função de transferência dum filtro só tem zeros, a equação diferença que descreve este sistema é não recursiva. Isto é, a saída é função da entrada actual e dum número finito de entradas anteriores, não dependendo de saídas anteriores. A implementação dum filtro num microprocessador pode ser portanto o algoritmo resultante desta equação diferença:

$$y(n) = h(0) x(n) + h(1) x(n-1) + \dots + h(N)x(n-N) \quad (2.42)$$

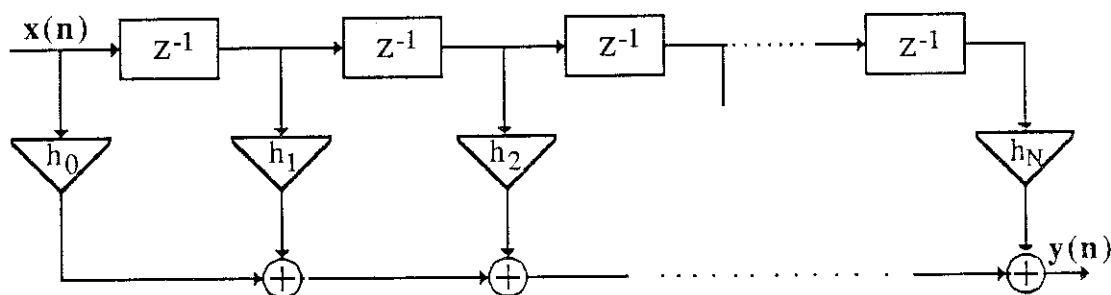


figura 2.5: Diagrama de blocos dum filtro FIR: implementação directa.

Esta implementação designa-se por directa, e é representável pelo diagrama de blocos da figura 2.5. Em alternativa a esta implementação, podemos escrever a função de transferência como um produto de factores (de 1ª ou 2ª ordem, conforme os zeros são reais ou complexos conjugados):

$$H(z) = \prod_{k=1}^{M=N/2} (a_k + b_k z^{-1} + c_k z^{-2}) \quad (2.43)$$

Esta estrutura, em que a saída de cada andar é a entrada do seguinte, designa-se por implementação cascata e está representada na figura seguinte.

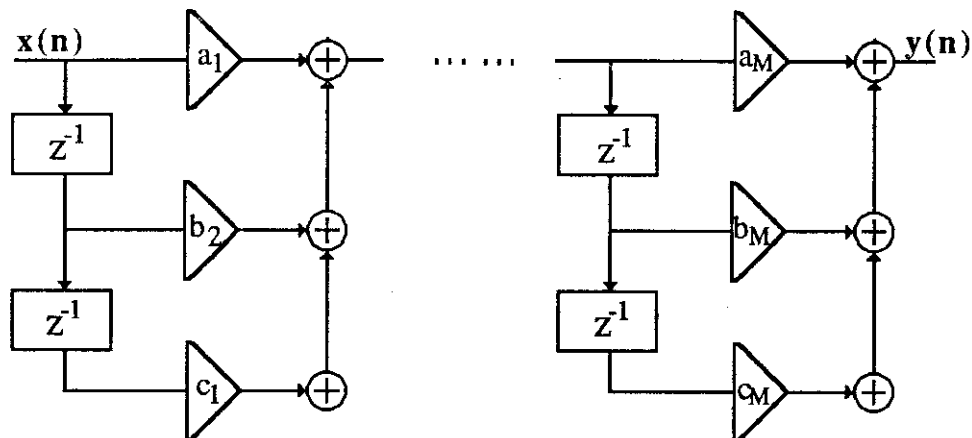


figura 2.6: Diagrama de blocos dum filtro FIR: implementação cascata.

As estruturas referidas são as mais comuns na implementação de filtros FIR (e IIR). No entanto, e no caso particular dos filtros FIR, além da estrutura paralela (resultante da transformação da função de transferência num somatório de funções parciais de 1ª e 2ª ordem), existem ainda outras que não têm contrapartida na implementação de filtros IIR. Por exemplo, a estrutura derivada da implementação de fórmulas de interpolação, como é o caso do polinómio Lagrange [Gold e Rabiner-1975].

7.1: Efeitos do comprimento de palavra finita nos filtros

O processamento digital de sinal é muitas vezes considerado como o estudo de sinais e sistemas discretos. Mas, devemos considerar que nos sistemas discretos os coeficientes dos filtros e os resultados das operações algébricas têm como

domínio o conjunto dos números reais. O que não é verdade quando se trata da implementação dum filtro em computador digital devido ao facto do comprimento de palavra ser finito. Assim, o filtro pode ter um comportamento bastante diferente do pretendido, pois os efeitos do comprimento de palavra finita contribuem para erros em várias fases do processamento digital: quantificação (sinal de entrada e coeficientes) e os erros resultantes das operações aritméticas (arredondamento ou truncatura nos produtos e "overflow" nas adições).

7.1.1: Erros de quantificação no sinal de entrada.

A quantificação do sinal de entrada é feita no conversor analógico/digital (ADC), no primeiro elemento da cadeia de processamento digital. No conversor, o sinal é amostrado periodicamente e codificado numa palavra de B bits, pelo que o sinal digital à saída do referido conversor só pode ter 2^B valores distintos, deste modo dois valores consecutivos diferem entre si de $\Delta = 2^{-B+1}$, se o sinal analógico estiver normalizado (com valores de amplitude compreendidos entre -1 e 1). Assim, os valores dos erros de quantificação ficam perfeitamente determinados se soubermos o código de codificação (complementos para um, complementos para dois, valor absoluto e sinal) e o tipo de quantificação (arredondamento e truncatura) utilizados. Supondo que se utiliza a notação em complementos para 2 e codificação por arredondamento, o sinal analógico fica codificado com o valor digital mais próximo, então o erro de quantificação $e(n)$ tem valores no seguinte intervalo :

$$\frac{-\Delta}{2} < e(n) < \frac{\Delta}{2} \quad (2.44)$$

Se a quantificação for por truncatura, o sinal analógico fica codificado com valor digital imediatamente inferior, sendo o erro de quantificação:

$$-\Delta < e(n) < 0 \quad (2.45)$$

Consideremos o erro como um ruído aditivo correspondente à realização dum processo estacionário aleatório em que as variáveis aleatórias são não correlacionadas (isto é, ruído branco), que a distribuição da probabilidade de erro é uniforme na gama de valores do erro de quantificação, e finalmente que o sinal é não correlacionado com o erro ². Com este modelo, e a partir das definições de média e variância de variáveis aleatórias podem calcular-se os valores respectivos. A média tem os seguintes valores:

$$\begin{aligned} m_e &= 0 && \text{arredondamento} \\ m_e &= -\frac{\Delta}{2} && \text{truncatura} \end{aligned} \quad (2.46)$$

A variância é igual em ambos os casos tendo o valor:

$$\sigma_e^2 = \frac{\Delta^2}{12} \quad (2.47)$$

Para analisar a influência relativa do erro no sinal é comum calcular-se a razão entre a potência do sinal e a do ruído, que se designa por razão sinal ruído (SNR- "signal noise ratio"):

² O que não é verdade se o sinal a codificar for uma constante ou a codificação do sinal for em complementos para um ou sinal e valor absoluto por truncatura [Oppenheim and Schaffer-1975].

$$\text{SNR} = 10 \log_{10} \frac{\sigma_x^2}{\sigma_e^2} = 6.02B + 10 \log_{10} \sigma_x^2 + K \quad (2.48)$$

daqui pode concluir-se que a relação sinal ruído melhora aproximadamente 6 db por cada bit aumentado ao comprimento de palavra na codificação.

Para estudar a influência do erro na saída podemos considerar o sinal de entrada (quantificado) como a soma do sinal discreto com o erro de quantificação. Sendo este sinal a entrada dum sistema linear, a saída também pode ser considerada a soma das respostas às duas entradas. Consequentemente a potência do erro à saída, vem afectada pela resposta impulsional

$$\sigma_f^2 = \sigma_e^2 \sum_{n=0}^{N-1} |h(n)|^2 \quad (2.49)$$

Ou, em alternativa, (teorema de Parseval), escrevendo a expressão em termos da resposta em frequência do sistema

$$\sigma_f^2 = \frac{\sigma_e^2}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 d\omega \quad (2.50)$$

7.1.2: Erros de quantificação nos coeficientes

Tendo em vista a aplicação e a implementação possível (dependente sobretudo da velocidade de cálculo) escolhe-se uma topologia. A partir das especificações calcula-se a função de transferência. Mas, mais uma vez, não podemos esquecer que os coeficientes precisam ser representados com comprimento de palavra finito, isto é, quantificados. Assim, a função de transferência após a quantificação dos coeficientes não tem os mesmos polos e

zeros. Consequentemente, a resposta em frequência tem características diferentes das pretendidas.

A implementação directa dos filtros é mais sensível aos erros de quantificação dos coeficientes que a implementação em cascata ou paralela [Chan e Rabiner-1973]. Para a implementação com funções de transferência de 2ª ordem (obtidas a partir de polos ou zeros complexos conjugados: $r^2z^{-2} + 2r\cos\theta z^{-1} + 1$) a localização possível dos polos ou zeros é dada pela intersecção de circunferências com o raio correspondente à quantificação de r^2 e linhas verticais correspondentes à quantificação de $2r\cos\theta$.

Como vimos nos filtros FIR de fase linear os zeros têm localização específica, devemos, por isso, estudar a influência da quantificação dos coeficientes na localização dos referidos zeros. Se a implementação fôr sob a forma canónica, é óbvio que não há qualquer influência; embora os zeros não sejam exactamente os pretendidos continuam a ser a imagem espelho um do outro, já que apesar da quantificação, se mantém a simetria na resposta impulsional (ou seja nos coeficientes). Se o sistema fôr implementado em cascata com factores de 1ª ou 2ª ordem com os zeros situados na circunferência unitária (como é o caso do tipo de filtros estudado), o sistema continua a ter fase linear após a quantificação, pois os factores de 1ª ordem são realizados exactamente e os de 2ª ordem continuam a ter zeros complexos conjugados naquela circunferência. No entanto deve assegurar-se que todos os zeros que ficam no interior da circunferência tenham a respectiva imagem no exterior. Isto pode ser facilmente conseguido pela implementação directa do polinómio de 4ª ordem.

$$H(z) = 1 + d_1z^{-1} + d_2z^{-2} + d_1z^{-3} + z^{-4} \quad (2.51)$$

Para a implementação em cascata com factores de 2ª ordem sabemos que os zeros respectivos devem ter os valores $z = r e^{\pm j\theta}$ e $z = (1/r)e^{\pm j\theta}$, resultando a função de transferência

$$H(z) = \frac{1}{r^2} (z^{-2} - 2r\cos\theta z^{-1} + r^2) (r^2 z^{-2} - 2r\cos\theta z^{-1} + 1) \quad (2.52)$$

daqui se conclui que, mesmo após a quantificação, cada um dos factores continua a ter zeros que são a imagem dos zeros do outro. Esta implementação é preferível, face à implementação directa do polinómio de 4ª ordem, uma vez que para uma grande maioria de filtros as características de frequência do filtro são extremamente sensíveis à quantificação [Oppenheim e Schaffer-1975].

Resumindo, e para o tipo de filtro estudado numa das secções anteriores, a quantificação dos coeficientes não é um factor determinante, visto que não altera a condição de fase linear por os zeros se localizarem na circunferência unitária. Além disso, a escolha desta estrutura com coeficientes aproximados (pela soma de potências de 2 com valor idêntico) é preferível a uma outra estrutura correspondente à aproximação da resposta em frequência dada pelos valores exactos destes coeficientes (muitas vezes impossíveis de representar), se aquela aproximação se traduzir em tempo de cálculo inferior e, isso será possível sempre que as multiplicações possam ser substituídas por um pequeno número de "shifts" e adições.

7.1.3: Erros Aritméticos

O número de bits do resultado de um produto é igual à soma dos bits dos operandos, assim se os operandos tiverem igual número de bits, o produto pode ter o dobro dos bits dos operandos. Consequentemente é necessário quantificar o

resultado após cada multiplicação, sobretudo se o filtro for recursivo pois o aumento do número de bits não pode continuar indefinidamente. Esta quantificação tem efeitos diferentes conforme o tipo de aritmética utilizada: vírgula fixa ou flutuante. Trataremos só o caso do processamento em vírgula fixa por se tratar dum algoritmo de execução mais simples num microprocessador comum (sem nenhuma unidade de processamento aritmético auxiliar).

Tal como o erro de quantificação na entrada, os efeitos do erro de truncatura nos produtos, são analisados considerando o filtro linear e o erro como um sinal aditivo não correlacionado com o sinal. Consequentemente o erro à saída é a resposta do erro introduzido na malha por cada um dos produtos. Existem no entanto casos para os quais este modelo não é válido. Um dos casos mais estudados designa-se por erro de "limit-cycle" que não ocorre em filtros FIR pois, em geral, não têm estrutura recursiva, isto é, o valor de cada amostra não depende de resultados anteriores. Não trataremos, por isso, este problema.

Outro problema associado às operações aritméticas é o "overflow". Este erro está associado às adições, e o único processo de o evitar é manter as entradas suficientemente pequenas de modo a que o resultado não ultrapasse o comprimento de palavra. Deste modo é necessário multiplicar as entradas por um valor, esta operação é designada por "scaling".

7.1.3.1: Erros de arredondamento

O modelo para o estudo dos erros de arredondamento é basicamente o mesmo que o utilizado para o estudo da quantificação na entrada. Contudo a distribuição das fontes de ruído depende da estrutura escolhida e da aritmética utilizada.

Se considerarmos o produto quantificado como a soma do produto com o erro de quantificação podemos saber a resposta do filtro ao erro. Para isso pressupõe-se, como no caso da quantificação do sinal de entrada, que $e_k(n)$ é ruído branco, com distribuição uniforme no intervalo de quantificação, não correlacionado entre si nem com o sinal de entrada; e que não há "overflow", isto é, não é necessário fazer o "scaling" do sinal de entrada. Na implementação canónica dos filtros FIR os produtos surgem imediatamente antes do nó de saída, o erro à saída é a soma de todos os erros de quantificação. A variância do sinal à saída é igual à soma das variâncias de todos os ruídos no nó de saída.

$$\sigma_f^2 = N \frac{\Delta^2}{12} \quad (2.53)$$

Neste caso, a variância do erro à saída não depende dos parâmetros do filtro uma vez que o erro é adicionado directamente na saída, no entanto, é dependente da ordem do filtro: ordem elevada traduz-se em variância elevada.

Para implementações em cascata de funções parciais de 2ª ordem, a variância do erro à saída de cada andar é igual à soma das variâncias do respectivo andar. Deste modo, podemos considerar o diagrama de blocos da figura como modelo de estudo do efeito deste erro.

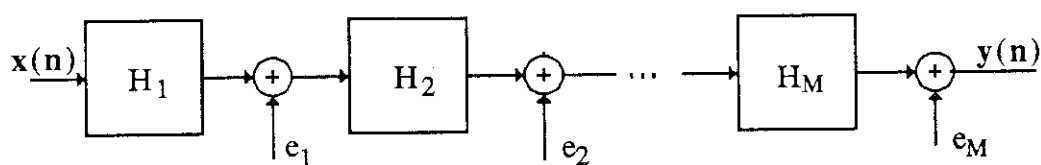


figura 2.7: Diagrama de blocos do modelo para o estudo do efeito dos erros de aritméticos num filtro em cascata.

Assim cada fonte de erro $e_i(n)$ vai ser filtrada pelos andares seguintes, portanto o erro à saída vai depender da ordem escolhida para a cadeia de funções de 2ª ordem. Deste modo, calculando resposta impulsional $g_i(n)$ a partir de cada uma das entradas de ruído e sendo k o número de multiplicações de cada andar, podemos calcular a variância do ruído à saída

$$\sigma_{e_i}^2 = k \frac{\Delta^2}{12} \sum_{n=0}^{N-2i} g_i^2(n) \quad (2.54)$$

A variância total à saída é a soma das M variâncias do ruído à saída provenientes das M fontes. Com M funções parciais diferentes existem $M!$ ordens possíveis. Para M elevados, a procura exaustiva da melhor ordem torna-se impraticável; no entanto, Chan e Rabiner [1973,1975] mostraram que a melhor ordem é aquela cujos picos das funções de transferência de cada uma das secções para a saída tem valores menores. Pois, calculando os módulos das funções de transferência de cada uma das secções para a saída, estes dependem dos zeros de cada uma dessas funções. Sabemos que os factores de 2ª ordem contribuem com picos no módulo da função de transferência em $\omega=\pi$ ou $\omega=0$ conforme os zeros respectivos sejam do lado direito ou do lado esquerdo do plano z ; portanto se tivermos uma implementação que agrupe, no final ou no princípio da estrutura, os zeros do mesmo lado vamos ter um pico muito elevado na resposta dessas secções para saída, o que naturalmente conduz a uma variância do erro dessas secções na saída. Assim, secções adjacentes na saída não devem ter zeros muito próximos, isto é os zeros devem estar distribuídos por todo o plano. Com estes argumentos os referidos autores, mostraram que uma vez que um filtro têm em geral zeros por todo o plano a grande maioria das ordens escolhidas tem bom comportamento face ao ruído. Consequentemente é indicado um algoritmo simples para a escolha da

ordem que conduz a variâncias de erro consideradas aceitáveis embora não seja necessariamente o valor mínimo. O algoritmo é o seguinte:

começando com $i=M$ (número de secções), escolher para a secção i aquela que conjuntamente com as já existentes tem menor valor de

$$\sum g_{i-1}^2(n) \quad (2.55)$$

7.1.3.2: Scaling

A maneira correcta utilizar a relação sinal ruído (SNR) máxima dum sistema digital é escolher o nível do sinal mais elevado, em cada ponto da estrutura. Isto porque numa determinada estrutura o passo (Δ) tem valor constante. Contudo, o aumento do valor do sinal tem limites pois a gama dinâmica do filtro (i.e. o comprimento de palavra) da aritmética em vírgula fixa não pode ser excedida sem haver "overflow" nas adições; por isso, é necessário implementar uma política de "scaling" que conserve o nível do sinal ao longo da estrutura no valor máximo possível sem no entanto haver "overflow". Como se sabe, se as operações forem executadas em complementos para 2 pode haver "overflow" nas adições parciais com resultado final correcto se o total for representável nesse comprimento de palavra. Consequentemente uma das entradas da adição pode ser incorrecta, isto é, pode ter havido "overflow" na multiplicação do sinal por um coeficiente superior à unidade. Assim numa estrutura há nodos onde pode ser necessário haver "scaling" e outros onde *a priori* não há qualquer interesse em efectuar essa operação. Os primeiros são naturalmente os nodos que precedem as multiplicações, e os outros são os que precedem as adições e os atrasos. No caso particular dos filtros FIR em cascata o "scaling" deve ser implementado entre as secções.

Em seguida coloca-se o problema do cálculo do valor das constantes de "scaling". Podemos supor que $x(n)$ é o sinal de entrada e $y_i(n)$ é o sinal de saída no nodo i onde se pretende efectuar a operação de "scaling". A partir da resposta impulsional $h_i(n)$ entre aqueles dois pontos da estrutura sabemos que

$$y_i(n) = \sum_{k=0}^{\infty} h_i(k)x(n-k) \quad (2.56)$$

Portanto, se o sinal $x(n)$ tiver valores na gama dinâmica, isto é $|x(n)| \leq M$, conclui-se que o somatório dos módulos dos valores resposta impulsional deve ser inferior à unidade; sendo assim a constante pode calcular-se a partir da desigualdade seguinte

$$c_i \leq \frac{1}{\sum_{k=0}^{\infty} |h_i(k)|} \quad (2.57)$$

Este valor é muito restritivo para a maior parte dos sinais em estudo, desperdiçando por isso grande parte do comprimento de palavra disponível. Em alternativa o valor da constante pode calcular-se a partir da resposta em frequência [Jackson-1986, Taylor-1983]. Assim, podem escrever-se equações impondo restrições na energia ou no espectro do sinal nos nodos onde se efectua a operação de "scaling". Esta abordagem não impede completamente o "overflow" como no caso anterior, mas torna-o bastante improvável, havendo uma melhor utilização da gama dinâmica [Jackson, 1986]. Assim, o cálculo da constante é feito a partir das normas L_p da transformada de Fourier. Um sinal com transformada $A(e^{j\omega})$ a norma L_p é dada por

$$\|A\|_p = \frac{1}{2\pi} \int_{-\pi}^{\pi} |A(e^{j\omega})|^p d\omega \quad (2.58)$$

Note-se que L_1 é o valor médio do valor absoluto da transformada e L_2 é o valor rms, e no caso de $A(e^{j\omega})$ ser uma função contínua (como é das transformadas de sinais amostrados) o valor de $\|A\|_p$ quando $p \rightarrow \infty$ coincide com o valor de pico de $A(e^{j\omega})$. Por outro lado também se demonstra que

$$\|A\|_1 \leq \|A\|_2 \leq \dots \leq \|A\|_\infty \quad (2.59)$$

Portanto, sendo $H_1(e^{j\omega})$ a transformada de Fourier de $h_1(n)$ e $X(e^{j\omega})$ a transformada do sinal de entrada vem

$$y_1(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_1(e^{j\omega}) X(e^{j\omega}) e^{j\omega} d\omega \quad (2.60)$$

Assim, o valor do sinal de saída é dado pela desigualdade seguinte:

$$y_1(n) \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_1(e^{j\omega}) (X(e^{j\omega}))| d\omega \quad (2.61)$$

Se considerarmos o sinal de entrada um sinal determinístico cuja transformada obedece à norma L_1 podemos escrever

$$y_1(n) \leq \|H_1\|_\infty \|X\|_1 \quad (2.62)$$

Concluiu-se que o cálculo da constante de "scaling" é efectuado a partir do valor de pico da resposta em frequência da estrutura entre os nodos considerados. Este

processo é o mais utilizado por ser bastante simples o cálculo da referida constante, embora seja unicamente apropriado para sinais de banda estreita com é o caso de sinusóides. Mas, a desigualdade anterior pode escrever-se genericamente do modo seguinte

$$y_i(n) \leq \| H_i \|_p \| X \|_q \quad \frac{1}{p} + \frac{1}{q} = 1 \quad (2.63)$$

onde p e q têm valores superiores à unidade. Esta desigualdade é conhecida por desigualdade de Holder e pode utilizar-se para o cálculo da referida constante de "scaling". Mas em geral só é utilizada para p=q=2 obtendo-se deste modo um valor menos conservativo para aquela constante. Estamos a supor que a energia do sinal de entrada (se for determinístico) é finita, e assim o valor rms de H_i deve ser inferior à unidade. Resumindo neste método o valor da constante é dado pela norma p da resposta em frequência entre dois nodos da estrutura

$$c_i \leq \frac{1}{\| H_i \|_p} \quad (2.64)$$

CAP 3: DETECÇÃO DE TRANSIENTES NO HIDRA

1: Introdução

A classificação de sono é a atribuição dum estágio de sono, com significado fisiológico, a cada minuto (época) do traçado poligráfico. Para isso, é necessário quantificar a existência das actividades básicas do EEG e os movimentos oculares do EOG. Esta informação não é retirada do mesmo eléctrodo, sendo necessárias pelo menos duas derivações: uma de EEG e outra de EOG. Mas, é opinião generalizada que há actividades do EEG mais visíveis determinada área: por exemplo alfa é mais visível na região occipital, fusos de sono (sigma) na região central, etc. Por outro lado, há também artefactos que são detectáveis por comparação do sinal entre canais: a influência dos canais de EEG no canal ocular e vice-versa. Por isso, parece vantajoso processar mais que um canal de EEG, simultaneamente com o canal ocular. Aliás, a análise simultânea e contextual de vários canais de EEG é procedimento normal e de rotina na análise visual destes sinais. Assim, e tal como referimos no capítulo anterior, detectaremos: beta, delta, teta e artefactos musculares num canal frontal; sigma num canal central; alfa num canal occipital; e finalmente os movimentos oculares no canal ocular.

Como referimos no capítulo anterior a detecção destas actividades é baseada no seguinte modelo de processamento: filtragem e reconhecimento de padrões (figura 3.1). O padrão pode ser simples, isto é, constituído unicamente por uma onda cujos parâmetros temporais (amplitude, período, etc) obedeçam a um determinado critério- decisão onda válida/não-válida; ou pode ser um padrão composto, isto é, constituído por um determinado número mínimo (grupo) de ondas válidas. Como foi referido, a operação de filtragem, facilita a determinação

dos parâmetros que conduzem à decisão de onda válida (ou não-válida): um simples algoritmo de determinação de picos e/ou passagens por zero.

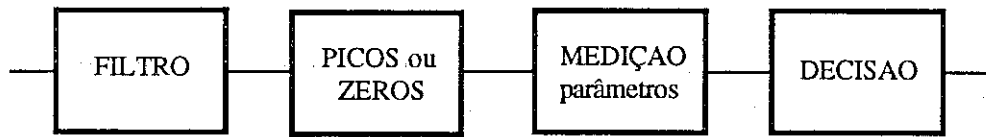


figura 3.1: Diagrama de blocos do modelo de processamento

Em resumo, e tal como é referido na literatura clínica, e mencionado no capítulo anterior: as actividades alfa, beta, sigma são definidas em "burst" (padrão composto), a actividade teta pode aparecer em grupo ou isolada, neste caso definimo-la como um padrão composto; e finalmente a actividade delta é definida tanto como uma actividade em grupo ou isolada, mas neste caso definimo-la como isolada (padrão simples); os movimentos oculares também são definidos como padrões simples. Para cada uma das actividades detectadas, além do instante de ocorrência propomos também armazenar os parâmetros que a definem para posterior análise. Por isso, no caso de actividades definidas como padrões simples serão armazenados os parâmetros característicos das ondas, no caso de actividades definidas por padrões compostos serão os valores médios (e eventualmente as variâncias) dos parâmetros, conjuntamente com a duração total.

Assim, podemos considerar como tarefas básicas, isto é, as que precisam ser executadas para todas as amostras do sinal de entrada: a aquisição, a filtragem e a integração da amostra no padrão simples (onda). As restantes tarefas são: decisão onda válida/não-válida no final de cada onda (ou seja após 2 picos do mesmo sinal ou duas passagens consecutivas por zero); no caso de padrões compostos devem também calcular-se os somatórios dos parâmetros (e somatório dos quadrados para o cálculo da variância); e finalmente quando for reconhecido o fim dum padrão

composto devem calcular-se as médias (e as variâncias) dos referidos parâmetros - cálculos estatísticos. Além disso, todos estes valores sendo detectados num preprocessador sem memória de massa, precisam ser enviados ao processador central que os armazena num ficheiro em disco, o que implica comunicação entre processadores.

2: Modelo de Processamento

Como estamos interessados em implementar estas tarefas em tempo real, o número de operações possível entre cada duas amostras consecutivas do sinal de entrada é um dado importante que condicionará a forma de implementação do modelo de processamento. Portanto, a escolha da frequência de amostragem além de condicionar o rigor das medidas temporais efectuadas no sinal, tal como já foi referido, também determina o modelo de cálculo. Das tarefas salientamos, por serem as mais morosas: a filtragem (pela possibilidade de envolver multiplicações e cálculos efectuados para todas as amostras do sinal de entrada) e os cálculos relacionados com a estatística que envolvem simultaneamente multiplicações e divisões (embora não sejam cálculos efectuados para todas as amostras do sinal de entrada).

Num processador com CPU 6809 da Motorola como o que foi referido no 1º capítulo, uma operação de multiplicação de 16x16 "bits" demora 120 ciclos, supondo um dos dados conhecidos (o coeficiente do filtro, por exemplo), endereçamento na página directa e truncagem para 16 "bits", uma operação de divisão de positivos de 48/16 "bits" com resultado em 24 (o quadrado do valor das amostras de entrada), demora 2600 ciclos pressupondo também endereçamento na página directa [Leventhal-1981, Davies-1979].

Por "hardware" podemos escolher uma das frequências de amostragem seguintes: 200, 250 e 1000Hz. A frequência de amostragem típica na análise do EEG é 200Hz, uma frequência aproximadamente igual à de Nyquist que não introduz erro de "aliasing" na zona de frequências do sinal EEG (aproximadamente 40Hz), se considerarmos que os artefactos podem ter componentes até 150Hz [Príncipe-1979b]. Por isso se escolhermos 200Hz, e quisermos detectar todas as actividades num só preprocessor temos aproximadamente 0.7ms para cada uma das actividades referidas anteriormente (num total de 7), isto é 700 ciclos relógio. Sendo assim, mesmo optando por filtros que não envolvam operações de multiplicação, a filtragem ocupa uma grande percentagem daquele tempo. Um filtro FIR de 10ª ordem demora 300 ciclos, com a estrutura apresentada no capítulo anterior e unicamente com 2 coeficientes diferentes da unidade. Resumindo, só a execução das tarefas básicas consomem praticamente todo o tempo de processamento, isto é, aproximadamente 500 ciclos para: filtragem, aquisição (com actualização do instante de ocorrência) e integração da amostra no padrão simples (comparação de 2 valores em memória e incremento dum contador). Com esta configuração torna-se de todo impossível efectuar qualquer estatística nos eventos detectados, ficariamos unicamente condicionados à detecção binária do eventos (pelo menos, nos padrões compostos), isto é poder saber se em determinado instante houve ou não um dado evento.

Mas a arquitectura do HIDRA apresentada no 1º capítulo sugere-nos naturalmente uma filosofia de processamento distribuído: isto é, podemos separar a análise dos diferentes canais de EEG/EOG por diferentes preprocessadores. Neste caso teremos 4 preprocessadores, cada um deles recebendo sinal dum eléctrodo diferente, portanto se considerarmos as mesmas actividades temos, tal como indica o diagrama de blocos da figura 3.2: detecção de 4 actividades (delta, teta, beta e artefactos musculares) no preprocessor 0 ligado ao eléctrodo frontal; 1 actividade

(sigma) no preprocessador 1 ligado ao eléctrodo central; 1 actividade (alfa) no preprocessador 2 ligado ao eléctrodo occipital; e, finalmente movimentos oculares no preprocessador 3 ligado ao eléctrodo ocular

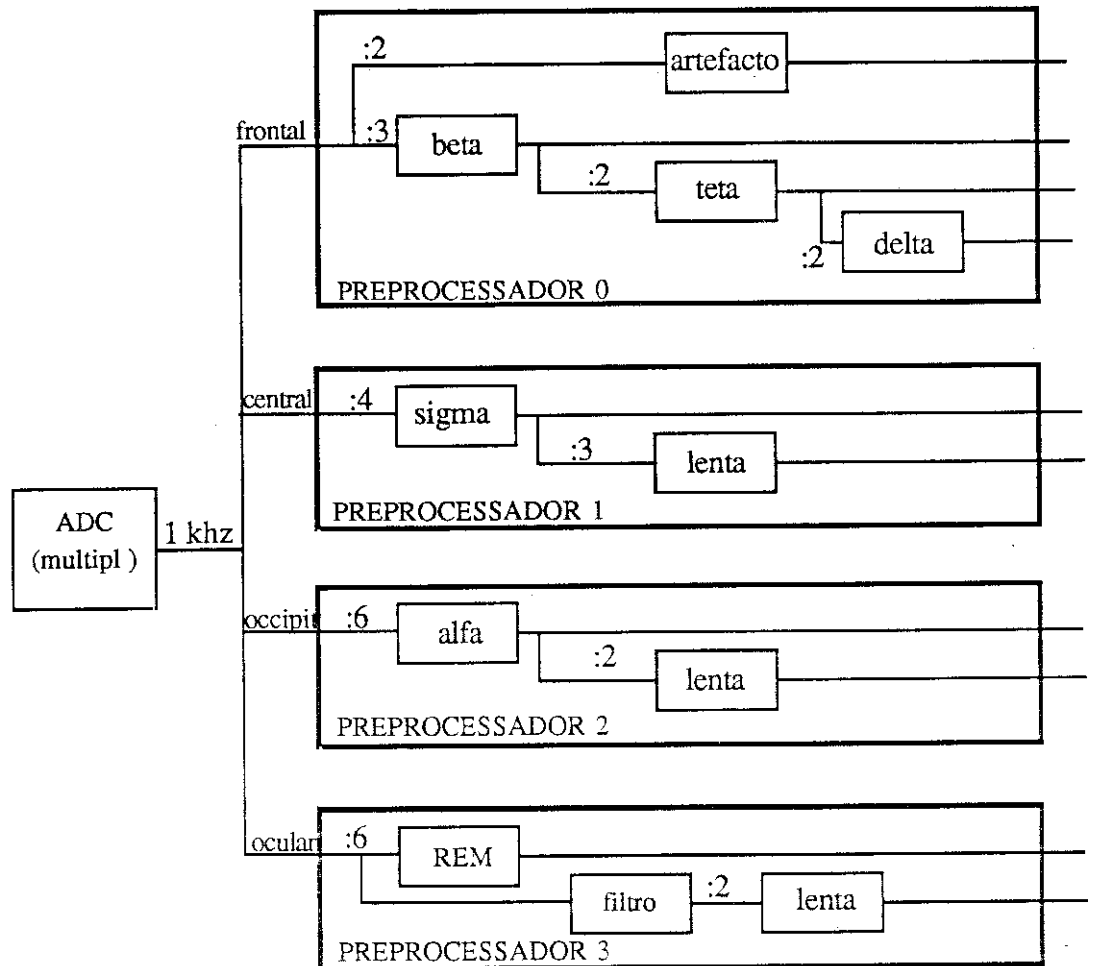


figura 3.2: diagrama de blocos do sistema de multiprocessamento com multifrequências de amostragem

Por isso, e para uma mesma frequência de amostragem (200Hz) teremos 1250 ciclos de relógio por tarefa no preprocessador 0, e 5000 nos restantes pre-processadores. Esta configuração (figura 3.2) torna possível o modelo de processamento em pelo menos 3 dos pre-processadores, pois 5000 ciclos seria suficiente para processar uma amostra e, no final de cada padrão composto, repartir

os cálculos estatísticos por alguns intervalos de processamento seguintes (dependente do número total de parâmetros, isto é do número de divisões necessárias). Além disso, tem-se ainda a hipótese de poder incluir nesses preprocessadores a detecção de ondas lentas que, como referimos pode ser útil na detecção de artefactos lentos e eventos que aparecem sincronamente em vários canais (os complexos K).

Mas antes de analisarmos este aspecto com maior detalhe, vamos deter-nos num aspecto também importante como é o da precisão nas medidas temporais efectuadas. Como esta frequência de amostragem (200Hz) e, tal como referimos no capítulo anterior, a percentagem de erro nas medidas temporais varia de: 15% (ondas beta) a 0.25% (ondas delta). Estes casos extremos sugerem-nos que a frequência de amostragem em beta deveria ser superior enquanto que em delta para manter uma percentagem de erro aceitável poderíamos amostrar a uma frequência inferior à frequência típica. Esta questão sugere-nos uma implementação com frequências de amostragem diferentes na análise de cada actividade: maior nas actividades da gama de frequência superior, e menor na gama inferior; no entanto, não podemos esquecer o erro de "aliasing", isto é, não deveremos permitir a influência deste erro na banda da actividade em estudo, o que implica colocar um filtro passa-baixo ("anti-aliasing") antes do bloco de processamento da actividade em questão, em vez de ler o sinal directamente da ADC. Por outro lado, teremos que implementar filtros passa-banda distintos, cada um centrado à sua frequência já que as actividades de EEG ocorrem a frequências diferentes. Contudo, se tivermos frequências de amostragem diferente, as especificações em frequência normalizada podem ser idênticas desde que se possa escolher a mesma largura de banda para todos eles. Portanto, este arranjo possibilita a utilização da mesma estrutura de programa (código e parâmetros) para mais do que uma actividade.

2.1: Encadeamento temporal das tarefas de processamento

Como vimos, a maneira prática de obter outras frequências de amostragem recorrendo só ao domínio digital é dividir ou multiplicar por um inteiro qualquer das frequências externas. Assim, a "slot" mínima de tempo deve ser escolhida de modo a que o algoritmo de detecção seja executável sem interrupção. Por outro lado, os cálculos têm que estar sincronizados com uma base de tempo externa responsável pela amostragem do sinal analógico. Por isso, resolveu-se utilizar o relógio de 1000Hz ("slots" de 1ms) para gerar "interrupts" o que cria intervalos de processamento de 1000 ciclos de relógio. Deste modo, em cada preprocessor os cálculos correspondentes às actividades do EEG têm que ser distribuídos por "slots" diferentes. Mas como em cada preprocessor são implementados vários detectores essa separação deve ser obtida naturalmente, isto é, sem grande sobrecarga de processamento. A solução adoptada atribui o processamento do detector A na "slot" i , o processamento do detector B na "slot" $i+1$, etc; conjuntamente devem existir "slots" livres onde, no final de cada onda válida, possam ser calculadas as estatísticas que não podem ser concluídas imediatamente após do padrão composto (figura 3.4). Resta agora seleccionar a periodicidade dos cálculos em cada detector, isto é, a frequência de amostragem para cada actividade, o que corresponde nesta implementação aos factores de decimação. Assim, no processador 1 escolhemos: 2, 3, 6, 12 para artefactos, beta, teta, e delta respectivamente; no preprocessor 2 : 4 e 12 para sigma e ondas lentas; no processador 3: 6 e 12 para alfa e ondas lentas; no processador 4: 6 e 12 para REM e ondas lentas. Em seguida, quando se descreverem os detectores, justificar-se-á esta escolha, que aliás não é única. O processador 0 está mais sobrecarregado em cálculos tendo por isso menor percentagem de "slots" livres que qualquer dos outros. No entanto esta solução permite distribuir uniformemente o tempo total de processamento facilitando

simultaneamente o encadeamento dos cálculos. De facto, quando surge uma amostra, o controle 'sabe' exactamente o que fazer pois cada "slot" está atribuída ao processamento, ou não tem qualquer processamento atribuído. No final do processamento de cada amostra ou nas "slots" livres, se existir alguma tarefa interrompida (estatística ou comunicação) o processador pode encarregar-se dela com o grau de prioridade que se pretender.

2.2: Organização das tarefas de processamento nos preprocessadores

Os valores das frequências de amostragem, da estrutura de processamento sugerida no diagrama da figura 3.2, para cada uma das actividades estão representadas na tabela 3.1, garantindo-se deste modo uma percentagem de erro $\leq 7\%$ na determinação das frequências das ondas. Na mesma tabela estão também representados parâmetros que caracterizam os respectivos padrões simples (com asterisco). Com esta escolha de frequências de amostragem, a relação entre a frequência central das actividades alfa, beta e sigma e as respectivas frequências de amostragem é aproximadamente a mesma. O que implica que pode ser utilizado o mesmo filtro digital. Note-se que em cada preprocessador é feito processamento a frequências de amostragem diferentes. Em princípio, este facto implicaria a implementação de filtros "anti-aliasing". Contudo, como à frequência mais elevada são necessários filtros passa-banda resolveu-se implementá-los com blocos passa-baixo e passa-banda em cascata, de modo que o bloco passa-baixo seja ao mesmo tempo filtro "anti-aliasing" dum detector com frequência de amostragem menor (figura 3.2). Assim, evitam-se os erros de "aliasing" sem qualquer sobrecarga no processo de cálculo. Nos casos delta, teta e REM, para terem a mesma frequência central normalizada, teríamos que diminuir a respectiva frequência de amostragem; no entanto isso implicaria filtros com valores de atenuação mais elevados a fim de

controlar o erro de "aliasing", pelo que resolvemos manter aqueles valores de frequência de amostragem. Como vemos na tabela, só a frequência de amostragem das actividades lenta e delta é inferior à frequência de amostragem típica na análise do EEG; pois embora os artefactos tenham componentes até 150Hz, com a frequência de amostragem de 166.6Hz não se introduz erro de "aliasing" nas zonas de frequência dos eventos alfa, lenta ou delta.

evento	amostragem (Hz)	frequência(Hz)		amplitude (μ v)	padrão	
		min	max		on	off
delta*	83.3	0.5	2.0	20.0	-	-
teta	166.6	3.75	7.5	15.0	3	3/1
alfa	166.6	8.0	12.0	5.0	6	6/3
sigma	250.0	11.75	16.0	5.0	6	6/3
beta	333.3	16.0	30.0	2.5	8	8/5
artefactos	500.0	35.0	120.0	10.0	-	-
REM *	166.6	0.5	5.0	20.0	-	-
lenta*	83.3	0.3	3.0	25.0	-	-

tabela 3.1: Definição dos parâmetros de detecção dos eventos. Actividades com * são definidas como padrões simples.

Resumindo, o modelo escolhido, para cada um dos preprocessadores, é uma estrutura de multifrequência em que frequência do "interrupt" é superior à frequência de amostragem para qualquer das actividades. Assim, o tempo total de processamento fica dividido em "slots" que serão distribuídas pelas diferentes tarefas: detecção (um detector em cada), estatísticas (dos padrões compostos) e comunicação (envio de dados ao processador central).

3: Descrição das tarefas de processamento

No processamento referido podemos distinguir 3 tarefas: detecção, estatísticas e comunicação com o processador central. A primeira (e simultaneamente de maior prioridade) é a detecção dos eventos, consequentemente sempre que haja amostra a processar, devem efectuar-se todos os cálculos relativos a essa amostra. Em seguida, no final de cada onda válida e se esta fizer parte dum padrão composto, devem efectuar-se cálculos estatísticos. Estes cálculos são demorados, levando mais tempo que o intervalo entre "interrupts". Por isso temos que repartir estes cálculos por "slots" de modo a que o sistema possa responder correctamente aos requisitos da tarefa de detecção com a frequência pretendida para cada uma das actividades. Por último a terceira tarefa (e última prioridade) consiste em enviar os dados para a memória central comum. Também neste caso o processo pode ser longo (o bus pode estar ocupado por outro preprocessor) sendo assim logo que surja o sinal de "interrupt", as informações relativas ao estado dos dados a enviar são guardadas, regressando ao programa de controlo de tarefas (programa principal).

Como foi referido, cada preprocessor recebe um "interrupt" com frequência de 1kHz. A resposta a este sinal no processador 6809 é vectorizada, isto é, logo que este sinal seja activado o processador vai executar um programa cujo endereço está contido numa palavra de memória. Neste caso, este programa faz simplesmente a leitura da ADC, guarda esse valor numa posição de memória, e actualiza o instante de ocorrência (um relógio: horas, minutos, segundos e um contador módulo 1000). No final da rotina o processador regressa ao local onde foi interrompido, executando a instrução RTI. Por estas razões, nós optamos por uma estrutura de controlo *a posteriori*, isto é, a tarefa interrompida é acabada e em seguida por consulta da unidade maior precisão do relógio (o contador módulo

1000) verifica-se se o sinal já esteve activo no passado recente. Outra alternativa seria garantir que o processamento em todas as "slots" seria sempre inferior a 1000 ciclos e, colocar o processador num ciclo de espera activa ("busy wait") à espera do referido sinal - instrução CWAI #XX. A primeira hipótese é mais eficiente sobretudo por a tarefa de comunicação ser assíncrona, sendo deste modo impossível atribuir-lhe uma determinada fatia de tempo. Além disso, a primeira tarefa pode ter duração muito variável, com este esquema o processador pode, após a conclusão da tarefa, passar à execução das outras tarefas até que surja o sinal de "interrupt".

Como foi mencionado na secção anterior, em cada preprocessor, é implementado mais do que um detector cujo processamento deve estar atribuído a "slots" de tempo distintas. No preprocessor 1 são detectados eventos com frequências de amostragem de 250Hz e 83.3Hz, isto é, respectivamente, de 4 em 4 e 12 em 12 "interrupts" externos. A sequência de cálculo pode ser também controlada por ordem de chegada dos "interrupts", como mostra a figura 3.3, por simples encadeamento de testes.

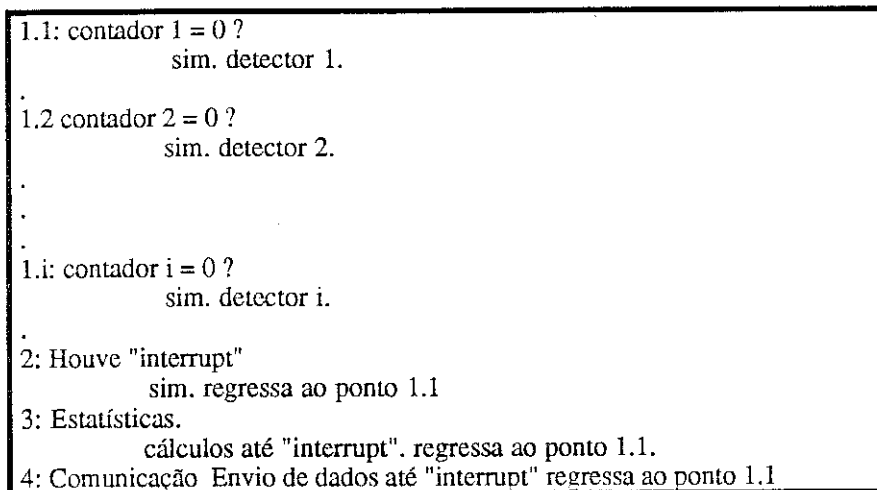


figura 3.3: Estrutura do programa principal com i detectores. Os contadores têm módulo igual aos factores de decimação e não se anulam ao mesmo tempo.

O processamento dos eventos pode efectuar-se por controle de dois contadores com a sequência seguinte: respectivamente $4n$ e $12n+2$, onde n é um número positivo. Como facilmente se verifica aquelas sequências (que não são únicas) nunca têm o mesmo valor, sendo o controlo do tempo de processamento também conseguido duma maneira muito simples. Nos preprocessadores 2 e 3 temos uma situação idêntica: dois eventos a detectar. Mas como a frequências de amostragem são diferentes a sequência de cálculo pode ser: $6n$ e $12n+2$. No preprocessador 0 tem 4 eventos a detectar, e um deles (artefactos) com frequência bastante elevada ($f_s = 500\text{Hz}$). Não é pois possível separar totalmente as detecções ao longo do tempo mas, excluindo os artefactos podemos calcular os outros eventos com a sequência seguinte: $3n$, $6n+1$, $12n+4$, respectivamente. Estes sequências que embora não sejam únicas nunca têm o mesmo valor, e os artefactos são calculados para a sequência $2n$ que, naturalmente, pode coincidir com qualquer das outras com valor par. A sequência de cálculo distribuída pelas "slots" de processamento está representada na figura seguinte.

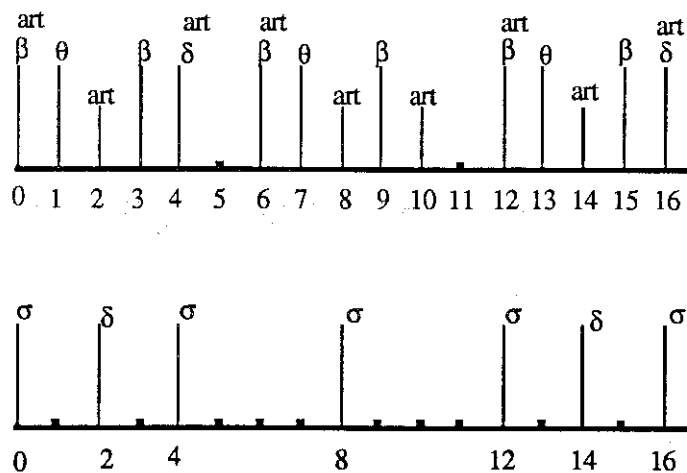


figura 3.4: Sequência de cálculo das amostras, nos preprocessadores 0 e 1, respectivamente. Nas "slots" livres (representadas por um ponto) podem ser atribuídas a qualquer das outra tarefas: estatísticas ou comunicação.

Resumindo, todo este processamento é controlado pelo decremento de contadores (com módulo igual ao coeficiente do n) e convenientemente inicializados. As outras tarefas (estatísticas e comunicação) são executadas em intervalos de tempo (de 1ms) em que não haja os cálculos relativos às amostras ou quando estes já foram finalizados sem ter havido "interrupt" (figura 3.4). Por isso nestas tarefas a leitura da "flag" de "interrupt" é feita a frequência muito mais elevada que a frequência daquele sinal externo, como veremos em seguida.

3.1: Detectores

Como tem sido referido, as decisões na detecção dos eventos dependem da definição de onda válida, isto é, de valores de parâmetros temporais da onda. Na tabela 3.1 estão indicados os valores dos referidos parâmetros para cada actividade (amplitude e período), excepto no caso de REM em que se considerou mais um parâmetro: o declive do 1º flanco da onda. A amplitude é, nos padrões compostos, a diferença entre 2 picos consecutivos de sinais contrários; enquanto que nos padrões simples a amplitude é o valor dum pico (positivo ou negativo). O período é nos padrões compostos o tempo que medeia dois picos positivos; enquanto que nos padrões simples é o tempo que medeia dois zeros consecutivos do sinal; e finalmente, o declive é o quociente do valor de pico da onda pelo tempo que medeia a passagem por zero até atingir o pico. Por isso, no caso de padrões compostos basta detectar os picos das ondas enquanto que nos outros casos devem também detectar-se as passagens por zero. Nos padrões compostos existe ainda um segundo critério que verifica a integração dessa onda num grupo de ondas. Esse critério, como veremos em seguida, é diferente na detecção de início e de fim de evento.

Para este processamento, e para cada uma das actividades, existem 2 blocos de programa distintos: PADRÃO e PICO. Em PICO faz-se a leitura dos referidos parâmetros; em PADRÃO tomam-se as decisões: verifica-se se os parâmetros definem uma onda válida (padrão simples) e se esta se integra num padrão composto considerando duas fase distintas (início e fim de evento). Portanto, para a maioria das amostras do sinal de entrada (filtrado) aquele processamento resume-se a um teste (a amostra continua a crescer?) e um incremento de contador (o período da onda). Deste modo, o bloco PADRÃO só necessita ser executado no início (inicialização dos parâmetros) e no fim (decisões onda válida/não-válida). Em PICO os parâmetros são medidos efectuada dividindo as instruções de acordo com a forma do sinal isto é: uma zona de código refere-se à zona positiva ascendente, outra à positiva descendente, etc. Deste modo evitam-se um conjunto de testes a "flags" que contenham a história passada da onda. Assim, a passagem do programa principal para o bloco PICO (e o inverso) é feita por endereço com a instrução EXG PC,Y como mostra o diagrama de blocos da figura seguinte

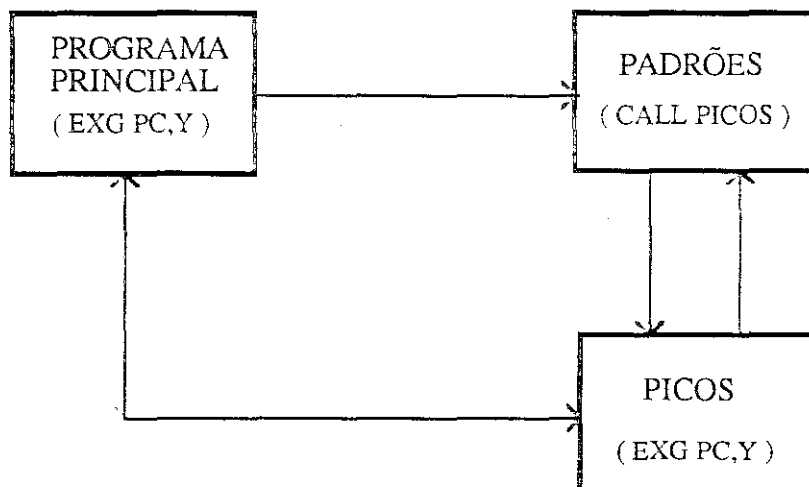


figura 3.5: diagrama de blocos dum detector, e estrutura de comunicação com programa principal.

Por isso, quando do regresso ao programa principal, o conteúdo do registo Y é o endereço da zona de código que integra o valor da amostra na onda (crescente positiva, por exemplo). No final de cada onda, na zona de código de PICO que descreve esta situação, existe a instrução RET que passa o controlo ao programa PADRÃO. À primeira vista pode parecer estranho passar o controlo para a rotina com a instrução RET em vez de CALL. Contudo este método é bastante usado na implementação do KERNEL de multiprocessamento exigindo zonas de "stack"/dados independentes. Como em cada preprocessor estão implementados vários detectores, esta técnica só resulta se houver uma zona de "stack" para cada detector. Isto porque a instrução RET pressupõe que no topo da referida zona esteja um endereço que neste caso tem que ser o de PADRÃO, existindo vários detectores (vários blocos PADRÃO) é impossível garanti-lo com uma única zona de "stack". Esta técnica traz sobretudo vantagens no caso de PADRÃO se referir a uma actividade descrita em padrão composto porque o regresso a este programa não se faz sempre para a mesma zona de código, isto é, pode ser feita em duas situações distintas: início e fim de actividade. Assim aquela instrução RET pode representar dois endereços diferentes, mais uma vez é a zona de código que indica a evolução da actividade e não o teste duma "flag". No caso de se tratar dum padrão simples, já poderia ser implementado de outro modo, pois o regresso (RET) faz-se sempre para o mesmo endereço. Neste caso poderia programar-se de modo convencional, isto é, a chamada a PADRÃO com CALL (em vez de RET), com a única diferença de que a instrução RET (agora neste bloco de código) seria executada no processamento da presente amostra portanto sem haver necessidade de preservar nenhum endereço em "stack". Por questões de uniformização resolvemos manter a primeira técnica (até porque não envolve sobrecarga no tempo de processamento, são executadas as mesmas instruções). Além disso, em qualquer dos casos, no

programa principal é sempre possível guardar o conteúdo do registo Y (endereço de PICO) no respectivo "stack", em vez de utilizar qualquer outro processo de guardar o referido endereço.

No final de cada onda válida, e no caso de pertencer a um padrão composto, o tempo de cálculo aumenta significativamente, pois devem calcular-se somas cumulativas de todos os parâmetros, e respectivos quadrados. O tempo de cálculo aumenta pois aquelas operações pressupõem: somas cumulativas de valores de 8 "bits" e respectivos quadrados (88 ciclos) para o período, cálculos idênticos para amplitude, valores de 12 "bits" (250 ciclos); o que conjuntamente com a validação dos referidos parâmetros e as tarefas básicas pode somar 1200 ciclos, se o código executado (da aquisição, por exemplo) fôr o do "loop" mais longo. Mas, como referimos, esta situação é pontual, corresponde ao processamento da última amostra do padrão simples (onda). Nesta estrutura não há qualquer problema se na "slot" de processamento seguinte se conseguir recuperar o atraso no processamento desta amostra. Nos preprocessadores 1 e 2 essa "slot" está livre, sem processamento de amostra de sinal, portanto este intervalo será atribuído à tarefa estatística (ver figura 3.4). No processador 0, esta estratégia não se aplica por duas razões: por o processamento dos artefactos poder coincidir com os cálculos da última amostra de onda beta (1500 ciclos), e para além disso na "slot" seguinte pode ocorrer o processamento da amostra doutro detector (teta ou delta)-figura 3.4. Note-se que no preprocessador 0 existe um número inferior de "slots" livres embora algumas delas tenham unicamente o processamento de artefactos (aproximadamente 250 ciclos) Por estes motivos, são unicamente calculadas médias dos parâmetros de beta, e para as outras actividades definidas em "burst" (teta e artefactos musculares) são unicamente calculadas as respectivas durações totais.

3.2: Estatísticas

No final de cada padrão composto (beta, sigma ou alfa), são calculadas médias e variâncias dos parâmetros das referidas ondas; a partir das somas cumulativas, actualizadas no final de cada onda válida do referido padrão, dos valores e dos quadrados dos mesmos valores. A média é o quociente da soma dos valores pelo número total de ondas. E a varância é quociente da soma dos quadrados pelo número de ondas menos o quadrado da média.

$$\begin{aligned} m_e &= \frac{1}{N} \sum x_i \\ \sigma &= \frac{1}{N} \sum x_i^2 - m_e^2 \end{aligned} \quad (3.1)$$

Como tem sido mencionado, a única particularidade deste algoritmo consiste no facto de os cálculos não poderem ser feitos consecutivamente. Assim, e porque o processamento da amostra de sinal anterior ocupa nalguns casos mais que os 1000 ciclos de relógio resolvemos neste bloco controlar o sinal de "interrupt" a uma taxa mais elevada que a própria ocorrência daquele sinal externo. Portanto, além de resolver a questão levantada na secção anterior, este processo permite também que estes cálculos ocorram não só nas "slots" de tempo livre mas também no tempo que sobra do processamento das amostras, já que como tem sido mencionado, para a maioria das amostras do sinal não é esgotado o respectivo tempo. Optamos por, em cada iteração da operação de divisão (isto é aproximadamente por cada 120 ciclos de relógio) saber se num passado recente houve um sinal de "interrupt" (consulta da unidade de maior precisão do relógio): Logo que este sinal seja detectado, este processamento é interrompido deixando os cálculos de modo a poderem ser retomados em próxima oportunidade. No entanto, os cálculos são suficientemente rápidos, pois mesmo no pior dos casos

(imediatamente após o fim de evento), sendo detectada uma nova onda do mesmo evento, a zona de memória onde as somas cumulativas são armazenadas pode ser re-inicializada. Embora não seja prática comum em programas de tempo real, a zona de memória das somas cumulativas é partilhada pelos programas PADRÃO e de cálculo das estatísticas, para que seja possível usar eficientemente a memória local dos preprocessadores. Segundo os nossos cálculos e prevendo a pior das hipóteses, isto é, dois "bursts" de actividade consecutivos, o tempo mínimo que os medeia é suficiente para o cálculo de todos os valores estatísticos referentes ao primeiro "burst". Senão vejamos, no caso do detector de SIGMA o tempo mínimo entre dois "bursts" é de 62ms, ou seja 62 "slots" temporais, portanto, neste intervalo existem pelo menos 40ms livres de qualquer processamento de amostras, pois durante este tempo são processadas 16 amostras sigma e 6 amostras de actividade lenta. Como foi referido, a divisão é a operação mais longa que no pior dos casos demora 2,6ms (2600 ciclos). Nestes cálculos, são necessárias 4 divisões que é sensivelmente um quarto do tempo disponível. No caso do processador 0, no detector BETA, só existem 2 divisões e embora não haja tantas "slots" livres (4) a maior parte são ocupadas com o processamento de artefactos (8), por isso o cálculo estará concluído antes de haver novamente detecção de nova onda.

3.3: Comunicação

Os parâmetros de todos os eventos são enviados para o processador central. Para saber se o "bus" está livre testa-se um "bit" numa posição de memória tal como se se tratasse dum periférico. Este teste pode ser feito para cada "byte" enviado, mas é necessário testar previamente a flag de "interrupt", se esta estiver activa a comunicação é interrompida e o controlo regressa ao processamento da amostra. Este teste é feito a uma taxa mais elevada (sensivelmente 10) que a taxa de

"interrupts", no caso de enviar um "byte" . Por isso, também esta tarefa pode ser executada logo que o processamento da amostra esteja concluído não ter surgido aquele sinal.

Os dados a enviar são sempre colocados num "buffer" por um dos blocos de processamento mencionados, a ESTATÍSTICA (padrões compostos) ou PADRÃO (padrão simples), pois como tem sido referido esta tarefa é assíncrona, não se pode saber à partida quando é que o processador tem acesso ao "bus". Assim, o algoritmo pressupõe a existência de dois "buffers" circulares: um na memória local do preprocessor e outro na memória central a que o mesmo processador tem acesso. O "buffer" da memória central está dividida em páginas de 256 "bytes", 255 para dados e 1 (o primeiro) que é uma "flag" onde o preprocessor escreve FF logo que a referida página esteja completa. Assim que o processador central detecte este facto guarda os dados em disco e inicializa a referida "flag". Evidentemente, que o processador local só volta escrever nessa página quando a "flag" estiver a zero. No "buffer" local não há qualquer controlo "buffer cheio", pressupõe-se que a transferência de dados para a memória central é suficientemente rápida para que essa situação nunca se verifique. Portanto, existindo neste "buffer" um ponteiro para escrita e outro para leitura, um teste entre ambos é condição suficiente para verificar se existem dados a enviar.

3.4: Estrutura dos ficheiros de dados

Cada ficheiro de dados é constituído por blocos de 256 "bytes", sendo o primeiro "byte" a informação de canal , o número do preprocessor, e os restantes 255 são os vectores de parâmetros referentes a cada uma das actividades. Estes vectores contêm a informação seguinte: instante de ocorrência (T_i , em hora, minuto e segundo, a partir do início da análise), valores dos parâmetros de detecção (média

e variância no caso das detecções em "burst") e uma palavra de 16 "bits" que identifica a actividade em questão.

T _i	duração	amplitude	amp. var.	período	per. var.	FFFF
(4 bytes)	(2 bytes)	(2 bytes)	(3 bytes)	(1 byte)	(2 bytes)	fim evento

figura 3.6: Estrutura de dados da actividade sigma no ficheiro.

Assim, para guardar a informação respeitante a eventos detectados em "burst" são necessários 16 "bytes" (excepto em beta e teta com 11 e 9, respectivamente) para ondas isoladas são necessários 9 "bytes". Na figura 3.6 está representada a estrutura de dados para actividade sigma.

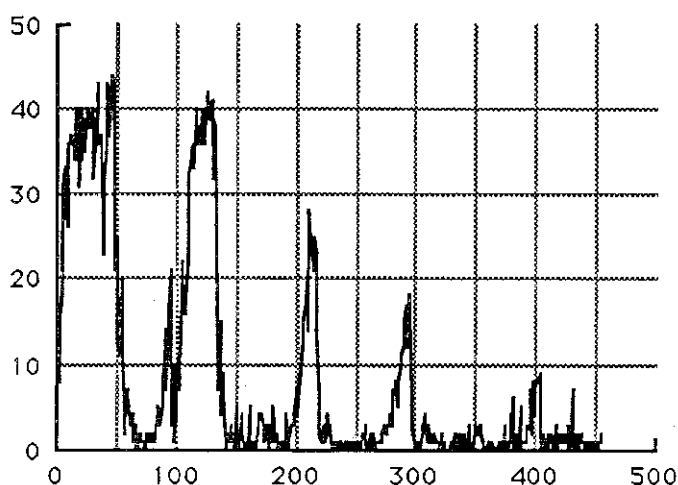


figura 3.7: Evolução da actividade delta ao longo da noite (segundos/minuto) em t4.

O ficheiro têm para cada caso tamanhos diferentes, dependendo naturalmente da quantidade de informação detectada. O tamanho é particularmente sensível à quantidade de actividade lenta de média e grande amplitude uma vez que esta banda está a ser detectada em todos os canais. Assim, são armazenados em média 30Kbytes/hora nuns casos, e noutros mais do dobro (70Kbytes/hora).

Foram desenvolvidos programas (em linguagem C) para tratar esta informação. Um programa para separar a informação por canal (1 ficheiro binário para cada preprocessor). O segundo programa descodifica os valores (período e duração em ms e amplitude em μv) das referidos ficheiros a partir das estrutura de dados para cada actividade e dos valores de calibração da amplitude e frequência de amostragem. Nesta fase são criados ficheiros ASCII, um para cada actividade, portanto um total de 10. Finalmente, o último programa tem por objectivo a apresentação de resultados ao utilizador tal como mostra a figura 3.7. Pode assim avaliar-se a evolução duma actividade ao longo da noite: segundos duma actividade/minuto, número de "bursts"/ minuto, etc. Podem fazer-se histogramas e estatísticas de qualquer parâmetro de detecção para qualquer actividade. É também a partir destes ficheiros que se calcula a descrição ao minuto do EEG/EOG, que serve de base à classificação do sono em estados e que será motivo de estudo do capítulo seguinte.

4: Especificação e teste dos filtros

As actividades a detectar têm frequências em intervalos diferentes na gama [0-30Hz], portanto com frequência central diferente, sendo a largura de banda é aproximadamente a mesma para algumas actividades (tabela 3.1). Esta é outra das vantagens de se utilizar uma estrutura em multifrequência, pois se a frequência central normalizada for a mesma em todas as actividades, e se a largura de banda do filtro se adaptar a todas as actividades, significa que podemos utilizar a mesma estrutura de filtragem e os mesmos coeficientes. Naturalmente que isso acontece se a relação entre as frequências centrais das actividades for idêntica à relação entre as respectivas frequências de amostragem. E de facto, isso acontece nas actividades da gama de frequência superior : alfa, sigma e beta (embora a largura de banda desta

actividade seja maior). Nas actividades da gama inferior (teta, lenta e delta) aquela relação já não se verifica, no entanto nestas actividades (de frequência baixa) estamos sobretudo interessados em eliminar as altas frequências que lhes estão sobrepostas. Assim, optámos por efectuar, nestas actividades, filtragem passa-baixo, com uma largura de banda normalizada que se adapte a qualquer das actividades. Consequentemente, as frequências normalizadas podem agrupar-se em dois intervalos, passa-banda e passa-baixo: as frequências de alfa, beta, e sigma no intervalo $[0.09\pi, 0.18\pi]$; e as actividades delta, teta e lenta no intervalo $[0.0, 0.09\pi]$, respectivamente. Por outro lado queremos aproveitar a filtragem dum detector como decimador de outro detector como foi referido anteriormente (figura 3.2). Isto é, o sinal filtrado para uma actividade de frequência superior pode ser a entrada doutro detector desde que aquela filtragem seja por um filtro passa-baixo. O filtro passa-banda com as características referidas deve ser implementado com dois blocos em cascata: um passa-baixo e outro passa-alto. Se o bloco passa-baixo tiver as características do outro grupo referido, conseguimos com uma única estrutura resolver completamente o problema de filtragem. Assim, as bandas de passagem (definidas a 3dB na arcada principal) dos blocos passa-banda e passa-baixo devem ajustar-se, respectivamente aqueles intervalos. A largura das bandas de passagem e transição é igual à largura da arcada principal, que no caso do filtro passa-baixo está dependente da colocação do primeiro zero na circunferência unitária. Por sua vez, os valores de atenuação na banda de atenuação são conseguidos pela localização dos zeros na referida circunferência. Por outro lado, essa colocação vai determinar os valores dos coeficientes do filtro que queremos representáveis em somas de potências de 2. Considerando ainda que o filtro passa-baixo é um decimador com factor de decimação M , sabemos que as zonas de frequência do sinal entrada que interferem com a banda base do sinal decimado estão centradas em $2\pi/M$. Neste caso, o factor de decimação tem valor 2 (ou 3), por isso devemos

ter valores de atenuação elevados na gama de frequências centrada em π (ou $2\pi/3$) e com largura igual à banda de passagem do sinal decimado. Para que a largura de banda a 3 dB seja 0.09π , a colocação do primeiro zero deve ser em $\pi/3$. Assim, para o bloco passa-baixo optámos pela seguinte colocação de zeros: $\pi/3$, $\pi/2$, $2\pi/3$, $5\pi/6$ e π . Para uma implementação em cascata obtemos a seguinte função de transferência

$$H_1(z) = (z^{-2} + z^{-1} + 1)(z^{-2} - z^{-1} + 1)(z^{-2} + 1.75z^{-1} + 1)(z^{-2} + 1)(z^{-1} + 1) \quad (3.2)$$

O módulo da função de transferência está indicado na figura 3.6, a banda de passagem a 3dB é $[0, 0.12\pi]$, e a atenuação tem valores de pelo menos 30dB no intervalo $[\pi/3, \pi/2]$ e pelo menos 40dB para valores superiores.

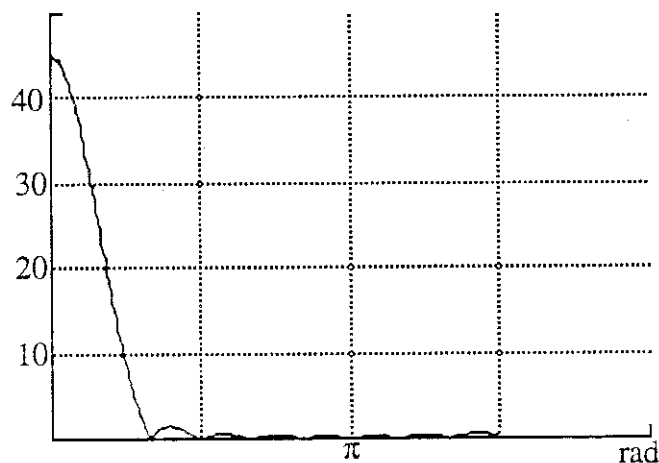


figura 3.8: Módulo da resposta em frequência do bloco passa-baixo

Embora, estes filtros tenham "ripple" elevado na banda de atenuação para $N=9$, o erro de "aliasing" resultante da decimação do sinal para frequências inferiores à de Nyquist pode ser desprezado. Senão vejamos o espectro dos sinal no preprocessador 0 (beta, teta e delta). Suponhamos que o sinal EEG e artefactos têm espectro sempre diferente de zero na banda de frequência 0 a 150Hz, o espectro do

sinal à saída do passa-baixo do bloco BETA está representado na figura 3.7 (a traço grosso). Após a decimação vai haver sobreposição, da primeira repetição do espectro, na banda base (na figura a traço fino), aproximadamente até ao valor de 17Hz.

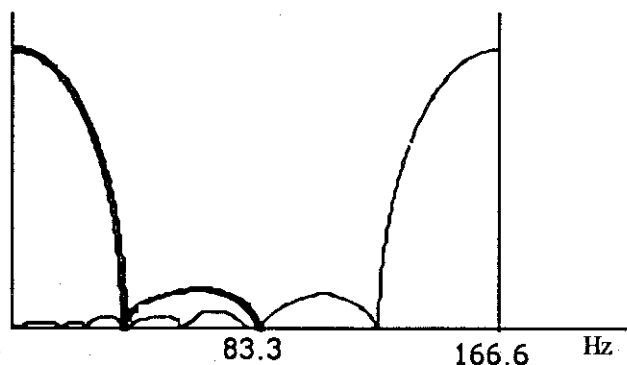


figura 3.9: Espectro do sinal à entrada do detector TETA.

Mas neste caso a banda de interesse do sinal de saída é agora de 0 a 10Hz (banda teta incluída), onde de qualquer modo o erro de "aliasing" é praticamente nulo por influência do zero colocado em 166.6Hz (π). Por sua vez este sinal vai ser novamente filtrado pela mesma estrutura com metade da frequência de amostragem, em que os zeros vão surgir para valores de frequência metade dos anteriores. O que vai anular zonas do espectro onde o erro de "aliasing" pode ser significativo (acima de 40Hz) e onde o filtro anterior não era tão selectivo (de 55.5 a 83.3Hz, por exemplo). A eliminação desta zona de frequência é importante para se poder voltar a decimar, como se torna necessário para implementar o detector DELTA. Idêntico raciocínio pode ser aplicado para este detector, assim como para os detectores dos outros preprocessadores.

Resumindo, mudando unicamente a frequência de amostragem e usando uma única estrutura de filtragem com características de atenuação fracas, podemos evitar o erro de "aliasing" ao efectuar operações de decimação, desde que a

colocação dos zeros seja estrategicamente escolhida. Ao mesmo tempo é uma maneira fácil de conseguir filtragem numa banda mais estreita, o que, com estes filtros, seria dificilmente concretizável com a mesma frequência de amostragem para todas as actividades (pelo menos com multiplicações substituídas por um pequeno número de adições e "shifts" e mantendo a ordem do filtro baixa)

O bloco passa-alto é implementado por:

$$H_2(z) = (z^{-1} - 1) \quad (3.3)$$

O módulo do filtro passa banda será a cascata $H(z) = H_1(z)H_2(z)$ e está representado na figura 3.8. A banda de passagem deste filtro, definida a 3dB, é $[0.07\pi, 0.2\pi]$. Além da anulação da componente d.c., a atenuação na banda superior de atenuação é de pelo menos 20dB.

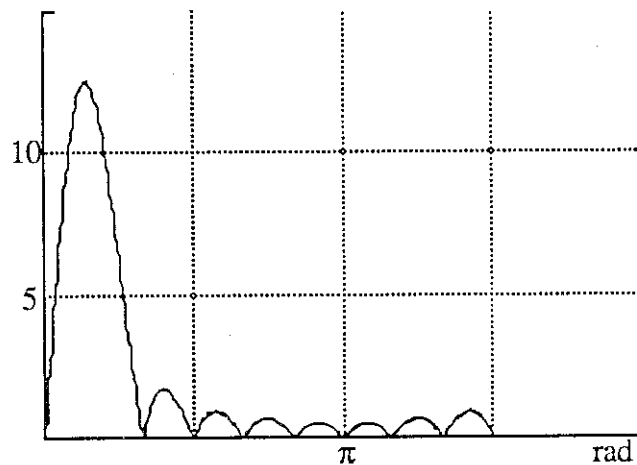


figura 3.10: Módulo da resposta em frequência do filtro passa-banda.

As respostas impulsionais das estruturas passa-baixo e passa-banda estão representados na figura 3.9. O bloco passa-baixo tem resposta impulsionial simétrica (portanto fase linear) e o passa-banda anti-simétrica (o que corresponde a fase "piecewise" linear).

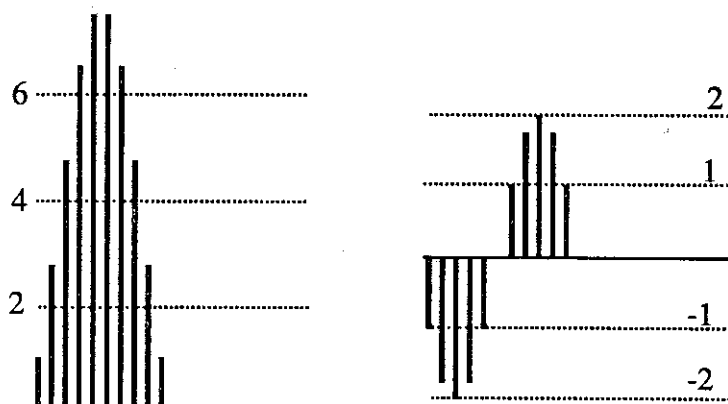


figura 3.11: Resposta impulsional dos filtros passa-baixo e passa-banda, respectivamente.

Como foi referido, estes filtros são implementados com várias frequências de amostragem, na tabela seguinte estão representados as características de frequência analógica destas estruturas de filtragem para cada uma das frequências de amostragem (e as respectivas actividades).

amostragem (Hz)	filtro	banda de passagem (Hz)	ordem
333.3	passa-baixo	0-20	9
	passa-banda (beta)	11-38	10
250.0	passa-baixo	0-15	9
	passa-banda (sigma)	9-28	10
166.6	passa-baixo (teta)	0-10	9
	passa-banda (alfa)	6-18	10
83.3	passa-baixo (delta)	0-5	9

tabela 3.2: Características dos filtros em frequência não normalizada

As saídas de filtros passa-banda (actividade alfa) e passa-baixo (actividade delta), para sinais dentro e fora da banda de passagem, estão representados nas figuras 3.13 e 3.14, respectivamente.

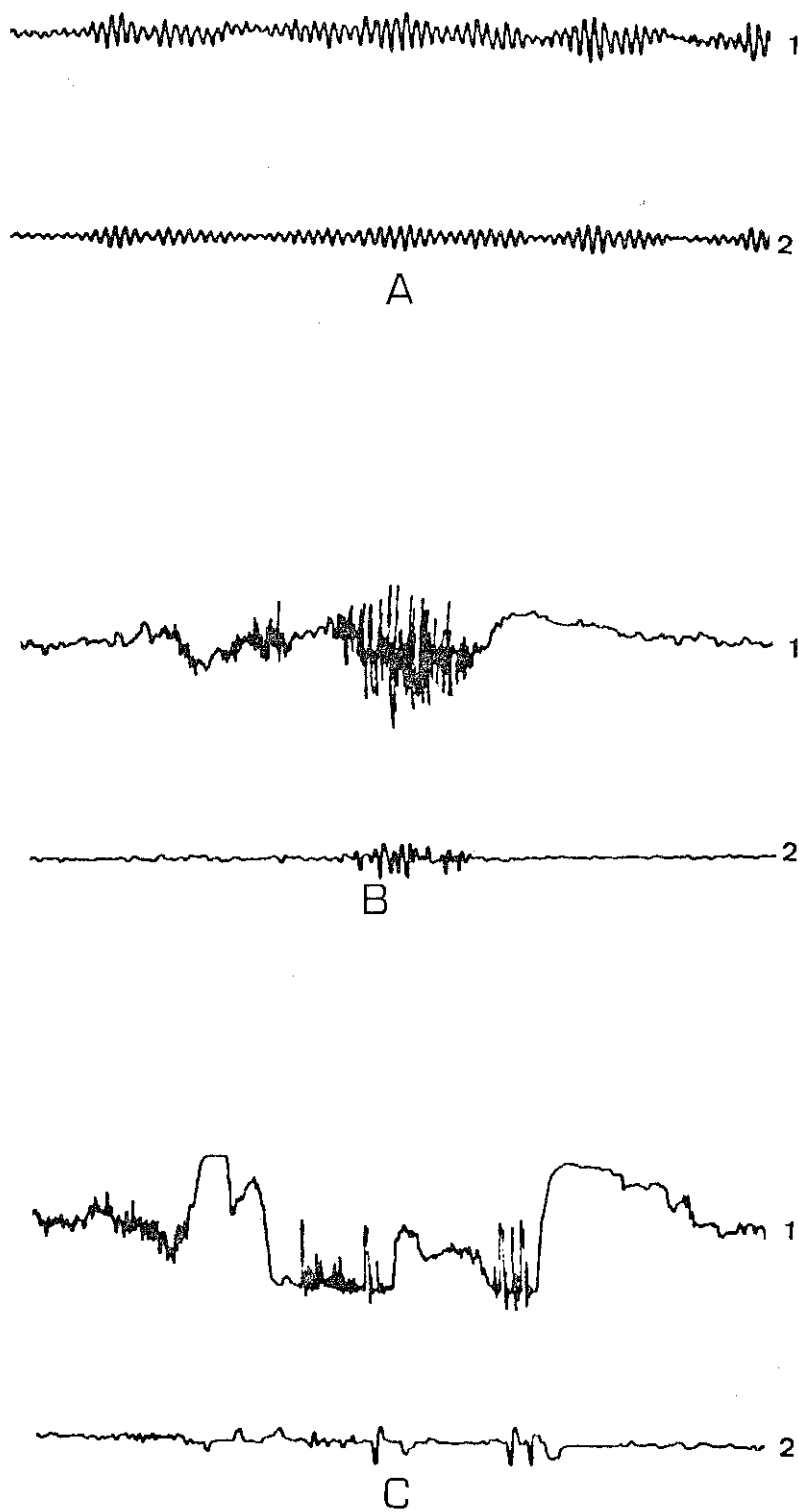


figura 3.12: Saída do filtro passa-banda (atividade alfa) para sinais de entrada: A- na banda do filtro. B- alta frequência. C- baixas e altas frequências. 1- sinal de entrada; 2-sinal de saída.

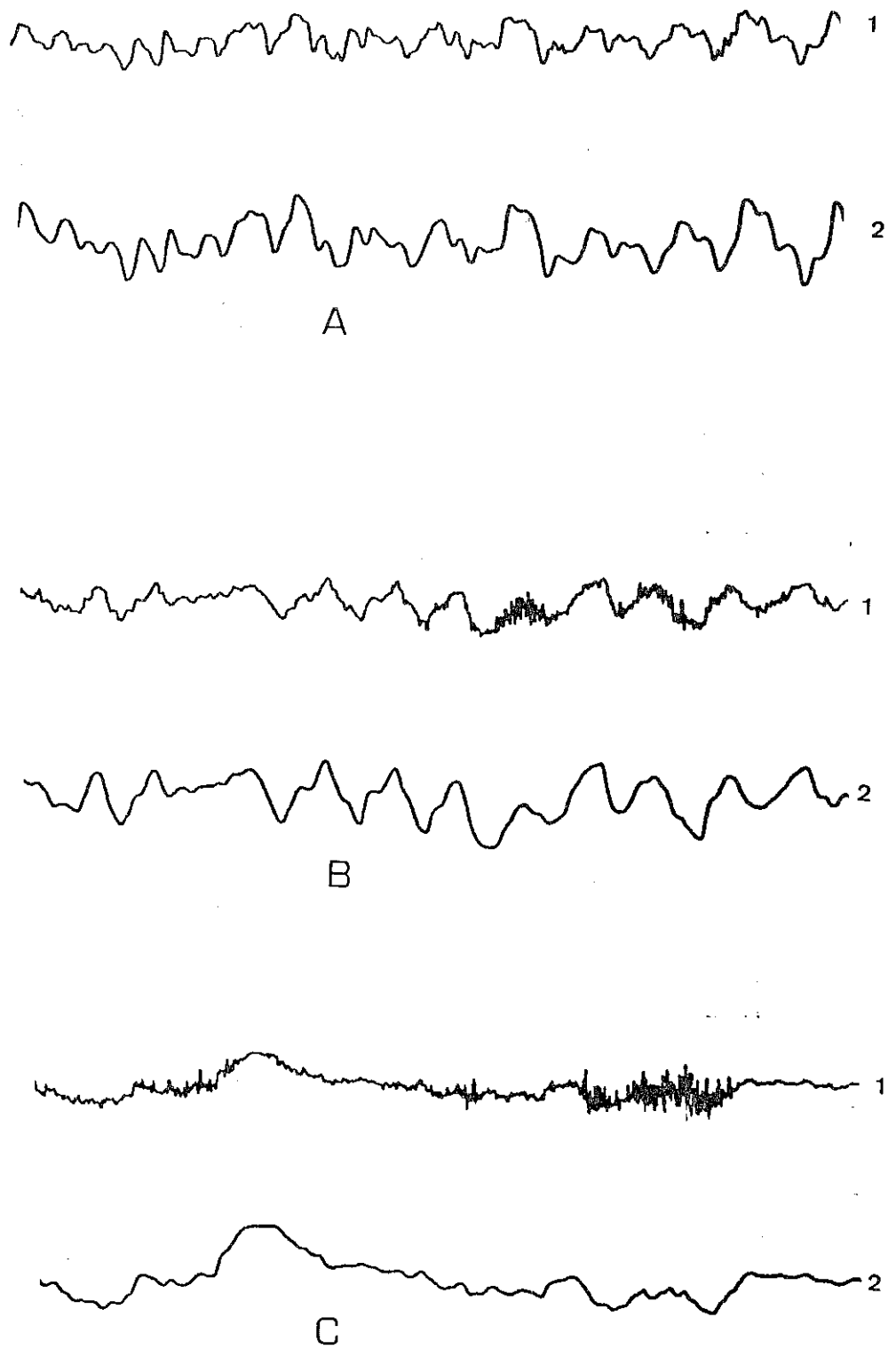


figura 3.13: Saída do filtro passa-baixo (atividade delta) para sinais de entrada: A/B- sinal na banda do filtro com atividade fora da banda sobreposta. C- altas frequências. 1- sinal de entrada; 2-sinal de saída.

Para a detecção dos artefactos musculares e da actividade REM escolhemos filtros com estruturas diferentes. O filtro para os artefactos é um diferenciador cuja função de transferência é

$$H_3(z) = z^{-4} - 2z^{-2} + 1 \quad (3.4)$$

Nas ondas REM não estamos interessados em fazer filtragem uma vez que queremos detectar flancos acentuados no sinal por isso o sinal passa unicamente por um bloco com um zero em $z = -1$.

4.1: Erros na implementação dos filtros.

Como já foi referido o sistema tem uma placa de aquisição com 12 "bits". Sendo assim, e seguindo, indicações presentes na literatura sobre o assunto [Príncipe-1979], resolvemos implementar os filtros com a entrada de 12 "bits" mas comprimento de palavra de 16 "bits". Além do erro de quantificação da entrada, e uma vez que se trata duma implementação em vírgula fixa, os erros têm origem nas multiplicações e nas adições se o "overflow" não for evitado. As operações são efectuadas em complemento para 2, como sabemos tratando-se de adições pode haver "overflow" nas somas parciais, no entanto o resultado final será correcto se for representável nesse comprimento de palavra. Então para garantir que não há "overflow" num filtro em cascata basta impor restrições na entrada e na saída dos blocos. Como facilmente se verifica pelo módulo da função de transferência, os filtros têm ganho elevado. No caso do bloco passa-baixo esse ganho ultrapassa mesmo o permitido pela diferença entre o comprimento de palavra da entrada e o do filtro (16). Devemos por isso implementar uma estrutura de "scaling". Partindo do princípio que o sinal EEG é um sinal quasi-sinusoidal escolhemos valores e

posições de "scaling" na estrutura do filtro relacionados com módulo da resposta em frequência dos blocos. Além disso, para facilitar a implementação escolhemos potências negativas de 2, isto é "shifts" à direita dos valores. Naturalmente que esta operação não deixa de ser uma multiplicação que introduz o respectivo erro de truncatura.

Para estudar o efeito dos erros de truncatura, vamos primeiro considerar só um andar de filtragem. Assim, e para o bloco passa-baixo $H_1(z)$, com a ordem e estrutura indicadas no parágrafo anterior, o "scaling" é efectuado à saída do 3º andar, onde o valor máximo da função de transferência é 11.25. As potências de 2 mais próximas daquele valor são 8 e 16, resolvemos escolher a primeira uma vez que o comprimento de palavra permite um ganho de 16. Na saída do bloco passa-baixo, $H_1(z)$, temos 3 situações distintas: nenhuma operação de "scaling", "scaling" por 4 ou por 8. O primeiro caso é implementado para a amostra que vai ser entrada do passa-alto; o segundo caso para a amostra de saída passa-baixo; e finalmente o terceiro para a amostra de saída passa-baixo/entrada de novo andar de filtragem. Portanto o filtro passa-banda temos duas possíveis fontes de erro: a multiplicação e a operação de "scaling" enquanto o filtro passa-baixo tem mais uma operação de "scaling". O erro de truncatura da multiplicação por 1.75 deixa de ter significado nesta estrutura, uma vez que esta operação se situa imediatamente antes duma divisão por 8. O "scaling" tem exactamente o mesmo resultado inteiro quer se considere a parte decimal do dividendo ou não. Como se sabe, se considerarmos os operandos da multiplicação inteiros, o erro manifesta-se pela impossibilidade de reter a parte decimal do resultado. Além disso a variância deste erro será afectada do "scaling", por isso podemos simplesmente considerar a variância do erro "scaling" $e_1(n)$. Resumido, podemos considerar duas fontes de erro: uma à saída do 3º bloco (da função de transferência 3.2) e outra à saída (só nos andares passa-baixo), como está indicado no diagrama de blocos da figura seguinte.

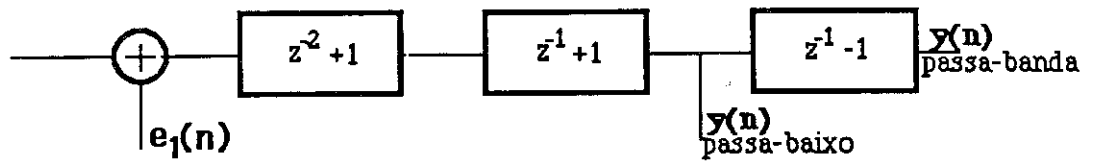


figura 3.14: Erros na estrutura de filtragem

Podemos calcular o valor deste erro considerando o modelo discutido no capítulo anterior, calculando a resposta impulsional das fontes de ruído para saída. O erro à saída do bloco passa-baixo introduzido pela fonte de erro $e_1(n)$ é dado por:

$$\sigma_f^2 = \frac{\Delta^2}{12} \sum h^2(n) = \frac{\Delta^2}{12} 4 \quad (3.5)$$

Por outro lado, também podemos calcular o máximo do valor absoluto do erro à saída do filtro passa-baixo. Podemos deste modo verificar se este erro pode afectar o resultado da divisão relacionada com a operação de "scaling" (por 4 ou 8) na saída. Isto é, se o resultado se altera quer consideremos apenas o sinal quantificado ou este adicionado com o valor absoluto do erro. O valor absoluto do erro à saída vem multiplicado pelo somatório do módulo da resposta impulsional ao quadrado (neste caso, também de valor 4). Portanto, no caso da divisão por 4 o resultado pode ter o bit menos significativo incorrecto se considerarmos só o sinal quantificado; se a divisão for por 8, o resultado será o mesmo em qualquer dos casos. Resumindo, quando o sinal se transmite para outros andares de filtragem podemos desprezar as fontes de erro dos blocos anteriores

A variância do erro à saída do passa-banda resulta simplesmente da fonte de erro no final do 3º andar

$$\sigma_f^2 = \frac{\Delta^2}{12} \quad (3.6)$$

O erro de quantificação da ADC também não é significativo na saída pois o somatório do módulo da resposta impulsional do filtro tem valor inferior à unidade, portanto em cada estrutura de filtragem basta considerarmos uma única fonte de erro.

CAP 4: CLASSIFICAÇÃO

1: Introdução

O problema de classificação pode ser formulado como a partição dum espaço multidimensional num conjunto de regiões que se designam por classes. Nesse espaço cada uma das dimensões corresponde a um dos parâmetros característicos extraídos do objecto a classificar. Portanto a classificação propriamente dita resume-se ao cálculo de funções de decisão que permitam atribuir ao conjunto de parâmetros do objecto a classificar (o vector de parâmetros característicos) uma determinada classe. A estratégia de cálculo das referidas funções pressupõe abordagens diferentes dependentes do tipo de dados. Podemos distinguir duas técnicas: estatística e determinística. Os métodos estatísticos baseiam-se fundamentalmente na aplicação da teoria de Bayes. Portanto consideram-se os dados como variáveis aleatórias que pertencem necessariamente a uma das classes em estudo, sendo conhecidas as probabilidades *a priori* de cada uma das classes em estudo e a função densidade de probabilidade condicional pode então calcular-se probabilidade *a posteriori* de cada classe. Assim, nos casos em que se tem conhecimento das propriedades estatísticas existem várias estruturas matemáticas para a obtenção de classificadores óptimos que a partir das probabilidades *a posteriori* calculam funções de risco ou custo [Duda e Hart-1973]. No entanto a situação mais usual é o desconhecimento das propriedades estatísticas dos dados. Também neste caso se podem seguir duas estratégias diferentes: uma em que se pressupõe a forma (em geral normal) das funções de distribuição estimando-se os parâmetros (média e covariância) que a caracterizam; a outra que consiste na estimativa das próprias funções. Para qualquer dos casos é necessário

um grande volume de dados para obter boas estimativas. Estes métodos conduzem muitas vezes a situações de cálculo bastante difíceis por se tratarem de operações num espaço multidimensional, pelo que são utilizadas conjuntamente técnicas de redução de dimensionalidade [Duda e Hart-1973].

Com os métodos determinísticos as funções de decisão são calculadas, a partir dos dados com classificação conhecida, não se pressupondo qualquer conhecimento estatístico sobre o problema, consideram-se as funções de decisão como uma combinação (por exemplo, linear) dos vectores de parâmetros característicos do objecto em estudo. Esta aproximação torna possível processos de aprendizagem iterativos pressupondo optimização segundo um determinado critério. Uma das concretizações desta estratégia é o perceptrão que, como veremos nos parágrafos seguintes, é um método de formulação e implementação bastante simples.

2: Redes Neurais

As redes neurais têm por motivação o sistema nervoso: um conjunto de elementos de processamento simples operando em paralelo. Estes elementos ou unidades de processamento estão interligados entre si, e o peso das ligações é variável. As entradas que convergem numa unidade de processamento são somadas e processadas pelo elemento, resultando deste modo o sinal de saída. Este sinal propaga-se às outras unidades de processamento através das respectivas ligações que ao mesmo tempo lhe atribuem maior ou menor importância conforme o peso da referida ligação. Na figura 4.1 está representado o elemento básico duma rede neuronal. Sendo o sinal à entrada da unidade de processamento de saída determinado pelo somatório seguinte:

$$\alpha = \sum w_i x_i \quad (4.1)$$

O sinal de saída y é determinado pelas características de processamento da unidade de processamento, podemos escrever que

$$y = f(\alpha - \theta) \quad (4.2)$$

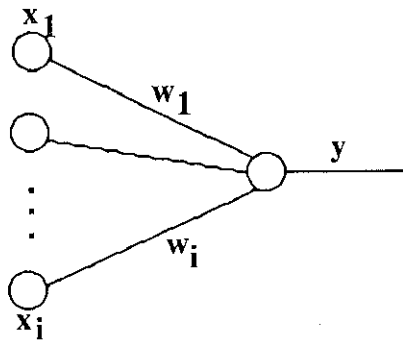


figura 4.1: Elemento básico duma rede neuronal

Portanto as características da unidade de processamento são a função f , designada de função de activação e o "offset" θ . A função de activação transforma o domínio (infinito) da entrada α numa gama de valores predefinida (0-1, por exemplo) ou apenas em 2 valores conforme se trate respectivamente duma função sigmóide (ou limiar linear), ou limiar.

A maneira de interligar as unidades de processamento constitui a definição da topologia da rede. Existem fundamentalmente dois tipos de topologia: recorrente e "feedforward". Na topologia recorrente há ramos de realimentação, isto é unidades de processamento recebendo sinal de entrada da própria saída ou das saídas de outras unidades de processamento, da mesma camada (lateral) ou de uma camada superior situadas (mais próximas da saída). Enquanto que na topologia "feedforward" cada unidade de processamento recebe unicamente sinal da camada

anterior (mais próxima da entrada) e envia sinal para a camada seguinte (mais próxima da saída). Para esta topologia ainda se faz outra divisão: redes de camada simples (percepção) e multicamada. As primeiras, têm uma camada de entrada cujas saídas têm valor igual ao sinal de entrada, e a camada de saída. Nas redes multicamada além das camadas já referidas existem uma ou várias camadas intermédias que não têm qualquer ligação ao exterior e que por isso se designam de "hidden".

No elemento básico de processamento da figura 4.1, com uma função de activação igual a um limiar, a saída pode ter dois valores, portanto, podemos considerar aquela rede com a capacidade de partição dum espaço de dimensão i (dos padrões de entrada) em duas regiões. Após a escolha duma função de activação o que determina o "mapping" entre a entrada e a saída são os valores dos pesos das ligações. A fase de ajuste dos pesos das ligações designa-se por fase de treino ou aprendizagem da rede.

2.1: Aprendizagem em Redes Neurais.

A fase que determina o funcionamento (ou a respectiva programação) é designada por aprendizagem, e a informação fica codificada no peso das ligações [Hinton-1985, Denker-1986, Anderson-1983]. A aprendizagem é um processo iterativo, isto é, os pesos vão sendo alterados até obter o resultado desejado. Esta fase pode ter duas estratégias diferentes: não-supervisionada e supervisionada. Na primeira estratégia, é unicamente fornecido um conjunto de padrões e espera-se que a rede lhe atribua uma classe, onde o número máximo de classes pode ser indicado pelo utilizador, isto é, "clustering" dos dados. Neste grupo estão incluídos os algoritmos de Kohonen e Grossberg [Lipman -1987]. A segunda estratégia consiste em apresentar um conjunto de padrões com classificação conhecida (o

conjunto de treino) e os pesos das ligações são iterativamente ajustados de modo a que a rede responda com a saída desejada para cada uma das amostras do referido conjunto entrada/saída. Neste grupo existe um conjunto de algoritmos de aprendizagem: aprendizagem estocástica em máquinas de Boltzmann [Hinton- 1986]; e as regras Delta e Delta Generalizada ("back -propagation") [Rumelhart et al- 1986]. Vamos limitar o nosso estudo às regras Delta e Delta Generalizada em redes "feedforward".

2.1.1: Perceptrão de camada simples

Como foi referido o perceptrão de camada simples é uma estrutura complementamente ligada, isto é, cada saída recebe sinal de todas as entradas. Onde, uma rede com um único nó de saída (figura 4.1) pode classificar as entradas em duas classes. Considerando os sinais de entrada analógicos e a função de activação um limiar, a saída pode ter dois valores: 1 e -1.

O algoritmo original para o treino destas redes foi desenvolvido por Rosenblatt. A partir da inicialização dos w_i e de θ com valores aleatórios muito pequenos, é calculado o primeiro valor de saída, em seguida os pesos das ligações vão sendo ajustados de acordo com a equação seguinte:

$$w_i(n+1) = w_i(n) + \epsilon [d(t) - y(t)] x_i(t) \quad (4.3)$$

onde $d(t)$ é a saída desejada e $y(t)$ é a saída efectiva que, neste caso, só podem ter dois valores, por isso a aplicação daquela equação pode resumir-se às regras seguintes:

se $y(t) = d(t)$, classificação é correcta o peso da ligação permanece inalterável.

se $y(t) = 1$ e devia ser -1 , deve decrementar-se w_i dum valor proporcional ao valor de entrada (ϵx_i).

se $y(t) = -1$ e devia ser 1 , deve incrementar-se w_i dum valor igual ao calculado para a regra anterior.

O valor ϵ é o coeficiente de aprendizagem que deve ter um valor no intervalo $(0,1)$. Os autores demonstraram que, se houver uma solução, o algoritmo converge num número finito de iterações. Se os padrões apresentados forem linearmente independentes, isto é classes linearmente separáveis, a solução existe.

Se retirarmos da unidade de processamento a não-linearidade considerando para função de activação a função sigmoide, a saída pode ter qualquer valor do intervalo $[0,1]$. Assim, a aplicação da expressão (4.3) ao ajuste do peso das ligações é feita para todas as iterações dum valor que é também proporcional à diferença da saída desejada pela saída obtida nessa iteração. Esta regra é conhecida por regra delta e é baseada na correcção do erro à saída da unidade de processamento, por isso, a regra delta pode ser considerada como uma implementação em gradiente descendente da minimização do erro quadrático (LMS) [Rumelhart et al- 1986].

Estas noções podem ser facilmente expandidas para classificação em M classes, bastando considerar M unidades de processamento na saída. A cada classe está associada uma unidade de processamento com saída 1 enquanto os restantes têm saída 0 . Na fase de teste, escolhe-se a saída com valor máximo e consequentemente a classe associada a esta unidade de processamento

O perceptrão unicamada é utilizado quando se pretende que entradas semelhantes produzam saídas semelhantes. Tem-se demonstrado a sua eficiência em aplicações em que se está interessado em eliminar o ruído de um conjunto de entradas. Em casos em que interessa uma classificação em que padrões da mesma

classe são os que têm menos características comuns, não é possível fazer a atribuição requerida com uma rede sem recodificação interna [Rumelhart-1986].

2.1.2: Perceptrão multicamada

Como já foi referido, estas redes contêm uma ou mais camadas entre a entrada e a saída que não têm qualquer ligação ao exterior. Assim o vector de entrada é apresentado à rede sendo activadas em paralelo todas as unidades da primeira camada tal como foi descrito para o perceptrão simples. Após o cálculo das saídas deste nível são activadas simultaneamente a unidades de processamento da camada seguinte e assim sucessivamente até à camada de saída. Resumindo, todas as camadas do mesmo nível são activadas em simultâneo, no entanto o funcionamento global é sequencial, começando pela entrada até à saída.

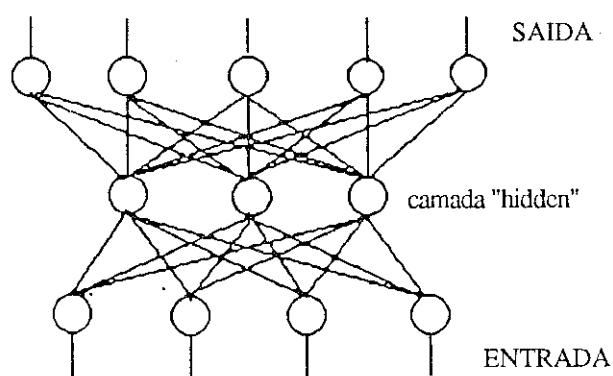


figura 4.2: Rede neuronal com três camadas: entrada, camada "hidden" e saída.

Para o processo de aprendizagem nestas redes, Rumelhart et al [1986] introduziram algumas modificações na regra delta, obtendo deste modo um novo algoritmo de aprendizagem que se designa por regra delta generalizada ou "back propagation". Como vimos na secção anterior, em que as entradas estavam

directamente ao elemento de saída, a regra de aprendizagem é bastante simples e consiste no ajuste dos pesos das ligações nas unidades de processamento de saída de modo a minimizar o gradiente do erro à saída. No caso das redes multicamada não podemos especificar para cada uma das camadas internas o valor de saída desejado. A regra delta generalizada resolve este problema uma vez que a informação de erro da saída é também propagada para as camadas internas atribuindo-lhe parcelas de erro proporcionais aos pesos das ligações.

Como foi referido na secção anterior, a regra delta pode considerar-se uma implementação do algoritmo LMS. Assim pode calcular-se a derivada do erro quadrático relativamente ao peso das ligações como o produto de duas componentes separadas: a derivada do erro relativamente às variações do sinal de entrada da unidade de processamento e a derivada do sinal de entrada da referida unidade de processamento relativamente às variações dos pesos nas ligações.

$$\frac{\partial E_p}{\partial w_{ji}} = \frac{\partial E_p}{\partial \alpha_{pj}} \frac{\partial \alpha_{pj}}{\partial w_{ji}} \quad (4.4)$$

Por outro lado o erro quadrático médio à saída, dado por:

$$E_p = \frac{1}{2} \sum_j (d_{pj} - y_{pj})^2 \quad (4.5)$$

representa o somatório do quadrado das diferenças entre as saídas e saídas desejadas em todas as unidades de processamento de saída, para cada padrão entrada/saída p.

Sendo o sinal de entrada em cada unidade de processamento dado pelo produto dos pesos das ligações por o sinal de saída da camada anterior, dado por:

$$\alpha_{pj} = \sum_i w_{ji} y_{pi} \quad (4.6)$$

a derivada desta expressão relativamente w_{ji} é diferente de zero para valores de $i=j$ e tem o valor

$$\frac{\partial \alpha_{pj}}{\partial w_{ji}} = y_{pi} \quad (4.7)$$

O cálculo da variação do erro em função das variações na entrada é também o produto de duas contribuições: uma reflectindo a variação do erro após uma mudança do sinal de saída e a outra reflectindo a variação do sinal de saída após variações na entrada.

$$\delta_{pj} = - \frac{\partial E_p}{\partial \alpha_{pj}} = - \frac{\partial E_p}{\partial y_{pj}} \frac{\partial y_{pj}}{\partial \alpha_{pj}} \quad (4.8)$$

O segundo termo do produto anterior é a derivada da função de activação $f(\alpha_{pj})$. O valor do primeiro termo depende da posição da unidade de processamento na rede. Para uma unidade de processamento da camada de saída é dado pela derivada da expressão (4.5). Sendo o erro à saída é dado por:

$$\delta_{pj} = f'(\alpha_{pj}) [d_{pj} - y_{pj}] \quad (4.9)$$

No caso de ser uma unidade de processamento da camada interna aquela expressão deve traduzir as variações de erro relativamente às variações do sinal de entrada na camada seguinte.

$$\sum_k \frac{\partial E_p}{\partial \alpha_{pk}} \frac{\partial \alpha_{pk}}{\partial y_{pj}} = \sum_k \delta_{pk} w_{kj} \quad (4.10)$$

Por esta razão o erro numa camada interior recebe a informação de erro da camada imediatamente superior, conseqüentemente o cálculo do erro nestas redes é feito duma forma recursiva partindo do erro da camada de saída:

$$\delta_{pj} = f'(\alpha_{pj}) \sum_k \delta_{pk} w_{kj} \quad (4.11)$$

sendo a derivada da função sigmoide relativamente ao sinal de entrada :

$$f(\alpha)[1-f(\alpha)] = y_{pj}(1-y_{pj}) \quad (4.12)$$

Finalmente, depois do erro calculado podem ajustar-se os pesos das ligações

$$\Delta w_{ji} = \epsilon \delta_{pj} y_{pi} \quad (4.13)$$

Esta é a forma básica do algoritmo de aprendizagem. No entanto, os autores sugerem que, para evitar oscilações no processo de convergência, se deve afectar o ajuste no peso das ligações com um valor que contenha também informação do ajuste da iteração anterior.

$$\Delta w_{ji} (n+1) = \epsilon \delta_{pj} y_{pi} + \mu \Delta w_{ji} (n) \quad (4.14)$$

Esta expressão pode ser utilizada de duas maneiras distintas: numa, cada padrão diferente na aprendizagem actualizam-se os pesos das ligações, na outra esta actualização só é feita após a apresentação de todos os padrões envolvidos na aprendizagem. Nesta última estratégia a primeira parcela da expressão anterior transforma-se num somatório que traduz o cálculo do erro para todos os padrões. Vogl et al [1988] consideram que esta técnica acelera o processo de convergência. Além disso, os mesmos autores sugerem um valor variável para o coeficiente de aprendizagem. Como este valor determina o passo de progressão ao longo da superfície de erro, a forma da curva conjuntamente com o passo são determinantes no processo de aprendizagem, isto é, se numa determinada fase de aprendizagem se estiver em zonas dum longo declive não se tem qualquer vantagem em ter valores baixos para o coeficiente de aprendizagem. Por outro lado, se se estiver numa zona plana com pequenas oscilações e se o coeficiente for elevado pode haver oscilações de sinais opostos no processo de aprendizagem. Então aqueles autores propõem que, decrescendo o erro de saída na iteração seguinte, se multiple o coeficiente de aprendizagem por um factor superior a um; se o erro aumentar (tipicamente de uma percentagem entre 1-5) se rejeitem todas as mudanças nos pesos das ligações e se multiplique o coeficiente de aprendizagem por outro valor inferior a um. Ao mesmo tempo faz-se $\mu=0$, e repete-se a iteração, quando o erro voltar a decrescer μ deve retomar o valor inicial. Estas modificações em μ significam que, se o processo estiver a seguir na direcção certa, a informação anterior pode ser útil caso contrário não traz qualquer vantagem. Outros autores referem que a velocidade de convergência será maior se o coeficiente de aprendizagem for diferente para cada ligação, e a partir desta ideia Silva [1990], implementa um algoritmo adaptativo que ajusta os coeficientes de aprendizagem pelo estudo das mudanças de sinal das derivadas parciais dos erros; se não houver mudança de sinal na derivada parcial do erro de determinada ligação o coeficiente aumenta, caso contrário deve diminuir.

Na regra delta, isto é, nas redes de camada simples, a superfície do erro de saída só tem um mínimo, portanto no processo de aprendizagem a solução encontrada é sempre ótima. Nas redes multicamada, a superfície pode ter vários mínimos, e a solução pode convergir para um mínimo local em vez do global como seria desejável, uma vez que esta solução implementa um gradiente descendente.

2.2: Redes Neurais em Classificação

O funcionamento das redes neurais em classificação tem dois modos distintos: aprendizagem e teste. Durante a aprendizagem são fornecidas à rede um par (ou conjunto de pares) entrada/saída e por tentativas os pesos das ligações entre unidades de processamento são ajustados até que o erro da saída efectiva da rede e o valor desejado atinja um valor aceitável. Na fase de teste são apresentadas novas entradas obtendo-se uma saída que corresponde à classificação da rede. No entanto, a estrutura da rede permanece inalterável qualquer que seja o modo de funcionamento. Portanto, o cálculo do sinal de saída é sempre feito do mesmo modo, os pesos das ligações é que são ajustados durante a aprendizagem permanecendo inalteráveis na fase de teste.

Redes neurais "feedforward" têm sido aplicadas em problemas de classificação e reconhecimento de padrões numa gama variada de domínios: voz [Lipmann-1989], reconhecimento de caracteres [Burr-1986, Lecun-1989], redução de ruído [Tamura-1988] etc. A aplicação de redes neurais em problemas de classificação aparecem como alternativa aos métodos clássicos, pois a sua utilização não pressupõe nenhum modelo matemático formal. Em vez disso, considera-se que um conjunto de treino, uma rede e um algoritmo de treino são suficientes para encontrar uma boa solução para o problema.

Portanto, a solução do problema depende da topologia escolhida, e do conjunto de treino. Lipmann [1988] e Burr [1988] sugerem uma interpretação geométrica para a influência da topologia num problema de classificação. Nesta interpretação, uma unidade de processamento corresponde a um hiperplano no espaço das medidas. Se a unidade de processamento for da primeira camada, o espaço referido é o espaço dos padrões de entrada. Portanto uma rede sem camadas internas só tem uma unidade de processamento por classe por isso só pode classificar classes linearmente separáveis. Quando se acrescenta uma camada interna, a unidade de processamento desta camada continua a representar um plano, mas os nodos da saída combinam aqueles nodos internos definindo limites nas referidas classes. Como esta combinação é uma operação de intersecção, definem-se assim regiões convexas para cada classe. Para criar classes com qualquer forma é necessário uma segunda camada "hidden", de modo a poder criar a negação da operação anterior, portanto classes não-convexas.

Mas, mesmo dada esta interpretação com é que se pode saber qual a capacidade de aprendizagem (capacidade) e como se comportará na classificação de novos dados (generalização) numa topologia. As respostas a estas perguntas têm sido o objectivo de diversos estudos com índoles distintas. Alguns trabalhos procuram a resposta destas questões na solução dum problema. Assim, a partir da amostragem aleatória no conjunto de padrões treinam uma rede variando o número de unidades "hidden" [Sejnowski-1986, Burr-1986, Gormann-1988]. Concluem, que o desempenho das redes com "hidden" é sempre melhor, qualquer que seja a propriedade considerada, que o de percepção unicamada. No caso de aumentarem o número de "hidden", o número de iterações necessários à aprendizagem diminuiu. Mas, a capacidade de generalização não aumenta indefinidamente com o número de unidades de processamento "hidden": aumenta, estabiliza, e finalmente degrada-se. Outro estudo, movimento dum braço de "robot" [Josin-1988] na

localização de pontos num plano x-y, a partir duma configuração fixa (2-32-2) variam o número de padrões de treino concluindo que com apenas 3/4 (do perímetro dum quadrado) padrões obtém 80% de concordância em 100 novos padrões (pontos duma circunferência inscrita no quadrado). Com 8/10 pontos (do área do quadrado) a generalização atinge o valor 100%. Enquanto Ahmad [1989] trata esta questão noutra problema: contar o número de 1 (uns) numa sequência tendo por saída 1 se aquele número exceder o número de 0 (zeros) da mesma sequência. Os autores concluem que para uma mesma topologia o erro em novas classificações decresce exponencialmente com o número de padrões apresentados. Por outro lado, concluem que a escolha de padrões pode ter bastante influência na capacidade de generalização, isto é pode haver grupos de padrões que consigam melhores generalizações que o mesmo número de padrões de outro grupo. Os autores chamaram aos primeiros padrões-limite (neste caso para sequências de comprimento $2k+1$, o número de 1 deve ser k e $k+1$).

Outros autores [Baum et al-1989, Baum-1990], para uma rede com uma camada "hidden" uma unidade de processamento na saída e funções de activação lineares (ou uma função lógica) procuram analisar quantos padrões duma dada distribuição são necessários para treinar essa rede e obter um grau de generalização aceitável em novos padrões da mesma distribuição. Demonstram um conjunto de teoremas que os autores consideram que justificam a seguinte regra heurística: se M exemplos aleatórios duma dada distribuição forem correctamente aprendidos por uma rede de W pesos haverá W/M erros em novas classificações. Portanto para 90% de generalização devemos ter um conjunto treino 10 vezes superior ao número de pesos da rede. Este número é substancialmente superior a qualquer dos exemplos referidos no parágrafo superior. Paralelamente Mitchison [1989] tratou o problema da capacidade de algumas redes, com uma camada "hidden" e entradas binárias, em que a configuração (n-h-s) segue a regra seguinte: $s \leq h \leq n$. Além disso,

com funções de activação lineares, a capacidade m é dada pela desigualdade seguinte $2n \leq m \leq nt \log t$ com $t = 1 + h/s$.

Resumindo, tudo parece indicar que estas propriedades (generalização e capacidade) são também dependentes do problema em questão (do grupo de dados), por isso se justifica a escolha de uma topologia e do conjunto de aprendizagem apropriados a cada caso.

3: Aplicação das Redes Neurais à Classificação de Sono.

Como referimos os detectores fornecem, em tempo real, ao processador central a informação relacionada com todos os eventos, e respectivo instante de ocorrência, que satisfazem um conjunto de critérios. Portanto, uma noite de sono fica caracterizada não só pela sequência dos eventos detectados mas também pelos parâmetros dos mesmos eventos. Mas, a classificação do sono em estádios é unicamente baseada na percentagem (ou número) de eventos detectados numa época (20, 30 ou 60 segundos) do traçado poligráfico. Essa informação está disponível nos dados armazenados, basta para cada época somar o período ou a duração (ou contar as presenças) dos eventos detectados naquele intervalo de tempo. Estes valores, calculados para cada uma das actividades, podem constituir os padrões de entrada duma rede neuronal, em que classificação visual dessas mesmas épocas são os valores desejados à saída. Portanto, e como em outros domínios temos um conjunto de padrões a classificar temos que escolher uma topologia para a rede, um conjunto de treino de modo a que se obtenha uma boa generalização. Por outro lado do ponto de vista da classificação em estádios de sono quais as actividades mais representativas. Este assunto será tratado no capítulo seguinte. Para isso, é utilizado um conjunto de programas de simulação duma rede neuronal [McClelland e Rumelhart -1988].

CAP 5: RESULTADOS

1: Introdução

Neste capítulo apresentaremos os resultados em duas partes seguindo a hierarquia de processamento. Na primeira parte apresentaremos os dados à saída dos detectores, isto é, os eventos do EEG detectados e respectivos parâmetros característicos com intuito de fazer a avaliação da "performance" dos detectores. Na segunda parte, utilizando uma parte dessa informação, apresentaremos a classificação em estádios de sono por uma rede neural. O método de classificação é primeiro validado com um conjunto de dados (eventos detectados por outro sistema) classificados por outros métodos [Príncipe et al - 1987, Chang - 1988], seguido da classificação dos dados obtidos pelo HIDRA.

1.1: Recolha de dados

A recolha de dados foi feita no Laboratório de Electroencefalografia do Hospital Geral de Santo António no Porto, usando uma máquina de EEG com filtragem analógica a 35Hz nos canais de EEG e EOG e a 120Hz no EMG e utilizando eléctrodos tipo concha (de cloreto de prata) colocados com colódio. Com um gravador de instrumentação Racal de 7 canais são gravadas, em fita magnética, as seguintes derivações do electroencefalograma: F₃-F₇, P_z-O₁ e C₃-A₂; um canal ocular em montagem bipolar (superior esquerdo/lateral direito - superior direito/lateral esquerdo); e o electromiograma, uma derivação submentoniana (tabela 5.1). Para este estudo foram utilizados seis indivíduos saudáveis com idades compreendidas entre 13 e 63 anos (tabela 5.1).

casos	idade	sexo	gravação
t4	14	femino	canal 1: F ₃ -F ₇ canal 2: C _Z -A ₂ canal 3: P _Z -O ₁ canal 4: EOG canal 5: EMG
t5	13	masculino	
t7	36	feminino	
t8	41	masculino	
t10	63	feminino	
t16	26	masculino	

tabela 5.1: Base de dados e configuração dos sinais gravados

2: Detectores

O HIDRA é configurado, tal como foi referido no capítulo 3, com 4 preprocessadores cada um deles recebendo o sinal de entrada dum eléctrodo diferente: 3 de EEG e 1 ocular (os quatro primeiros sinais indicados na tabela 5.1). Como tem sido referido são detectadas as actividades seguintes: alfa, beta, teta, sigma, delta, artefactos musculares e REM. Durante este processamento é feito um traçado poligráfico em que conjuntamente com os sinais de entrada são também registadas marcas que assinalam as diferentes detecções (figuras I a III) ou os sinais de entrada filtrados nas bandas das diferentes actividades.

Cada 60s deste registo em papel são classificados visualmente em estádios de sono por um especialista, que desconhece os códigos das marcas de detecção. A análise visual pode ser relacionada com a informação dos detectores, contida num ficheiro binário durante o processamento em tempo real, desde que se consiga uma sincronização do papel com o tempo 'interno' da detecção, isto é, medido pelo relógio do HIDRA. Como é habitual referir, à velocidade de 15mm/s correspondem

3 páginas de papel por minuto, o que não é exacto por duas razões: por um lado a velocidade de arrastamento do papel não é constante, por outro, mesmo considerando a velocidade constante, três páginas correspondem a um valor ligeiramente superior ao minuto. Assim, optou-se neste estudo pela sincronização através da inclusão na saída de um dos detectores duma marca com a sinalização do minuto e/ou a inclusão na montagem dum Time-Marker¹. Portanto, se a classificação visual for feita entre estas marcas do minuto, a classificação visual fica sincronizada com o tempo interno dos detectores. O Time-Marker pode ser útil se o registo for interrompido sem parar a análise (mudar papel, encravamento no papel). Com a informação deste relógio podemos saber o tempo total da interrupção do registo e, conseqüentemente, qual a informação armazenada pelos detectores que não está classificada visualmente. Naturalmente que também se pode utilizar o Time-Marker para marcar o registo, basta saber a diferença inicial entre o tempo marcado por este e o relógio interno. E mesmo que esse valor não seja um valor exacto o erro não é acumulado ao longo da análise, permanecendo constante.

Como tem sido referido, no final da análise em tempo real é criado um ficheiro binário com a informação dos detectores. Este é separado num conjunto de ficheiros ASCII: um para cada actividade que contém toda a informação dos respectivos parâmetros descodificada (valores de amplitude em μv e de período em ms). A partir destes podemos tratar graficamente toda a informação aí contida: por exemplo, saber a evolução média das actividades ao longo da noite; também se podem efectuar estatísticas simples sobre os parâmetros detectados: médias, variâncias e histogramas. Estas estatísticas podem ser calculadas para o tempo total

¹Aparelho que permite um registo de relógio no traçado poligráfico, isto é, a escrita dos números do relógio (horas, minutos e segundos) por uma das penas da máquina de EEG.

da análise (toda a noite), por um período de tempo ou por estágio (classificação visual ou da máquina).

2.1: Avaliação global dos detectores

Dada a grande quantidade de eventos é impossível pedir a um especialista que marque todos os eventos numa actividade de modo a poder comparar-se este resultado com as saídas dos detectores e fazer o estudo tradicional de falsos positivos/ falsos negativos. Além disso, a marcação dos eventos por um especialista é uma tarefa de fiabilidade duvidosa, como já foi referido noutros estudos de detecção de eventos no EEG [Guedes de Oliveira-1981].

Como tem sido referido, durante a análise em tempo real é feito um traçado poligráfico em que se registam conjuntamente com os sinais de entrada as marcas de detecção. A observação deste registo, durante várias horas de funcionamento do sistema, permite-nos uma avaliação qualitativa do processo de detecção, pelo menos no que se refere às detecções de eventos definidos em "burst" ou de grande amplitude. Assim, podemos dizer que sempre que as actividades sigma, alfa, beta, artefactos musculares e delta são claramente visíveis, também são correctamente assinaladas. Como podemos ver na figura I, na parte esquerda, correspondente a estágio 0 (depois do despertar), as marcas de detecção assinalam a presença de artefactos musculares, um pouco de alfa e beta, e também ondas delta (artefactos neste caso); na parte direita, correspondente a estágio 1, os detectores reconhecem a presença predominante de beta, e dois "bursts" de teta. Na figura II na parte esquerda, correspondente a estágio 2, os detectores reconhecem a actividade sigma claramente presente no canal central e algumas ondas lentas; na parte da direita, correspondente a sono lento, são assinaladas a actividade delta e lenta nos canais frontal e central, e um fuso de sono que está possivelmente mascarado pela

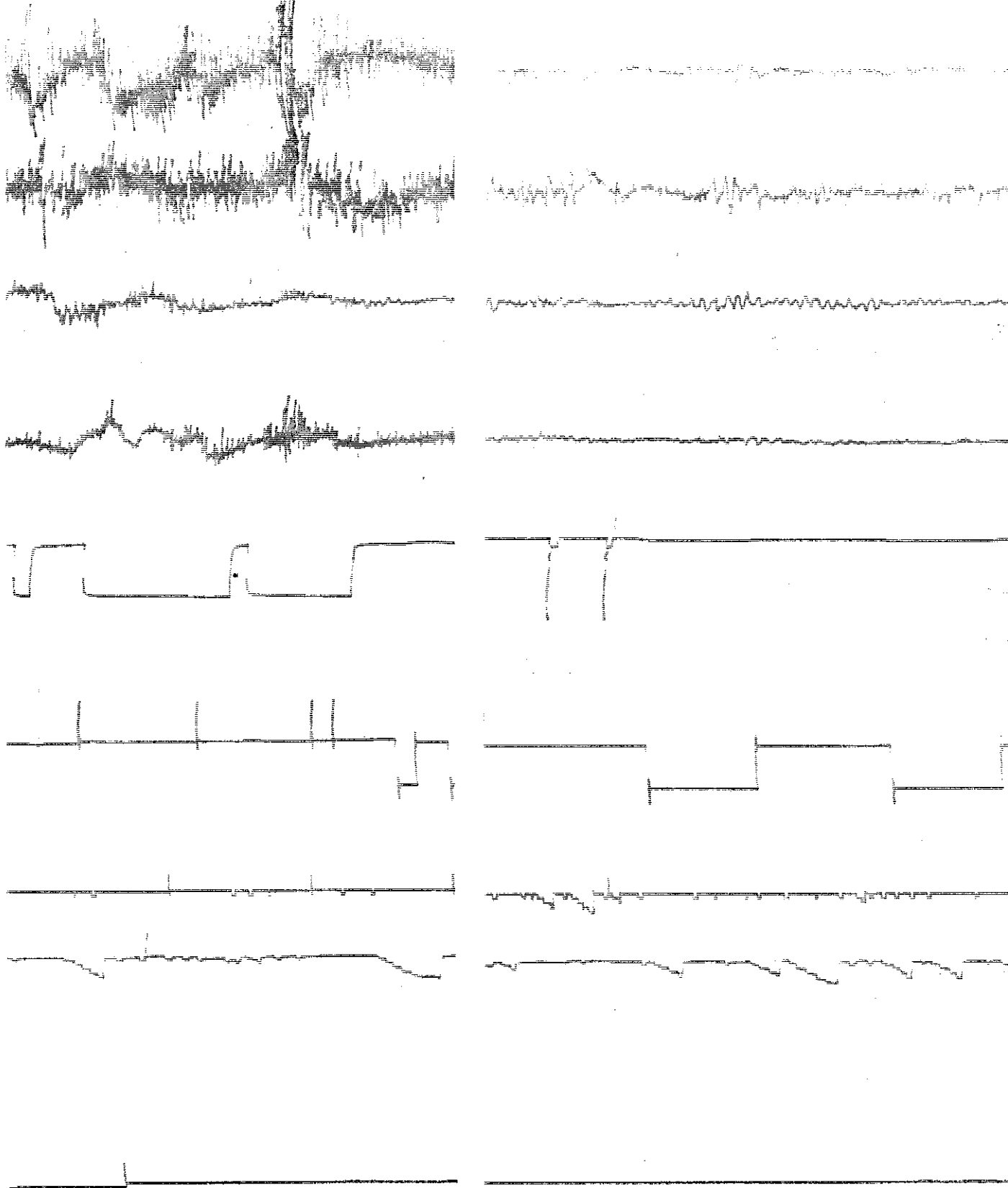


figura I: Registo poligráfico: estádios 0 e 1. Entradas: F3- F7, C3- A2, Pz- O1, EOG em 1, 2, 3 e 4 respectivamente. Saídas 5- 9 marcas de detecção : 5- teta (↑) e artefactos(⌊); 6 - beta (⌊) e delta (↑); 7 - sigma (⌊) e lenta (↑); 8- alfa (⌊) e lenta (↑); 9 -REM (↑) e lenta (↓).

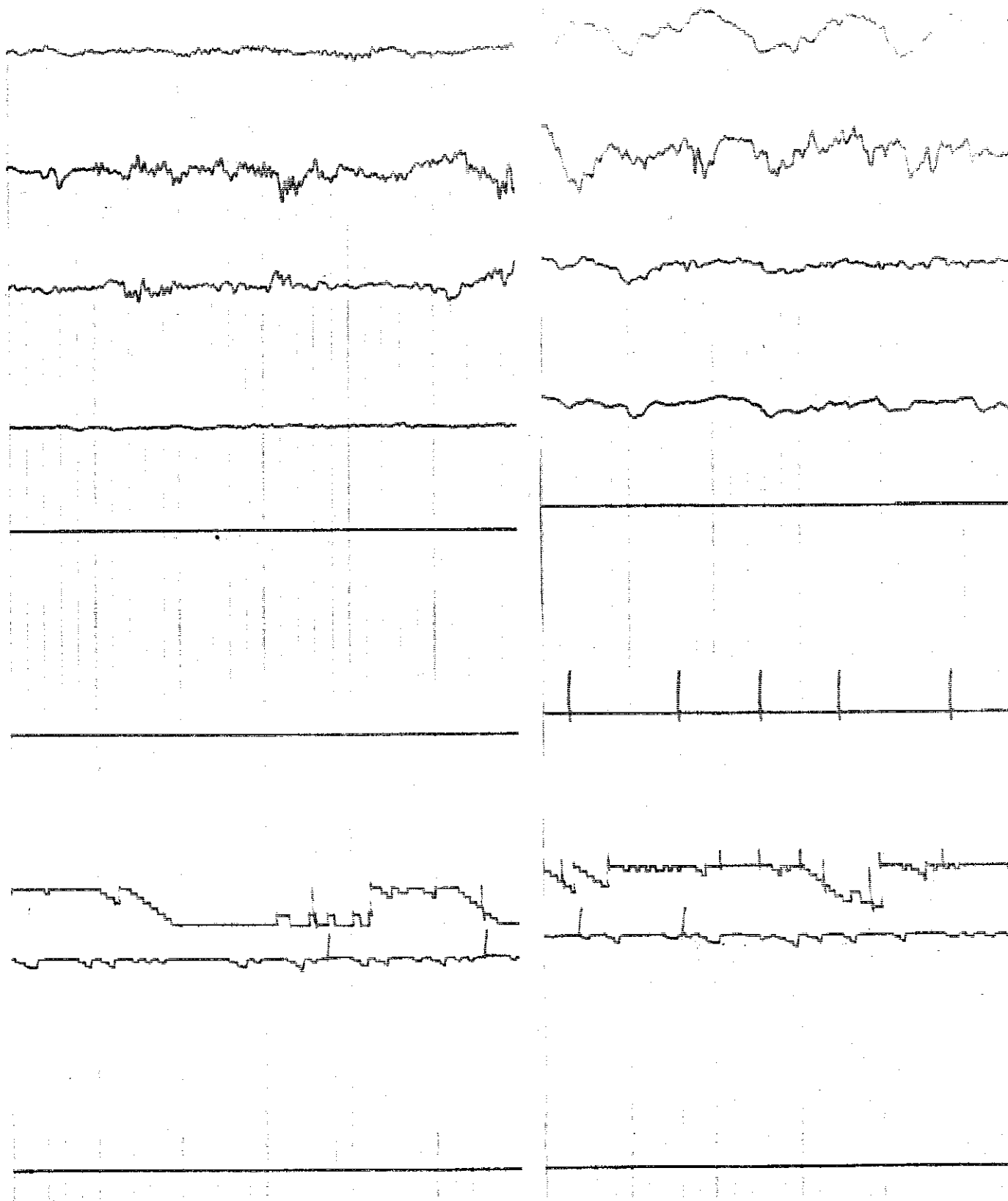


figura II: Registo poligráfico: estádios 2 e 3. Entradas: F3- F7, C3- A2, Pz- O1, EOG em 1, 2, 3 e 4 respectivamente. Saídas 5- 9 marcas de detecção : 5- teta (\uparrow) e artefactos (\square); 6 - beta (\square) e delta (\uparrow); 7 - sigma (\square) e lenta (\uparrow); 8- alfa (\square) e lenta (\uparrow); 9 -REM (\uparrow) e lenta (\downarrow).

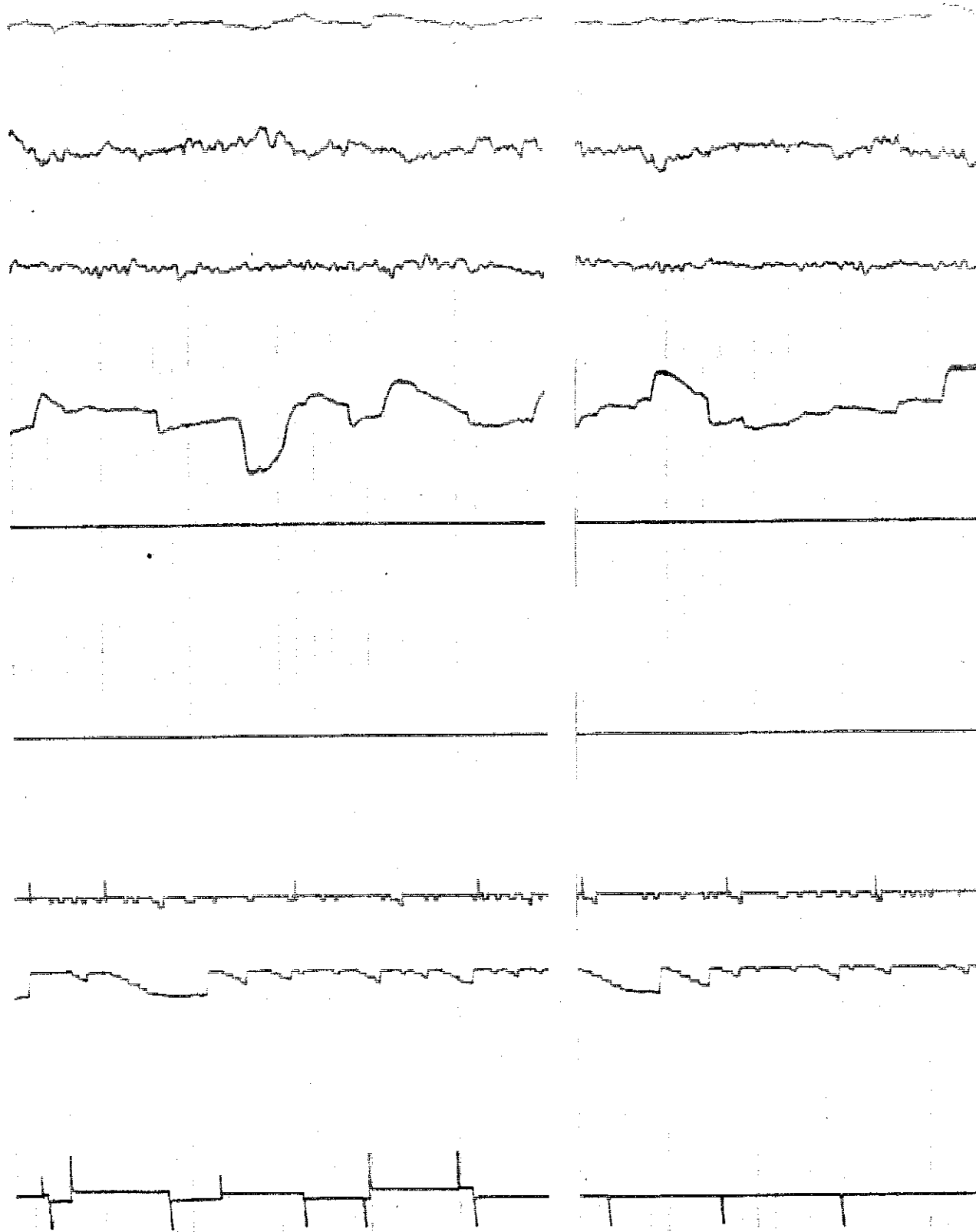


figura III: Registo poligráfico: estágio REM. Entradas: F3- F7, C3- A2, Pz- O1, EOG em 1, 2, 3 e 4 respectivamente. Saídas 5- 9 marcas de detecção : 5- teta (\uparrow) e artefactos (\square); 6 - beta (\square) e delta (\uparrow); 7 - sigma (\square) e lenta (\uparrow); 8- alfa (\square) e lenta (\uparrow); 9 -REM (\uparrow) e lenta (\downarrow).

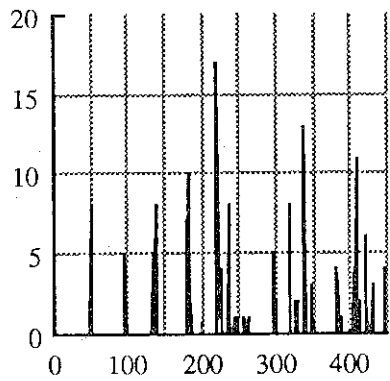
actividade delta. O REM tem muitos falsos negativos como podemos ver na figura III, na parte da esquerda o detector reconhece as ondas REM, mas na da direita, ondas que parecem semelhantes não são assinaladas, note-se que os outros detectores estão silenciosos, excepto o detector de ondas lentas dos canais ocular e central e o detector alfa com aproximadamente 1s de actividade. Em nossa opinião o detector tem muitos falsos negativos por o sinal de entrada estar muitas vezes saturado, nessas condições as ondas não satisfazem o critério de declive definido no processo de validação deste evento.

Os parâmetros armazenados permitem saber a evolução média de cada actividade ao longo da noite. Nas figuras 5.1 a 5.6 estão representados graficamente o tempo total, em segundos por minuto, para todas actividades (excepto no sigma que é o número de "bursts"), conjuntamente com o hipnograma obtido por classificação visual para os indivíduos deste estudo. A relação destas actividades fornece informação quantitativa sobre resultados qualitativos existentes na área de análise do sono. Como facilmente se verifica, os artefactos musculares estão relacionados com o estágio acordado tal como é referido por Smith [1977]. A actividade beta está relacionada com os períodos de REM, mais visível nos casos t8 e t10 (os do nível etário superior) o que confirma resultados anteriores [Smith-1977]. A actividade teta não apresenta nenhuma distribuição em particular e a percentagem é pouco significativa (excepto em t16). Mas, pode ser por duas razões: o ser detectada em "burst" (pelo menos 3 ondas consecutivas) e o efeito de primeira noite ao qual segundo Hasan [1982] esta actividade é particularmente sensível. Este mesmo autor refere que apenas esta actividade demonstra variabilidade inter-noites quando se considera o mesmo indivíduo. Por outro lado Hasan refere que teta é particularmente importante para detectar o estágio 0, dado que teta não tem percentagem significativa neste estágio quando comparada com os valores dos outros estádios. Em todos os casos deste estudo, a actividade alfa é

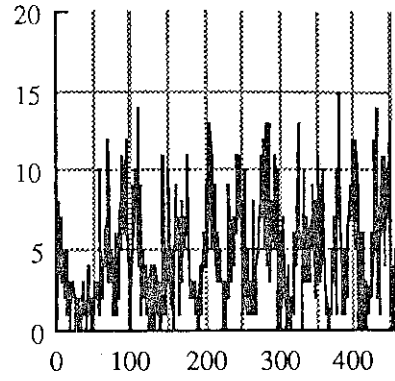
pouco relevante porque a gravação foi sempre iniciada depois do indivíduo ter começado a dormir (estádio 1 ou estágio 2). Assim, a actividade alfa nunca atinge níveis significativos (em alguns casos é sempre inferior a 5s), está relacionada com REM ou segue os artefactos musculares pelo que está também relacionada com os movimentos "arousal". A actividade delta atinge os níveis mais elevados nos estádios 3/4 e é mais elevada na primeira parte da noite (nos dois primeiros ciclos) diminuindo a sua percentagem ao longo da noite. Além disso as percentagens mais elevadas e durante maior período de tempo ocorrem nos casos t4 e t5 (os mais jovens) como já foi referido por Smith[1977], Feinberg[1974] e Liard[1979]. A percentagem de actividade sigma apresenta como era de esperar picos, no estágio 2, cujo valor não depende da idade [Príncipe - 1981]. Esta actividade apresenta grande variabilidade de minuto para minuto, e em alguns casos (por exemplo t4 e t8) a percentagem mínima de actividade por minuto não tem valor nulo, como seria de esperar em períodos de REM.

Concluindo, a variação média, para todas as actividades, evolui entre um máximo e mínimo, apresentando-se tendencialmente em "clusters" (existe em alguns períodos de tempo e desaparece nos outros) como seria de esperar neste fenómeno. Neste grupo de indivíduos as actividades igualmente predominantes (com maior percentagem no minuto) são os ritmos sigma e delta. Mas, em qualquer das actividades consideradas existe grande variabilidade inter-indivíduos. Isto é, a percentagem relativa das actividades, os valores de pico e os períodos de tempo em que as actividades são dominantes são bastante variáveis de indivíduo para indivíduo. Estes resultados são concordantes com resultados anteriormente publicados, o que reforça a nossa convicção de que os detectores EEG implementados no HIDRA produzem resultados plausíveis e concordantes com a informação visual retirada do EEG.

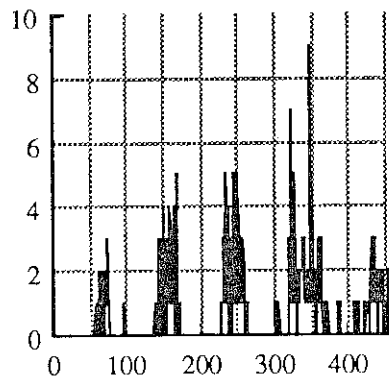
artefactos- segundos/minuto



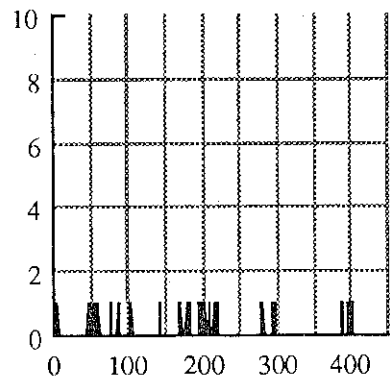
alfa- segundos/minuto



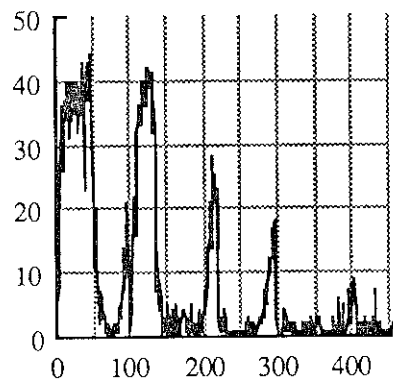
beta- segundos/minuto



teta- segundos/minuto



delta -segundos/minuto



sigma- nº de bursts/minuto

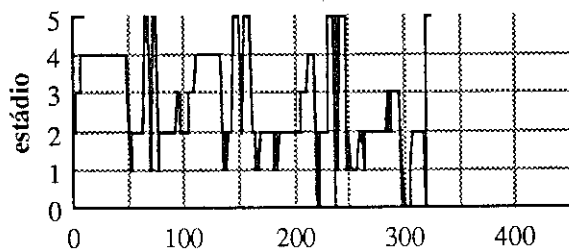
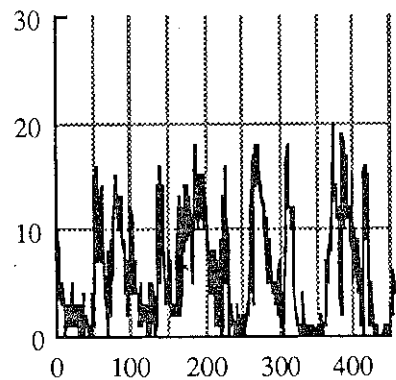
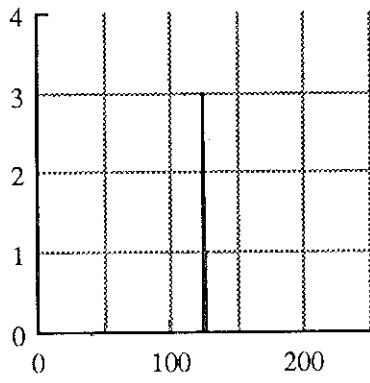
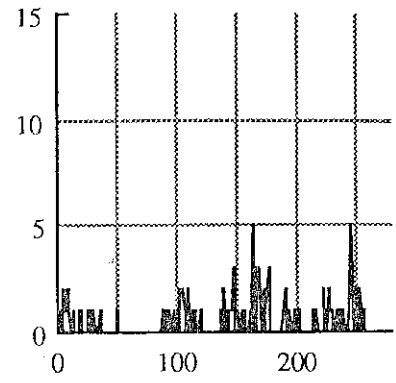


figura 5.1: Evolução das actividades do EEG ao longo da noite e classificação visual em t4

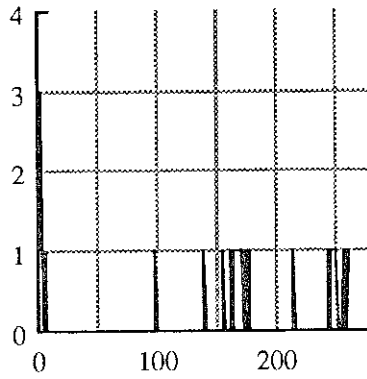
artefactos- segundos/minuto



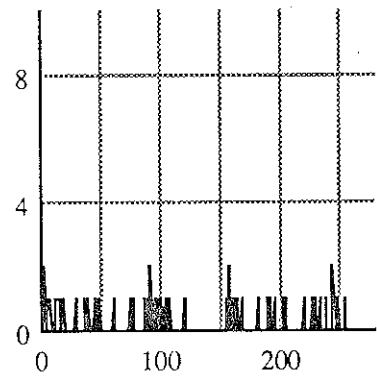
alfa- segundos/minuto



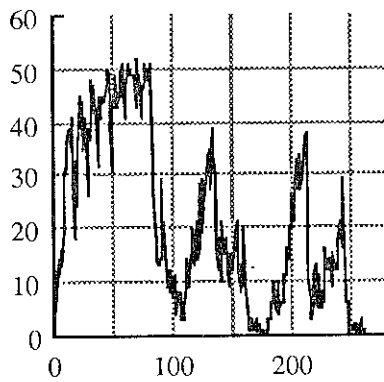
beta- segundos/minuto



teta- segundos/minuto



delta- segundos/minuto



sigma- nº de bursts/minuto

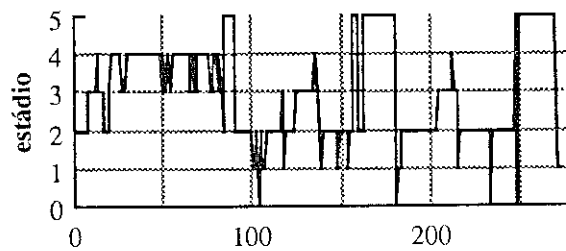
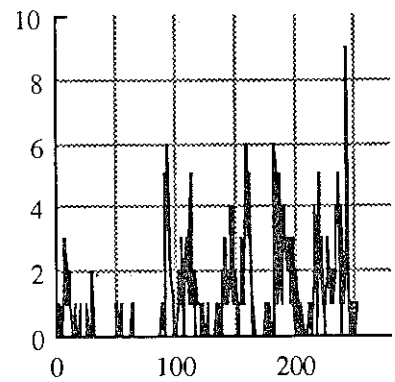
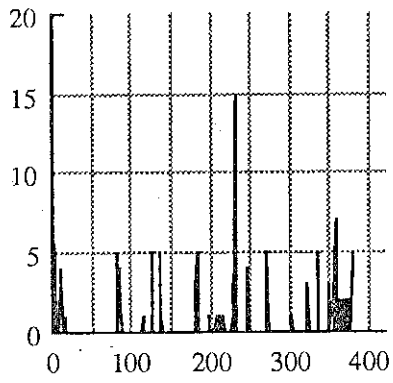
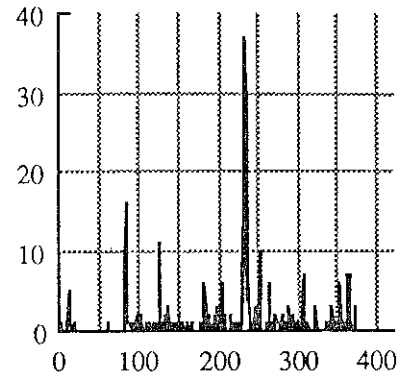


figura 5.2: Evolução das actividades do EEG ao longo da noite e classificação visual de t5

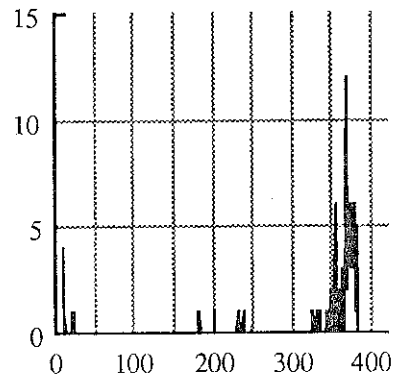
artefectos- segundos/minuto



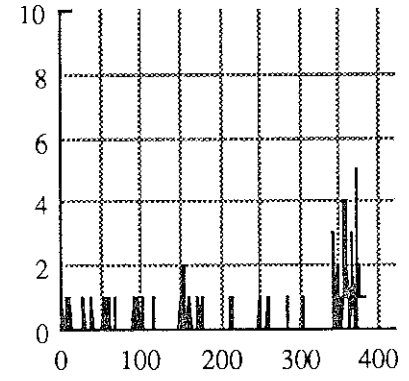
alfa- segundos/minuto



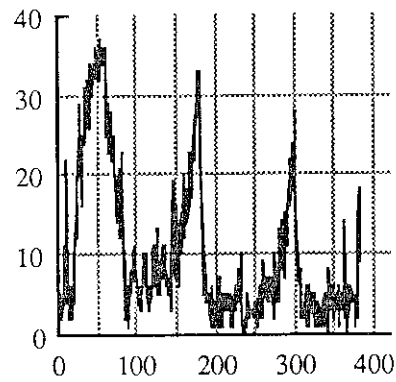
beta- segundos/minuto



teta- segundos/minuto



delta- segundos/minuto



sigma- nº de bursts/minuto

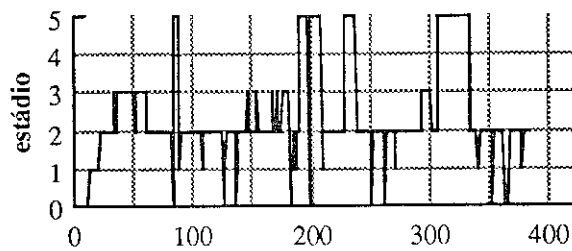
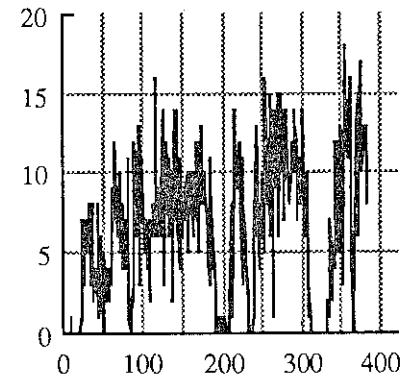
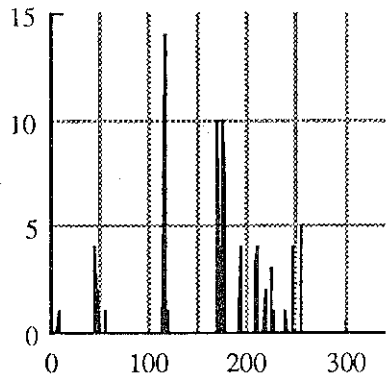
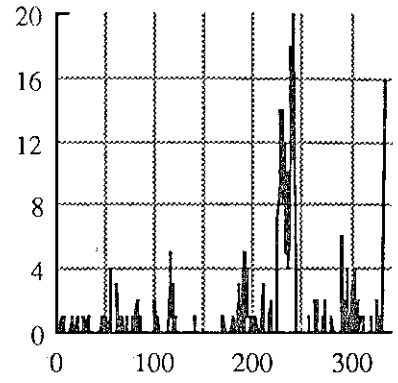


figura 5.3: Evolução das actividades de EEG ao longo da noite e classificação visual de 17

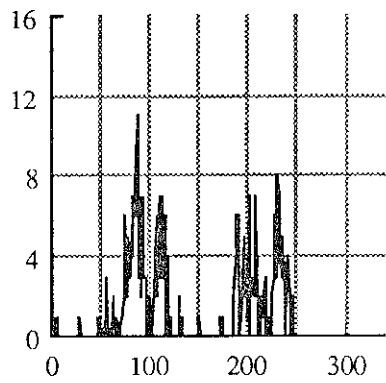
artefactos- segundos/minuto



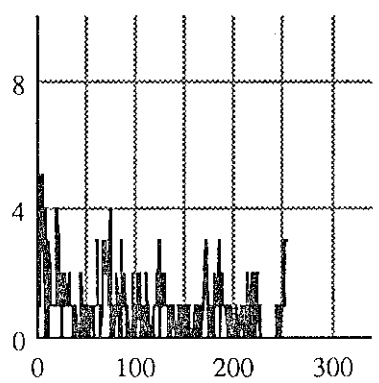
alfa- segundos/minuto



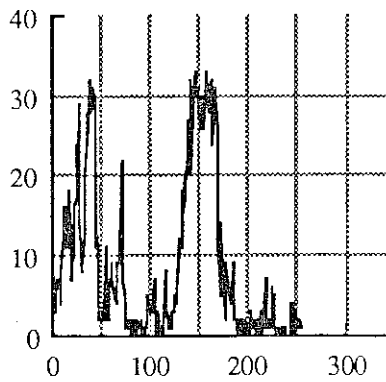
beta- segundos/minuto



teta- segundos/minuto



delta- segundos/minuto



sigma- nº de bursts/minuto

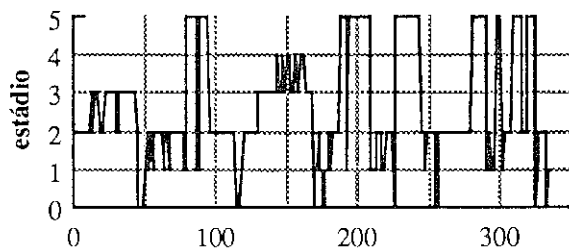
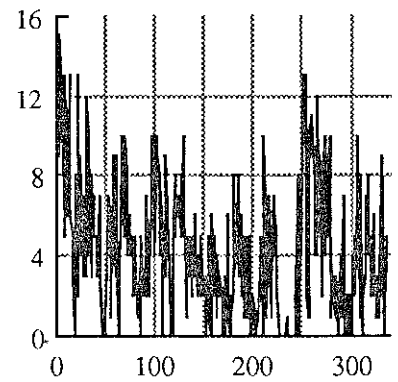
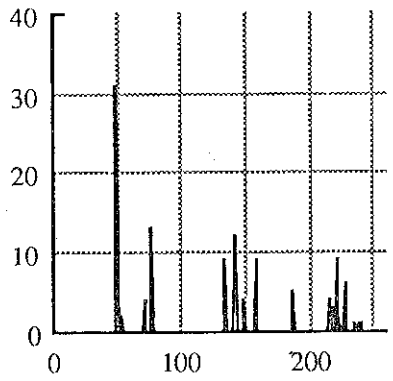
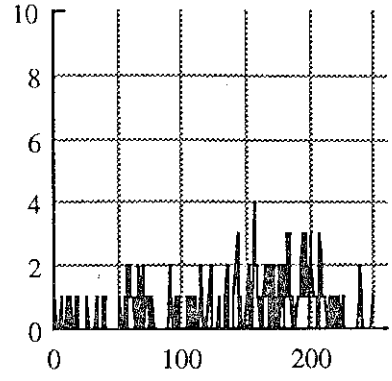


figura 5.4: Evolução das actividades de EEG ao longo da noite e classificação visual de t8

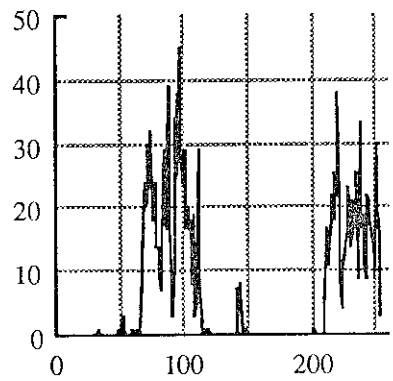
artefactos- segundos/minuto



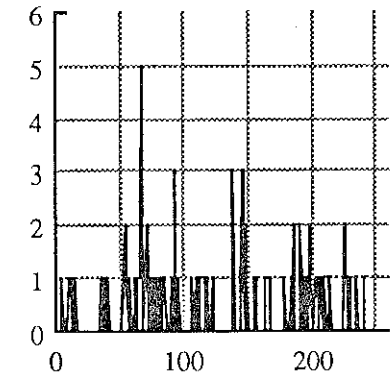
alfa- segundos/minuto



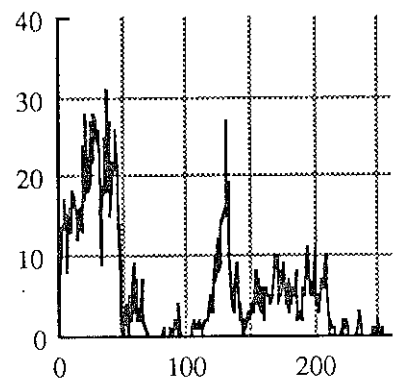
beta- segundos/minuto



teta- segundos/minuto



delta- segundos/minuto



sigma- n° de bursts/minuto

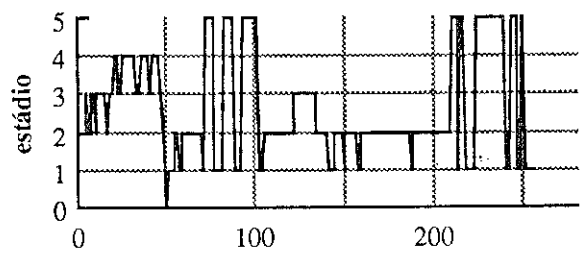
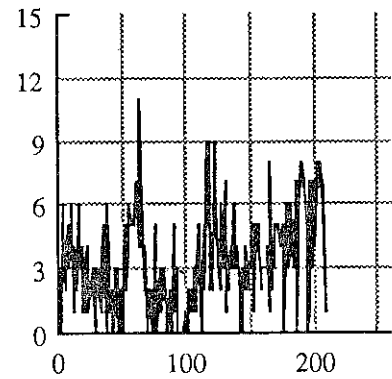
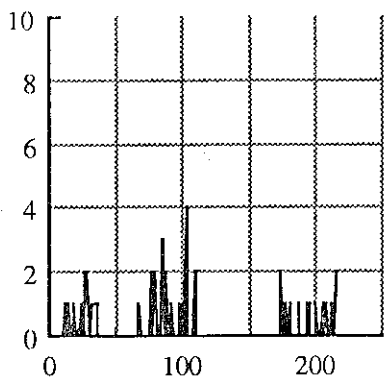
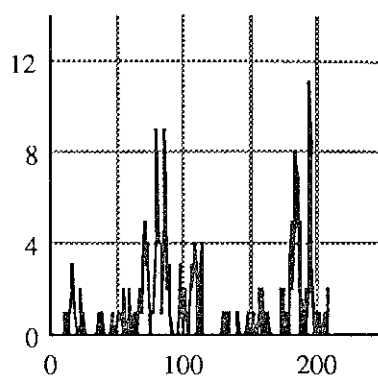


figura 5.5: Evolução das actividades do EEG ao longo da noite e classificação visual de t10

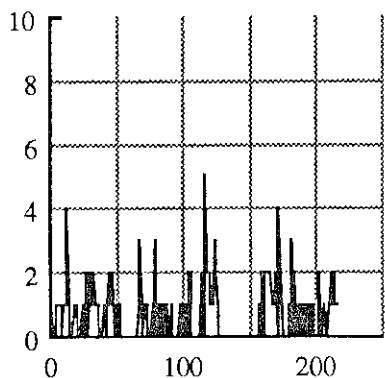
artefactos-segundos/minuto



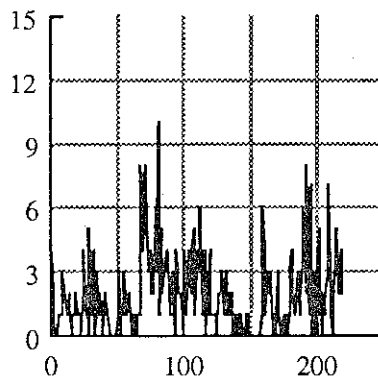
alfa- segundos/minuto



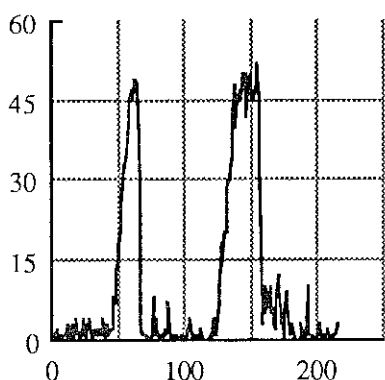
beta- segundos/minuto



teta- segundos/minuto



delta- segundos/minuto



sigma- n° de bursts/minuto

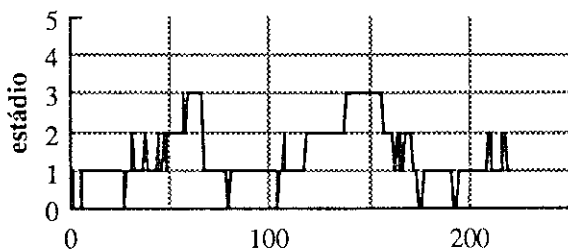
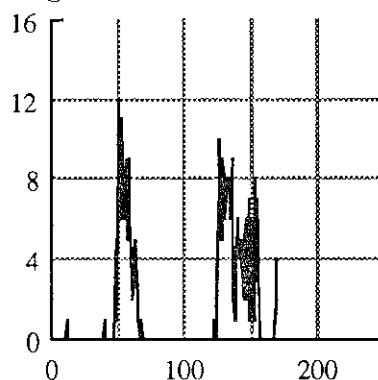


figura 5.6: Evolução ao longo da noite das actividades do EEG.e classificação visual de t16

2.2: Estudo estatístico das actividades detectadas

Com este estudo temos em vista dois objectivos: a microanálise das actividades dominantes e, eventualmente, após conhecimento das estatísticas dos parâmetros para determinado indivíduo, a refinação do processo de detecção por utilização médias e das variâncias. A ideia é excluir as detecções para as quais os parâmetros são suficientemente diferentes dos valores médios obtidos. Ainda que esta fase só se consiga fazer no final da noite, pode melhorar a concordância do sistema com a análise visual. Assim, resolvemos utilizar a classificação visual em estádios de sono para dividir os eventos detectados em grupos e fazer um estudo estatístico inter e intra-grupos dos respectivos parâmetros. Este método pode enfermar de erros já que a classificação visual pode ser baseada em informação diferente da recolhida pelos detectores (isto é insensível à existência de falsos positivos/falsos negativos). Mas, se considerarmos o desempenho dos detectores elevado, e dado que as actividades dominantes aparecem concentradas num estágio de sono, a influência dos falsos positivos nas estatísticas das actividades nesses estádios pode considerar-se pequena. Os eventos detectados em estádios em que a actividade não é dominante são, em geral, menosprezadas pelo especialista, ou podem mesmo ser falsos positivos. Logo o estudo tem validade aproximada quando aplicado às actividades dominantes em cada estágio. É evidente que esta metodologia pressupõe que os falsos negativos são desprezáveis. Assim, utilizaremos os agrupamentos naturais dos eventos em estádios com características semelhantes: estágio 3/4 (actividade delta dominante), estágio 2 (actividade sigma), estágio 0 (actividade alfa, pouco significativa no grupo de dados utilizado) e estágio 1/5 (beta ou ausência de qualquer actividade do EEG). Limitaremos este estudo às actividades sigma e delta. Em quatro dos casos (t4, t8, t10, t16) faremos um teste t-Student em que se testa a hipótese nula, isto é, a igualdade de médias entre as

detecções dos diferentes estádios. Para cada um dos casos apresentaremos os histogramas dos parâmetros detectados para as actividades sigma e delta, referindo qualitativamente a tendência dos dados.

2.2.1: Actividade Sigma

Esta actividade é fundamental para a classificação do estádio 2. Assim, neste caso calcularemos os valores médios e o desvio padrão de todos os parâmetros dos sigma detectados comparando a média dos parâmetros dos eventos detectados em estádio 2 com os mesmos valores durante os outros estádios.

Na tabela seguinte estão representados o número de épocas de cada um dos estádios. Como vemos, o número de eventos sigma é maior no estádio 2 que em qualquer dos outros estádios. As estimativas do valor médios e do desvio padrão dos parâmetros dos eventos detectados em estádios diferentes estão representadas na tabela 5.3.

CASOS	ESTADIO 2		ESTADIO 3/4		ESTADIO 1/5	
	épocas	eventos	épocas	eventos	épocas	eventos
t4	134	1588	101	290	69	244
t5	106	263	99	26	55	30
t7	129	2055	33	262	38	98
t8	135	882	62	260	120	274
t10	107	472	50	152	92	79
t16	49	165	28	121	128	-

tabela 5.2: Número de eventos sigma por estádio

Nos casos t4, t8, t10 e t16 foi efectuado um teste t-student para todos os parâmetros entre as detecções ocorridas em estádio 2 e as detecções nos outros grupos (estádios 3/4 e estádios 1/5). Rejeitou-se a hipótese nula (ou da igualdade

de médias) com nível de significância 0.01 em todos os parâmetros, excepto a amplitude de t8 e t10 dos estádios 1/5 e 2. Note-se pela tabela de valores apresentados que a duração média e a amplitude média dos "bursts" em estádio 2 é sempre maior que a duração média dos eventos nos outros estádios. O período já não apresenta esta tendência: em quatro dos casos em estádio 3/4 o período é maior que nos outros estádios, e nos restantes em estádio 2 é maior que nos outros. No entanto, a estimativa do período para estádio 1/5 é sempre inferior aos valores estimados para estádio 2.

CASOS	ESTADIO 2		ESTADIO 3/4		ESTADIO 1/5		TOTAL		
	\bar{x}	Sd	\bar{x}	Sd	\bar{x}	Sd	\bar{x}	Sd	
t4	dur	1056.6	466.8	831.1	241.2	840.1	306.4	1005.9	453.5
	per	74.4	3.9	73.4	4.1	72.8	4.4	73.9	4.0
	amp	19.8	5.7	17.1	4.1	18.7	6.3	19.4	5.7
t5	dur	818.9	292.2	773.9	210.0	778.3	296.4	815.8	288.6
	per	72.7	3.0	71.3	3.1	72.0	3.3	72.7	3.0
	amp	20.5	5.2	18.9	4.1	17.3	5.1	20.4	5.3
t7	dur	1015.5	365.1	887.8	343.1	894.8	322.3	995.3	362.2
	per	70.0	2.8	70.1	2.7	67.4	2.8	69.9	2.9
	amp	20.4	5.9	20.3	5.9	18.6	5.3	20.4	6.0
t8	dur	824.0	278.0	728.0	279.0	740.0	216.0	790.7	265.6
	per	71.2	3.5	73.2	3.3	69.9	3.3	71.2	3.6
	amp	16.6	4.3	14.0	3.2	16.3	4.1	16.0	4.2
t10	dur	782.0	230.0	745.0	182.0	700.0	245.0	768.7	224.4
	per	71.6	3.4	73.4	3.4	70.7	3.9	71.9	3.5
	amp	18.8	4.8	16.7	3.8	18.1	4.3	18.3	4.6
t16	dur	847.3	243.0	780.0	206.0	-	-	816.4	230.6
	per	72.4	3.0	72.5	2.9	-	-	72.4	2.9
	amp	19.2	5.7	16.1	4.0	-	-	17.8	5.3

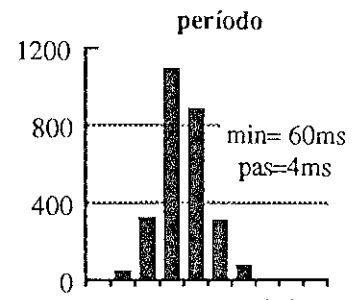
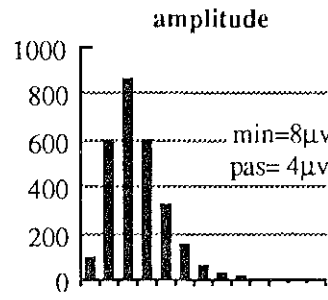
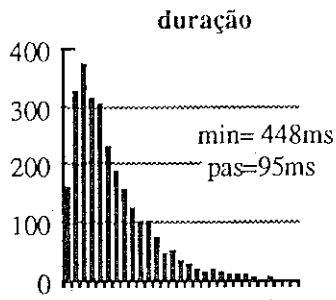
tabela 5.3: Estatísticas nos parâmetros de sigma por estádio (classificação visual) e total (ao longo da noite): dur e per em ms, e amp em μv .

Gondeck [1974] e Príncipe [1981] referem que a frequência dos fusos de sono em estágio 2 aumenta com a idade (a partir dos 13 anos). Na tabela 5.3 pode verificar-se esta tendência (embora com valores médios diferentes): t5, t4 e t16 têm frequências inferiores (13.7, 13.4 e 13.8) aos outros casos (14.2, 14, 13.9). Esta tendência é visível nos histogramas do período (figuras 5.7 a 5.12) que apresenta o número máximo de eventos em 72ms com um outro pico em 68 e 76 conforme a idade, excepto em t7 que apresenta sensivelmente o mesmo número de eventos em 68 e 72ms. Note-se que nestes casos o histograma é aproximadamente simétrico, isto é a mediana está muito próxima da média; o mesmo acontece com o histograma das amplitudes, ao contrário do histograma da duração (figuras 5.7 a 5.12) que é assimétrico sobretudo nos casos t4 e t7, em que a média da duração tem valores mais elevado. Em qualquer dos casos, existe uma grande percentagem dos fusos com duração compreendida entre 500 e 850 ms o que coincide com os valores publicados. No entanto, existem casos em que a duração ultrapassa 1.5s o que não coincide com os valores publicados nos estudos referidos, mas aqui o critério de final de evento é diferente: 3 ondas (em 6) que não obedeçam ao critério onda válida, enquanto que nos outros estudos bastava uma. Note-se no entanto que em qualquer dos estudos as ondas não-válidas não contribuem para o cálculo da duração.

2.2.2: Actividade Delta

Esta actividade é característica do sono profundo, também conhecido por sono lento: os estádios 3/4. A classificação destes estádios baseia-se na percentagem de tempo ocupada por esta actividade numa época (deve ser superior a 20%).

actividade sigma



actividade delta

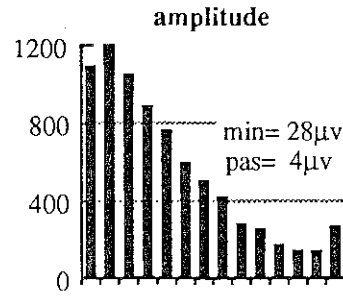
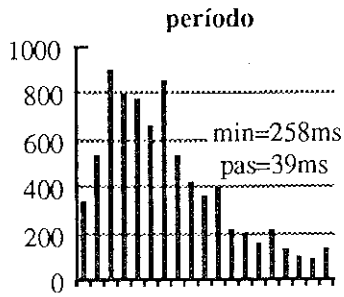
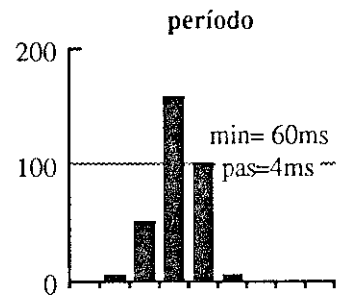
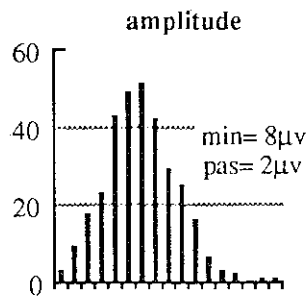
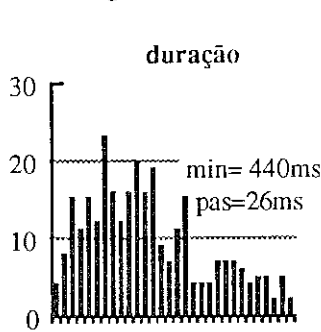


figura 5.7: Histogramas dos parâmetros das actividades sigma e delta detectadas em t4.

actividade sigma



actividade delta

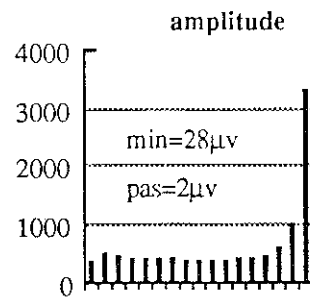
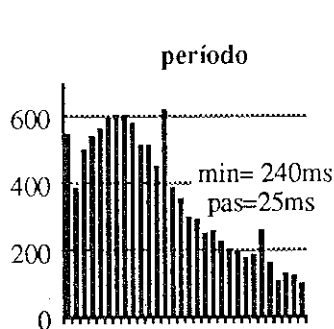
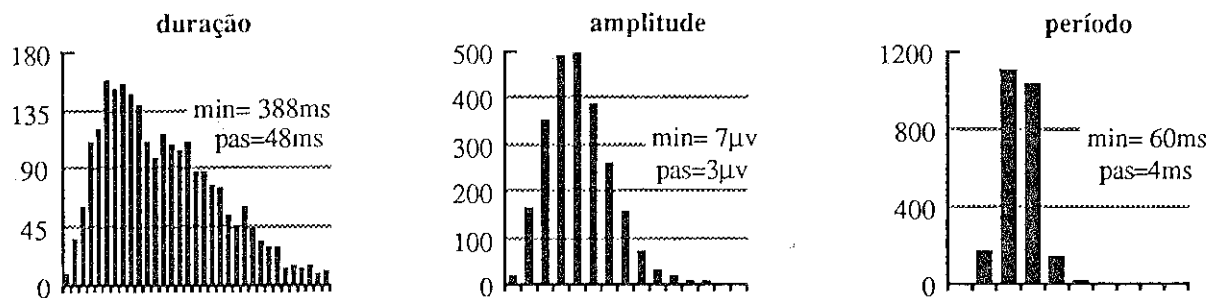


figura 5.8: Histogramas dos parâmetros das actividades sigma e delta detectadas em t5.

actividade sigma



actividade delta

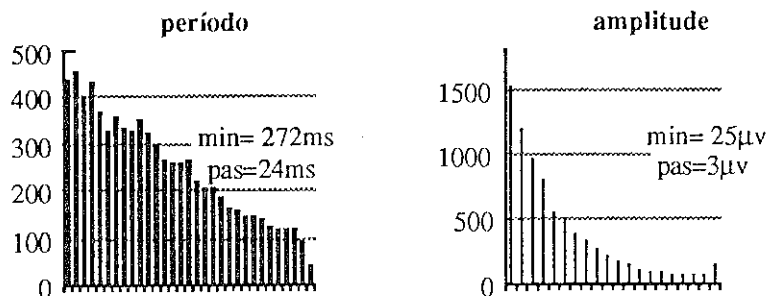
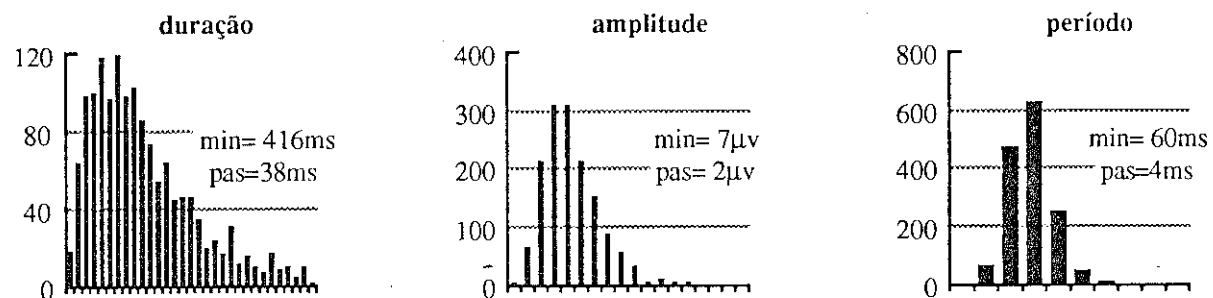


figura 5.9: Histogramas dos parâmetros das actividades delta e sigma detectados em t7

actividade sigma



actividade delta

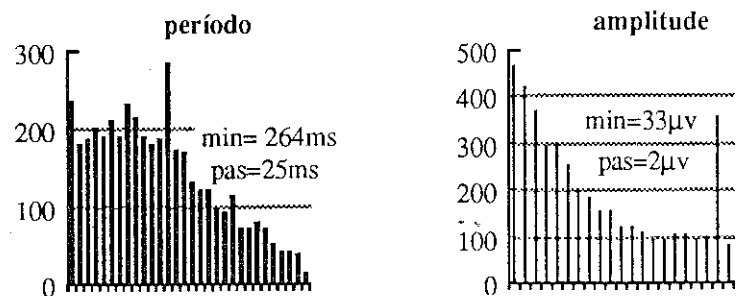
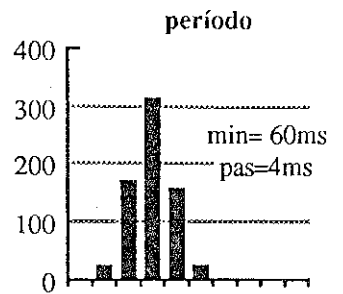
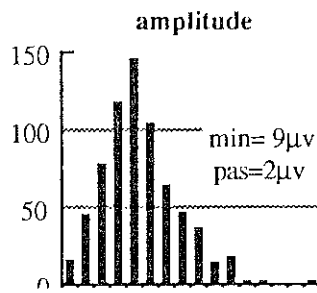
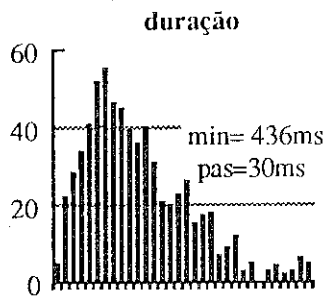


figura 5.10: Histogramas dos parâmetros das actividades delta e sigma detectadas em t8

actividade sigma



actividade delta

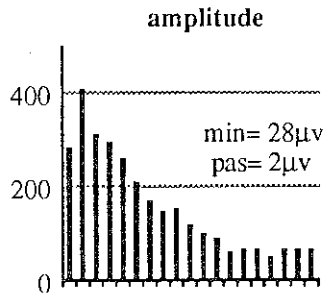
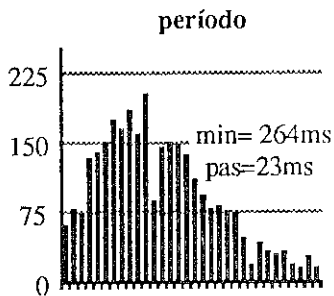
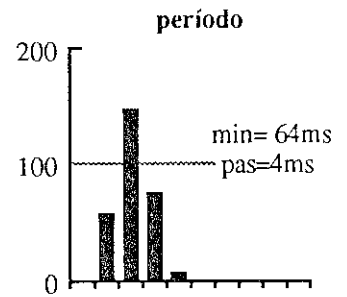
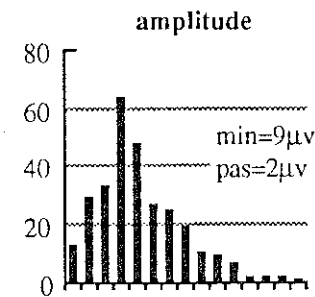
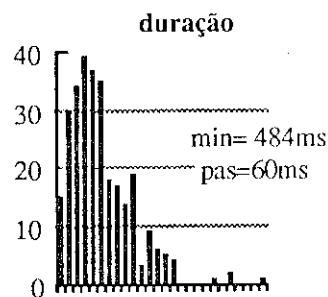


figura 5.11: Histogramas dos parâmetros das actividades delta e sigma detectadas em t10.

actividade sigma



actividade delta

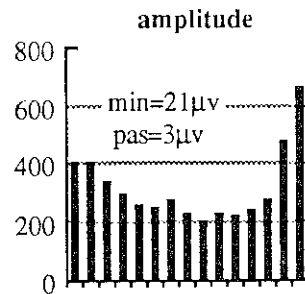
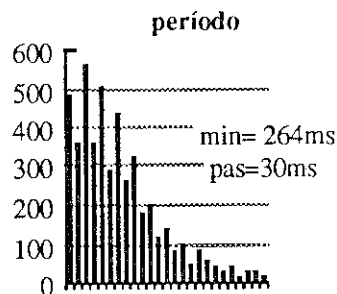


figura 5.12: Histogramas dos parâmetros das actividades delta e sigma detectadas em t16

CASOS	ESTADIO 2		ESTADIO 3/4		ESTADIO 0		ESTADIO 1/5	
	épocas	eventos	épocas	eventos	épocas	eventos	épocas	eventos
t4	134	766	101	6524	13	20	69	70
t5	106	2070	99	8900	4	6	55	282
t7	129	3769	33	2077	12	157	38	364
t8	135	972	62	2812	15	94	120	250
t10	107	1097	50	1763	1	-	92	-
t16	49	1403	28	2994	11	68	128	305

tabela 5.4: Número de eventos delta por estádio

Na tabela anterior estão representados o número de épocas de cada estádio e o número de eventos por estádio. O número de eventos por época de estádio 3/4 é sempre maior, excepto em t7 em que número de eventos delta é máximo em estádio 2. Mas neste caso, o número de épocas de estádio 2 é quatro vezes superior ao número de épocas 3/4 (aliás classificadas visualmente como 3).

CASO	ESTADIO 2		ESTADIO 3/4		ESTADIO 0		ESTADIO 1/5		TOTAL	
	\bar{x}	s _d	\bar{x}	s _d	\bar{x}	s _d	\bar{x}	s _d	\bar{x}	s _d
t4 per	582.3	188.0	493.0	155.6	680.4	145.7	633.7	143.0	512.5	167.6
	amp	42.2	13.3	46.0	13.8	42.2	10.2	40.5	12.3	45.3
t5 per	585.9	194.6	474.1	164.7	-	-	605.9	199.3	518.4	185.8
	amp	46.1	11.3	51.2	9.7	-	-	43.4	11.3	49.6
t7 per	525.3	196.6	496.6	182.9	565.1	194.5	612.3	208.1	520.7	194.0
	amp	37.2	13.3	41.1	13.5	46.5	21.2	36.2	13.6	38.7
t8 per	507.1	179.8	551.0	177.1	571.4	221.6	570.3	193.9	542.1	180.8
	amp	46.7	12.8	48.7	12.6	59.1	14.2	46.1	12.4	48.3
t10 per	514.7	154.3	551.3	157.7	-	-	585.9	201.9	539.5	160.4
	amp	39.9	10.1	39.6	9.5	-	-	39.5	10.3	39.7
t16 per	451.9	151.3	443.7	133.7	619.4	197.3	597.9	211.1	458.9	154.3
	amp	37.8	13.9	48.1	13.3	33.9	12.9	32.4	11.9	43.8

tabela 5.5: Estatísticas nos parâmetros de delta para todos os estádios (classificação visual) e durante a noite: per em ms e amp em μ v

As estimativas dos valores médios e do desvio padrão dos eventos delta detectados nos diferentes estádios estão representados na tabela seguinte. O teste t-student efectuado (nos casos t4, t16, t8 e t10) entre as detecções em estádio 3/4 e cada um dos outros estádios rejeita a hipótese nula (igualdade das médias) com nível de significância 0.01 para todas os casos quando se trata do parâmetro período. Para o caso da amplitude tem-se a mesma resposta, excepto para o caso t10. As estimativas da média da amplitude têm, em qualquer dos casos, valores superiores nos estádio 3/4 do que em qualquer dos outros estádios, excepto em alguns estádio 0, o que é aqui justificável com a presença de movimentos "arousal" que o detector identifica com delta; mas o período das ondas em estádio 0 é bastante superior ao do estádio 3/4, e esta tendência é também visível no estádio 1/5. O período das ondas em estádio 2 é inferior aos valores de período em estádio 1/5 e 0, e com valor mais próximo (superior ou inferior) do valor estimado para estádio 3/4. Um dos critérios de detecção da onda delta foi considerar como actividade válida só ondas com valor de amplitude superior a $20\mu\text{v}$ e, como vemos o valor médio das ondas detectadas em estádio 3/4 é bastante mais elevado (pelo menos o dobro) e, os histogramas (figuras 5.7 a 5.12) revelam uma saturação em dois dos casos (t5 e t16). Smith [1977] refere que o valor da amplitude desta actividade decresce com a idade, mas esta tendência não é visível nestes dados. Mas, Smith escolhe para valor mínimo de amplitude o valor de $5\mu\text{v}$ considerando, tal como Johnson [1977], que esta actividade está presente em todos os estádios só que a classificação visual tende a menosprezá-la quando o seu valor de amplitude é baixo. Smith não encontra nenhuma tendência relacionável com a idade para os valores de frequência das ondas com amplitudes superiores a $20\mu\text{v}$. Este resultado é também confirmado por estes dados, em que os histogramas do período das ondas são assimétricos com a maioria dos eventos entre 0.25s e 0.6s.

2.2.3: Conclusões

O estudo estatístico dos parâmetros em cada um dos indivíduos veio confirmar, de forma quantitativa, a variabilidade inter-indivíduos dos valores das actividades do EEG. A variabilidade inter-indivíduos pode se confirmada com um teste F, em que se testa a hipótese nula, ou de igualdade de médias. Esta hipótese é rejeitada com nível de significância 0.05 para todos os parâmetros da actividades delta e sigma. Contudo, dada a definição das ondas, esta diferença pode ser devida a falsos positivos. A microanálise permite também caracterizar quantitativamente as actividades do EEG, uma vez que esta informação estatística dos parâmetros é directamente relacionável com as actividades. Pode, portanto, ser utilizada em estudos clínicos (por exemplo influência de drogas) completando o conhecimento qualitativo dos especialistas sobre os sinais.

3: Validação do método de classificação

A classificação em estádios de sono é feita, nestes sistemas, com base nas regras de Rechtschaffen e Kales (R+K), portanto deve utilizar-se a informação percentual das actividades presentes na época (neste caso, o minuto). Deste modo, a entrada da rede neuronal será a informação percentual das actividades consideradas importantes na classificação, sendo a saída o respectivo estágio de sono. Naturalmente que o objectivo será que a rede tenha a mesma resposta, para a classificação de cada minuto, que a obtida pelo especialista na classificação visual. As redes neuronais nunca foram utilizadas em classificação de sono, por isso se justifica tentar apurar a qualidade deste novo método. Com este objectivo utilizaram-se dados já classificados por outros métodos de classificação: um

"Expert System" e um "Belief Automaton" [Príncipe et al - 1989] em que a percentagem de concordância foi respectivamente 86.3% e 84.9%. Estes dados foram recolhidos por um "Front-End" desenvolvido na Universidade da Florida [Chang et al - 1988].

Por informação a fornecer à rede optamos pelas actividades do EEG: alfa, beta, delta, sigma e artefactos musculares, classificando as épocas em quatro estádios: 0, 1/5, 2, 3/4. Portanto, para cada época (minuto), existe um vector que contém o número de segundos de alfa, beta, delta, artefactos musculares e, o número de "bursts" de sigma ("spindles"). Esta informação pode ser a entrada directa da rede (designada por entrada analógica) ou codificada em binário (entrada binária) com quatro níveis tendo a possibilidade de distinguir, para cada actividade, entre percentagens nulas, baixas, médias e elevadas.

3.1: Entradas Binárias

Em vez de se utilizar uma divisão uniforme das actividades pelos quatro níveis, a codificação das actividades em 2 "bits" reflecte o conhecimento das regras de R+K para a classificação do sono. Para isso, atribuímos a cada nível de codificação a percentagem de actividade que é relevante na classificação de R+K.

saída	0	1	2	3
alfa (seg)	0-5	6-10	11-30	>30
sigma (nº)	0	1	2-3	>3
delta (seg)	0-5	6-12	>12	-
beta (seg)	0-5	6-10	11-30	>30
artefactos (seg)	0-5	6-10	11-20	>20

tabela 5.6: Codificação da quantidade de cada actividade em 4 níveis.

Por exemplo, quantidades de alfa (ou qualquer das outras actividades) inferiores a 5s podem ser consideradas sem qualquer importância; enquanto uma quantidade de alfa superior a 30s deve ter o valor máximo da codificação; por sua vez sempre que delta tenha valores superiores a 12s deve atingir o nível máximo. Na tabela 5.6 apresentamos os valores de codificação para cada um das actividades. Os ficheiros de dados, e naturalmente os padrões de aprendizagem são codificados tal como o indicado nesta tabela. Consequentemente, tratando-se de informação em binário o perceptrão tem 10 entradas (2 por actividade) e quatro saídas (uma para cada um dos estádios considerados). Além do perceptrão escolhemos também outras configurações com camadas "hidden": com uma camada (5 unidades de processamento) e duas camadas (5+3).

3.1.1: Aprendizagem

O treino numa rede é, como vimos no capítulo anterior, uma fase determinante no funcionamento das redes neuronais. Tem por objectivo descobrir o "mapping" entre o espaço das entradas e o espaço das saídas, logo o poder de generalização, a propriedade mais importante das redes em classificação, fica dependente da fase de treino. Como referimos, a topologia também é determinante neste processo e, embora existam algumas indicações de índole geral na literatura sobre o comportamento das topologias em classificação, a escolha tem sido um processo experimental. Isto é, para a solução dum problema escolhe-se a topologia, variando o número de camadas e/ou unidades de processamento, que apresente melhores resultados. Nesta aplicação, vamos utilizar as três topologias usualmente referidas em classificação: perceptrão, uma camada e duas camadas "hidden".

Padrão I (II)	alfa	beta	delta	sigma	artefacto
ESTAGIO 0	2	0	0	0	2
	2	2	0	0	3
	3	0	0	0	0
	2	0	0	0	0
(II)	2	2	0	0	0
(II)	3	1	1	1	1
ESTAGIO 1/5	0	3	0	0	0
	0	2	0	0	0
	0	0	0	0	0
	0	1	0	0	0
(II)	1	3	0	0	1
(II)	1	2	1	1	1
ESTAGIO 2	0	0	0	1	0
	0	0	0	2	0
	0	0	0	3	0
	0	0	1	2	1
(II)	1	1	1	2	2
(II)	0	2	0	3	1
ESTAGIO 3/4	0	0	2	1	0
	0	0	2	1	0
	0	0	2	3	2
	0	0	2	0	0
(II)	0	0	2	3	0
(II)	1	1	2	3	1

tabela 5.7: Padrões de aprendizagem: I e II.

Quanto à escolha do conjunto de treino tem, na maioria dos casos, seguido uma única regra, escolha aleatória e do maior número possível de padrões no conjunto de teste. Esta aplicação sugere-nos outra alternativa, traduzir para um conjunto de vectores o conhecimento que o especialista tem sobre o problema de classificação do sono. Destes factos resultam três grupos de vectores treino, conforme a metodologia seguida na respectiva escolha, que designaremos por estratégias de aprendizagem: uma que representa o conhecimento *a priori* do especialista sobre a classificação de sono, a segunda um dos ficheiros completos (grande número de vectores) e, finalmente escolha aleatória de um pequeno grupo de vectores em cada ficheiro (o mesmo número do primeiro grupo). Assim, tendo em atenção o processo de codificação dos dados, para a primeira estratégia escrevemos dois grupos de vectores (tabela 5.7), um que contém unicamente exemplos típicos e bem definidos de cada estágio (padrão I - 16 vectores) e outro que, para além dos

anteriores, contém também um grupo de vectores marginais (padrão II - 24 vectores). A segunda estratégia baseia-se na aprendizagem de um ficheiro completo de dados, utilizando esta rede para a classificação dos outros ficheiros. Na terceira, os padrões aleatórios são recolhidos nos ficheiros de dados os quatro primeiros padrões distintos de cada estádio.

3.1.2: Resultados

Tal como foi referido no capítulo anterior, para a aprendizagem aplica-se o algoritmo de “backpropagation”, sendo feita a simulação com um conjunto de programas do grupo PDP [McClelland et al-1988]. A cada uma destas redes foram apresentados os 4 grupos de padrões, referidos na secção anterior, efectuando a aprendizagem até que o erro quadrático médio, depois de decrescer bastante, estabilize. Naturalmente o valor do erro e o número de iterações difere de caso para caso.

Na tabela 5.8 apresentamos a percentagem de concordância (entre a classificação pelo especialista e a rede neuronal) para cada uma das topologias referidas, após a aprendizagem dos padrões I e II referidos na secção anterior.

CASOS pad. treino	10067		10109		11623		11717		11740	
	I	II	I	II	I	II	I	II	I	II
perceptrão	81.5	81.7	86.5	86.5	96.9	96.4	89.5	90.9	84.3	84.5
1 camada	82.0	80.8	86.3	86.7	96.9	95.7	91.0	89.5	83.6	81.1
2 camadas	82.0	78.6	86.7	86.1	96.9	96.2	87.6	89.3	82.9	81.6

tabela 5.8: Resultados de classificação com três topologias e dois padrões de aprendizagem.

Estes resultados são bastante idênticos quer do ponto de vista da topologia, quer do ponto de vista do padrão de aprendizagem. Os resultados do perceptrão são

sempre muitos próximos do perceptrão de uma camada: o caso com grau de concordância (11623) mais elevado tem o mesmo valor, em 3 casos os resultados do perceptrão são sensivelmente melhores (0.3%), e no restante (11740) a rede de uma camada é ligeiramente melhor (0.9%). Os resultados para a rede de duas camadas são sempre (excepto num caso) inferiores aos do perceptrão, o que é de certo modo contraditório pois com estas redes é possível uma divisão do espaço em classes com configuração arbitrária. Mas isso, poderia significar que neste caso o conhecimento sobre o conjunto de dados deveria ser aumentado durante a fase de treino. Esta hipótese também não parece comprovada porque a diferença entre a classificação desta topologia após a aprendizagem dos padrões I e II nem sempre tem melhores resultados com a classificação após aprendizagem de II (mais padrões). Contudo a diferença do número de padrões entre os dois grupos de padrões é escassa.

Após a aprendizagem de 16 padrões aleatórios diferentes (4 de cada estágio) de cada um dos ficheiros pelo perceptrão obtiveram-se os resultados da tabela 5.9.

Aprendizagem	10067	10109	11623	11717	11740
Aleatória	52.6	72.0	83.4	90	63.6
Toda a noite	82.0	87.8	96.9	93.1	89.8

tabela 5.9: Aprendizagem dum padrão aleatório(extraído do ficheiro) ou todo o ficheiro com classificação no próprio ficheiro

Na mesma tabela está também representado, o grau de concordância após a aprendizagem (até o erro quadrático médio convergir) do ficheiro completo. No caso da aprendizagem do padrão aleatório os resultados são bastante piores que os anteriores (aprendizagem dos padrões I e II). Na aprendizagem do ficheiro completo os resultados são sensivelmente os mesmos do perceptrão após a aprendizagem com padrão I, excepto o caso 11717 (3% superior)

Após o treino do perceptrão e da rede de uma camada com um ficheiro completo (11623) obtiveram-se os resultados da tabela 5.10. Como vemos os resultados são sensivelmente melhores no perceptrão que na rede de uma camada, embora não se possa dizer que estes resultados sejam melhores que os obtidos pela mesma topologia quando treinada com o padrão I; houve melhoria em 10109 (0.7%), piorou em 2 casos (de 6.8% no caso 11740), e finalmente em 11740 manteve o mesmo grau de concordância. Note-se ainda que a aprendizagem de 11623 obteve o mesmo grau de concordância que o teste da mesmo ficheiro após a aprendizagem do padrão I.

	10067	10109	11623	11717	11740
perceptrão	80.8	87.2	96.9	90.9	82.9
1 camada	80.8	86.5	97.7	90.9	76.3

tabela 5.10: Aprendizagem de 11623 e classificação em todos os ficheiros com as duas topologias

Pode assim perguntar-se se os padrões I e II são suficientemente representativos dos padrões de sono, isto é, atendendo a que existe uma codificação dos ficheiros de entrada quantos vectores são realmente distintos nesses dados, e por outro lado quantos é que coincidem com os padrões de aprendizagem referidos. Na tabela 5.11 apresentamos essa informação: quantos são bem classificados (pelo perceptrão), quantos coincidem com os padrões de aprendizagem e, conseqüentemente, quantos são inferidos. Com a codificação utilizada poderíamos pensar à partida em 1024 vectores possíveis e, como vemos, só existe um pequeno conjunto daqueles vectores em cada um dos ficheiros, o que é naturalmente característica da estrutura intrínseca do sono, que torna algumas combinações impossíveis. É no entanto de salientar o poder de generalização do perceptrão, pois em todos os casos o número de padrões inferidos é sempre maior que o número de padrões coincidentes com os padrões de aprendizagem (pelo

menos, 2.5 vezes superior). O número de coincidências dos ficheiros com os padrões de aprendizagem é, em geral, maior no padrão II mas, como vimos, isso não se traduz num aumento da percentagem de concordância na classificação dos mesmos ficheiros (nem na classificação dos padrões distintos), excepto no caso 11740.

CASOS	10067		10109		11623		11717		11740	
vectores ≠s	53		39		35		43		48	
treino	I	II	I	II	I	II	I	II	I	II
concordante	34	34	23	23	30	29	27	26	30	33
coincidentes	12	12	9	11	9	11	6	8	10	11
inferidos	22	22	14	12	21	18	21	18	20	22

tabela 5.11 : Vectores distintos em cada ficheiro e coincidências com cada um dos padrões considerados

3.2: Entradas analógicas

Pode perguntar-se se o processo de quantificação dos dados de entrada da rede neuronal (a codificação em binário) limita os resultados da classificação, por isso resolvemos fazer experiências idênticas com entradas 'analógicas'. Por dado analógico queremos dizer os resultados, dados directamente pelo primeiro nível de processamento, em segundos de cada actividade no minuto (ou número de "bursts" no caso dos fusos). Estes valores precisam ser normalizados para o intervalo [0,1]. A normalização das actividades quantificadas em segundos é facilmente obtida dividindo aquele valor do tempo total em segundos da época por 60. Para o número de fusos é preciso arbitrar um valor, decidimos que fosse 10. Este valor é bastante superior ao número médio de fusos/época em cada um dos casos (varia entre 0.7 e

4.9) e, por outro lado, só um dos casos é que apresenta épocas com mais de 10 fusos.

Treino → Teste ↓	10067	10109	11623	11717	11740
10067	82.1	71.9	80.8	59.6	71.7
10109	88.2	91.8	86.7	64.3	65.3
11623	95.7	91.4	96.2	83.1	65.2
11717	79.8	83.8	85.9	90.7	80.9
11740	77.2	81.8	80.4	85.7	91.0

tabela 5.12: Percepção aprendizagem e teste em ficheiros completos

Claro que a topologia da rede tem que ser diferente pois o número de entradas (5 unidades de processamento) é também diferente: percepção de 5 entradas e 4 saídas e rede uma camada com 10 unidades de processamento na camada intermédia e o mesmo número de entradas e saídas. Mais uma vez a fase de treino decorreu até estabilizar o erro quadrático médio, depois de ter diminuído significativamente, tendo como padrão de treino um ficheiro completo, o que permite comparação imediata com os resultados anteriores. Além disso, fez-se um teste de classificação em todos os outros ficheiros. Se compararmos os resultados deste teste com os anteriores quando a aprendizagem e a classificação são feitas no mesmo ficheiro (diagonal da tabela 5.12 e 2ª linha de 5.9) vemos que não podemos tirar conclusões unilaterais, isto é, com entradas analógicas, dois casos (10109 e 11740) melhoram de 4% no máximo, outros têm ligeiras diferenças e um caso (11717) piora de 2.5%. No caso da aprendizagem de 11623, os resultados obtidos com dados analógicos (3ª coluna da tabela 5.12) são aproximadamente iguais em 3 dos casos e piores nos restantes (11717 e 11740), que os do mesmo tipo de rede com entradas binárias (tabela 5.10). Além disso, podemos dizer que, quando se utiliza a aprendizagem dum ficheiro na classificação dos outros, os resultados não

são bons mas a degradação não é geral, isto é, não há um teste que seja igualmente mau em todos os ficheiros, excepto no caso de 11740 mas este caso não tem sono profundo (estádio 3/4). Em geral, se compararmos as colunas da tabela anterior vemos que têm melhores resultados com um grupo de ficheiros (concordâncias muito próximas das obtidas com a aprendizagem no próprio ficheiro), piorando bastante com outros. Deste modo tudo parece indicar que a variabilidade inter-indivíduo também se reflecte na aprendizagem da redes formando um conjunto diferente de atractores no espaço multidimensional .

Quando a rede de uma camada é treinada e testada no mesmo ficheiro os resultados melhoram (tabela 5.13) relativamente ao perceptrão de entradas analógicas (tabela 5.12). No entanto quando a aprendizagem é feita num dos ficheiros e a classificação nos outros ficheiros, uns casos são melhores na multicamada e outros pioram. A tendência geral da aprendizagem de ficheiros completos com perceptrão mantém-se, isto é, a degradação na classificação dos outros ficheiros também não é geral. Se comparamos o teste com aprendizagem de 11623, no mesmo tipo de rede com entradas analógicas e binárias (tabelas 5.13 e 5.10) vemos que, nuns casos a concordância é melhor com entrada analógico, noutros é o inverso, sendo a diferença máxima de 2.7%.

Treino → Teste ↓	10067	10109	11623	11717	11740
10067	85.3	71.4	78.8	62.5	66.8
10109	86.5	91.4	83.9	63.8	60.2
11623	95.9	86.3	98.9	82.3	66.2
11717	86.4	91.2	91.9	94.5	81.2
11740	82.9	82.7	79.0	82.2	92.1

tabela 5.13: Rede multicamada (uma camada): aprendizagem e teste em ficheiros completos

Outro dos aspectos estudados diz respeito a importância relativa das actividades para a classificação do sono. As regras de R+K dão mais importância às actividades: alfa, sigma e delta; e no caso de se querer separar os estádios 1/5 à presença de REM; todas as outras actividades aparecem mencionadas como possíveis. Assim consideramos apenas aquelas actividades e treinamos um perceptrão e uma rede de uma camada com os ficheiros completos, obtendo os resultados da tabela 5.14. Naturalmente que as topologias das redes também se modificam, existindo apenas três entradas, por isso o perceptrão tem 3 unidades de processamento de entrada e 4 de saída (3+4) e consideraremos 6 unidades na camada "hidden" da rede de uma camada (3+6+4).

Teste e Treino	10067	10109	11623	11717	11740
perceptrão	81.5	91.0	97.2	93.0	85.0
uma camada	81.9	89.9	97.5	90.6	85.1

tabela 5.14: Aprendizagem e teste dos ficheiros completos considerando apenas três actividades: alfa, sigma e delta.

Os resultados do perceptrão com três actividades são melhores em dois casos (10109 e 11623), ligeiramente inferiores (0.8%) em outros dois casos e piores (6%) no restante (11740), que os obtidos pelo mesmo tipo de rede com cinco actividades. A rede multicamada é sempre melhor (2% a 6%) quando se consideram as cinco actividades. Agora comparando os resultados obtidos com as duas topologias com as três entradas também não temos resultados unilaterais, isto é, há resultados melhores no perceptrão que na rede unicamada e vice-versa; ao contrário dos resultados com cinco actividades que era sempre melhor para a rede de uma camada.

Também podemos perguntar, qual é o efeito na classificação de escolher padrões que contenham unicamente a informação das actividades dominantes. Com

entradas analógicas considerando apenas três actividades é muito simples escrever um grupo de vectores que representem o conhecimento do especialista sobre o sono. Assim, o padrão III foi escrito tendo unicamente em atenção que a presença de determinada actividade, com um valor mínimo, implica determinado estágio. Podemos no entanto pensar em outras combinações possíveis também contempladas nas regras de R+K: por exemplo, a presença de actividade sigma em estágio 3/4 e a de delta em estágio 2, e ainda uma percentagem baixa de alfa que muitas vezes aparece em REM. Obtivemos deste modo o padrão IV: o conjunto de 7 vectores representados na tabela 5.15.

saída	alfa	delta	sigma
estádio 0	0.2	0.0	0.0
estádio 1/5	0.0	0.0	0.0
estádio 2	0.0	0.0	0.1
estádio 3/4	0.0	0.2	0.0
estádio 1	0.1	0.0	0.0
estádio 2	0.0	0.15	0.1
estádio 3/4	0.0	0.25	0.2

tabela 5.15: Padrões de aprendizagem III e IV

Os resultados da classificação, após a aprendizagem do padrão III, obtidos pelas topologias já referidas estão representados na tabela 5.16. Como vemos os resultados nalguns casos são ligeiramente melhores no perceptrão, em 11717 a diferença é máxima (6%). No entanto, a diferença entre os resultados do perceptrão com a aprendizagem deste padrão e da noite completa chega a atingir aproximadamente 10 % (em 11717) enquanto que nas mesmas circunstâncias com entradas binárias a diferença máxima era de aproximadamente 5% (tabelas 5.8 e 5.9).

Teste	10067	10109	11623	11717	11740
perceptrão	79.5	86.9	94.7	83.6	80.9
uma camada	79.5	85.7	94.4	77.6	80.5

tabela 5.16: Teste do perceptrão e da rede uma camada com aprendizagem do padrão III.

A aprendizagem do padrão IV, não foi totalmente conseguida pelo perceptrão, o erro convergiu para um valor superior à unidade, não aprendendo o último vector referente a estágio 3/4, a cujo valor máximo na saída correspondia a estágio 2. O referido padrão foi completamente aprendido pela rede com uma camada "hidden". Os resultados da classificação com ambas as topologias estão representados na tabela 5.17.

Teste	10067	10109	11623	11717	11740
perceptrão	80.8	88.4	94.4	81.7	84.1
uma camada	81.8	89.5	96.4	88.8	83.9

tabela 5.17: Perceptrão e rede unicamada após aprendizagem de padrão IV

Os resultados obtidos pela rede de uma camada nesta experiência são sempre melhores que os do perceptrão, e sempre melhores que os obtidos com a aprendizagem do padrão III qualquer que seja a topologia considerada. A diferença entre a concordância obtida pelo padrão IV na rede de uma camada e os resultados obtidos pela aprendizagem da rede completa diminui para aproximadamente 2% (valor idêntico ao obtido com entradas binárias).

No caso de entradas analógicas podemos também questionar quantos vectores distintos existem nos ficheiros, destes quantos são coincidentes com os padrões de amostragem, e quantos são inferidos pela rede. Na tabela 5.17 estão representados estes valores conjuntamente com as classificações correctas do perceptrão e da rede de uma camada com a aprendizagem dos padrões III e IV, respectivamente.

CASOS	10067		10109		11623		11717		11740	
vectores ≠s	149		216		163		186		132	
treino	III	IV	III	IV	III	IV	III	IV	III	IV
concordante	102	107	168	170	148	152	128	146	91	95
coincidentes	3	4	3	4	2	3	2	4	2	4
inferidos	99	103	165	166	145	149	126	142	89	91

tabela 5.18: Entradas analógicas: padrões distintos em cada um dos ficheiros, concordantes com os padrões de aprendizagem e inferências.

Quando se consideram as entradas analógicas, o número de vectores distintos (apesar do número de actividades ser inferior) em cada um dos ficheiros, aumenta consideravelmente (pelo menos o dobro). No entanto a percentagem de vectores distintos bem classificados é sempre maior com este tipo de entradas que com as entradas binárias (pelo menos 5% superior), embora, em qualquer dos casos estas percentagens de concordância sejam sempre inferiores às percentagens obtidas para os ficheiros completos.

3.3: Conclusões

Resumindo, podemos dizer que, quando se transmite à rede neuronal o conhecimento do especialista sobre a classificação do sono, se obtêm melhores resultados do que com as outras estratégias de treino. Foram feitas experiências com os dois tipos de codificação, entrada binária (5 actividades - tabela 5.8) e entrada analógica (3 actividades- tabelas 5.16 e 5.17). As percentagens de concordância obtidas nestes testes tiveram valores idênticos aos outros métodos já referidos. Se considerarmos os melhores resultados em ambas as experiências, padrão I/II com percepção e padrão IV com rede de uma camada, e os compararmos com os resultados de aprendizagem e teste de ficheiros completos das

mesmas topologias vemos que, na maioria dos casos não há diferenças significativas; com entradas binária/perceptrão a diferença varia entre 1 e 5 %; com entrada analógica a diferença é da ordem de 1%. Quando se considera aprendizagem num ficheiro e teste nos outros os resultados são invariavelmente piores, mesmo com a aprendizagem de 11623, o ficheiro onde se obtêm percentagens de concordância sempre elevadas; com a aprendizagem noutros ficheiros a degradação acentua-se e não é unilateral, isto é, não há casos igualmente maus ou bons (tabelas 5.12 e 5.13). Somos, por estas razões, levados a pensar que a variabilidade inter-indivíduos se reflecte na aprendizagem. Quando se utilizam poucos padrões aleatórios os resultados podem ser bastante piores, mesmo quando retirados do mesmo ficheiro onde são feitos os testes (tabela 5.9). Assim, podemos concluir que para haver aprendizagem com padrões aleatórios deveremos ter um grande número de vectores, o que nesta aplicação não traz qualquer vantagem dada a variabilidade inter-indivíduo. Consequentemente, nem sempre por se apresentar mais quantidade de informação na fase de aprendizagem se tem melhor generalização, pois como verificamos há também dependência do tipo de informação apresentada. No entanto, como toda a informação fica codificada, não temos processo de saber qual é a informação que devemos acrescentar para melhorar o grau de concordância da classificação.

No caso de entradas binárias, os resultados com redes multicamada não foram animadores pois o grau de concordância não melhorou com a utilização destas redes (embora sejam resultados não apresentados, o erro quadrático médio total em cada teste diminui para estas redes). Com entradas analógicas e padrão III na aprendizagem, os resultados também não melhoram com as topologias multicamada; já com o padrão IV a rede de uma camada teve, sem dúvida, melhores resultados, mas aqui o perceptrão não conseguiu aprender o conjunto de padrões. Assim, também podemos concluir que, com esta estratégia de treino, sempre que

um perceptrão é capaz de codificar toda a informação presente nos vectores de aprendizagem não há qualquer vantagem na utilização de redes multicamada.

A quantificação prévia dos dados (a codificação binária) também não tem influência significativa nos resultados obtidos. Se compararmos os resultados da aprendizagem e teste do perceptrão com ficheiros completos com ambos os processos de codificação (tabelas 5.9 e 5.12 - diagonal) vemos que as diferenças são no máximo de 3% e podem ser favoráveis aos dados analógicos ou aos digitais. Em contrapartida, o período de aprendizagem nos padrões analógicos é muito superior ao período de aprendizagem para entradas binárias. Por exemplo, para a aprendizagem, pelo perceptrão, dos padrões I/II bastam 200 iterações, enquanto que o padrão III leva à volta de 16000. A diferença entre as duas formas de codificação reside no grau de generalização em padrões distintos que é superior no caso analógico. Em qualquer dos casos, o grau de generalização é elevado quando a aprendizagem é feita, como referimos, com padrões criados a partir do conhecimento existente sobre o assunto, mesmo quando essa informação é reduzida ao mínimo (padrão III).

Quando se comparam os resultados, da aprendizagem e teste em ficheiros completos com os dados analógicos, utilizando 3 e 5 actividades também não podemos tirar conclusões claramente favoráveis a um dos processos. Com o perceptrão (tabela 5.12 e 5.14) temos dois casos idênticos, 2 casos favoráveis à rede com 3 actividades na entrada (2.3% de diferença máxima), e um favorável à rede com 5 actividades (6%); os casos que, com rede dimensionada para 3 actividades, têm concordância elevada (acima de 90%) produzem valores inferiores com a rede dimensionada para 5 actividades. Com a rede de uma camada os resultados são sempre melhores para a rede com 5 actividades na entrada, com a diferença compreendida entre 1.5% a 6%. Podemos por isso concluir que as regras de R+K são um conjunto incompleto, isto é, sempre que uma época não tem uma

das actividades fundamentais (alfa, sigma ou delta) francamente dominante, ou ausência dessas actividades, o especialista humano comporta-se subjectivamente. Por este motivo é difícil escrever um conjunto de vectores que melhorem os resultados da classificação em todos os casos. Aliás, note-se que, quer com a aprendizagem em ficheiro completo, quer com a aprendizagem em padrões escolhidos os resultados não são uniformes, isto é, a percentagem de concordância mínima e máxima diferem sempre de aproximadamente 14%, com um valor ligeiramente inferior (0.5%) na rede uma camada com 5 actividades na entrada e aprendizagem em ficheiro completo.

4: Classificação dos dados obtidos pelo HIDRA

Uma vez que o grau de concordância global na classificação dos dados binários e dos analógicos não tem diferenças significativas mas a classificação dos padrões distintos em cada um dos casos é melhor nas entradas analógicas resolvemos considerar só este tipo de informação à entrada da rede. Portanto, a normalização dos dados é efectuada tal como foi referido: alfa e delta dividindo por 60. Nestes casos o número médio de fusos/época é superior ao número médio calculado para os outros casos (aproximadamente 1.6 vezes superior). Como em dois casos o número de fusos na época atinge valores superiores a 20 resolvemos arbitrar a normalização para 20.

Utilizando um perceptrão (3+4) e um rede de uma camada (3+6+4) treinados com os padrões III e IV, respectivamente, classificaram-se o conjunto de casos obtidos pelo HIDRA. Nas tabelas 5.19 e 5.20 apresentamos os resultados em percentagem de concordância repartida pelos diferentes estádios.

Estádios	t4 (316)	t5 (263)	t7 (371)	t8 (330)	t10 (251)	t16 (216)
0	7.6	0.0	0	6.6	0.0	0.0
1/5	2.8	58.1	60.4	5.2	61.2	95.0
2	98.5	36.1	94.4	98.5	94.4	26.5
3/4	80.1	100.0	62.7	80.9	75.5	100.0
total	68.1	64.5	80.5	58.6	78.4	72.6

tabela 5.19: Percepção após aprendizagem do padrão III

As percentagens de concordância totais são claramente inferiores às obtidas com os dados anteriores. Em geral, estes números reflectem péssima concordância no estágio 0 mas este caso está justificado pela natureza destes dados (número reduzido de épocas neste estágio que se caracterizam por baixa percentagem de alfa e movimentos "arousal").

Estádios	t4	t5	t7	t8	t10	t16
0	0.0	0.0	0.0	0.0	0.0	0.0
1/5	8.6	58.1	65.4	22.4	75.2	97.6
2	98.5	39.0	84.7	97.0	89.8	18.3
3/4	94.1	100.0	72.0	93.5	85.7	100.0
total	73.7	65.6	76.5	65.9	83.6	74.5

tabela 5.20: Rede unicamada após aprendizagem de padrão IV.

Na maioria dos estágio 1/5 a concordância é também muito baixa, das matrizes de confusão verifica-se que nesses casos o estágio 1/5 é classificado como estágio 2 o que, pelo critério utilizado, pressupõe a existência de sigma naqueles estádios. Os estádios 2 e 3/4 têm em geral elevada percentagem de concordância, excepto nos casos t5 e t16 em que o estágio 2 apresenta percentagem baixa de concordância: no caso de t5 os estádios 2 ficam classificados como estágio 3/4 e em t16 estes repartem-se igualmente por 3/4 e 1/5. Estes resultados são paralelamente

confirmados quando a rede de uma camada é treinada com o padrão IV. A percentagem global de concordância aumenta em todos os casos, mas os problemas permanecem: baixa concordância no estágio 1/5 e em alguns casos de estágio 2.

CASOS	t4		t5		t7		t8		t10		t16	
vectores ≠s	274		142		207		190		135		90	
treino	III	IV	III	IV	III	IV	III	IV	III	IV	III	IV
concordante	189	207	72	75	146	137	113	123	103	106	49	52
coincidentes	0	1	1	2	1	1	1	3	1	2	1	3
inferidos	189	206	71	73	145	136	112	120	102	104	48	49

tabela 5.21: Vectores distintos nos ficheiros de dados do HIDRA. Coincidências com os padrões de aprendizagem e correctamente classificados pelo perceptrão e unicamada.

O número de vectores distintos nestes ficheiros é mais elevado que nos dados anteriores, tal como se verifica na tabela 5.21. Nestes casos as percentagens de concordância nos vectores distintos variam de 53% a 79% (valores também inferiores às concordância nos ficheiros completos), enquanto nos dados anteriores variavam de 72% a 93%. Também, o número de vectores coincidentes com os padrões de aprendizagem é mais baixo que nos casos anteriores. Neste casos, os padrões têm muito mais ruído que os dados anteriores. Em geral, os vectores dum estágio contêm, além da informação da actividade característica desse estágio, informação doutras actividades. Consequentemente, quando a classificação depende da presença predominante duma actividade, como é o caso dos estádios 2 e 3/4, esse ruído não tem grande influência. Mas quando não existe uma actividade dominante, como é o caso de 1/5, a concordância baixa consideravelmente. O caso t4 é um exemplo desse facto onde, apesar das coincidências com o padrão de aprendizagem serem nulas, a concordância em estádios 2 e 3/4 são elevadas.

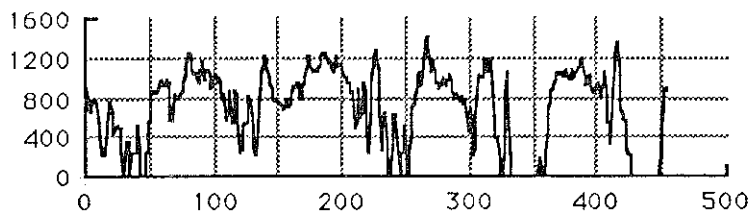
De acordo com o conhecimento sobre o sono, as actividades alfa, delta e sigma deviam aparecer em "cluster". Como referimos, na discussão sobre avaliação

dos detectores em alguns casos, isso não acontece (por exemplo, os fusos em estágio REM), o que implica que alguns eventos detectados serão falsos positivos. Podemos por isso pensar em utilizar as estatísticas para melhorar a classificação.

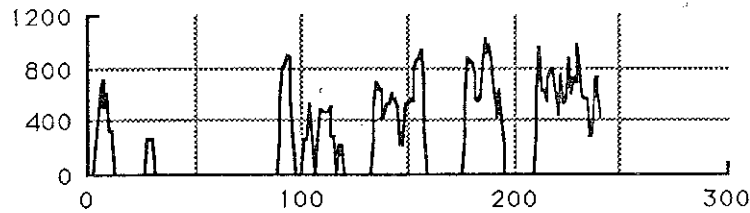
4.1: Evolução das actividades ao longo da noite

O sono tem características cíclicas, isto é, uma noite é um conjunto de intervalos caracterizados pela evolução de estádios de sono leves a sono profundo com um período de REM. Sendo cada um daqueles estádios caracterizado por uma das actividades referidas, a evolução destas actividades também deve ser cíclica. Por outro lado, o estudo estatístico das actividades detectadas revelou propriedades estatísticas diferentes em estádios diferentes, o que nos leva a pensar que, adaptando as definições dos parâmetros das actividades às estatísticas, podemos melhorar a concordância com a análise visual.

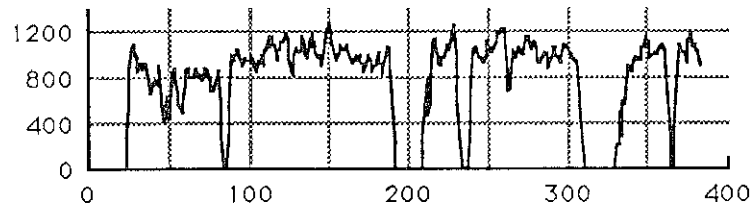
A actividade sigma, como é referido por Hasan [1983] e confirmado pelas estatísticas e evolução da duração média/ minuto ao longo da noite, revela uma variação com um máximo que deve ocorrer em estágio 2 (figura 5.13). Além disso é também visível nesses gráficos que, em alguns casos (t5, t7, t16) o sigma ocorre em "cluster", enquanto que nos restantes (t4, t8, t10) há sempre eventos detectados. É no último grupo que a concordância em 1/5 tem valores mais baixos (e muitas vezes classificados como 2), por isso podemos pensar em utilizar o conhecimento estatístico da duração para calcular um novo critério na validação dos eventos. Vamos redefinir o limiar de duração dos eventos como o valor médio menos metade do desvio padrão, isto é, um valor com igual diferença da média que o valor de pico. Estes valores são: 780, 658 e 656 respectivamente para t4, t8 e t10. Portanto, considerando como elementos válidos só aqueles cuja duração excede aquele valor mínimo. Deste modo, grande parte dos eventos detectados vão ser



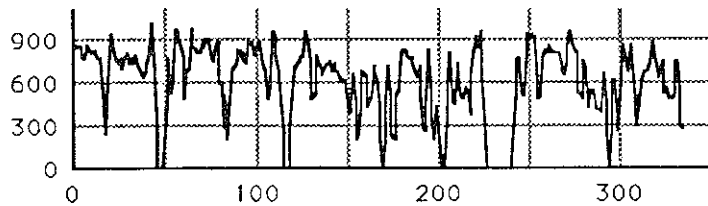
A: Duração média/minuto dos fusos de sono em t4. Duração média total 1005.9 ms e desvio padrão 453.5 ms.



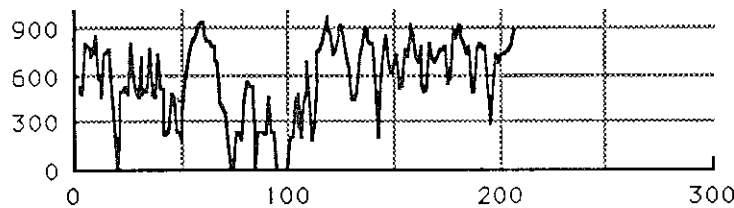
B: Duração média/minuto dos fusos de sono em t5. Duração média total 815.8ms e desvio padrão 288.6 ms.



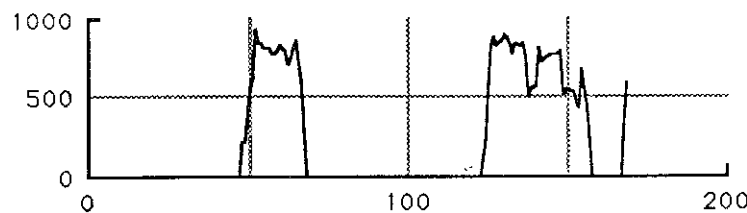
C: Duração média/minuto dos fusos de sono em t7. Duração média total 995.3ms e desvio padrão 362.2 ms.



D: Duração média/minuto dos fusos de sono em t8. Duração média total 790.7 ms e desvio padrão 265.6 ms.

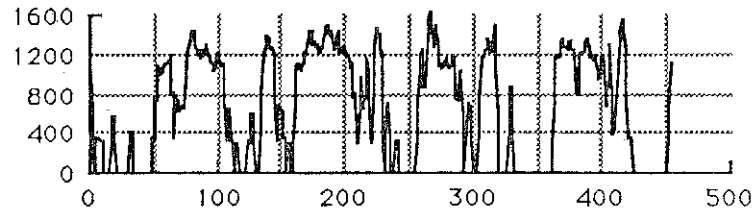


E: Duração média/minuto dos fusos de sono em t10. Duração média total 768.7 ms e desvio padrão 224.4 ms.

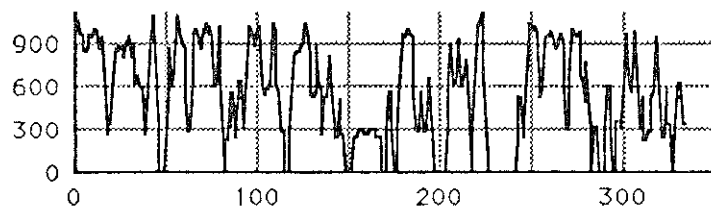


F: Duração média/minuto dos fusos de sono em t16. Duração média total 816.4 ms e desvio padrão 230.6 ms.

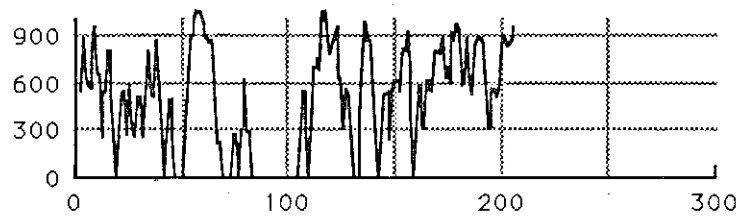
figura 5.13 Evolução da duração média dos fusos de sono (após aplicação dum filtro de mediana de 3 pontos) de t4, t5, t7, t8, t10, t16.



A: Duração média/minuto dos fusos de sono em t4 após validação (fusos com duração superior a 780 ms)



B: Duração média/minuto dos fusos de sono em t8 após validação (fusos com duração superior a 658 ms)



C: Duração média/minuto dos fusos de sono em t10 após validação (fusos com duração superior a 656 ms)

figura 5.14: Evolução da duração média dos fusos em t4, t8, t10 após validação intermédia (depois da aplicação dum filtro de mediana)

eliminados, pois como referimos, os histogramas revelam que uma parte significativa dos eventos tem valores de duração daquela ordem de grandeza. De facto, como prova a figura 5.14, a evolução da duração média/minuto, após o processo de validação, não aumentou significativamente o valor dos picos mas aumentou os períodos em que a duração média é nula, o que melhorou o "cluster".

Nos casos de t5 e t16, em que muitos estádios 2 são classificados como estádio 3/4, os histogramas da amplitude destas actividades (figuras 5.7 - 5.12) revelaram grande quantidade de delta com valores de amplitude muito elevados, mesmo com os valores máximos permitidos pela gama dinâmica do processamento (possivelmente, saturação), ao mesmo tempo que existem muitos valores de amplitude baixa. Por isso, como a média dos eventos é aproximadamente igual ao pico de amplitude menos o desvio padrão, escolhemos para segundo critério de validação o valor dado pela média menos o desvio padrão. Esses valores são 29 μv e 39 μv para t16 e t15, respectivamente.

Com os novos dados efectuamos nova classificação com a rede multicamada (3+6+4) após a aprendizagem do padrão IV. Estes resultados estão representados na tabela seguinte.

Estádios	t4	t5	t8	t10	t16
1/5	36.2	60.0	43.0	91.3	98.4
2	93.3	74.2	89.7	86.1	32.6
3/4	95.0	98.9	95.1	89.7	100.0
total	78.7	79.7	70.5	92.0	78.7

tabela 5.22: Classificação após validação intermédia

Se compararmos estes resultados com os anteriores obtidos pela mesma rede, vemos que há uma melhoria significativa do grau de concordância total e dos estádio 1/5 nos três casos (t4,t8,t10), e uma ligeira melhoria no estádio 3/4,.

embora essas melhorias sejam feitas à custa de uma diminuição na concordância de estágio 2. A percentagem total de concordância aumenta também entre 5% e 15%.

CAP 6: CONCLUSÕES

As metodologias de análise do EEG de sono têm, como é natural, sido influenciadas pelo desenvolvimento tecnológico dos computadores. Sobretudo têm sido particularmente sensíveis ao binómio velocidade de cálculo/capacidade de armazenamento de informação. Assim, os sistemas testados em maior número de casos foram os sistemas híbridos por fazerem processamento analógico, em tempo real, e terem como filosofia a redução da quantidade de informação a armazenar. O aparecimento de sistemas digitais baseados na filosofia dos híbridos (nomeadamente o HIDRA) foi uma evolução natural que se traduziu em maior flexibilidade e reprodutibilidade na definição do processo de detecção. Mas os sistemas híbridos foram unicamente direccionados para a classificação de estádios de sono, limitando, portanto, a recolha de informação à presença/ausência dos eventos de EEG. Nesta tese propusemos uma metodologia mais completa que engloba também a microanálise, isto é, a descrição das actividades detectadas com um conjunto de parâmetros. Assim, as várias análises de sono podem ser encaradas num sistema integrado pois, no final da análise, além do hipnograma, o utilizador dispõe naturalmente dum conjunto de informação fundamental para alguns estudos nesta área, tais como influência de drogas ou idade no sono. Portanto, com o HIDRA, a filosofia de detecção não precisa ser modificada ao contrário dos sistemas anteriores em que estes estudos eram feitos separadamente da classificação em estádios.

O modelo de processamento escolhido tem uma estrutura hierárquica em que na primeira camada se identificam as actividades básicas e na segunda se faz a decisão sobre o estágio de sono a atribuir a cada época. A identificação das actividades básicas

definidas como um padrão, que se destaca da actividade de fundo, constituído por ondas quase-sinusoidais (isoladas ou em grupo) que individualmente se caracterizam por medidas temporais, é feita em tempo real por algoritmos de reconhecimento de padrões. As medidas temporais, por exemplo período e amplitude, são os parâmetros do processo de decisão padrão válido/não-válido de uma actividade. Além disso, como tem sido referido, as actividades do EEG não são generalizadas, isto é, não são igualmente visíveis em todos os eléctrodos, algumas estão melhor definidas numa área (grupo de eléctrodos), enquanto outras se destacam noutra área, por conseguinte pretendeu-se utilizar mais que um canal de EEG. A nossa aproximação consistiu no "mapping" desta estratégia do processamento sinal num sistema de multiprocessamento, atribuindo a cada preprocessor, cujos sinais de entrada provêm de eléctrodos distintos, a detecção das actividades mais visíveis no respectivo sinal de entrada; e os parâmetros característicos de cada actividade são enviados à memória central comum ("mailbox"), de onde podem ser lidos e armazenados em ficheiro pelo processador central [Guedes de Oliveira- 1987].

O modelo de detecção pressupõe uma operação de filtragem, na banda da actividade em estudo, que tem por objectivo facilitar as medições temporais em que se baseia o processo de detecção padrão válido/não-válido. Em cada preprocessor são detectadas pelo menos duas actividades (caracterizadas por gamas de frequência distintas), assim a filtragem (passa-baixo) da actividade de frequência mais elevada pode também ser filtragem "anti-aliasing" da outra actividade. Esta estratégia permite baixar a frequência de amostragem na análise das actividades da gama de frequência inferior, sem sobrecarga no tempo de processamento. A atribuição de tempo de processamento às tarefas de detecção (os detectores) é implementada com uma estrutura de "multitasking" fixo controlado por ordem de chegada do "interrupt" externo com

frequência de 1KHz, a partir do qual se obtêm simultaneamente as frequências de amostragem para as actividades (por decimação) [Príncipe et al - 1985]. Esta metodologia, além de permitir distribuir automaticamente "slots" de 1ms por cada um dos detectores e controlar a respectiva frequência de amostragem, garante "slots" livres para outras tarefas. No HIDRA uma das tarefas também obrigatória é a comunicação, isto é, o envio de dados aos processador central; outra tarefa é a estatística, cálculo de médias (e variâncias) dos parâmetros das actividades definidas em "burst". Qualquer destas actividades tem caracter assíncrono, isto é, só há estatísticas após a detecção de um evento, ou comunicação quando há dados ou o "bus" está livre. No entanto, tratando-se de processamento em tempo real, todas as tarefas têm que estar sincronizadas com o "interrupt" dispondo, portanto, de 1ms para o respectivo processamento. Nos detectores, os cálculos dividem-se basicamente em duas componentes: o processamento da amostra e a decisão. A primeira está intimamente ligada com o processamento aritmético da operação de filtragem que não é eficientemente implementado num microprocessador comum. Consequentemente, aplicaram-se técnicas que permitem minimizar o processamento aritmético: operações em vírgula fixa e estruturas com poucas multiplicações. Estas opções acarretam, por um lado, a escolha de estruturas de filtragem com coeficientes que permitem a substituição das operações de multiplicação por um pequeno número de "shifts" e adições. Por outro lado, com as operações efectadas em vírgula fixa, são necessários estudos adicionais para a implementação duma estratégia de "scaling" que permita o uso eficiente da gama dinâmica do filtro (o comprimento de palavra). De qualquer modo, num microprocessador comum as estruturas referidas, com operações em vírgula fixa, são muito mais rápidas e não comprometem necessariamente a qualidade da análise. Estes filtros, para ordens baixas, não têm atenuações elevadas, mas neste

modelo não são elementos isolados na cadeia de decisão, além disso têm fase linear o que é fundamental na detecção de transientes. Os cálculos estatísticos envolvem operações de multiplicação e divisão, ultrapassando por isso o tempo de 1ms, mas este processamento só precisa ser concluído quando fôr reconhecido o início de novo padrão. Deste modo estes cálculos são repartidos pelas "slots" livres ou no tempo que resta do processamento nos detectores. Esta metodologia permite otimizar a capacidade de cálculo efectuado nos preprocessadores, embora os cálculos estatísticos sejam frequentemente intercalados com a leitura da "flag" de "interrupt" (todos os "loops" da divisão). Note-se, no entanto, que mesmo com a frequência de amostragem típica (200Hz) estes cálculos não poderiam ser efectuados entre dois "interrupts" consecutivos. Desta metodologia com multifrequência de amostragem, resulta um "kernel" de controlo de tarefas pouco pesado (com um número reduzido de testes), o que aliado à optimização do processamento aritmético e à implementação dos blocos de decisão evitando testes a "flags" de controlo, permite a utilização de um microprocessador comum a 1MHz.

A metodologia referida pressupõe o envio de dados do primeiro nível para o segundo nível de processamento. No HIDRA os níveis de processamento são fisicamente distintos, estabelecendo-se a comunicação através de uma fatia da memória central partilhada por ambos, cujo protocolo de acesso é facilitado pela estrutura de interrupts da Motorola e um código de semáforos [Guedas de Oliveira et al - 1987]. As comunicações entre os dois níveis de processamento (preprocessadores e o processador central) têm lugar em duas fases distintas do processo: "set-up" e processamento em tempo real. A primeira contribui para a configuração do sistema: instalação de programas e parâmetros constantes nos preprocessadores, e início da execução dos referidos programas. A segunda consta do envio dos parâmetros dos

eventos detectados durante o processamento em tempo real. O "software" desenvolvido para os preprocessadores consiste basicamente na leitura e escrita da memória central comum ao processador central e ao respectivo preprocessador. O desencadear destas operações pode ser local (no preprocessador) com a utilização de dois comandos escritos para o efeito ou, no caso da leitura, pode ser por activação do "fast interrupt" (FIRQ) pelo processador central. No processador central são desenvolvidos um conjunto de programas dos quais destacamos: transferência de informação binária para o preprocessador, isto é, escrita na memória central seguida da activação do FIRQ desencadeando operações de leitura nos preprocessadores; programação do canal de leitura da ADC para determinado preprocessador; execução de programas [Martins et al 1990]. Estes comandos são, no entanto, bastante herméticos para um utilizador comum, tendo sido desenvolvida uma estrutura que configura o sistema para esta aplicação. No sistema com quatro preprocessadores, cada um recebendo sinal dum canal diferente da ADC (supondo esta ligada às derivações frontal, central e occipital do EEG e EOG), são primeiro carregados os programas de calibração, procedendo-se à leitura dum sinal sinusoidal com amplitude de $50\mu\text{v}$; a partir deste valor são calculados os critérios de amplitude para cada evento que conjuntamente com os critérios de frequência (período) é gerada uma tabela com quatro linhas e um número variável de colunas; em seguida os programas e as linhas da tabela são colocadas na memória central correspondente a cada preprocessador, procedendo-se com o desencadear duma operação de leitura para cada preprocessador. Logo que todos os preprocessadores tenham terminado a leitura, pode desencadear-se o início simultâneo de todos os detectores. Finalmente, e se o utilizador o desejar, pode dar início à operação de escrita em disco dos parâmetros dos eventos detectados. Naturalmente que este "interface" não tira qualquer partido da versatilidade do "hardware" do sistema, permite unicamente

facilitar a utilização do sistema neste grupo de testes. Por isso, o "interface" precisa ser melhorado de modo a poder ser configurado por um utilizador comum, quer na escolha de programas quer de parâmetros de detecção

O sistema está em fase de instalação no Hospital Geral de Santo António, com alguns problemas de fiabilidade. Neste momento é ainda necessário verificar se os programas instalados nos preprocessadores (os detectores) entram de facto em execução, pois algumas vezes na inicialização a comunicação parece não funcionar, outras vezes há mesmo bloqueio do "bus" por parte de um dos preprocessadores. Este facto, dificulta o manuseamento do sistema por um utilizador comum, impedindo a sua utilização sistemática e em rotina. Por isso, de momento podemos dizer que o sistema está em fase de teste. Em paralelo e, face às necessidades criadas no Laboratório de Neurofisiologia pela existência de vários equipamentos funcionando independentemente e tendo como objectivo análises dos sinais que podem ser consideradas complementares, pretende-se ligar, por linha série, o HIDRA a um PC [Cunha et al -1990]. Esta estratégia traz vantagens no processo de integração de informação dos vários sistemas e no desenvolvimento duma "interface" com as características já referidas pois, podem utilizar-se as ferramentas comerciais existentes para a manipulação de écrans e menus. Além disso, o "software" desenvolvido em C para o tratamento de dados do primeiro nível de processamento (a microanálise) pode ser uma facilidade do "interface" ficando deste modo disponível para uso clínico.

Em estudos de sono, o que em última análise permite avaliar a "performance" do sistema são os valores de concordância da classificação automática em estádios de sono com a classificação visual de um especialista. Neste estudo, utilizámos um novo método para a classificação automática, constituído por uma rede neuronal tendo por entradas as percentagens das actividades presentes em cada minuto e, por saída o

no espaço multidimensional da entrada, com a rede de uma camada "hidden" as classes são regiões convexas e, finalmente, com a duas camadas "hidden", as classes podem ter qualquer forma. Quanto ao conjunto de treino, temos a possibilidade de comparar várias estratégias: conhecimento *a priori*, isto é, com padrões criados pelo utilizador (16 ou 7 padrões); aprendizagem exaustiva, isto é, o treino com um ficheiro completo (aproximadamente 400 padrões); e aprendizagem aleatória, isto é, 16 padrões retirados ao acaso de cada ficheiro. Estas experiências permitiram-nos concluir que a primeira estratégia de treino é a mais eficiente para esta aplicação [Príncipe et al - 1989]. Aliás, com a aprendizagem em ficheiros completos obtêm-se resultados piores que os obtidos com o treino de padrões escolhidos, excepto quando o teste é feito no próprio ficheiro. Por isso, podemos concluir que a variabilidade inter-indivíduo se reflecte na aprendizagem da rede formando um conjunto de atractores diferente no espaço multidimensional. Além disso, nem sempre número elevado de padrões se traduz em capacidade de generalização elevada. O valor de concordância, de 88%, foi obtido com a estratégia de treino referida, mas a rede com topologias diferentes e padrões de entrada com dimensões e codificação diferentes: perceptrão com vectores de entrada contendo a informação de cinco actividades codificadas em binário, e rede com uma camada "hidden" com vectores de entrada com a informação de três actividades com valores no intervalo [0,1]. No último caso, o perceptrão não aprendeu o conjunto de treino escolhido (com 7 padrões), por isso podemos dizer que, para esta estratégia de treino sempre que um perceptrão é capaz de codificar toda a informação contida nos padrões de treino, não há qualquer vantagem em recorrer a outra topologia. Em qualquer das experiências, a gama de variação da percentagem de concordância é bastante alargada, isto é, a diferença entre o valor mínimo e máximo é de aproximadamente 14%. Este valor mantém-se quando a aprendizagem e teste são feitas

no mesmo ficheiro qualquer que seja a codificação dos dados de entrada ou o número de actividades na entrada da rede neuronal. Por outro lado, a diferença entre as percentagens de concordância obtidas por uma rede unicamada, aprendizagem e testes nos ficheiros completos, com 5 actividades na entrada e 3 actividades não ultrapassa 6%; e com o perceptrão essa diferença é ainda menor mas nem sempre favorável à rede com 5 actividades na entrada. Deste modo, podemos concluir que na realidade as regras de R+K são um conjunto incompleto e, sempre que não há actividades dominantes, o especialista comporta-se subjectivamente. Por esta razão é difícil escrever um conjunto de vectores com informação de 5 actividades que melhorem os resultados da classificação em todos os casos.

Com os dados do HIDRA a percentagem de concordância média é mais baixa, aproximadamente 79.3%. Nestes dados o número de vectores distintos em cada ficheiro é superior aos existentes nos ficheiros anteriores, também o número de coincidências com o padrão de treino é inferior. De facto, estes padrões contêm mais ruído que os anteriores, isto é, os padrões de entrada da rede além da actividade dominante do minuto contêm informação das outras actividades. Somos, por isso, levados a pensar que o processo de detecção no HIDRA funciona pior que no "front-end" dos dados da Universidade da Florida. Em contrapartida, o estudo da evolução das actividades ao longo da noite veio corroborar resultados anteriores: ciclicidade das actividades, maior percentagem de delta nos dois primeiros ciclos de sono. Por isso, pensamos que a razão não está na metodologia de detecção, que é basicamente a mesma nos dois sistemas, mas no processo de recolha de dados. Com os dados do HIDRA houve alguns problemas, nomeadamente de saturação do sinal de entrada, enquanto que os outros dados percentecem a um conjunto de dados normativos cujo processo de recolha foi extremamente cuidado de modo a obter um sinal com boa qualidade. E

qualquer processo de detecção automática funciona mal quando o sinal de entrada tem má qualidade, sobretudo neste modelo hierárquico de processamento em que pequenos erros no primeiro nível de processamento se reflectem fortemente no segundo nível. Esta é a maior desvantagem de um modelo hierárquico, para que a classificação tenha bons resultados é necessário que o primeiro nível funcione a 100%. Embora no HIDRA se possa ter *a posteriori* algum poder crítico das detecções, recorrendo às estatísticas, o estudo estatístico dos parâmetros das actividades detectadas confirmam quantitativamente as diferenças inter-indivíduos e inter-estádios (classificação visual) nos valores dos mesmos parâmetros. Deste modo, foi possível calcular novos limiares de decisão (para um dos parâmetros das actividades) adaptados a cada um indivíduo e assim melhorar a classificação global em 6%. Apesar do método utilizado nesse cálculo ser completamente empírico, somos levados a pensar que este assunto não foi ainda esgotado, e que neste campo se abrem perspectivas para novos estudos. Mas de facto, um dos pontos fracos desta filosofia hierárquica é a falta de interacção entre os dois níveis de processamento, de modo a que o processo de detecção ficasse menos dependente da qualidade dos sinais de entrada. O que, dada a imprevisibilidade dos sinais biológicos, é difícil de implementar, em tempo real, uma abordagem diferente, isto é, que perante uma hipótese global (neste caso, um estágio de sono) se procure nos sinais os eventos que a definem a menos que se procurem outras características nos sinais. Apesar de tudo, a metodologia utilizada na detecção permite tirar maior partido dos sistemas automáticos pois, além de extrair dos sinais informação tendo como único objectivo poupar trabalho e tempo a um especialista, também permite obter características dos eventos, impossíveis de obter visualmente, aumentando o conhecimento do especialista e, conseqüentemente, abrir novas perspectivas sobre este fenómeno fisiológico.

O sistema foi testado numa gama restrita de indivíduos normais e, além disso nunca foi abordado o problema da separação dos estádios 3/4 e 1/5. A separação dos estádios 3/4 depende unicamente dum limiar na percentagem de delta presente no minuto. A separação de 1/5 depende da presença de REM no minuto, e não só porque as regras de R+K referem que podem haver épocas que devem ser classificadas como REM desde que sejam contíguas as épocas em que esta actividade está presente. Portanto, à saída da rede neuronal pode ser aplicado um algoritmo que traduza a informação contextual necessária à separação de 1/5 e a definição de um limiar para a separação de 3/4. Esta experiência foi feita nos dados da Universidade da Florida, baixando a percentagem média em aproximadamente 2%. Nos dados do HIDRA esta separação nunca foi tentada dado o mau funcionamento do detector REM. Aliás o funcionamento deste detector precisa ser cuidadosamente estudado: mudar unicamente a filosofia de detecção ou estudar um processo de ter sinais diferentes (mudança de montagem dos eléctrodos ou nível do sinal na gravação). Isto porque neste sistema também não foi convenientemente explorada a possibilidade de relacionar a informação inter-canal, por exemplo fazer um estudo de REM relacionando a informação das ondas lentas em todos os canais.

Os sistemas automáticos tiveram, desde o início da investigação nesta área, um papel muito importante. No entanto, qualquer dos sistemas descritos (o HIDRA incluído) foram sempre limitados ao âmbito da investigação, isto é, nenhum deles foi incluído em rotina clínica. Isto porque estes sistemas nunca tiveram um "interface" fácil com o utilizador. Neste contexto "interface" significa comparação rápida e fácil duma decisão da máquina em estádio de sono com o sinal original com o qual o especialista está habituado a lidar. E de facto, é esta a metodologia que os sistemas comerciais recentemente disponíveis exploram (por exemplo, Sleepscan da New Bio-Logic, o

sistema da CNS, etc). Naturalmente que a maioria dos sistemas surgiram no mercado com a divulgação do disco óptico (grande capacidade de armazenamento em sistemas de pequeno porte) e/ou com o écran de alta resolução (visualização do sinal dispensando o registo em papel). Nestes sistemas o utilizador pode obter um hipnograma automaticamente e, em paralelo, relacionar essa informação com o sinal original. Em estudos de apneias o hipnograma e o tratamento de variáveis respiratórias podem ser suficientes mas noutros estudos (drogas, por exemplo) estes resultados podem ser bastante limitativos. Em alguns estudos nesta área, a ocorrência e persistência das actividades são também importantes. Por exemplo, "bursts" de teta ou paroxismos delta são mais frequentes em doentes depressivos que na população normal; ou as possíveis relações das actividades rítmicas com o tipo de epilepsia, nomeadamente as relações temporais dos padrões epiletiformes (ponta/onda) com alguns ritmos de sono (os fusos), com ou sem influência de drogas. Estas perguntas só podem ser formuladas em sistemas que forneçam também informação sobre as actividades rítmicas. Resumindo, a ocorrência e as relações temporais entre as várias actividades, assim como os respectivos parâmetros podem estabelecer diferenças entre a população normal e o sono patológico, contribuindo simultaneamente para o melhor entendimento e descrição deste fenómeno fisiológico.

Actualmente qualquer sistema de análise de sono deve ser alargado às outras variáveis fisiológicas como é o caso das variáveis respiratórias, além da possibilidade de alargar a análise a padrões de EEG anormais (ponta/onda). Uma das características mais importante do HIDRA é a facilidade de configuração: por um lado o número de processadores periféricos instalados (num máximo de 16), por outro lado a possibilidade de escolher os algoritmos a instalar em cada um deles. Estes factos reforçam a ideia, anteriormente manifestada, de que é necessário desenvolver

"software" que, a este nível, permita interactuar com o utilizador não-especialista (isto é, que não conheça o "hardware" do sistema) de modo a tirar maior partido do sistema. O HIDRA pode ser um instrumento muito útil e versátil em ambientes onde os algoritmos de processamento são caracterizados por baixa percentagem de processamento aritmético e decisões simples, portanto implementáveis em tempo real, como é o caso do reconhecimento de padrões em alguns sinais biológicos (EEG e ECG, por exemplo). Actualmente o armazenamento não é um problema tão crítico, mas de qualquer modo, sobretudo quando se trata da análise de várias horas de sinal, são necessários métodos que retirem do sinal informação facilmente relacionável com o conhecimento do especialista sobre os sinais. No entanto, deve ser também explorada a versatilidade em termos de "interface" que os sistemas comerciais oferecem. Aliás, a possibilidade de poder repetir um teste sem haver nova classificação visual foi um dos problemas sentidos durante o desenvolvimento deste sistema. Este problema é bastante importante na fase de teste, apesar dos algoritmos poderem ser testados individualmente, pois só deste modo e perante a filosofia actual (utilização das regras de R+K), se pode constituir uma base de dados. Esta base de dados pode ser utilizada tanto na aferição de métodos de detecção como na troca de informação inter-laboratório. Mas este problema só pode ser resolvido se houver algum processo de sincronização com o sinal original, o que é impossível de conseguir com a gravação analógica dos sinais.

REFERÊNCIAS

AHMAD S, GERALD T. 1989. "Scaling and Generalization in Neural Networks: a case of study". Advances in Neural Information Processing Systems 1-Eds. David S. Touretky.

AGNEW H.W. Jr. 1973. "Integrator analysis of the sleep electroencephalogram". Electroenceph. Clin. Neurophysiol. 34:391-397.

AGNEW H.W.Jr, PARKER J.C.,WEBB W. B., WILLIAMS R.L. 1967. "Amplitude measurements of sleep electroencephalogram". Electroenceph. Clin. Neurophysiol 22: 84-86.

ANDERSON JAMES A. 1983. "Cognitive and Psychological Computation with Neural Models". IEEE Transactions on Systems, Man and Cybernetics. SMC-13 (5): 79-815.

BAAS, BOURNE J.R. 1984. "A Rule-Based Microcomputer System for Electroencephalogram Evaluation". IEEE Transactions on Biomedical Engineering BME 31-10:660-664.

BAUME., HAUSSLER 1989. "What Size Gives Valid Generalization?". Advances in Neural Information Processing Systems 1-Eds. David S. Touretky.

BAUM 1990. "When are the k-Nearest Neighbor and Back-Propagation Accurate for Feasible Sized Sets of Examples?". Neural Networks Proceedings Eurasip Sesimbra 1990.

BICKFORD, BRIMM, BERGER, AUNG. 1973. "Application of Compressed Spectral Array in Clinical EEG". Automation and Clinical Electroencephalography, Raven Press. New-York.

BREMER G., SMITH J.R.,KARACAN 1970. "Automatic Detection of K-Complex in Sleep Electroencephalograms". IEEE Transactions on Biomedical Engineering.BME-17 (4): 314-323.

BURR D.J. 1986. "A Neural Network Digit Recognizer" Proceedings International Conference on Systems,Man and Cybernetics, Atlanta-1986.

CHAN D., RABINER L.R. 1973. "Analysis of Quantization Errors in the Direct Form or Finite Impulse Response Digital Filters". IEEE Transactions on Audio and Electroacoustics AU-21 4:354-366

CHAN D., RABINER L.R. 1973. "Theory of Roundoff Noise in Cascade Realizations of Finite Impulse Response Digital Filters". The Bell System Technical Journal vol.51 3:329-345.

CHAN D., RABINER L.R. 1973. "An Algorithm for Minimizing Roundoff Noise in Cascade Realizations of Finite Impulse Response Digital Filters". The Bell System Technical Journal vol.51 3:347-385

CHANG, SMITH, PRINCIPE 1988. "A Layered Processing Model for Knowledge-Based Contextual Interpretation of Multichannel EEG/EOG Signals". Proc. of Instrumentation Society of America. Colorado 1988.

CHURCH, MARCH, HIBI et al 1975. "Changes in Frequency and Amplitude of Delta Activity During Sleep". Electroenc. and Clinical Neuroph.39: 1-7.

CROCHIERE E., RABINER L. 1981. "Interpolation and Decimation of Digital Signals- A Tutorial Review". Proceedings of the IEEE. Vol.69 (3):300-331.

CROCHIERE E., RABINER L. 1983 Multirate Digital Signal Processing. Prentice-Hall, Inc., New-Jersey, 1983.

CRUZ A.N. 1990. Multiprocessador e Processamento em Tempo Real. Tese-Departamento de Electrónica e Telecomunicações- Universidade de Aveiro.

CUNHA J. P., MARTINS DA SILVA, VAZ F., GUEDES DE OLOVEIRA, CUNHA M.B., TOMÉ A.M., CRUZ A., PRINCIPE. 1990. "LINE: Laboratório Integrado de Neurofisiologia". Proceedings do 2º Congresso de Engenharia Biomédica. Aveiro 1990

DAVIES 1979. "Trade-Offs in Fixed-Point Multiplication Algorithms for Microcomputers". Computers and Digital Techniques 2-3:105-112.

DENKER JOHN S. 1986. "Neural Network Models of Learning and Adaptation" Physica 22d: 216-222.

DUDA AND HART 1973. Pattern Classification And Scene Anaysis. John Wiley & Sons-1973

FEINBERG 1974. "Changes in Sleep Cycle Patterns with Age". J. Psychiat. Res. vol 10: 283-306.

GAILLARD J.M., TISSOT R. 1972. "Analyse Automatique du Sommeil par un Systeme Hybride: Nouveaux Resultats". *Electroencephalography and Clinical Neurophysiology* 33 : 403-410.

GAILLARD J.M., TISSOT R. 1973. "Principles of Automatic Analysis of Sleep Records with an Hybrid System". *Computer and Biomedical Research* 6: 1-13.

GAILLARD J.M., AUBERT C. 1975. "Specificity of Benzodiazepine Action on Human Sleep Confirmed. Another Contribution of Automatic Analysis of Polygraph Recordings". *Biological Psychiatry* 10-2: 185-197

GALA SUNIT, 1986. Automaton for Sleep Stage Scoring Based on the Theory of Evidence. Tese - Universidade da Florida- 1986

GONDECK A.R., SMITH J. R. 1974. "Dynamics of Human Sigma Spindles". *Electroenc. and Clinical Neuroph.* 37:293-297.

GORMAN P., SEJNOWSKI. 1988. "Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets". *Neural Networks* 1: 75-90.

GUEDES DE OLIVEIRA P., PRINCIPE J.C., CRUZ A.N., 1985. "Automated Event Detection and Characterization In EEG Monitoring- Part I: Hardware Support". In: *Biorhythms and Epilepsy* . eds A. Martins da Silva et al. Raven Press- New York 1985..

GUEDES DE OLIVEIRA P., PRINCIPE J.C., CRUZ A.N., TOMÉ ANA.M. 1987. "HIDRA: a Hierarchical Instrument for Distributed Real-Time Analysis of Biological Signals". *IEEE Transactions on Biomedical Engineering* BME-34 12:921-927.

GUEDES DE OLIVEIRA. 1981. Reconhecimento Automático de Ondas Características no EEG de Doentes Epilépticos: Realização baseada num microcomputador. Tese- Universidade de Aveiro.

HASAN J. 1983. Differentiation of Normal and Disturbed Sleep by Automatic Analysis. *Acta Physiologica Scandinavica. Supplementum* 526.

HASAN J., VAAHTORANTA, SAARIKIVI et al. 1982, "Inter-Individual and Inter-Night Variability Study of EEG Waveforms Studied with a Hybrid Sleep Analyser". Sleep 1982.6th Eur. Congr. Sleep Res. Zurich.

HINTON GEOFFREY 1985. "Learning in Parallel Networks". Byte (April): 265-273.

HINTON, SEJNOWSKI 1986. "Learning and Relearning in Boltzman Machines". In: Parallel Distributed Processing: Explorations in the Microstruture of Cognition-vol.1: Foundations. Eds: Rumehart, McClelland and the PDP Group. Cambridge MIT Press.

INOUE K., KUMAMARU K., SAGARA S., MATSUOKA S.1982." Pattern Recognition Approach to Human Sleep EEG Analysis and Determination of Sleep Stages" Memoirs of the Faculty of Engineering, Kyushu University, 42-3:177-195.

JAGANNATHAN V., BOURN J.R., JANSE B.H., WAR J.W. 1982. "Artificial Intelligence Methods in Quantitative Electroencephalogram Analysis". Computer Programs in Biomedicine 15: 249-258.

JACKSON LELAND B. 1986. Digital Filters and Signal Processing. Kluver Academic Publishers. Norwell.

JANSEN B.H., HASMAN A., LENTEN R. 1981. "Piece-Wise EEG Analysis: An Objective Evaluation". Int. J. Bio-Medical Computing 12: 17-27.

JANSEN B.H, DAWANT B. M. 1987. "Knowledge-Based Approach to Sleep EEG Analysis- An Application to Sleep EEG Analysis". IEEE/ Ninth Annual Conference of the Engineering in Medicine and Biology Society.

JANSEN B.H, DAWANT B. M. 1989. "Knowledge-Based Approach to Sleep EEG Analysis- A Feasibility Study". IEEE Transactions on Biomedical Engineering. BME-36 (5): 510-518.

JOHNSON L.C. 1977. "The EEG During Sleep as Viewed by a Computer. EEG Informatics". A Didactic Review of Methods and Applications of EEG. (ed. A. Remond). Elsevier /North Holland Biomedical Press.

JOHNSON, SEALES, NAITOH, CHURCH. 1979. "The effects of Flurazepam Hydrochloride on Brain Electrical Activity during Sleep". Elentroenc. and Clinical Neuroph. 47: 309-321.

- JOSIN G. 1988. "Neural-Space Generalization of a Topological Transformation". *Biological Cybernetics* 59: 283-290.
- KEMP B, JASPERS, FRAZEN, JANSSEN.1985 "An Optimal Monitor of the Electroencephalographic Sigma Sleep State". *Biological Cybernetics* 51: 263-270.
- KEMP B.1986 "An Optimal Monitor of the Rapid-Eye-Movement Brain State". *Biological Cybernetics* 54: 133-139.
- KEMP B, KAMPHUISSEN.1986 "Simulation of Human Hyponograms Using Markov Chain Model". *Sleep* 9: 405-414..
- KEMP B.1987: A Model-Based Monitor Of Human Sleep Stages. Thesis.University of Twente.
- KTONAS 1984 . "Automated Analysis of Abnormal Electroencephalograms". *CRC Critical Reviews in Bioeng.* 9: 39-97.
- KUMAR A. 1977. "A Real-Time System for Pattern Recognition of Human Sleep Stages by Fuzzy Sets Analysis". *Pattern Recognition* 9: 43-46.
- LECUN, BOSER, DENKER, HENDERSON, HOWARD, HUBBARD, JACKEL. 1989. "Backpropagation Applied to Handwritten Zip Code Recognition". *Neural Computation* 1: 541-551.
- LEUNG, ZUE 1988. "Some Phonetic Recognition Experiments Using Artificial Neural Nets". *Proceeding IEEE Conference*.
- LEVENTHAL LANCE 1981. 6809 *Assembly Language Programming* Osborne-McGraw-Hill 1981
- LIPPMANN R.P.1987. "An Introduction to Computer With Neural Nets". *IEEE ASSP Magazine*.
- LIPPMANN R.P.1988. "Neural Nets for Computing" *Proceedings of IEEE Conference*.
- LIPPMANN R.P.1989. "Review of Neural Networks for speech Recognition". *Neural Computation* 1: 1-38.

LIARD, BURGER 1979 "System For Studying The Delta Rhythm During Sleep And Its Topographical Amplitude Distribution". *Electroencephalography and Clinical Neurophysiology* 47: 115-118.

MARTIN W.B., JOHNSON L.C., VIGLIONE S.S., NAITOH P., JOSEPH R.D., MOSES J.D. 1972. "Pattern Recognition of EEG-EOG as a Technique for All-Night Sleep Stage Scoring". *Electroenceph. Clin. Neurophysiol.* 32:471-427.

MARTINERIE J., JOSEPH J.P., NAILLON M., 1980. "Computerized Detection of Rapid Eye Movements During Paradoxical Sleep". *Int. J. Bio-Medical Computing* (11): 163-171.

MARTINS E., TOME ANA, CRUZ A., GUEDES DE OLIVEIRA, JOSE PRINCIPE. 1990 "Managing the Automatic EEG Signal on a Multiprocessor System" .*Proceedings 2º Congresso de Português de Engenharia Biomédica - Aveiro.*

McCLELLAND J. , RUMELHART D. 1988. *Explorations in Parallel distributed Processing. A handbook of models, programs and exercises.* MIT Press-Cambridge.

MINARD J.G. , KRAUSMAN D. 1971. "Rapid Eye Movement Defenition and Count : an on line detector". *Electroenceph. Clin. Neurophysiol* 31: 99-102.

OPPENHEIM A.V., SCHAFER.1975. *Digital Signal Processing.* Prentice-Hall, Inc., New-Jersey, 1975.

PRINCIPE J.C. et al 1979. "Microcomputer-Based Digital Filters for EEG Processing". *IEEE Transactions on Acoustic, Speech, and Signal processing ASSP-27* 6: 697-705

PRINCIPE. 1979b. *Automatic Detection and Quantification of Petit Mal Seizures in the Electroencephalogram.* Thesis.University of Florida. Gainesville.

PRINCIPE J.C., SMITH J.R. 1982. "Sleep Spindle Characteristic as a Function of Age". *Sleep* 5(1): 73-84.

PRINCIPE J.C.,GUEDES DE OLIVEIRA P. F. VAZ, ANA TOME. 1985. *Automated "Event Detection and Characterization In EEG Monitoring- Part II: Signal Processing"* In: *Biorhythms and Epilepsy* . eds A. Martins da Silva et al. Raven Press- New York 1985.

PRINCIPE J.C., SMITH J.R. 1986. "Design and Implementation of Linear Phase FIR Filters for Biological Signal Processing". IEEE Transactions on Biomedical Engineering BME-33 6:550-559

PRINCIPE J.C., SMITH J.R. 1986. "SAMICOS- A Sleep Analyzing Microcomputer System". IEEE Transactions Biomedical Engineering BME-33 10:935-941.

PRINCIPE J.C., GALA K.S., SMITH 1987. "Sleep Staging Automaton Based on the Theory of Evidence". IEEE/ Ninth Annual Conference of the Engineering in Medicine and Biology Society.

PRINCIPE J.C., GALA K.S., CHANG T.G. 1988. "Sleep Staging Automaton Based on the Theory of Evidence". IEEE Transactions on Biomedical Engineering. BME-36 (5):503-509.

PRINCIPE J.C., ANA MARIA TOME. 1989. "Performance and Training Strategies in Feedforward Neural Networks: An Application to Sleep Scoring". IJCCN, Washington 1989.

PRINCIPE J.C., GALA K.S., CHANG, T.G., ANA MARIA TOME 1989. "Information Processing Models for Automatic Sleep Scoring". Proceedings IEEE BME, Seattle 1989.

RABINER, GOLD. 1975. Theory and Application of Digital Signal Processing. Prentice-Hall, Inc, New-Jersey, 1975.

RAY S.R., LEE W.D., MORGAN C.D., AIRTH-KINGREE W. 1986. "Computer Sleep Stage Scoring- An Expert Approach". Intern. Journal Bio-Medical Computing 19: 43-61.

ROSA A.C. 1989. Detecção Optima de Complexos K com Base num Modelo Estocático do EEG de Sono. Tese- Instituto Superior Técnico -Lisboa

RUMELHART, HINTON, WILLIAMS 1986. "Learning Internal Representations by Error Propagation". In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition- vol.1: Foundations. Eds: Rumelhart, McClelland and the PDP Group. Cambridge MIT Press.

RUMELHART, HINTON, WILLIAMS 1986. "Learning Internal Representations by Error Propagation". Nature 323 (9): 533-536.

- SALINSKY, GOINS, SUTULA, ROSCOE, WEBER 1988. "Comparison of Sleep Staging by Polygraph and Color Density Array". *Sleep* 11(2):131-138.
- SHIRAKAWA S., SMITH J.R., AZUMI K. 1987. "Filter Transient Response to EEG Waveforms". *Int. J. Bio-Medical Computing* 20:87-95
- SILVA F., ALMEIDA L. 1990 "Acceleration Techniques for the Backpropagation Algorithm". *Neural Networks Proceedings Eurasip Sesimbra-1990*.
- SJENOWSKI, ROSENBERG 1986. "NETtalk a Parallel Network that Learns to Read Aloud". The Johns Hopkins University Electrical Engineering and Computer Science Technical Report. In: *Neurocomputing Foundations of Research*. eds Anderson and Rosenfeld. MIT Press. Cambridge.
- SMITH J.R. 1978. *Computers in Sleep Research*. CRC Critical Reviews in Bioeng. 3: 93-148.
- SMITH J.R., KARACAN I. 1971. "EEG Sleep Stage Scoring with an Hybrid System". *Electroenceph. Clin. Neurophysiol.* 31:231-237.
- SMITH J.R., KARACAN I. 1972. "Quantification of the Effect of a Hypnotic-like Drug on Slow-Wave Sleep". 1st Europ. Congr. Sleep Res. Basel, 504-508.
- SMITH J.R., FUNKE W.F., YEO W.C., AMBUEHL R.A. 1975. "Detection of Human Sleep EEG Waveforms". *Electroenceph. Clin. Neurophysiol.* 38:435-437.
- SMITH J.R., KARACAN I., KEANE B.P., YANG M. 1976. "Automated Sleep EEG Analysis Applied to the Evaluation of Drugs: Illustration by Study of Clorazepate Dipotassium". *Electroenceph. Clin. Neurophysiol.* 41: 587-594.
- SMITH J.R., KARACAN I., YANG M. 1977. "EEG Ontogeny Of Delta Activity During Human Sleep". *Electroenceph. Clin. Neurophysiol.* 43: 229-237.
- SMITH J.R., KARACAN I., YANG M. 1978. "Automated Analysis of the Human Sleep EEG". *Waking and Sleeping* 2: 75-82.
- SMITH J.R., KARACAN I., YANG M. 1979. "Automated Measurement of Alpha, Beta, Sigma and Theta Bursts Characteristics". *Sleep* 1: 435-443.
- SMITH J.R. 1980 "EEG Waveform Detection". *IEEE 1980 Frontiers of Engineering In Health Care*

SMITH J.R. 1985. "Ultradian Rhythm in Sleep Electroencephalogram" In: Biorhythms and Epilepsy . eds A. Martins da Silva et al. Raven Press- New York 1985.

SMITH J.R., PRINCIPE J. 1988. "Computer Analysis of Human Sleep". Enciclopédia: Medical Devices and Instrumentation, John Wiley and Sons.

SPEHLMAN: EEG Primer 1981, Elsevier / North-Holland Biomedical Press- Amsterdam.

STANUS E, LACROIX B., KERKHOFS M., MENDLEWICZ J. 1987. "Automated Sleep Scoring: a Comparative Reliability of two Algorithms". Electroencep.and Clinical Neurophys. 66: 448-456.

TAMURA, WAIBEL 1988. "Noise Reduction Using Connectionist Models". Proceedings IEEE Conference.

TAYLOR FRED J. 1983. Digital Filters Design Handbook. Marcel Dekker, Inc. New York.

VOGLT.P., MANGIS J.K., RIGLER A.K., ZINK, ALKON 1988. "Accelerating the Convergence of Back-Propagation Method". Byol. Cyben. 59: 257-263.

WAUQUIER, CLINKE, DECLERCK.1985. "Sleep Alterations by Seizures and Anticonvulsants". Biorhythms and Epilepsy . eds A. Martins da Silva et al. Raven Press- New York 1985.

WOLF, RODER-WANNER, BREDE et al. 1985. "Influences of Antiepileptic Drugs on Sleep". Biorhythms and Epilepsy . eds A. Martins da Silva et al. Raven Press- New York 1985.

WONG A. 1988 "Recognition of General Patterns Using Neural Networks". Biological Cybernetics. 58: 361-372.