

MODELO NEUROCOMPUTACIONAL DE ELEMENTOS CENTRALIZADORES COM
INTERAÇÕES LATERAIS APLICADO AO RECONHECIMENTO DE PADRÕES
TEMPORAIS

Cláudio Chauke Nehme

TESE SUBMETIDA AO CORPO DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

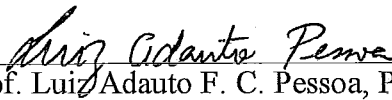
Aprovada por:



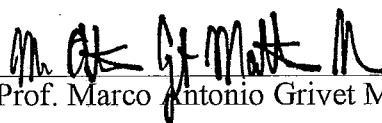
Prof. Luís Alfredo Vidal de Carvalho, D. Sc.
(Presidente)



Prof. Sueli Bandeira Teixeira Mendes, Ph. D.



Prof. Luiz Adauto F. C. Pessoa, Ph. D.



Prof. Marco Antonio Grivet M. Maia, Ph. D.



Pesq. Eutiquio Torres Calazans Junior, Ph. D.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 1996

NEHME, CLÁUDIO CHAUKE

Modelo Neurocomputacional de Elementos Centralizadores com Interações Laterais Aplicado ao Reconhecimento de Padrões Temporais [Rio de Janeiro] 1996.

X, 132 p. 29,7 cm (COPPE/UFRJ, D. Sc., Engenharia de Sistemas e Computação, 1996)

Tese - Universidade Federal do Rio de Janeiro, COPPE

1. Reconhecimento de Padrões 2- Redes Neurais

I. COPPE/UFRJ II. Título (série).

Às Minhas Amadas
Lais e Denise

Agradecimentos

À minha orientadora Sueli Mendes meus sinceros agradecimentos pela amizade e compreensão nos momentos mais difíceis.

Ao Professor Luís Alfredo pela confiança, apoio e amizade.

Ao Instituto de Pesquisas da Marinha, em especial ao Comandante Lopes, encarregado do Grupo de Acústica Submarina, pelo apoio nos últimos seis meses desta tese.

À Cláudia da Secretaria do Programa de Sistemas, pela sua maneira sempre gentil de ajudar.

E à Denise, pelo amor, compreensão e paciência.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências (D. Sc.).

MODELO NEUROCOMPUTACIONAL DE ELEMENTOS CENTRALIZADORES
COM INTERAÇÕES LATERAIS APLICADO AO RECONHECIMENTO DE
PADRÕES TEMPORAIS

Cláudio Chauke Nehme

MARÇO, 1996

Orientadores: Prof. Luís Alfredo Vidal de Carvalho e Sueli Bandeira Teixeira Mendes

Programa: Engenharia de Sistemas e Computação

O Problema de Reconhecimento de Padrões Temporais é debatido nesta tese sob o enfoque neurocomputacional.

O modelo matemático que melhor representa, teoricamente, a capacidade de reconhecer padrões pelo cérebro é o **Adaptive Resonance Theory (ART)**. A partir deste modelo introduziu-se a dinâmica para torná-lo apto a manipular padrões temporais preservando a memória adaptativa no interior do neurônio. A extensão do **ART** foi avaliada quanto a estabilidade e convergência do modelo dinâmico. Finalmente foi sugerido uma implementação em redes neuronais estruturada sob a arquitetura **ART Temporal** para obter resultados da proposta realizada nesta tese.

Abstract of Thesis presented to COPPE/UFRJ as partial fulfillment of the requirements for the degree of Doctor of Science (D. Sc.).

NEUROCOMPUTATION MODEL OF OFF-CENTER ELEMENTS WITH
SURROUNDING INTERACTIONS APPLIED TO SPATIOTEMPORAL PATTERN
RECOGNITION

Cláudio Chauke Nehme

MARCH, 1996

Thesis Supervisors: Prof. Luís Alfredo Vidal de Carvalho and Sueli Bandeira Teixeira
Mendes

Department: Computer Science and System Engineering

The Spatiotemporal Pattern Recognition Problem is discussed in this thesis under the neurocomputation viewpoint.

The best mathematics model which, theoretically, represents the ability of recognize patterns in the brain is the Temporal Adaptive Resonance Theory (Temporal ART), able to handle spatiotemporal patterns keeping adaptive memory in the neural body. The dynamic stability and convergence were evaluated for this extension of the ART. Finally, it is suggested a neural network structure under the Temporal ART to achieve the expect results.

Índice

I - Introdução.	1
II - O Problema do Reconhecimento de Padrões.	3
III - Estrutura Comportamental do Modelo Neuronal.	6
3.1 - Introdução.	6
3.2 - Capacidades de Estruturas Cerebrais do Tipo Centralização com Envolvimento Lateral Recorrente.	7
3.3 - O Modelo da Estrutura CELR.	8
3.3.1 - O Modelo do Neurônio da Estrutura CELR.	8
3.3.2 - O Efeito da Aprendizagem nos Neurônios da Estrutura CELR.	11
3.3.3 - O Mecanismo de Aprendizagem para uma Rede da Estrutura CELR.	14
3.3.4 - Características das Funções de Realimentações dos Neurônios da Estrutura CELR.	16
3.3.5 - O Efeito da Combinação de Entradas no Armazenamento em STM.	18
IV - A Memória Temporal em um Modelo Neuronal.	22
4.1- Introdução.	22
4.2 - Concentração de Informações Temporais.	22
4.2.1 - Comportamento de Padrões Temporais.	23
4.2.2 - Concentração de Informações Contidas em Padrões Temporais.	24
4.3 - Alguns Modelos de STM para Representação de Memória Temporal.	28
4.3.1 - STM no Perceptron Multi-Camadas.	28
4.3.2 - STM no Modelo Aditivo com Realimentação Positiva.	29
4.3.3 - STM por Convolução.	30
4.3.4 - STM com a Aplicação do Filtro ARMA.	33
4.3.5 - STM com a Aplicação do Filtro Gamma.	38
V - Modelo de um Neurônio da Estrutura CELR para Padrões Temporais.	43
5.1 - Introdução.	43
5.2 - Modelo da STM-CELR Temporal.	43
5.3 - Introdução à Teoria de Liapunov.	49
5.4 - Análise da Estabilidade da Dinâmica e da Convergência do Modelo STM-CELR Temporal.	55

5.5 - O Modelo STM-CELR Temporal com Filtro Gamma.	63
VI - A Arquitetura ART-Temporal com o Filtro Gamma.	65
6.1 - Introdução ao Modelo ART.	65
6.1.1 - O Modelo e a Arquitetura ART.	65
6.2 - O Modelo ART2.	71
6.3 - O Modelo ART2-Gamma.	76
6.4 - Testes com o ART2-Gamma.	79
6.4.1 - Seleção e Pré-Processamento dos Padrões de Entrada.	79
6.4.2 - Resultados das Classificações com o ART2-Gamma.	80
VII - Conclusões.	92
Bibliografia.	94
Anexo I.	102
Anexo II.	104

I - Introdução

Este trabalho de tese tem como proposta desenvolver um modelo neurocomputacional que preserve informações temporais de padrões não-estacionários na estrutura do neurônio.

É uma preocupação atual tratar deste tema devido a falta de métodos eficientes no reconhecimento de padrões não-estacionários. Na área de Reconhecimento de Padrões, a maioria dos problemas resolvidos são para o caso estático [Bock, 88], [Carpenter & Grossberg, 87a, 87b, 88, 90], [Carpenter, 89], [Fu, 80], [Gonzalez & Thomason, 78]. Para este caso, tanto os métodos estatísticos quanto os neurocomputacionais já alcançaram níveis de satisfações que podem ser verificados na solução de problemas usuais [Tank & Hopfield, 87a], [Chauke, 89], [Chauke & de Carvalho, 93].

A dificuldade do tema proposto é justamente observar, analisar e manter adaptativamente as informações variantes no espaço temporal. Algumas soluções tentam reproduzir todo o espaço de padrões a cada instante de tempo [Hecht-Nielsen, 87], mas este procedimento é caro e ineficiente, como visto em [Chauke & de Carvalho, 93].

O tipo de solução eficiente e teoricamente viável reduz-se no acompanhamento adaptativo das características peculiares do padrão como se elas fossem sinais temporais. O acompanhamento então recai em um problema de filtragem adaptativa cujas soluções são bem fundamentadas na área de Processamento de Sinais. Desta forma, o modelo de um filtro adaptativo traduz o conhecimento de uma característica do padrão, restando apenas o reconhecimento desse modelo para que a tarefa de classificação de padrões se realize.

Como classificador neuronal, escolheu-se adaptar a arquitetura ART (**Adaptive Resonance Theory**) [Grossberg, 76, 88], [Carpenter & Grossberg, 87a, 87b, 88] através da introdução dos filtros adaptativos nos seus neurônios. Esta extensão do ART capacita o classificador a trabalhar com sinal temporal e tomar uma decisão após alcançar um real conhecimento do padrão temporal.

A contribuição principal desta tese é a elaboração de uma família de sistemas dinâmicos capaz de representar o comportamento dos neurônios da arquitetura ART dedicada a análise de padrões temporais. Como forma de validação desta proposta, foram

desenvolvidos e provados alguns teoremas que garantem sua estabilidade e sua convergência.

O Capítulo 3 descreve o porquê da escolha do ART. O Capítulo 4 apresenta como deve-se comportar um modelo neurocomputacional para capacitar-se a trabalhar no espaço temporal. O Capítulo 5 descreve o modelo proposto nesta tese e avalia suas condições de estabilidade e convergência. E no Capítulo 6 encontram-se a implementação e alguns resultados desta proposta.

II - O Problema do Reconhecimento de Padrões

Definir Reconhecimento de Padrões não é uma tarefa simples, há várias abordagens e aplicações em que metodologias diferentes entram em cena. É mais simples relacionar o seu vasto campo de atividades e as áreas da ciência que contribuem na formação das técnicas utilizadas para solucionar o problema.

O Reconhecimento de Padrões está intimamente ligado aos sentidos básicos do homem, como a audição, a visão, o tato e também o paladar. Nós temos em mente as características físicas e abstratas do mundo, onde modelos de representação são naturalmente criados pelo nosso sistema neurológico permitindo comparar entidades externas e diferenciá-las de forma que decisões precisas sejam tomadas nas tarefas do cotidiano. Este mecanismo é o Reconhecimento Natural de Padrões, sendo seus sensores são os olhos, as mãos, a boca, etc..

As soluções dos problemas em Reconhecimento empregam teorias específicas de áreas bem diversificadas:

- Teoria da Decisão;
- Teoria da Informação;
- Teoria das Probabilidades;
- Modelagens Estatísticas;
- Teoria da Computação;
- Teoria de Autômatos;
- Modelagens Lógicas;
- Sistemas Neurocomputacionais, etc...

Resumindo, Reconhecimento de Padrões é um processo que discrimina os elementos de um domínio de análise, caracterizando-os em classes identificáveis; classificando-os em grupos que conservam valores semelhantes dos atributos pré-definidos que representam o arcabouço de um padrão genérico.

O problema em si origina-se justamente no entendimento do padrão. Define-se um padrão como sendo uma lista de características fundamentais de um objeto, ou entidade, com o objetivo de expressar a sua essência.

As características fundamentais - **features** - são resultados de algumas medidas sobre o objeto (ou entidade). Tais medidas são selecionadas com o intuito de melhor distinguir objetos diferentes e aproximar os semelhantes.

Estruturalmente, um padrão pode ser representado por um vetor multi-dimensional sobre o Espaço de **Features**.

As medidas podem ser efetuadas por transformações que diminuem a dimensionalidade da informação do objeto mantendo ainda as suas características principais.

A ligação entre o padrão a ser analisado e a sua classe é dada pelo Classificador. É uma função de decisão, mapeando pontos do espaço de **features** em um conjunto representando as diversas classes pré-definidas.

O método de reconhecimento é efetuado por um processo efetivo que determina uma seqüência de operações lógicas obtendo um dos resultados: pertence a classe C; não pertence a alguma classe conhecida.

Quando um padrão não é reconhecido pelo método de reconhecimento, diz-se que ele pertence à Classe de Rejeição.

Geralmente, tenta-se monitorar a falta de reconhecimento. Um critério estabelecido para se escolher as regras de decisões que melhor se adaptam ao ambiente dos padrões é a taxa de erro de classificação. Por exemplo, a regra de decisão que tiver a menor taxa de erro para uma etapa do reconhecimento é a candidata natural para participar, naquela etapa, do classificador.

Alguns fatores colaboram com os erros de reconhecimento. O ambiente de extração das **features** freqüentemente apresenta ruídos, perturbações, distorções, etc.. Para diminuí-los, tratamentos para a pré-extração são efetuados: filtragens, eliminação de ruídos conhecidos, e outros.

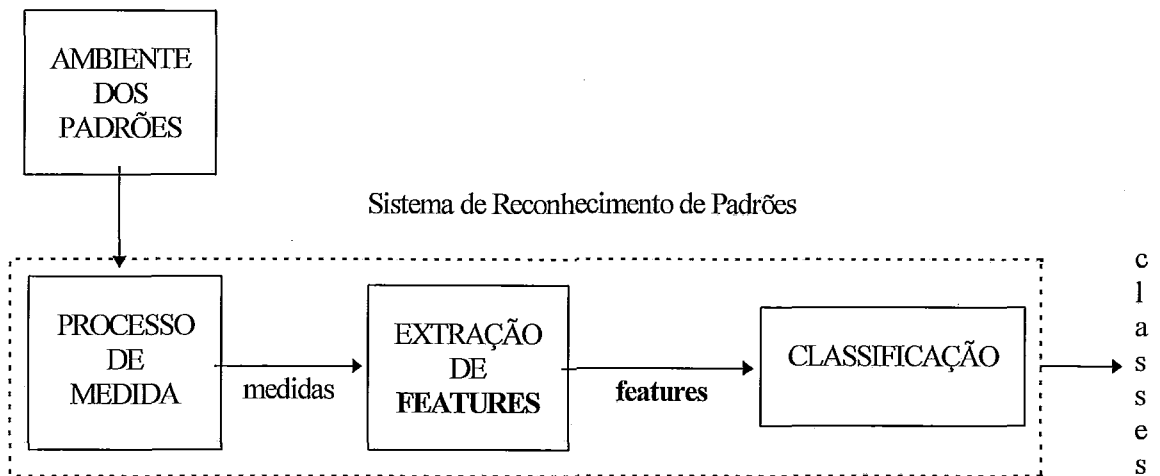


Figura 2.1: Sistema de Reconhecimento de Padrões.

Este modelo foi o precursor na Teoria de Reconhecimento de Padrões, enfocando a modelagem estatística na extração das **features** e nas regras de decisões dos classificadores.

Nos meados da década de 1950, surgiu o interesse na determinação de propriedades que caracterizam os padrões e de como representá-los. O Enfoque Lingüístico (Sintático) é o responsável por esta tarefa, definindo linguagens de representação de classes e autômatos das mesmas linguagens.

O problema da análise do padrão no espaço temporal é fortalecido pelo aumento do ruído e pela variação das suas características fundamentais. Ou seja, os sinais (padrões) são variantes no tempo e as **features** têm codificações que nem sempre são estacionárias. Este fato é o que fundamentalmente o diferencia do caso estático.

Um outro problema no caso temporal é a etapa de segmentação do padrão em primitivas, problema relacionado às variações comentadas no parágrafo acima. Vários estudos foram feitos com a metodologia convencional e nem sempre convém aplicá-lo [Chauke, 89], [Chauke & de Carvalho, 93], [Gutfreund & Mezard, 88], [Guyon, Personnaz, Nadal & Dreyfus, 88], [Hecht-Nielsen, 87], [Miller, 87], [de Vries & Principe, 92], [Wang & Arbib, 90]. Este tema não será abordado.

Os Tópicos 4.1 e 4.2 desta tese abordam os problemas do comportamento e da análise dos padrões temporais. Achou-se melhor discutir o aspecto temporal naqueles tópicos, pois permite um melhor entendimento do tema do Capítulo 4.

III - Estrutura Comportamental do Modelo Neuronal

3.1 - Introdução

Os estudos no desenvolvimento de modelos matemáticos para simular o funcionamento do cérebro foram iniciados para criar um novo enfoque de pesquisas na área de reconhecimento de padrões [Attneave,54], [Farley & Clark, 54], [Hebb, 49], [McCulloch & Pitts,43], [Rosenblatt, 58]. Existiam apenas os enfoques estatístico e lingüístico [Chauke & Maia, 92]. Com a introdução de modelos neurocomputacionais, criou-se uma linha de pesquisa mais próxima ao reconhecimento natural de padrões.

Os modelos neurocomputacionais possuem a associação, a memorização e a aprendizagem como características naturais. No cérebro há várias estruturas, cada qual responsável pelo processamento de informações de alguma específica região, tal como o tálamo, o córtex, o neocórtex, etc.. As regiões estão em contato com os estímulos externos (visão, audição, tato, etc.) e/ou com a manipulação do conhecimento (pensamento, fala, interpretação, etc.).

Apesar do cérebro possuir várias capacidades, como a semântica, a representação do conhecimento, a dedução, o planejamento, etc., neste trabalho apenas a capacidade de reconhecer padrões é considerada como objeto de estudo.

Stephen Grossberg é um dos pesquisadores que mais tem contribuído no desenvolvimento e na pesquisa de modelos neurocomputacionais dedicados ao reconhecimento de padrões. Dentre alguns desses modelos podem ser citados o **Masking Fields**, o **Instar**, o **Outstar** e o **ART (Adaptive Resonance Theory)** como os mais conhecidos, [Grossberg, 76, 88], [Carpenter & Grossberg, 88], [Carpenter, 89].

Neste capítulo será introduzido o modelo **Shunting Cooperative-Competitive Network** para mostrar-se a estrutura "ideal" de um modelo neuronal capaz de reconhecer padrões estáticos. A arquitetura **ART** possui os neurônios deste modelo. Referências sobre o **shunting cooperative-competitive network** são [Chauke & de Carvalho, 93], [Grossberg, 73, 87, 88], [Cohen & Grossberg, 87], [Wilson & Cowan, 72], [von der Malsburg, 73] e [Amari, 77].

3.2 - Capacidades de Estruturas Cerebrais do Tipo Centralização com Envolvimento Lateral e Recorrente

O modelo **shunting cooperative-competitive network** foi inicialmente vislumbrado na estrutura anatômica e funcional da região cerebral do hipocampo. Há uma ligação tenue das células do hipocampo com uma organização estruturada em um mecanismo centralizador com envolvimento lateral e recorrente. Nessa estrutura, o neurônio principal envia sinais (alguns excitatórios, outros inibitórios) para os seus neurônios vizinhos no hipocampo. Alguns dos neurônios realimentam outros e continua-se o processo até que o sinal de realimentação não excite mais qualquer neurônio dentro da região.

A grande vantagem da estrutura CELR (Central com Envolvimento Lateral e Recorrente) é a capacidade de aumentar em intensidade e em contraste as regiões de informações sensoriais [Ratcliff, 65]. Com este destaque, as informações sensoriais são beneficiadas diretamente ao terem que conviver com ruídos existentes e inevitáveis em vários sistemas neuronais. Uma outra vantagem é a capacidade de conviver com estímulos externos controlando o nível de saturação neuronal. Uma característica, em geral, dos neurônios é o nível da saturação finito em resposta a estímulos externos.

O que pode ocorrer com uma população de neurônios que não interage entre si (neurônios não-interativos)? Considere o processamento de sinais de alguns padrões de entrada em uma população de neurônios não-interativos. Se as intensidades destes sinais forem pequenas, eles poderão se perder em presença de ruído. Se, de outra forma, suas intensidades forem muito grandes, eles poderão saturar seus respectivos neurônios (associadores), criando um padrão uniforme de excitação na população e destruindo todas as informações sobre esses padrões de entrada. Tal situação é conhecida como **noise-saturation effect** [Grossberg, 73]. Esta rápida observação demonstra a ocorrência dessas duas situações extremas e indesejáveis numa população de neurônios não-interativos. O processamento de padrões neste caso fica restrito a faixas controladas e estreitas das intensidades dos sinais de entrada.

A flexibilidade de processamento de padrões arbitrários é característica da estrutura CELR. As interações entre os neurônios desta estrutura solucionam os problemas de saturação e perda de detecção de informação (nível de ruído).

A memória de curta duração (**STM - Short Term Memory**) reflete o aprendizado novo (recente). Seu conteúdo é volátil e, segundo [Grossberg, 68, 73, 88], depende de reforço para torná-lo duradouro na sinapse (**LTM - Long Term Memory**).

Uma outra característica da estrutura CELR é a capacidade de armazenar STM. Através da reverberação de um padrão de atividade distribuído sobre uma população de neurônios, a estrutura torna tal padrão como sendo um conhecimento novo para a população e ao longo da reverberação (tempo indefinido) armazena esse conhecimento na STM de alguns neurônios. Esta reverberação pode ser subitamente cortada por entradas inibitórias referentes a um novo padrão submetido à estrutura. Desta forma, os processos de reverberação podem manter representações de seqüências de eventos no interior da STM até a última contribuição ou uma punição (por entradas inibitórias) ocorrer e transferir a memória dessas seqüências para as sinapses (LTM) [Grossberg, 73, 87, 88].

3.3 - O Modelo da Estrutura CELR

3.3.1 - O Modelo de um Neurônio da Estrutura CELR

Existem dois modelos numa rede neuronal: o do funcionamento interno de um neurônio e o das interações entre neurônios. Os parâmetros importantes para a modelagem interna de um neurônio são:

- a) o evento E_p (padrão de entrada e outras entradas);
- b) o neurônio N_i da rede como ponto isolado independente do tempo;
- c) e a função temporal $X_i(t)$ atrelada a N_i . É um processo ocorrendo em N_i .

A função temporal $X_i(t)$ é a ativação média do neurônio N_i no instante t . Sua formulação define todo o comportamento do neurônio, ao longo do tempo, referente ao evento E_p . Como o evento E_p é passageiro, é pressuposto que $X_i(t)$ tenha um decaimento gradual e suave. Este decaimento permite que $X_i(t)$ seja considerada continuamente diferenciável.

A função $X_i(t)$ pode também ser traduzida como a reação a um experimento psicológico onde ela registra a ocorrência do evento E_p . É natural quantificar a não ocorrência de algo pelo valor zero. Portanto se E_p nunca foi apresentado a rede, então $X_i(t) \equiv 0$. Supondo que E_p tenha sido apresentado a rede, pela primeira vez, em $t = t(0)$

então $X_i(t)=0$ para $t \leq t^{(0)}$. Mas $X_i(t)$ não poderá manter-se nulo para $t > t^{(0)}$, pois $X_i(t)$ foi introduzido para representar a ocorrência de E_p . Desta forma supõe-se que $X_i(t) > 0$ para $t > t^{(0)}$.

Como t cresce, o efeito da ocorrência de E_p em $t > t^{(0)}$ gradualmente e suavemente diminui. Portanto, $X_i(t)$ deve gradualmente e suavemente retornar ao nível nulo.

Grossberg [68, 87] traduziu a ocorrência de um evento E_p num instante $t = t^{(0)}$ através de uma formulação matemática. Isto é, considere uma entrada $I_i(t)$ que perturbe $X_i(t)$ no instante $t = t^{(0)}$. $X_i(t)$ cresce mais rapidamente quando a entrada $I_i(t)$ for mais forte e decai a zero quando $I_i(t)$ for nulo. Assim,

$$\dot{X}_i(t) = -\alpha X_i(t) + I_i(t), \quad (3.1)$$

onde α é uma constante positiva e $X_i(0)$ é não - negativo.

A equação (3.1) esboça a situação onde o crescimento de $I_i(t)$ provoca um aumento na taxa de $X_i(t)$, isto é $X_i(t)$ cresce rapidamente. E quando $I_i(t) = 0$, $X_i(t)$ decresce na taxa $-\alpha X_i(t)$ até atingir o valor nulo.

Como a função $X_i(t)$ tem a responsabilidade de medir o grau de ativação do neurônio N_i , é razoável supor que seus valores sejam não-negativos. É também razoável supor que $I_i(t)$ seja não-negativa e contínua, pois $I_i(t)$ é a forma de apresentação do evento E_p ao neurônio N_i . Se E_p não se manifesta a N_i , $I_i(t)$ é nula. Porém, se E_p estiver relacionado a N_i então $I_i(t)$ retrata positivamente a presença de E_p .

É pressuposto que a entrada $I_i(t)$ seja limitada em tempo e que ocorra inicialmente no instante $t = t^{(0)}$. Assim, $I_i(t)$ torna-se positiva quando $t > t^{(0)}$ e retorna a zero após ultrapassar o intervalo finito de sua apresentação.

Analisando-se a equação (3.1), verifica-se que $X_i(t)$ pode tornar-se tão grande quanto $I_i(t)$ possa alcançar. No Tópico 3.2 observou-se que se $X_i(t)$ acompanhasse na íntegra os valores de $I_i(t)$ ocorreriam perdas de detecções de informações e problemas de saturação que provocariam destruições de informações dos padrões de entrada. Para eliminar estes problemas deve-se limitar $X_i(t)$:

$$\dot{X}_i(t) = -\alpha X_i(t) + (M_i - X_i(t)) I_i(t). \quad (3.2)$$

Na equação (3.2), [Grossberg, 68] supõe que $X_i(t)$ seja limitado por M_i , onde $0 \leq X_i(0) \leq M_i$. Em situações extremas $X_i(t)$ poderá ultrapassar M_i . Mas como $X_i(t)$,

$\dot{X}_i(t)$ e $I_i(t)$ são contínuas, M_i em (3.2) controla o ganho da entrada permitindo que $X_i(t)$ esteja em uma vizinhança bem pequena de M_i nestas situações e ao passar um curto instante de tempo $X_i(t)$ torna-se menor que M_i . Para o problema proposto não há perda de generalidade esta consideração feita por Grossberg. Portanto, como $X_i(t) \leq M_i$ para todo $t \geq 0$, independente do valor de $I_i(t)$, $X_i(t)$ satura em M_i , não influenciando os outros neurônios da rede.

O evento E_p pode ocorrer em vários instantes. Supondo-se que E_p ocorra nos instantes t_1, t_2, \dots, t_{N^i} , onde $t_j < t_{j+1}$, $j = 1, \dots, N^i - 1$, que $I_i(t)$ torne-se momentaneamente forte em todos os instantes $t = t_j$, $j = 1, \dots, N^i$ e que $J_i(t)$ seja uma função contínua, não-negativa e positiva no intervalo $(0, \lambda_j)$, $\lambda_j > 0$, então:

$$I_i(t) = \sum_{k=1}^{N^i} J_i(t - t_k). \quad (3.3)$$

A entrada $I_i(t)$ em (3.3) é uma sucessão de pulsos de entradas $J_i(t - t_k)$ no instante $t = t_k$. A entrada $J_i(t)$ é uma representação do evento E_p no instante t . A forma de onda desta representação é a "assinatura" do evento E_p para a rede neuronal.

A reverberação do neurônio N_i , ou auto-excitação, é modelada por uma função f_i que transmite um sinal de realimentação $f_i(X_i)$ para a função de ativação X_i de N_i . Como desejam-se evitar os problemas de saturação e perdas de detecções de informações, o sinal de realimentação também deve ser limitado. Desta forma, (3.2) torna-se:

$$\dot{X}_i(t) = -\alpha X_i(t) + (M_i - X_i(t)) [I_i(t) + f_i(X_i)], \quad (3.4)$$

onde $I_i(t)$ é dado pela equação (3.3) e $i = 1, \dots, M$, sendo M o número de neurônios da rede neuronal.

O modelo do neurônio da estrutura CELR é mais complexo do que foi apresentado pela equação (3.4). A estrutura CELR prevê interações entre seus neurônios. Essas interações só ocorrem a partir da etapa de aprendizagem. Após a aprendizagem, a função de ativação do neurônio N_i observa as influências de interação causadas por neurônios vizinhos.

3.3.2 - O Efeito da Aprendizagem nos Neurônios da Estrutura CELR

A aprendizagem tem um papel fundamental na formação das sinapses de uma rede neuronal. A partir de um conjunto de padrões de treinamento, é esperado que as sinapses se adaptem aos conhecimentos introduzidos pelos padrões e sejam capazes de definir hiper-regiões de categorias identificáveis. Uma vez a rede treinada, é suposto que ela possua a capacidade de extrapolação, isto é, reconhecer padrões quaisquer de entrada pertencentes ao domínio inicialmente gerado pelos padrões de treinamento.

No Capítulo 2 mencionou-se que um padrão (ou evento) pode ser separado (dividido) em componentes atômicos. Tais componentes formam um padrão específico se forem organizadas numa estrutura de lista e suas posições na lista não se alterarem. Várias aplicações tratam com padrões que são formados por listas de componentes. As redes neuronais quando são treinadas para reconhecer tais padrões executam, na maioria dos casos, um processo de reconhecimento dos componentes e da seqüência de ocorrência. Em algumas situações elas são capazes de predizer os componentes respeitando a ordem da seqüência. Supondo que uma rede tenha sido treinada para reconhecer padrões desta natureza, realizando predição, e que o padrão $P=[AB]$ é reconhecível pela rede, onde $[AB]$ é uma lista cujos componentes A e B ocupam respectivamente a primeira e a segunda posições, então ao se apresentar o componente A à rede espera-se que ela responda B num curto período de tempo após receber A.

A formalização do efeito esperado no parágrafo anterior é dada considerando que a apresentação de A à rede foi efetuada no instante $t = t^0$ por uma entrada $J_A(t - t^0)$ enviada pelo evento E_p ao neurônio N_A . A rede em seguida responde ao evento E_p através de uma saída do neurônio N_B . Esta saída torna-se forte em N_B no instante $t = t^0 + \tau_{AB}$, onde τ_{AB} é alguma reação de tempo positiva [Grossberg, 68].

O evento E_p age somente no neurônio N_A causando um aumento de sua ativação descrita por $X_A(t)$. O mecanismo de associação da rede provoca o aumento da ativação do neurônio N_B , registrado em $X_B(t)$, τ_{AB} unidades de tempo após t^0 . O aumento do valor de $X_B(t)$ ocorreu somente porque $X_A(t)$ tornou-se alto. A associação é então resumida pelo seguinte fluxo: um sinal de N_A deve ser enviado a N_B no instante $t^0 + \tau_{AB}$, $X_B(t)$

deve tornar-se grande em valor alimentando E_p por uma saída. Esta associação deve ocorrer toda vez que $X_A(t)$ tornar-se grande após $P=[AB]$ tiver sido aprendido.

Observe que o sinal de entrada para N_B enviado por N_A no instante t será forte se, e somente se, o valor de $X_A(t - \tau_{AB})$ for grande. Então, [Grossberg, 68] assume por simplicidade que este sinal é proporcional a $X_A(t - \tau_{AB})$, onde a constante de proporcionalidade é C_{AB} . Esta constante é positiva e representa o fluxo excitatório do neurônio N_A para o N_B

Em termos matemáticos deve-se considerar que o sinal de N_A para N_B é uma entrada para N_B da mesma forma que $I_B(t)$ é tratada. Portanto, de (3.4)

$$\dot{X}_B(t) = -\alpha X_B(t) + (M_B - X_B(t)) [I_B(t) + f_B(X_B) + C_{AB} X_A(t - \tau_{AB})]. \quad (3.5)$$

No Tópico 3.3.4 será verificado que o sinal de excitação do neurônio N_A para o neurônio N_B é representado por uma função $f_A(X_A)$ tal como em (3.4). Portanto, (3.5) é reescrita como:

$$\dot{X}_B(t) = -\alpha X_B(t) + (M_B - X_B(t)) (I_B(t) + f_B(X_B) + C_{AB} f_A(X_A)). \quad (3.6)$$

Para um padrão $P = [r_i r_j]$ genérico, (3.6) torna-se:

$$\dot{X}_j(t) = -\alpha X_j(t) + (M_j - X_j(t)) [I_j(t) + f_j(X_j) + C_{ij} f_i(X_i)]. \quad (3.7)$$

Se for impossível para a rede aprender a lista $[r_i r_j]$, então não haverá sinal de N_i que possa alcançar N_j e, portanto, $C_{ij} = 0$.

Nenhum fato obriga que o neurônio N_j receba excitação apenas por sinais oriundos de N_i . Ou seja, a rede pode aprender padrões $P = [r_i r_j]$ e $T = [r_k r_j]$ de forma independente. Assim, o neurônio N_j recebe sinais excitatórios dos neurônios N_i e N_k . A equação (3.7) é então transformada em:

$$\dot{X}_j(t) = -\alpha X_j(t) + (M_j - X_j(t)) [I_j(t) + f_j(X_j) + C_{ij} f_i(X_i) + C_{kj} f_k(X_k)]. \quad (3.8)$$

A expressão arbitrária referente aos sinais excitatórios que N_j possa receber é dado por:

$$\dot{X}_j(t) = -\alpha X_j(t) + (M_j - X_j(t)) \left[I_j + \sum_{i=1}^M C_{ij} f_i(X_i) \right]. \quad (3.9)$$

Observe que o somatório da expressão (3.9) engloba a função $f_j(X_j)$, com $C_{jj} = 1$, e pressupõe que N_j possa receber sinais excitatórios de todos os neurônios da rede. Se não for estabelecida a conexão excitatória do neurônio N_i para o N_j então $C_{ij} = 0$, não prejudicando a veracidade de (3.9). O valor de C_{ij} não precisa ser a unidade e C_{ij} representa uma via para auto-excitação.

No Tópico 3.2 foi citado que na estrutura CELR ocorrem interações inibitórias e existem entradas inibitórias para os seus neurônios. As funções que representam os sinais inibitórios e as entradas inibitórias são bem similares às funções do caso excitatório. O efeito causado num neurônio por entradas e sinais inibitórios é a diminuição da ativação do neurônio. A introdução dos efeitos inibitórios sofridos pelo neurônio N_j na função de ativação $X_j(t)$ é verificada na seguinte equação:

$$\begin{aligned} \dot{X}_j(t) = & -\alpha X_j(t) + (M_j - X_j(t)) \left[I_j + \sum_{i=1}^M C_{ij} f_i(X_i) \right] \\ & - (X_j(t) + Q_j) \left[K_j + \sum_{i=1}^M E_{ij} g_i(X_i) \right]. \end{aligned} \quad (3.10)$$

Na Equação (3.10), X_j pode variar no intervalo finito $[-Q_j, M_j]$, $Q_j \geq 0$, em resposta as entradas excitatórias I_j , as inibitórias K_j , aos sinais de realimentações excitatórias $f_i(X_i)$ e aos sinais de realimentações inibitórias $g_i(X_i)$. Os termos $(M_j - X_j(t))$ e $(X_j(t) + Q_j)$ são controladores de ganho para evitar-se os problemas de saturação e perdas de detecções de informações. A constante de proporcionalidade E_{ij} representa o fluxo inibitório do neurônio N_i para o N_j , seu valor também é positivo e quando não há sinal inibitório de N_i para N_j , $E_{ij} = 0$. Como na equação (3.9), para o caso de envio de sinais excitatórios, não é obrigatório que todos os neurônios enviem sinais de realimentações inibitórios para o neurônio N_j , preservando a generalidade da equação (3.10), pois quando não há envio de sinais, a constante de proporcionalidade é nula, como deve ser o caso de E_{ij} . A função de realimentação inibitória $g_i(X_i)$ possui características semelhantes da função $f_i(X_i)$, conforme será visto no Tópico 3.3.4.

3.3.3 - O Mecanismo de Aprendizagem para uma Rede da Estrutura CELR

No Tópico 3.3.2 citou-se o comportamento desejado e esperado de uma rede neuronal após a etapa de aprendizagem. Por exemplo, se o padrão $P = [r_i r_j]$ tiver sido aprendido pela rede então após a aprendizagem espera-se que a apresentação de r_i à rede provoque o envio de um sinal de realimentação excitatório de N_i somente para N_j e N_j responda a esta entrada através de um sinal que esteja associado ao componente r_j . Porém, antes da aprendizagem, N_i deve ser capaz de enviar sinais (excitatórios ou inibitórios) para todos os neurônios N_k que correspondam aos símbolos r_k que possam ser aprendidos. A definição do tipo de sinal que N_i enviará para N_k é obtida ao longo do processo de aprendizagem. Se a seqüência $r_i r_k$ estiver contida em alguma lista a ser aprendida então N_i enviará sinais excitatórios para N_k fortalecendo este elo. Se a seqüência correta a ser aprendida for $r_i r_j$ mas o neurônio N_k estiver sendo ativado quando N_i envia um sinal excitatório para N_j e N_k estimular a associação de outro componente que não seja r_j então é provável que N_i envie um sinal inibitório para N_k fortalecendo esta inibição ao longo da aprendizagem de $r_i r_j$. Ao desenrolar o processo de aprendizagem de $r_i r_j$, N_i tende a eliminar o envio de sinais excitatórios e inibitórios para os neurônios que não influenciam o elo entre N_i e N_j .

O efeito que a aprendizagem de $r_i r_j$ provoca sobre a rede pela sua apresentação várias vezes, com intervalo de tempo ω entre r_i e r_j , é tornar os valores de $X_i(t-\omega)$ e $X_j(t)$ grandes durante a duração das fortes entradas $I_i(t-\omega)$ e $I_j(t)$, respectivamente, e torná-los pequenos com a perda destas entradas. O aumento da freqüência de apresentação de $r_i r_j$ garante que $X_i(t-\omega)$ e $X_j(t)$ terão valores grandes com maior freqüência. Se de outra forma, $r_i r_k$ tiver sido apresentado, $X_i(t-\omega)$ poderá ocasionalmente possuir valores grandes mas $X_j(t)$ deverá sempre ter valores muito pequenos. Se somente r_j for apresentado, $X_j(t)$ poderá ocasionalmente possuir valores grandes mas $X_i(t-\omega)$ deverá sempre ter valores muito pequenos. E se não for apresentada qualquer lista durante um período de tempo, então $X_i(t-\omega)$ e $X_j(t)$ devem ser sempre muito pequenos nesse período. Destes fatos conclui-se que a aprendizagem de $r_i r_j$ ocorrerá se, e somente se, o produto

$$X_i(t-\omega)X_j(t) \quad (3.11)$$

for freqüentemente grande e todos os outros produtos $X_i(t-\omega)X_k(t)$, onde $k=1, \dots, M$ e $k \neq j$, permanecerem pequenos.

Para a rede neuronal calcular e avaliar tais produtos é necessária a introdução de algum processo a mais no modelo neuronal. [Grossberg, 68] postulou a existência do processo $Z_{ij}(t)$ que cresce somente quando $X_i(t-\omega)X_j(t)$ atinge valores altos. A grosso modo, $Z_{ij}(t)$ encontra-se na ligação entre um axônio de N_i e um dendrito de N_j , pois ele é o único local onde o sinal $C_{ij}f_i(X_i(t-\tau_{ij}))$ de N_i e a ativação $X_j(t)$ de N_j são simultaneamente acessados. Portanto, o produto (3.11) deve ser substituído por

$$C_{ij}f_i(X_i(t-\tau_{ij}))f_j(X_j(t)), \quad (3.12)$$

e diz-se que $Z_{ij}(t)$ cresce se, e somente se, (3.12) alcançar valores grandes. [Grossberg, 68, 87, 88] expressou este fato matematicamente dizendo que $Z_{ij}(t)$ cresce numa taxa igual a (3.12) menos, talvez, um decaimento espontâneo (ou "esquecimento") $uZ_{ij}(t)$. Isto é:

$$\dot{Z}_{ij}(t) = -uZ_{ij}(t) + C_{ij}f_i(X_i(t-\tau_{ij}))f_j(X_j(t)), \quad (3.13)$$

onde $u > 0$, $C_{ij} \geq 0$, $T_{ij} > 0$ e $Z_{ij}(0) \geq 0$.

Por analogia da consideração de limitação de $X_i(t)$ em (3.2), se Z_{ij} possuir um máximo finito fixado M_{ij} , então (3.13) torna-se

$$\dot{Z}_{ij}(t) = -uZ_{ij}(t) + C_{ij}(M_{ij} - Z_{ij}(t))f_i(X_i(t-\tau_{ij}))f_j(X_j(t)). \quad (3.14)$$

Em resumo, as funções $Z_{ij}(t)$ existem nas ligações sinápticas porque este é o único local aonde a ativação passada de N_i e a presente de N_j coexistem. Além disto, $Z_{ij}(t)$ serve para armazenar o aprendizado da seqüência $r_i r_j$ ao longo das suas várias apresentações à rede. A função $Z_{ij}(t)$ implementa o mecanismo de armazenamento da memória de longa duração (LTM).

Supondo-se que $r_i r_j$ tenha sido aprendida, $Z_{ij}(\omega)$ terá valores altos nos instantes onde $X_i(t)$ e $X_j(t+\tau_{ij})$ forem grandes. Como $X_j(t+\tau_{ij})$ tornará grande somente se o sinal enviado por N_i a N_j for forte, [Grossberg, 68] conclui que os requisitos heurísticos mostram a influência de $Z_{ij}(\omega)$ na intensidade do sinal $C_{ij}f_i(X_i(t))$ enquanto estiver

sendo transferido na via de N_i para N_j . Ou seja, a função de $Z_{ij}(t)$ é controlar a passagem do sinal enviado por um axônio de N_i para um dendrito de N_j . Os requisitos heurísticos também implicam que $X_j(t + \tau_{ij})$ torna-se grande se $C_{ij}f_i(X_i(t))$ e $Z_{ij}(t + \tau_{ij})$ forem altos, ou somente se o produto

$$C_{ij}f_i(X_i(t)) Z_{ij}(t + \tau_{ij}) \quad (3.15)$$

tiver um valor alto.

Observe que $Z_{ij}(t)$ aplica-se tanto a sinapses excitatórias quanto a sinapses inibitórias. O fato de propor-se o aprendizado da lista r_j à rede não implica em obter-se apenas sinapse de N_i para N_j . Existem outras sinapses associadas a efeitos excitatórios e inibitórios definidas pelo mecanismo de aprendizagem da rede. O tratamento funcional de todas as sinapses é único. Portanto, convencionou-se $Z_{ij}^+(t)$ aplicável a sinapses excitatórias e $Z_{ij}^-(t)$ aplicável a inibitórias.

O modelo de estrutura neuronal torna-se completamente definido ao adaptar-se o termo (3.15) à equação (3.10)

$$\begin{aligned} \dot{X}_j(t) = & -\alpha X_j(t) + (M_j - X_j(t)) \left[I_j + \sum_{i=1}^M C_{ij} Z_{ij}^+(t) f_i(X_i) \right] \\ & - (X_j(t) + Q_j) \left[K_j + \sum_{i=1}^M E_{ij} Z_{ij}^-(t) g_i(X_i) \right]. \end{aligned} \quad (3.16)$$

3.3.4 - Características das Funções de Realimentações dos Neurônios da Estrutura CELR

As funções de realimentação têm um papel importante para as redes neuronais durante a etapa de aprendizagem. Elas são responsáveis por aumentar o contraste das informações entre si e delas em relação ao ruído. Se as informações fossem representadas por um mapa de relevo colorido, nele poderia observar-se que existiriam contrastes de relevo e cores separando ruído de informações e informações entre si.

Grossberg [73] analisou várias funções para efetuar a realimentação neuronal como desejado. Suas pesquisas levaram-no às funções do tipo sigmóide. Essas funções servem tanto para f_i quanto para g_i na equação (3.16).

A título de exemplificação, considere as seguintes possibilidades para uma função f de realimentação:

- 1) $f(\omega)$ linear ou não-linear que cresça menos que a identidade;
- 2) $f(\omega)$ não-linear que cresça mais do que a identidade.

Na possibilidade 1), por exemplo $f(\omega) = \omega(1+\omega)^{-1}$, o ruído será amplificado e reverberado. Note que se $f(\omega)$ for linear, então não haverá aumento brusco de contraste, ou seja não será definido um contorno para realçar a região de informação.

Na possibilidade 2), por exemplo $f(\omega) = \omega^2$, o problema do realce de contorno da informação é evitado, e ruídos com valores suficientemente pequenos irão dissipar-se ao longo do tempo.

Considere agora que um padrão de entrada bem perceptível seja adicionado ao ruído por um curto período de tempo. Neste caso, duas propriedades são observadas:

1ª) todos os neurônios que receberam uma forte entrada referente ao padrão de entrada irão suprimir as ativações de todos os outros neurônios, inclusive o ruído;

2ª) ocorrerá normalização: a ativação total $X(t) = \sum_{i=1}^M X_i(t)$ de todos os neurônios será aproximada a um limite positivo ao longo do tempo.

A primeira propriedade mostra que ocorrerá um extremo aumento da intensidade do contorno. Somente as informações do padrão de entrada perdurarão. Se um neurônio se ativar mais com a entrada do que os outros, então a rede extinguirá todos os outros escolhendo-o como sobrevivente.

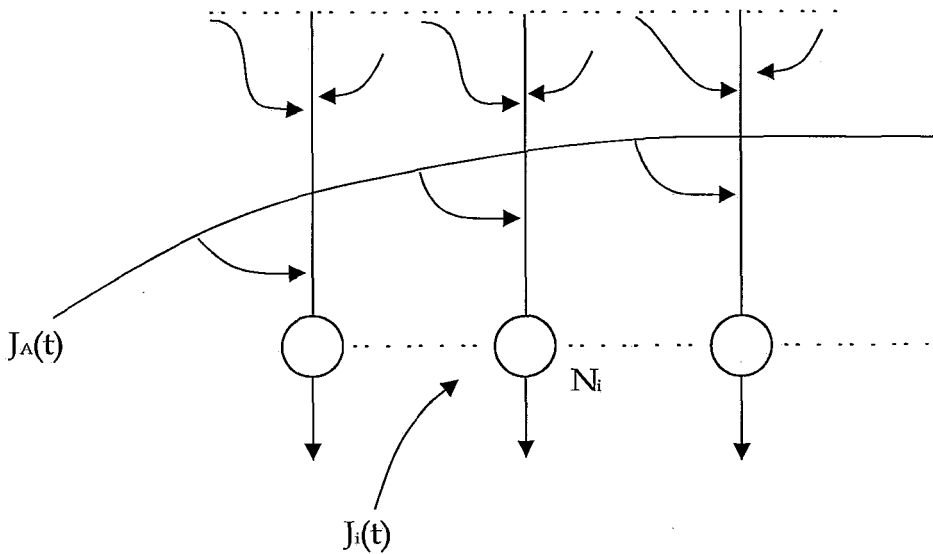
A segunda propriedade mostra que a rede neuronal regula precisamente sua ativação total podendo, inclusive, armazenar a ativação de certos neurônios indefinidamente na STM pela reverberação de suas ativações ao longo do ciclo excitatório recorrente [Grossberg, 73].

Observe que a primeira propriedade é muito forte, muitos padrões serão eliminados na tentativa de suprimir o ruído. Uma maneira de evitar este problema é escolher $f(\omega)$ de forma que cresça mais rápido que a identidade para pequenos valores de ω e aproximadamente linear para grandes valores de ω . Assim, o ruído dissipa-se e surge um limiar de extinção. Isto significa que, dado um padrão de entrada com energia suficiente para distinguir-se do ruído, as ativações dos neurônios que estiverem abaixo do

limiar serão destruídas (incluindo o ruído), enquanto aquelas que estiverem acima do limiar aumentarão a intensidade das regiões de informações e serão armazenadas na STM.

3.3.5 - O Efeito da Combinação de Entradas no Armazenamento em STM

A determinação dos neurônios a serem excluídos do armazenamento em STM é dependente do total de entradas $I = \sum_{k=1}^M I_k$ na rede. Grossberg [73] considera que há uma combinação de entradas que rege essa determinação. Suponha que uma entrada não específica J_A seja utilizada para perturbar o sistema, sendo também combinada com entradas específicas J_i para o neurônio N_i (Figura 3.1).



**Figura 3.1: Entrada $J_A(t)$ perturbadora do sistema.
Entrada $J_i(t)$ específica para o neurônio N_i .**

Essa combinação de entradas gera dois casos importantes. O primeiro é uma combinação multiplicativa entre $J_A(t)$ e $J_i(t)$ influenciando a ativação X_i . A entrada $J_A(t)$ é dita ser uma derivação da excitação. No segundo, $J_A(t)$ e $J_i(t)$ combinam aditivamente influenciando a ativação X_i .

No primeiro caso, a entrada $J_A(t)$ não altera os níveis relativos de entradas para os neurônios (Figuras 3.2 e 3.3). Os padrões de entrada para a rede mantêm-se inalterados quanto a forma, modificando-se apenas os valores de intensidade.

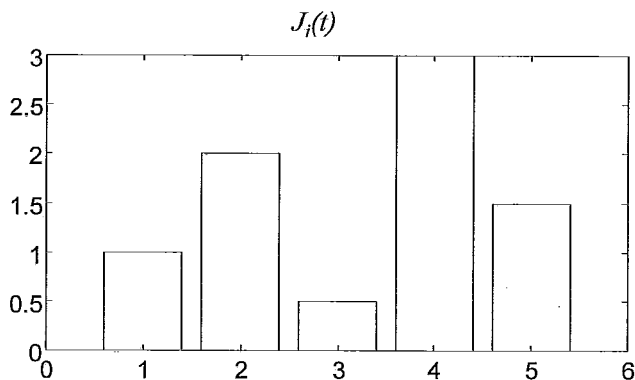


Figura 3.2: Exemplo dos padrões de entrada $J_i(t)$ para os neurônios N_i .
O valor de $M = 5$.

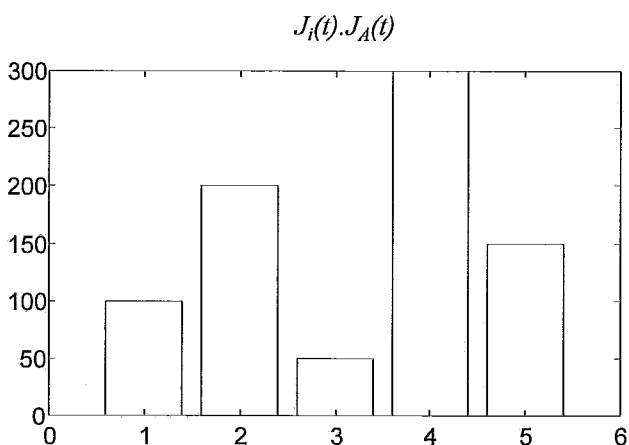


Figura 3.3: A combinação multiplicativa da entrada $J_A(t)$ com os padrões de entrada $J_i(t)$ da Figura 3.2.

Para o segundo caso, uma entrada intensa de $J_A(t)$ tende a uniformizar as entradas $J_i(t)$ conforme exemplifica a Figura 3.4.

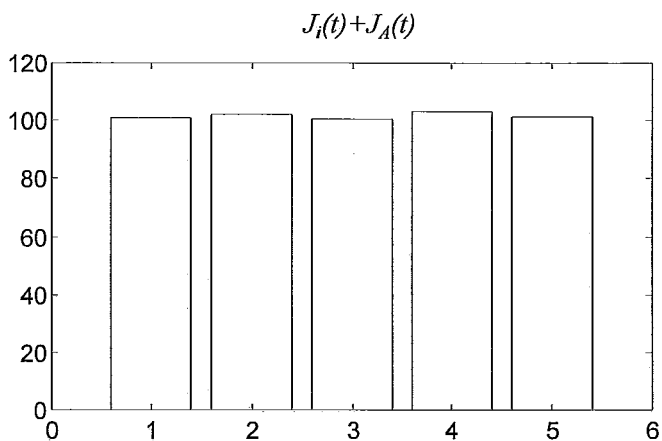


Figura 3.4: A combinação aditiva da entrada $J_A(t)$ com os padrões de entrada $J_i(t)$ da Figura 3.2.

Na Figura 3.5 (a), a entrada inibitória previne o despertar da ativação $X_i(t)$, porém a entrada excitatória recorrente colateral foge da inibição. Portanto, se o neurônio N_i já estiver reverberando, ele continuará até que a entrada inibitória alcance-o. Este tipo de inibição previne a transferência de novas informações para a STM e permite armazenar as informações anteriores na mesma STM. Na Figura 3.5 (b), a entrada inibitória previne a ativação $X_i(t)$ de um despertar mas evita reverberações anteriores.

VI - A Memória Temporal em um Modelo Neuronal

4.1 - Introdução

A análise de padrões no espaço temporal torna-se mais difícil do que no caso de padrões estáticos pois há um aumento de ruído e há variação temporal das características fundamentais dos padrões. No caso temporal os padrões são sinais temporais cujas características fundamentais têm codificações que nem sempre são estacionárias. Este fato diferencia fundamentalmente o comportamento dos padrões temporais em relação aos estáticos.

O objetivo de pesquisar-se uma representação para a memória temporal é definir uma estrutura que se adapte às variações não-estacionárias dos padrões temporais mantendo-se sensível às peculiaridades de tais padrões. Em consequência, após a estrutura adaptar-se ao comportamento das peculiaridades (ou características fundamentais) de um padrão temporal, deverá ser possível associar o seu conteúdo com tal padrão quando este estiver presente.

O mecanismo neuronal mais apropriado para representar as variações de comportamento de um padrão é o próprio neurônio. Esta afirmação é considerada por ser o neurônio responsável pelo armazenamento da memória de curta duração (STM), cuja estrutura possui percepção mais dinâmica do que as sinapses. Conclui-se portanto que um histórico do padrão temporal deve ser armazenado num modelo apropriado da STM.

Neste capítulo serão estudados alguns modelos de STM que atendam as necessidades de representação da memória temporal. A princípio não é verificada a necessidade de modificações nos modelos de representações de sinapses e de aprendizagem.

Antes de abordar tais modelos, verifica-se a necessidade de ser exposto a análise de informações temporais que levou ao caminho seguido por esses modelos.

4.2 - Concentração de Informações Temporais

Há dois temas que devem ser discutidos neste tópico:

- 1) comportamento de padrões temporais (sinais não-estacionários) ;

2) concentração de informações contidas em padrões temporais.

A análise do comportamento ajuda a conhecer os limite e as rotas das informações ao longo do tempo. E a elaboração de uma forma de concentrar as informações é o passo inicial para descrever um modelo para uma STM.

4.2.1 - Comportamento de Padrões Temporais

O comportamento de padrões temporais será determinado sobre os sinais não-estacionários desses padrões. Considere então que o sinal $X(s,t)$ não-estacionário seja definido de $t_0 \leq t \leq t_\lambda$ no espaço temporal e de $S_0 \leq s \leq S_\lambda$ no espaço de medidas de padrões. A Figura 4.1 esquematiza a evolução do sinal.

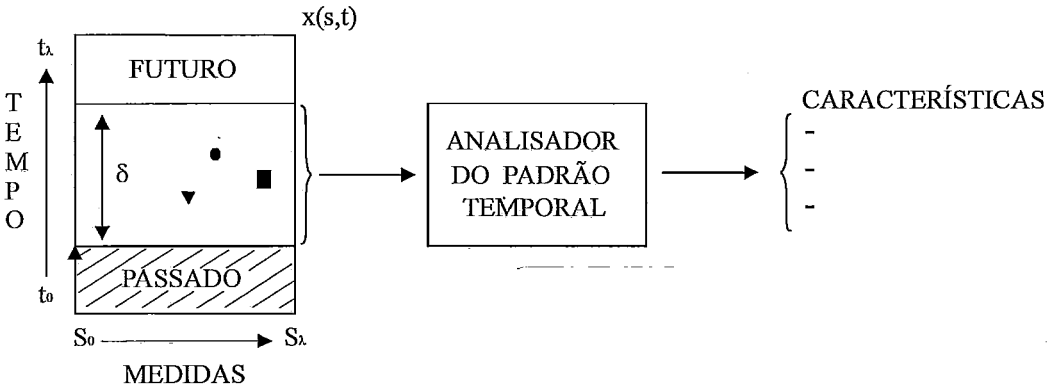


Figura 4.1: Esquema da evolução do sinal no tempo e no espaço de medidas.

O intervalo δ apresentado na Figura 4.1, $0 < \delta < t_\lambda - t_0$, é o tamanho de uma janela, dita deslizante, que acompanha uma faixa dinâmica do sinal, isto é, move-se com a evolução temporal do sinal. A análise do padrão é realizada sobre a parte do sinal que está no interior dessa janela.

Em algumas técnicas de reconhecimento de padrões estacionários, existem as etapas de segmentação de um padrão (sinal) em componentes independentes (primitivas do padrão) e de identificação dessas componentes. A segmentação divide um padrão mantendo garantias que qualquer parte segmentada possui um conjunto de características que define um único componente e que este componente não faz parte de qualquer outra parte segmentada. Para sinais não-estacionários o problema é mais complicado, pois as

características têm suas propriedades estatísticas variantes ao longo do tempo. A segmentação neste caso, quando necessária, deve ser adaptativa, já que nem sempre ocorrem variações bruscas do sinal quando ele passa de um componente para outro. Isto é, as características não-estacionárias possuem trajetórias variantes e estabelecer fronteiras torna-se uma tarefa difícil. Observe que foi citada a existência da segmentação por análise do sinal e a segmentação por análise das características. A diferença entre estas duas formas de segmentação é que as características são extraídas do sinal, sendo necessária a utilização de técnicas de processamento de sinais para serem detectadas e acompanhadas na evolução do sinal.

Concluindo, o comportamento de padrões temporais não-estacionários é avaliado pela análise adaptativa do sinal ou pela sua filtragem adaptativa para acompanhar às características dos componentes do padrão.

4.2.2 - Concentração de Informações Contidas em Padrões Temporais

Muitos problemas de percepção são tratados a partir de comparações de informações sensoriais obtidas em um extenso período de tempo. Em várias aplicações de engenharia é comum a existência de informações que se manifestam em pequenos intervalos de tempos com uma certa distribuição de frequências. O objetivo deste tópico é sugerir um método de concentração de informações que permita detectar o instante de surgimento de uma informação mantendo memória das outras informações já surgidas.

Para tornar o texto mais compreensivo, serão explicados temas da área de reconhecimento de voz tal como [Tank & Hopfield, 87a] fizeram didaticamente. Um estudo mais abrangente desta área é encontrado em [Chauke, 89].

Um processo de reconhecimento de uma locução é apresentado na Figura 4.2. A primeira etapa é o processamento acústico da locução que tem por finalidade segmentá-la em fonemas utilizando uma base fonética. Essa base armazena conhecimentos sobre a variedade de fonemas utilizados num certo vocabulário. O decodificador lingüístico analisa a seqüência de fonemas produzida pelo processador acústico e, consultando o modelo de linguagem junto com o vocabulário definido para o universo de discurso, procura a sentença que melhor represente a locução original [Chauke, 89].

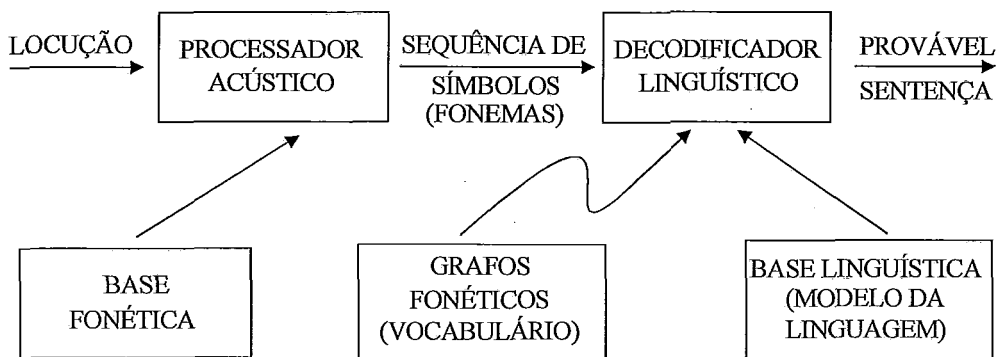


Figura 4.2: Processo de Reconhecimento.

A Figura 4.3 exemplifica alguns problemas relacionados a segmentação e ao reconhecimento da seqüência de fonemas.

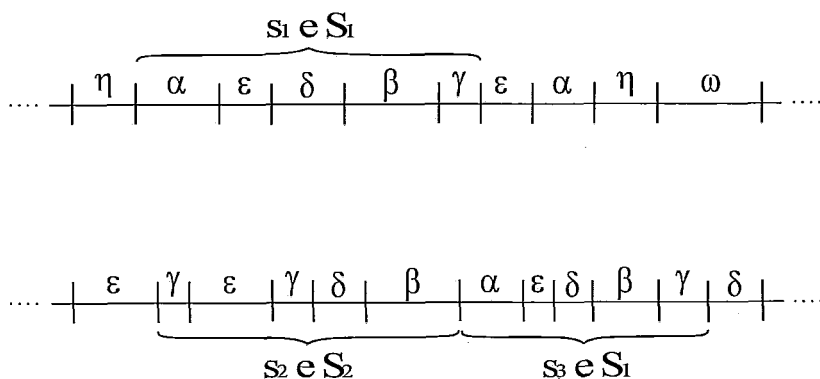


Figura 4.3: Dois segmentos de locuções distintas.

Nesta figura, as distâncias entre as barras verticais denotam as durações temporais dos vários fonemas ($\alpha, \beta, \delta, \dots$). O problema exposto reduz-se à identificação das seqüências s_1 e s_3 como elementos da categoria S^1 e à identificação da seqüência s_2 como um membro de uma diferente categoria S^2 . Observe que existem situações onde o intervalo de tempo de um mesmo fonema é variável e que mesmo assim deve-se reconhecer todos os fonemas corretamente mantendo informações em memória para identificar seqüências como membros de categorias.

O primeiro passo para manter informações em memória é conhecer o comportamento de cada fonema das seqüências de uma mesma categoria. Obtêm-se as

distribuições de probabilidades para cada fonema, considerando o instante inicial é a duração de sua elocução. A Figura 4.4 exemplifica este fato.

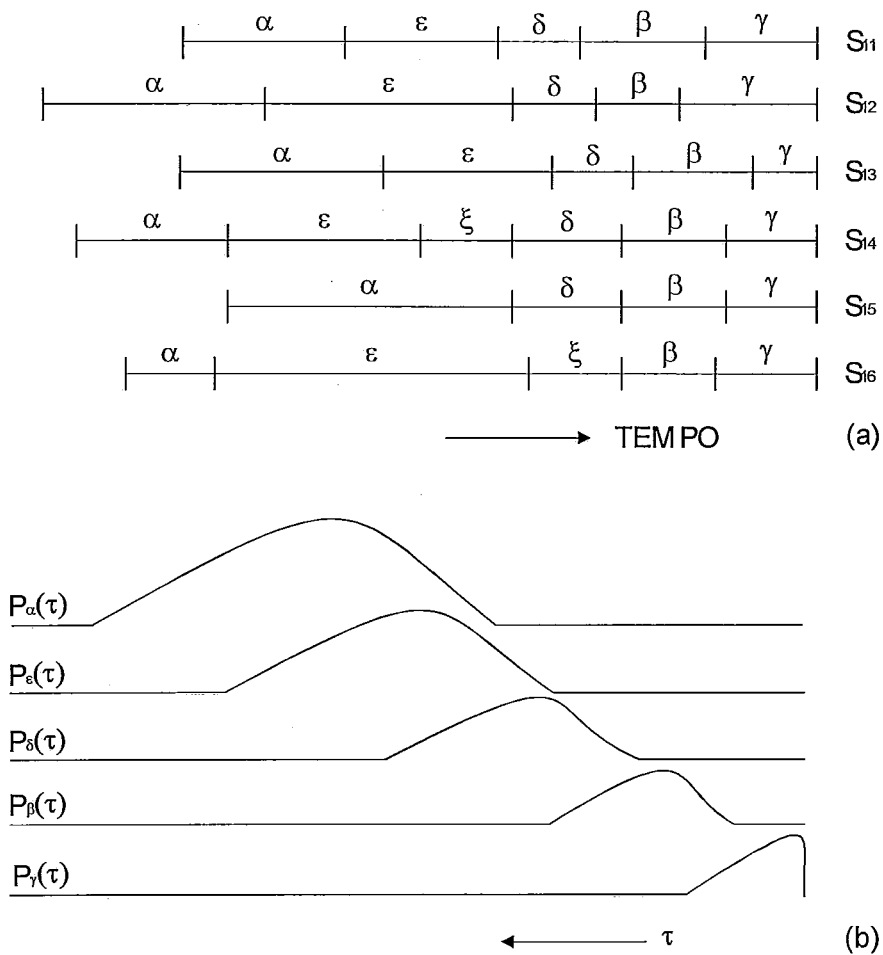


Figura 4.4: (a) $s_1^1, s_2^1, s_3^1, s_4^1, s_5^1$ e s_6^1 pertencem a categoria S^1 . (b) distribuições de probabilidades das seqüências de (a).

Na Figura 4.4 encontram-se seis seqüências de fonemas de uma única categoria S^1 . Não há elocuições equivalentes e são distintas umas das outras por motivos do posicionamento temporal de fonemas e pela existência de alguns fonemas não pertencentes a todas as seqüências. Na parte (b) da figura estão as distribuições de probabilidades dos fonemas que formam elementos da categoria S^1 . As distribuições foram definidas considerando as posições dos fonemas e o intervalo τ entre o final de cada fonema e o tempo esperado para término de uma seqüência. As faixas largas das distribuições são causadas pelo efeito das variações de faixas temporais de elocuições dos fonemas.

Definidas as distribuições de probabilidades, pode-se agora construir um mecanismo para reconhecimento de seqüências mantendo memória de cada símbolo detectado. A Figura 4.5 ilustra esta idéia.

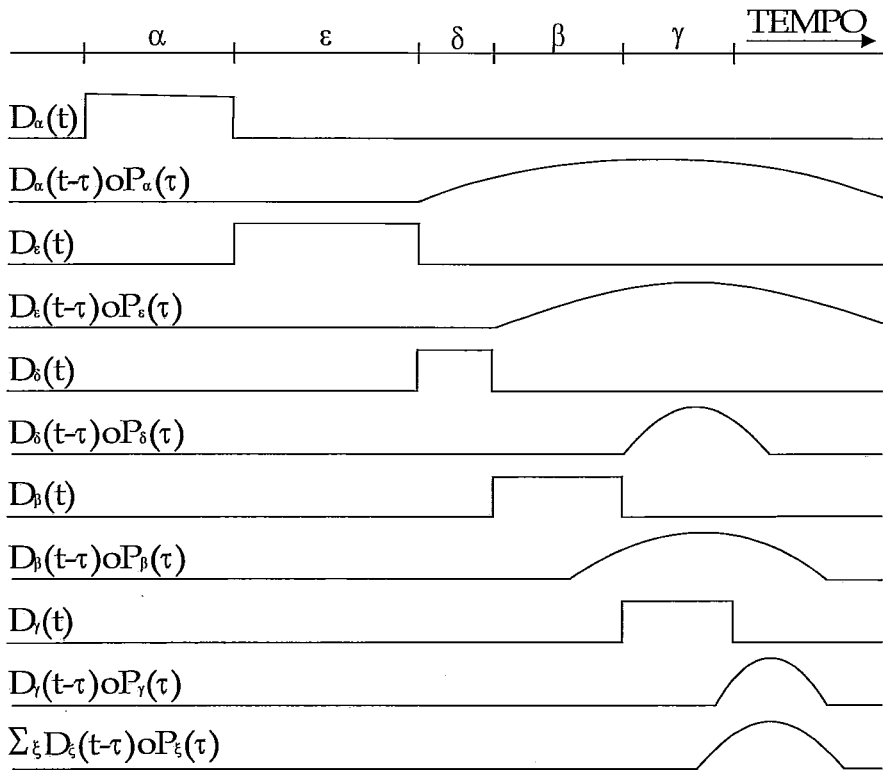


Figura 4.5: Mecanismo de reconhecimento de seqüências fonéticas.

A probabilidade de detectar-se o Fonema β , na seqüência da Figura 4.5, τ instantes de tempo antes do término da elocução de todos os fonemas será $P_\beta(\tau)$, se tal seqüência pertencer à categoria S^1 . Para preservar em memória futura a detecção do fonema β como elemento de uma seqüência da categoria S^1 , deve-se propagar a frente no tempo esta informação através de um sinal representativo da distribuição de probabilidade de β até o instante de decisão de reconhecimento da seqüência analisada. No exemplo da Figura 4.5, o detector para β têm uma saída $D_\beta(t)$. Dai a propagação do sinal representativo da distribuição $P_\beta(\tau)$ terá a forma $D_\beta(t-\tau) \circ P_\beta(\tau)$, onde o operador "o" é a função convolução (ref.). O resultado da aplicação da função convolução sobre $D_\beta(t)$ e $P_\beta(\tau)$ é a obtenção da resposta do mecanismo de detecção do fonema β pertencente a alguma seqüência da categoria S^1 frente a uma entrada da seqüência

analisada. Esta resposta fornece um grau de pertinência do fonema para seqüências da categoria S^1 e mantém em memória tal resposta até o instante final do reconhecimento da seqüência completa. Os sinais de outras detecções formados por convoluções similares são também resultados de propagações dos respectivos fonemas α, δ, ϵ , etc.. No Tópico 4.3.3 será definida formalmente a função convolução e explicada a sua aplicação.

O reconhecimento de uma seqüência s_i pertencente a categoria S^1 é efetuado se a soma de todos os sinais convoluídos ultrapassar um limiar pré-estabelecidos e estes sinais estiverem em fase com a categoria. Isto significa que há uma grande evidência acumulativa que s_i pertença a S^1 . Na parte baixa da Figura 4.5 é ilustrado este fato. Se s_i não pertence à categoria S^1 , então os sinais convoluídos não introduzirão informações, sendo observada pela soma de baixa amplitude ou pela defasagem entre os sinais e a categoria.

4.3 - Alguns Modelos de STM para Representação de Memória Temporal

No Tópico 4.2.2 verificou-se que o mecanismo de detecção e reconhecimento de seqüências fonéticas pode ser baseado na função convolução para armazenamento de memória temporal. Neste tópico serão apresentados inicialmente os modelos de STM do perceptron multi-camadas e do modelo aditivo com realimentação positiva. Esses modelos são simples e bem limitados, foram precursores na tentativa de introduzir-se memória temporal. Aprofundando-se nos exemplos, serão apresentados os modelos de STM por convolução, pelo filtro ARMA e pelo modelo Gamma. São modelos mais eficientes que se aproximam das necessidades contidas nesta tese.

4.3.1 - STM no Perceptron Multi-Camadas

O perceptron foi historicamente o modelo precursor para classificar padrões estáticos. A estrutura de armazenamento da STM possui como ativação dos neurônios a fórmula

$$x_i = \sigma \left(\sum_{j < i} w_{ij} x_j \right) + I_i \quad (4.1)$$

onde x_i é a ativação do neurônio i , w_{ij} é o peso associado à sinapse do neurônio j para o neurônio i , σ é uma função de transferência do sistema de ativação e I_i representa uma entrada externa.

Sua característica fundamental de processamento de padrões estáticos é visualizada pela avaliação apenas da ativação do corrente instante.

Uma forma simples para adicionar dinâmica no modelo é introduzir um termo que registra a capacidade de memória:

$$\tau \frac{d x_i}{d t} = - x_i + \sigma \left(\sum_j w_{ij} x_j \right) + I_i \quad (4.2)$$

O modelo resultante (4.2) é conhecido como modelo aditivo introduzido por Grossberg [Chauke & de Carvalho, 93].

O modelo aditivo pode portanto lembrar sobre o passado através da constante de tempo τ . Biologicamente, τ é aproximadamente igual a 5ms para a equação de membrana celular [Shen, 89]. Porém, várias das aplicações em engenharia necessitam lembrar alguns segundos de sinal.

Aparentemente o cérebro é capaz de modelar a resolução temporal e a profundidade da STM, utilizando unidades de processamento com pequenos tempos constantes [de Vries & Principe, 92]. Existem duas variações do modelo aditivo para estender a capacidade da STM. Elas possuem a realimentação positiva e o retardo explícito respectivamente.

4.3.2 - STM no Modelo Aditivo com Realimentação Positiva

O modelo aditivo que suporta realimentação positiva é caracterizado pela fórmula

$$\tau \frac{d x_i}{d t} = - x_i + \sigma (net_i) + I_i + kx_i \quad (4.3)$$

onde net_i é uma representação de $\sum_j w_{ij} x_j$ e k é uma constante positiva. A equação (4.3) é conhecida como modelo de "Auto-Excitação".

Simplificando (4.3), obtém-se:

$$\left(\frac{\tau}{1 - k} \right) \frac{d x_i}{d t} = - x_i + \hat{\sigma} (net_i) + \hat{I}_i \quad (4.4)$$

$$\text{Onde } \hat{\sigma}(net_i) = \frac{\sigma(net_i)}{1-k} \quad \text{e} \quad \hat{I}_i = \frac{I_i}{(1-k)}$$

Verifique como o modelo da equação (4.4) possui a forma do modelo aditivo da equação (4.2). Isto é, a equação (4.3) realmente é um modelo aditivo que respeita características do funcionamento biológico. A diferença é que agora existe uma realimentação de memória (kx_i), cujo controle é realizado pelo sistema neuronal. Por exemplo, sabe-se que $\tau = 5$ ms. Se $k = 0.995$, será possível ter memória correspondente a $\tau = 1$ s, onde $\hat{\tau} = \tau/(1-k)$.

Apesar da simplicidade do modelo aditivo com realimentação positiva, existem dois problemas marcantes envolvidos [de Vries & Principe, 92]:

- 1) A nova constante de tempo é muito sensível a k . Por exemplo, o acréscimo de 0,5% em k (i.é, $k = 1$) faz com que o modelo perca a estabilidade. Visto que, para este espaço de tempo é possível que os sinais analisados possuam a característica de variação temporal dos seus parâmetros relevantes. Como neste caso, para um longo período de tempo a auto-excitação poderá levar à instabilidade;
- 2) A segunda desvantagem é dada pelo modelo ainda a ser governado pela dinâmica de primeira ordem. Como resultado, a determinação dos pesos no domínio temporal é limitada a um gradiente com influências dos acontecimentos mais recentes. Isto é, os itens mais recentes carregam um grande peso em relação aos mais antigos. Observe que para uma rede composta de N neurônios, o número de pesos no domínio espacial é $O(N^2)$, enquanto o domínio temporal é governado somente por τ . Isso implica, portanto, numa limitação sobre a estrutura de representação do passado na rede.

4.3.3 - STM por Convolução

O mecanismo de armazenamento de memória a ser apresentado será representado por convoluções temporais.

A função convolução contínua é por definição:

$$y(t) = \int_{-\infty}^{+\infty} x(\tau) h(t - \tau) d\tau \quad (4.5)$$

onde $x(\cdot)$ é o sinal de entrada, $y(\cdot)$ é a saída e $h(\cdot)$ é a função de transferência do sistema. No caso discreto,

$$y(n) = \sum_{k=-\infty}^{+\infty} x(k) h(n - k) \quad (4.6)$$

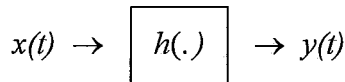


Figura 4.6 : Representação de uma convolução.

Observe que a função de transferência $h(\cdot)$ é responsável por manifestar o comportamento do sistema em relação a alguma entrada. Na aplicação específica de armazenamento de memória, sua função é guardar o comportamento do sinal durante toda a sua duração.

Com isto, o modelo (aditivo) da STM é dado por

$$\tau \frac{d x_i}{dt} = - x_i + \sigma \left(\sum_j \int_0^t w_{ij}(t - s) x_j(s) ds \right) + I_i \quad (4.7)$$

Observe que o modelo de convolução mantém todas as informações temporais desde o início da manifestação do sinal. Por este motivo ele também é conhecido como modelo da STM com retardo.

No caso específico, a função convolução contínua é dada por:

$$n e t_i (t) = \sum_j \int_0^t w_{ij}(t - s) x_j(s) ds \quad (4.8)$$

e a discreta por:

$$n e t_i (t) = \sum_j \sum_{n=0}^t w_{ij}(t - n) x_j(n) ds \quad (4.9)$$

onde $w_{ij}(\cdot)$ é a função de transferência.

Segundo de Vries e Principe [92], existe um amplo suporte biológico no que se refere à substituição dos pesos constantes w pelos variantes no tempo $w(t)$.

A análise sobre a eficiência da utilização de retardo foi verificada por Miller [87]. Suas conclusões levaram-no ao fato de que as conexões axoniais na região do cortex impõe a existência de faixas para condução de sinal com um certo retardo, no intuito de permitir a convergência temporal das informações de um único neurônio. Essas faixas são da ordem de 100 a 200 ms.

Nota-se que o modelo de convolução na formulação aqui apresentada é mais flexível no domínio temporal do que o aditivo de 1ª ordem (visto nos itens anteriores), pois a determinação de pesos correspondentes ao passado não está restrita a um gradiente que pondera com maior valor as entradas mais recentes. Porém, esta flexibilidade é considerada bastante custosa, visto que [de Vries & Principe, 92]:

1 - Análise: o modelo de convolução requer a solução de equações diferenciais funcionais, mais complicada do que a solução de equações diferenciais ordinárias (modelo aditivo);

2 - Carga Computacional: o modelo de convolução N-dimensional exige uma quantidade de operações da ordem $O(N^2T)$ para calcular o próximo estado do conjunto de equações diferenciais funcionais, onde T é o número de passos de tempo necessário para realizar a integral da convolução. Num modelo aditivo N-dimensional, esta grandeza é da ordem $O(N^2)$;

3 - Aprendizagem: no modelo de convolução, os pesos são os parâmetros $W(t)$ variantes no tempo. Com isto, a dimensionalidade dos parâmetros cresce linearmente com o tempo. Fato este prejudicial à aprendizagem de segmentos temporais longos.

O problema 3 acima apresentado é facilmente verificado na equação (4.8), onde o número de pesos w_{ij} é igual a t . Como é muito custoso atender a esta necessidade, a modelagem fica prejudicada, pois o número de parâmetros certamente não acompanhará a profundidade da memória do sistema.

Apesar do modelo de convolução ser poderoso na modelagem biológica, os problemas citados obrigam à procura de uma solução aplicável na prática que ainda preserve o mesmo fundamento. Uma das soluções abordadas restringe por aproximação o núcleo variante no tempo $w(t)$ num conjunto de dimensão fixa de pesos constantes.

4.3.4 - STM com a Aplicação do Filtro ARMA

No Tópico (4.3.3) chegou-se à necessidade de se restringir o núcleo de funções $w(t)$ num conjunto de dimensão fixa de pesos constantes. Uma forma encontrada para efetuar esta restrição foi a aproximação de uma função $w(t)$ pela aplicação do Modelo Regressão ARMA.

Antes de se definir o modelo ARMA (**Autoregressive Moving Average**) serão apresentados os modelos AR (**AutoRegressive**) e o MA (**Moving Average**).

Definição 4.1 : Diz-se que uma série temporal $x(n-1)$, $x(n-m)$ representa a realização de um processo estacionário autorregressivo (AR) de ordem M se ela satisfizer a equação linear

$$x(n) = \sum_{k=1}^M a_k x(n-k) + e(n) \quad (4.10)$$

onde a_1, \dots, a_M são constantes, conhecidas como parâmetros do modelo AR, e $\{e(n)\}$ é um processo estocástico de ruído.

Observe que o presente valor do processo, isto é $x(n)$, é igual a uma combinação linear finita dos valores passados do processo, mais um erro de regressão $e(n)$.

Se a equação (4.10) for reescrita da seguinte forma,

$$e(n) = \sum_{k=0}^M -a_k x(n-k), \quad (4.11)$$

onde $a_0 = -1$, então $e(n)$ é o resultado da convolução discreta e finita da seqüência de entrada $\{x(n)\}$ com a seqüência de parâmetros $\{-a_k\}$.

Definição 4.2: No modelo de média móvel (MA), o processo resultante $\{x(n)\}$ é descrito pela seguinte equação:

$$x(n) = e(n) + b_1 e(n-1) + \dots + b_k e(n-k), \quad (4.12)$$

onde b_1, \dots, b_k são constantes, conhecidas como parâmetros do modelo MA, e $\{e(n)\}$ é um processo estocástico de ruído branco com média zero e variância σ_e^2 .

Neste caso, dada uma realização completa de $\{e(n)\}$, pode-se calcular $x(n)$ através de uma média ponderada de $e(n)$, $e(n-1)$, ..., $e(n-k)$.

Definição 4.3: Diz-se que um modelo é autorregressivo com média móvel (ARMA) se ele for uma combinação dos modelos AR e MA. Ou seja, dado o ruído branco $\{e_n\}$ como entrada do sistema o processo $\{x_n\}$ do modelo ARMA é produzido na sua saída segundo a seguinte equação:

$$x(n) = \sum_{k=1}^M a_k x(n-k) + \sum_{l=1}^N b_l e(n-l) + e(n) \quad (4.13)$$

Verificou-se que no modelo de convolução existe um problema ligado ao grande número de parâmetros das funções peso, pois este número varia linearmente com a evolução do tempo. Explicitamente, o problema está ligado à função convolução. No caso contínuo, observou-se que

$$net(t) = \int_0^t w(t-s)x(s)ds, \quad (4.14)$$

e no caso discreto,

$$net(t) = \sum_{n=0}^t w(t-n)x(n). \quad (4.15)$$

Logo, neste item, investigar-se-á as condições necessárias para que a equação (4.14) possa ser reduzida a um modelo de memória capaz de preservar as características do modelo de convolução, onde o número de parâmetros seja fixo e eles não sejam dependentes do tempo.

Teorema 4.1 [de Vries & Principe, 92]:

A equação integral

$$net(t) = \int_0^t w(t-s)x(s)ds, \quad (4.16)$$

pode ser reduzida a um sistema de equações diferenciais ordinárias K-dimensional, com coeficientes constantes, se (e somente se) $w(t)$ for uma solução de

$$\frac{d^K w}{dt^K}(t) = \sum_{i=0}^{K-1} a_i \frac{d^i w}{dt^i}(t), \quad (4.17)$$

onde a_0, a_1, \dots, a_{k-1} são constantes.

A prova a ser apresentada será apenas no sentido: se $w(t)$ for solução de (4.17) então a equação (4.16) poderá ser reduzida no sistema de equações diferenciais citado no teorema. Note que este é o sentido procurado no enfoque do texto.

Prova:

Considere por hipótese que $w(t)$ satisfaça à equação (4.17) com os coeficientes a_0, a_1, \dots, a_{k-1} constantes.

As condições iniciais para a equação (4.17) são:

$$\hat{w} \equiv \frac{d^i w}{dt^i}(0), \text{ para } i = 0, \dots, k-1 \quad (4.18)$$

$$\text{Define-se que } w_i(t) = \frac{d^i w}{dt^i}(t) \quad (4.19)$$

De (4.19), se pode reescrever a equação (4.17) como um conjunto de K equações diferenciais de primeira ordem:

$$\frac{dw_i}{dt}(t) = w_{i+1}(t), \text{ para } i = 0, \dots, K-2 \quad (4.20)$$

e

$$\frac{dw_{K-1}}{dt}(t) = \sum_{i=0}^{K-1} a_i w_i(t). \quad (4.21)$$

Obs.: A equação (4.21) pode ser verificada através da indução matemática.

Neste ponto introduzir-se-ão algumas variáveis auxiliares com o objetivo de passar (4.17) para (4.16), pois $w(t)$ é uma solução. Com isto, (4.16) será reduzida a um sistema de equações diferenciais ordinárias k-dimensional com coeficientes constantes.

Sejam as variáveis $x_i(t)$ definidas como:

$$x_i(t) \equiv \int_0^t w_i(t-s)x(s)ds, \quad (4.22)$$

$$i=0, \dots, K-1.$$

Observe que a saída do sistema é

$$net(t) = \int_0^t w_0(t-s)x(s)ds = x_0(t). \quad (4.23)$$

Diferenciando-se (4.22) em relação a t e utilizando-se a regra de Leibniz (Anexo I), tem-se:

$$\frac{dx_i}{dt} = \int_0^t \frac{\partial w_i}{\partial t}(t-s)x(s)ds + w_i(0)x(t). \quad (4.24)$$

Substituindo-se as relações de recorrência das equações (4.20) e (4.21) na equação (4.24), obter-se-á $x_i(t)$ recursivamente:

$$\begin{aligned} \frac{dx_i}{dt}(t) &= \int_0^t w_{i+1}(t-s)x(s)ds + \hat{w}_i x(t) \\ &= x_{i+1}(t) + \hat{w}_i x(t), \end{aligned} \quad (4.25)$$

$$i=0, \dots, K-2.$$

e

$$\begin{aligned} \frac{dx_{K-1}}{dt}(t) &= \int_0^t \left(\sum_{i=0}^{K-1} a_i w_i(t-s) \right) x(s) ds + \hat{w}_{K-1} x(t) \\ &= \sum_{i=0}^{K-1} a_i \int_0^t w_i(t-s)x(s)ds + \hat{w}_{K-1} x(t) \\ &= \sum_{i=0}^{K-1} a_i x_i(t) + \hat{w}_{K-1} x(t) \end{aligned} \quad (4.26)$$

Portanto, desta forma, demonstra-se o teorema proposto. Ou seja, $net(t)$ é uma solução do sistema de equações diferenciais dado por (4.25) e (4.26).

O modelo ARMA para representar a STM é a combinação das equações (4.23), (4.25) e (4.26). A memória do modelo encontra-se nas variáveis de estado $x_i(t)$.

É interessante observar a relação entre os parâmetros do modelo ARMA e o modelo de convolução. Os parâmetros autorregressivos a_i são os coeficientes da relação de recorrência da equação (4.17), e os parâmetros MA \hat{w}_i são iguais as condições iniciais da mesma equação.

O próximo teorema esclarece mais detalhadamente como são o comportamento e a forma da família de soluções do sistema de equações descrito na equação (4.17).

Teorema 4.2 :

As soluções do sistema $\frac{d^K w}{dt^K}(t) = \sum_{i=0}^{K-1} a_i \left(\frac{d^i w}{dt^i} \right)(t)$ podem ser escritas como uma combinação linear das funções $t^{i_j} e^{\mu_j t}$, onde $1 \leq j \leq m$, $0 \leq i_j < K_j$ e $\sum_{j=1}^m K_j = K$.

No Teorema 4.2 μ_j 's são os autovalores do sistema. Em particular, os μ_j 's são as soluções da equação característica de (4.17).

$$\mu^{K_j} - \sum_{i=0}^{K_j-1} a_i \mu^i = 0 \quad (4.27)$$

Na equação (4.27), μ refere-se ao número de diferentes autovalores e K_j a multiplicidade dos autovalores μ_j . As funções $t^{i_j} e^{\mu_j t}$ são as autofunções de (4.17), e j enumera os vários modos do sistema [de Vries & Principe, 92]. Este Teorema é provado em vários livros textos clássicos de equações diferenciais ordinárias, como por exemplo em [Braun, 83].

Notas:

- 1 - A análise dos autovalores de um sistema permite descrever o seu comportamento de estabilidade.
- 2 - As autofunções descrevem as formas das soluções do sistema, ou seja, as possíveis trajetórias do sistema.
- 3 - Os modos do sistema são nomeações dadas aos autovalores e as autofunções associados. Geralmente referencia-se a um modo do sistema por um valor inteiro. Por exemplo: 1º modo, 2º modo, ..., k-ésimo modo.

Com os resultados obtidos até agora, verifica-se que o modelo ARMA é estruturalmente equivalente ao modelo de convolução, obviamente considerando-se as restrições impostas. O próximo passo a ser analisado será a dinâmica do ARMA para o armazenamento de memória.

Na visão "top-down" do modelo ARMA encontram-se os esquemas globais de recorrência para a representação do passado, causadores de problemas no controle da estabilidade do sistema. Segundo [de Vries & Principe, 92], os modelos de redes neuronais que utilizam o ARMA não o implementam por completo. E ainda, o problema se agrava quando os parâmetros AR a_i são adaptativos. Alguns modelos que utilizam subestruturas do ARMA e que conseguiram fácil controle de estabilidade são o "Feedforward Tapped Delay Line" - um modelo MA - e o "The First Order Autoregressive Model" (ou "Context Unit") - um modelo AR de primeira ordem [de Vries & Principe, 92].

A proposta do modelo Gamma é exatamente a de permitir o fácil controle da estabilidade preservando as características estruturais do ARMA e do modelo de convolução. O Gamma é uma variação do ARMA que implementa um modelo de memória em uma cascata de estruturas localmente recursivas.

4.3.5 - STM com a Aplicação do Filtro Gamma

Como o Modelo Gamma é uma aproximação do ARMA, então é razoável que ele possua um subconjunto de funções que satisfaça a equação (4.17). Ou melhor, que este subconjunto seja uma classe de funções do conjunto de todas as combinações lineares das autofunções $t^i e^{-\mu t}$ obtidas no Teorema 4.2.

Portanto, considere

$$w(t) = \sum_{i=1}^K w_i g_i^\mu(t), \quad (4.28)$$

onde

$$g_i(t) \equiv \frac{\mu^i}{(i-1)!} t^{i-1} e^{-\mu t}, \quad (4.29)$$

$$i = 1, \dots, K \quad \text{e} \quad \mu > 0.$$

Observe que $g_i(t)$ é uma combinação linear das autofunções do sistema. Logo $g_i(t)$ é uma solução para a equação (4.17).

O nome gamma dado ao modelo é originário da Função Gamma $\Gamma(x) \equiv \int_0^\infty t^{x-1} e^{-t} dt$, pelo fato das funções $g_i(t)$ serem os seus integrandos.

Comparando-se as funções $g_i(t)$ com as soluções da equação (4.17), nota-se que os autovalores no modelo gamma são iguais, isto é, $\mu_j = \mu$. Daí, [de Vries & Principe, 92] concluem que as funções $g_i(t)$ são as autofunções da seguinte relação de recorrência:

$$\left(\frac{d}{dt} + \mu\right)^K g(t) = 0 \quad (4.30)$$

Torna-se neste instante possível gerar um modelo ARMA para representar $net(t)$ considerando-se a função $w(t)$ descrita na equação (4.28). Este procedimento é bem parecido com a construção utilizada na prova do Teorema 4.1.

De início, escreve-se $g_i(t)$ como um conjunto de equações diferenciais de primeira-ordem:

$$\frac{dg_1(t)}{dt} = -\mu g_1(t) \quad (4.31)$$

$$\frac{dg_i(t)}{dt} = -\mu g_i(t) + \mu g_{i-1}(t), \quad (4.32)$$

$i = 2, \dots, K$.

Substituindo-se a função $w(t)$ de (4.28) em (4.15), tem-se:

$$\begin{aligned} net(t) &= \int_0^t \left(\sum_{i=1}^K w_i g_i(t-s) \right) x(s) ds \\ &= \sum_{i=1}^K w_i x_i(t), \end{aligned} \quad (4.33)$$

onde

$$x_i(t) = \int_0^t g_i(t-s) x(s) ds, \quad (4.34)$$

$i = 1, \dots, K$

são as variáveis de estado gamma. Estas variáveis guardam vestígios de memória dos estados neuronais $x(t)$.

O cálculo para obtenção das variáveis $x_i(t)$ é recursivo, visto que:

$$\frac{dx_i(t)}{dt} = \int_0^t \frac{\partial g_i}{\partial t}(t-s) x(s) ds + g_i(0) x(t) \quad (4.35)$$

Mas $g_i(0) = 0$, para $i \geq 2$ e $g_1(0) = \mu e^{-\mu 0} = \mu$, logo

$$\begin{aligned} \frac{dx_i}{dt}(t) &= \int_0^t (-\mu g_i(t-s) + \mu g_{i-1}(t-s)) x(s) ds \\ &= -\mu x_i(t) + \mu x_{i-1}(t), \end{aligned} \tag{4.36}$$

$i = 1, \dots, K$

onde, por definição, $x_0(t) \equiv x(t)$.

As condições iniciais para a equação (4.36), obtidas através da equação (4.34) onde $x_i(0) = g_i(0) x(0)$, são:

$$\begin{aligned} x_0(0) &= x(0), \quad x_1(0) = \mu x(0) \\ x_i(0) &= 0, \quad \text{para } i = 2, \dots, K. \end{aligned} \tag{4.37}$$

Portanto, quando $w(t)$ satisfizer a condições da equação (4.28) então $net(t)$ poderá ser obtido pelo sistema de equações diferenciais ordinárias K -dimensional descrito pela equação (4.36).

Mostrou-se até agora a existência de uma combinação linear das funções gamma que é solução da função $w(t)$. É importante saber então se a seleção arbitrária de funções gamma afeta a generalidade de $w(t)$, e quais são as distorções causadas por esta seleção.

Teorema 4.3 [de Vries & Principe, 92]:

Para todo $w(t) \in L_2[0, \infty]$, para todo $\varepsilon > 0$, existe um conjunto de parâmetros K e w_i , $i = 1, \dots, K$, tal que

$$\int_0^\infty \left| w(t) - \sum_{i=1}^K w_i g_i(t) \right|^2 dt < \varepsilon. \tag{4.38}$$

A prova deste teorema é baseada na completude dos polinômios de Langerre e pode ser encontrada em [Szego, 39].

O termo $L_2[0, \infty]$ é o conjunto de todas as funções complexas f em $0 \leq t \leq \infty$ que são Lebesgue mensuráveis e que $\int_0^\infty |f| dt < \infty$, onde a integral é a de Lebesgue.

Observe que a integral da equação (4.38) permite fazer uma aproximação tão fina quanto se queira. Ou seja, uma seleção arbitrária de funções gamma permite ainda a convergência da aproximação para $w(t)$.

Teorema 4.4 [de Vries & Principe, 92]:

A memória de convolução descrita por

$$net(t) = \int_0^t w(t-s) x(s) ds, \quad (4.39)$$

pode ser aproximada de forma arbitrária e fechada pelo seguinte sistema:

$$net(t) = \sum_{i=1}^K w_i x_i(t), \quad (4.40)$$

onde $x_0(t) = x(t)$, e

$$\frac{dx_i}{dt}(t) = -\mu x_i(t) + \mu x_{i-1}(t), \quad (4.41)$$

$i = 1, \dots, K$ e $\mu > 0$

(as condições iniciais são como especificadas em (4.37)).

NOTA: A equação (4.41) é considerada como Memória Gamma, utilizada para representar a estrutura de retardo.

Para o caso discreto, [de Vries & Principe, 92] escolheram a diferença de primeira ordem como aproximação da derivada em (4.41), isto é:

$$\frac{dx_i}{dt}(t) \approx x_i(t+1) - x_i(t). \quad (4.42)$$

A aplicação de (4.42) leva às seguintes relações de recorrências para a memória gamma (temporal) discreta:

$$\begin{aligned} x_0(t) &= x(t) \\ x_i(t) &= (1 - \mu) x_i(t-1) + \mu x_{i-1}(t-1), \\ & i = 1, \dots, K. \end{aligned} \quad (4.43)$$

Em [Principe & de Vries, 93], as propriedades da estrutura da memória gamma são analisadas detalhadamente, e conclui-se que no caso discreto o sistema é estável para $0 < \mu < 2$. Mostra-se que, para uma estrutura de K-ésima ordem, a profundidade da memória é aproximadamente igual a $D = K/\mu$ e a resolução igual a $R = \mu$. Portanto, para uma memória gamma de K-ésima ordem, há uma importante relação que caracteriza o equilíbrio entre a profundidade e a resolução da memória:

$$K = D \times R. \tag{4.44}$$

Como μ é um parâmetro adaptativo, as características de memória podem ser otimizadas em relação a um índice de eficiência.

O Modelo Neuronal Gamma também é um modelo aditivo onde os neurônios têm a capacidade de armazenar seus históricos de ativação numa estrutura de memória gamma adaptativa de ordem K e resolução μ_i [de Vries & Principe, 92]. A Figura 4.7 apresenta um esquema do Modelo Neuronal Gamma Aditivo.

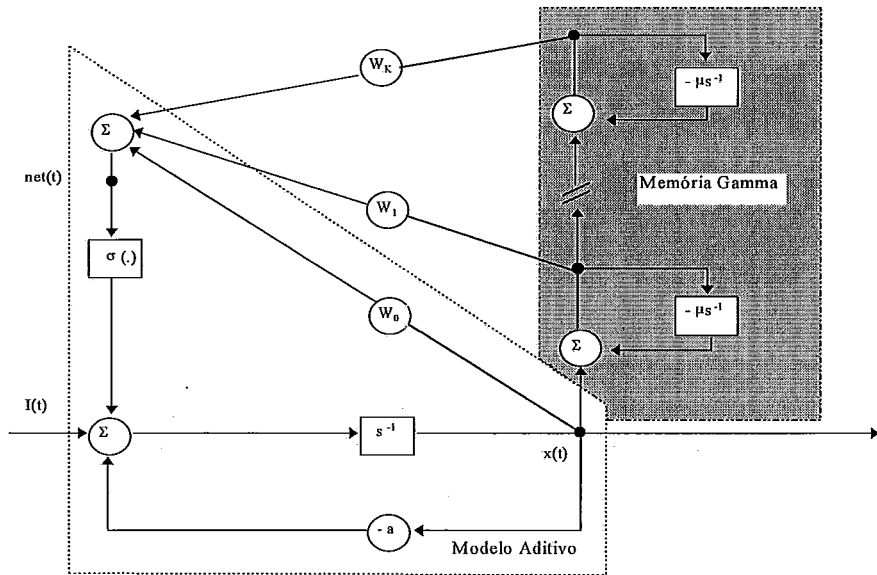


Figura 4.7: Modelo Neuronal Gamma Aditivo

V - Modelo de um Neurônio da Estrutura CELR para Padrões Temporais

5.1 - Introdução

Neste capítulo discute-se como um neurônio da estrutura CELR pode ser expandido para suportar características temporais dinâmicas. Considerou-se como modelo base aquele definido no Capítulo 3 pela equação (3.16) e a expansão foi fruto da introdução da função convolução como responsável pelo acúmulo temporal de informação.

O modelo descrito pela equação (3.16) é completo quanto à representação das interações excitatórias e inibitórias. Uma restrição deste modelo, permitindo apenas interações inibitórias e a auto excitação, é um caso particular do seguinte sistema:

$$\dot{x}_i = a_i(x_i) \left[b_i(x_i) - \sum_{k=1}^n c_{ik} d_k(x_k) \right] \quad (5.1)$$

onde a matriz $C = (c_{ij})$ é simétrica.

Cohen e Grossberg [83] provaram que (5.1) é um sistema estável. A partir do uso da Teoria de Liapunov, definiram a função de Liapunov.

$$V(x) = -\sum_{i=1}^n \int_0^{x_i} b_i(\xi_i) d'_i(\xi_i) d\xi_i + \frac{1}{2} \sum_{j,k=1}^n c_{jk} d_j(x_j) d_k(x_k) \quad (5.2)$$

para obterem o resultado desejado.

No Tópico 5.2 é apresentado o modelo proposto de um neurônio da estrutura CELR para padrões temporais. No Tópico 5.3 introduz-se a Teoria de Liapunov para facilitar o entendimento da análise da estabilidade do modelo STM-CELR temporal (Tópico 5.4).

5.2 - Modelo da STM-CELR Temporal

No Capítulo 3 obteve-se a formulação completa de uma STM da estrutura CELR representada por

$$\begin{aligned} \dot{x}_i(t) = & -A_i x_i(t) + (B_i - x_i(t)) \left[I_i + w_{ii}(t) f_i(x_i(t)) \right] \\ & - (x_i(t) + C_i) \left[J_i + \sum_{j=1}^n D_{ij}(t) g_j(x_j(t)) \right] \end{aligned} \quad (5.4)$$

onde ocorrem apenas a auto-excitação e as interações inibitórias.

No Tópico 4.2.2 foi abordado como concentram-se informações de padrões temporais nas funções convolução. A transformação do modelo (5.4) para o espaço temporal através de funções convolução resulta no seguinte sistema:

$$\begin{aligned} \dot{x}_i(t) = & -A_i x_i(t) + (B_i - x_i(t)) \left[I_i + \zeta(w_{ii}(t), f_i(x_i(t))) \right] \\ & - (x_i(t) + C_i) \left[J_i + \sum_{j=1}^n \zeta(D_{ij}(t), g_j(x_j(t))) \right] \end{aligned} \quad (5.5)$$

onde $\zeta(\alpha(t), \beta(t))$ é a convolução das funções temporais $\alpha(t)$ e $\beta(t)$.

Para atender o objetivo de concentrar informações temporais numa STM é necessário identificar quais são os pontos do modelo que podem canalizar as informações dos padrões de entrada e suas reverberações, como também pontos que representem o conhecimento temporal dos padrões já aprendidos. Ao aplicar-se a função convolução sobre o par de funções (padrões de entrada-reverberação, conhecimento de padrões aprendidos), o resultado será o grau de similaridade entre o par de funções. Esse grau representa o quanto a STM é “sensível” ao padrão de entrada e suas reverberações. Os pares de funções que servem como base para esta medida são $(f_i(x_i(t)), w_{ii}(t))$ e $(g_j(x_j(t)), D_{ij}(t))$.

Aplicando-se a definição da convolução contínua (ver Tópico 4.3.3), a equação (5.5) torna-se:

$$\begin{aligned} \dot{x}_i(t) = & -A_i x_i + (B_i - x_i) \left[I_i + \int_0^t w_{ii}(t-\tau) f_i(x_i(\tau)) d\tau \right] \\ & - (x_i + C_i) \left[J_i + \sum_{j=1}^n \int_0^t D_{ij}(t-\tau) g_j(x_j(\tau)) d\tau \right]. \end{aligned} \quad (5.6)$$

A equação (5.6) é na integra a formulação da STM-CELR temporal. Porém, para garantir a aplicabilidade deste modelo é necessário avaliar a sua estabilidade e convergência. Observe que a análise de (5.6) recai na solução de equações diferenciais funcionais e que o número de sinapses w_{ii} e D_{ij} é da ordem da variação de t . Estes fatos

tornam a aplicabilidade de (5.6) muito incômoda. Mas há uma solução bem mais confortável para estes problemas. No Tópico 4.3.4 encontra-se o seguinte teorema que torna um modelo STM por convolução equivalente a um modelo STM pelo filtro ARMA:

Teorema 5.1:

A equação integral

$$net(t) = \int_0^t w(t-s)x(s)ds$$

pode ser reduzida a um sistema de equações diferenciais ordinárias K-dimensional, com coeficientes constantes, se e somente se $w(t)$ for uma solução de

$$\frac{d^K w}{dt^K}(t) = \sum_{i=0}^{K-1} a_i \frac{d^i w}{dt^i}(t), \quad (5.7)$$

onde a_0, a_1, \dots, a_{K-1} são constantes.

A aplicação do Teorema 5.1 ao modelo da STM dado pela equação (5.6) transformará a convolução em um sistema mais simples que (5.6) e equivalente ao atual modelo no que refere-se à parte biológica e matemática.

Por questões de definição, convenção e simplicidade, sejam (do Teorema 5.1):

$$x_{il}(t) \equiv \int_0^t w_{iil}(t-s)f_i(x_i(s))ds; \quad (5.8)$$

$$x_{jl}(t) \equiv \int_0^t w_{ijl}(t-s)g_j(x_j(s))ds; \quad (5.9)$$

$$w_{ijl}(t) = \frac{d^l w_{ij}}{dt^l}(t), \quad l = 0, \dots, K-1; \quad (5.10)$$

$$\frac{dw_{ijl}}{dt}(t) = w_{ij(l+1)}(t), \quad l = 0, \dots, K-2; \quad (5.11)$$

$$\frac{dw_{ij(K-1)}}{dt}(t) = \sum_{l=0}^{K-1} a_{ijl}w_{ijl}(t), \quad a_{ijl} \text{ constantes.} \quad (5.12)$$

As equações (5.8) a (5.12) são adaptações do Teorema 5.1 ao ambiente da equação (5.6). A modificação mais forte foi definir $w_{ij}(t)$ como $D_{ij}(t)$ para as sinapses inibitórias, porém não há perda de consistência.

Portanto, aplicando o Teorema 5.1 e as equações (5.8) a (5.12) no modelo descrito por (5.6) têm-se:

$$net_i(t) = \sum_{l=0}^{K-1} a_{iil} x_{il}(t) + \sum_{l=0}^{K-1} \hat{w}_{iil} f_i(x_i(t)); \quad (5.13)$$

$$net_j(t) = \sum_{l=0}^{K-1} a_{ijl} x_{jl}(t) + \sum_{l=0}^{K-1} \hat{w}_{ijl} g_j(x_j(t)); \quad (5.14)$$

onde \hat{w}_{iil} e \hat{w}_{ijl} são as condições iniciais do sistema de equações diferenciais ordinárias do Teorema 5.1. As equações (5.13) e (5.14) são modelos ARMA cujos parâmetros AR são as constantes a_{iil} e a_{ijl} , e os parâmetros MA são condições iniciais do Teorema 5.1 (\hat{w}_{iil} e \hat{w}_{ijl}).

De (5.13) e (5.14), (5.6) torna-se:

$$\begin{aligned} \dot{x}_i(t) = & -A_i x_i + (B_i - x_i) \left[I_i + net_i(t) \right] \\ & - (x_i + C_i) \left[J_i + \sum_{j=1}^n net_j(t) \right], \end{aligned} \quad (5.15)$$

ou seja,

$$\begin{aligned} \dot{x}_i(t) = & -A_i x_i + (B_i - x_i) \left[I_i + \sum_{l=0}^{K-1} (a_{iil} x_{il}(t) + \hat{w}_{iil} f_i(x_i(t))) \right] \\ & - (x_i + C_i) \left[J_i + \sum_{j=1}^n \sum_{l=0}^{K-1} (a_{ijl} x_{jl}(t) + \hat{w}_{ijl} g_j(x_j(t))) \right]. \end{aligned} \quad (5.16)$$

Portanto, pelo Teorema 5.1, os sistemas (5.6) e (5.16) são equivalentes.

Se a equação (5.16) puder ser reescrita de forma que se transforme em algo similar a equação (5.1), atendendo as condições desta equação, então poderá provar-se pela aplicação do Teorema de Cohen-Grossberg [83] que (5.16) converge e é estável. Logo, fazendo $y_i = x_i + C_i$ em (5.16), tem-se que $\dot{y}_i = \dot{x}_i$ e portanto:

$$\begin{aligned} \dot{y}_i(t) = & -A_i (y_i - C_i) + \\ & + (B_i + C_i - y_i) \left[I_i + \sum_{l=0}^{K-1} (a_{iil} y_{il}(t) + \hat{w}_{iil} f_i(y_i(t) - C_i)) \right] \\ & - y_i \left[J_i + \sum_{j=1}^n \sum_{l=0}^{K-1} (a_{ijl} y_{jl}(t) + \hat{w}_{ijl} g_j(y_j(t) - C_i)) \right], \end{aligned} \quad (5.17)$$

onde

$$y_{il}(t) = \int_0^t w_{iil}(t-s) f_i(y_i(s) - C_i) ds \quad (5.18)$$

$$e \quad y_{jl}(t) = \int_0^t w_{ijl}(t-s) g_j(y_j(s) - C_j) ds \quad (5.19)$$

Desta forma:

$$\begin{aligned} \dot{y}_i(t) = & A_i C_i - (A_i + J_i) y_i + \\ & + (B_i + C_i - y_i) \left[I_i + \sum_{l=0}^{K-1} (a_{iil} y_{il}(t) + \hat{w}_{iil} f_i(y_i(t) - C_i)) \right] \\ & - y_i \sum_{j=1}^n \sum_{l=0}^{K-1} (a_{ijl} y_{jl}(t) + \hat{w}_{ijl} g_j(y_j(t) - C_j)) . \end{aligned} \quad (5.20)$$

ou seja:

$$\begin{aligned} \dot{y}_i(t) = & y_i \left[\frac{A_i C_i}{y_i} - (A_i + J_i) + \right. \\ & + \frac{(B_i + C_i - y_i)}{y_i} \left[I_i + \sum_{l=0}^{K-1} (a_{iil} y_{il}(t) + \hat{w}_{iil} f_i(y_i(t) - C_i)) \right] \\ & \left. - \sum_{j=1}^n \sum_{l=0}^{K-1} (a_{ijl} y_{jl}(t) + \hat{w}_{ijl} g_j(y_j(t) - C_j)) \right] . \end{aligned} \quad (5.21)$$

Logo:

$$a_i(y_i) = y_i \quad (5.22)$$

$$\begin{aligned} b_i(t) = & \frac{1}{y_i} \left[A_i C_i - (A_i + J_i) y_i + \right. \\ & \left. + (B_i + C_i - y_i) \left[I_i + \sum_{l=0}^{K-1} (a_{iil} y_{il}(t) + \hat{w}_{iil} f_i(y_i(t) - C_i)) \right] \right] \end{aligned} \quad (5.23)$$

$$c_{ijl} = a_{ijl} \quad (5.24)$$

$$d_j(y_j(t)) = g_j(y_j(t) - C_j) \quad (5.25)$$

e

$$\dot{y}_i(t) = a_i(y_i) \left[b_i(y_i) - \sum_{j=1}^n \sum_{l=0}^{K-1} (c_{ijl} y_{jl}(t) + \hat{w}_{ijl} d_j(y_j(t))) \right]. \quad (5.26)$$

Comparando-se os sistemas (5.1) e (5.26) é fácil verificar que não são equivalentes. Em (5.26) há um somatório em l , $l = 0, \dots, K-1$, que diferencia de (5.1). Além disto, os argumentos dos somatórios de (5.26) não podem, a priori, sofrer as mesmas restrições de positividade e simetria impostas para os coeficientes de (5.1). Com o objetivo de esclarecer (5.1), segue-se o Teorema de Cohen-Grossberg:

Teorema 5.2 (Cohen-Grossberg):

Sistemas da forma

$$\dot{x}_i(t) = a_i(t) \left[b_i(t) - \sum_{j=1}^n c_{ij} d_j(x_j) \right] \quad (5.27)$$

aditem a função de Liapunov para a análise global do equilíbrio.

$$V(x) = -\sum_{i=1}^n \int_0^{x_i} b_i(\xi_i) d'_i(\xi_i) d\xi_i + \frac{1}{2} \sum_{j,k=1}^n c_{jk} d_j(x_j) d_k(x_k) \quad (5.28)$$

se $C = (c_{ij})$ for uma matriz de coeficientes e as funções a_i , b_i e d_j obedecerem as seguintes condições

:

1) Simetria

$$c_{ij} = c_{ji}; \quad (5.29)$$

2) Positividade

$$a_i(x)_i \geq 0; \quad (5.30)$$

3) Monotonicidade

$$d'_j(x_j) \geq 0. \quad (5.31)$$

O Teorema 5.2 é muito poderoso, pois garante a estabilidade do sistema (5.27). Apesar do sistema (5.26) não encaixar-se no Teorema Cohen-Grossberg, no Tópico 5.4 será apresentada a prova de sua estabilidade e convergência. Considerou-se a Teoria de

Liapunov para a análise de (5.26) e foram definidos e provados novos teoremas referentes ao Teorema Cohen-Grossberg para alcançar o objetivo deste trabalho.

5.3 - Introdução à Teoria de Liapunov.

A Teoria de Liapunov foi desenvolvida para analisar o comportamento de sistemas de equações diferenciais cujas soluções não sejam necessariamente conhecidas. A teoria aqui estudada é conhecida como Segundo Método de Liapunov ou Método Direto de Liapunov.

Pelo método direto de Liapunov obtêm-se conclusões sobre a estabilidade ou instabilidade de um ponto crítico do sistema a partir de uma função auxiliar (função de Liapunov) aplicada sobre as trajetórias dinâmicas do sistema de equações diferenciais. A técnica é muito poderosa e fornece um tipo de informação global; como por exemplo, uma estimativa da extensão da região de estabilidade assintótica de um ponto crítico.

O segundo método de Liapunov é uma generalização dos princípios físicos quanto a um sistema conservativo, onde:

- uma posição de repouso será estável se a energia potencial for um mínimo local, de outra forma será instável; e
- a energia total será uma constante durante qualquer movimento.

O objetivo principal deste tópico é apresentar geometricamente como os teoremas solucionam os problemas de estabilidade. Não serão estudados todos os teoremas aplicados no próximo tópico (5.4), porém as noções aqui adquiridas deverão ser suficientes para o entendimento das provas lá necessárias.

As condições de estabilidade de um sistema dinâmico podem ser:

- estável;
- assintoticamente estável;
- instável.

Se um sistema dinâmico for representado pela equação

$$\dot{x} = D(x, t) \tag{5.32}$$

então o estado do sistema no instante t será $x(t)$. Se $x(t) \in \mathfrak{R}^n$ então $x(t)$ produzirá uma trajetória g no espaço \mathfrak{R}^n . O comportamento da trajetória g define a condição de

estabilidade. A Figura 5.1 ilustrará geometricamente as definições, a seguir, destas condições:

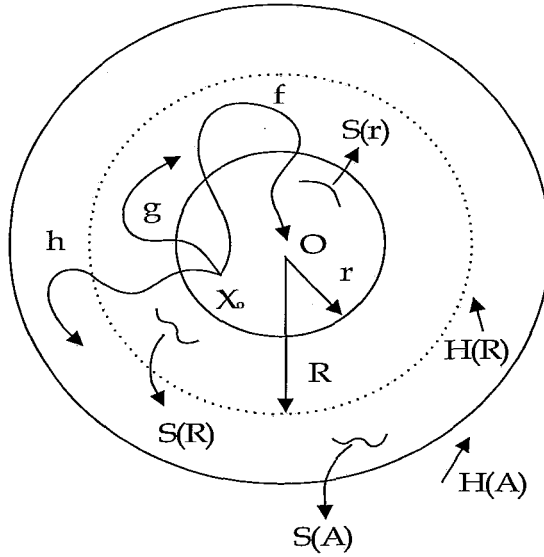


Figura 5.1: Visão geométrica das condições de estabilidade.

Definição 5.1: Sistema Estável

Sejam $R, A, r \in \mathfrak{R}^+$ raios de regiões esféricas do espaço \mathfrak{R}^n tal que $R < A$ e $r \leq R$. Se uma trajetória g iniciar em $x_0 \in S(r)$ (região esférica de raio r) então g permanecerá em $S(r)$, para algum $r \leq R$. Isto é, uma trajetória começando em $S(r)$ nunca alcançará a fronteira (ou borda) $H(R)$ da região esférica $S(R)$.

Definição 5.2: Sistema Assintoticamente Estável

O sistema deve ser estável e toda trajetória f iniciada dentro de alguma região esférica $S(r)$ tende a origem com o crescimento indefinido do tempo.

Definição 5.3: Sistema Instável

Para algum R e algum r , mesmo que r seja infinitamente pequeno, sempre há numa região esférica $S(r)$ um ponto x tal que a trajetória h , ao longo de x , alcança a fronteira $H(R)$ de $S(R)$.

Definição 5.4: Função Definida Positiva

Sejam $x \in \mathfrak{R}^n$ e $V(x)$ uma função escalar. $V(x)$ é dita ser definida positiva se:

- $V(x)$ é contínua, com as 1^{as} derivadas parciais contínuas numa região aberta Ω que contém a origem;
- $V(0) = 0$;
- Para $x \neq 0$, $x \in \Omega$, tem-se que $V(x)$ é positiva.

Se a função V for positiva definida, então ela possui um gradiente. Se a função escalar V for aplicada a uma trajetória g do sistema (5.32), então

$$\dot{V}(x) = \langle \dot{x}, \text{grad}V \rangle, \quad (5.33)$$

onde a função $\langle \cdot, \cdot \rangle$ é o produto escalar e $\text{grad} V$ é o gradiente da função V .

Definição 5.5: Função de Liapunov

Se V for uma função definida positiva e $\dot{V} \leq 0$ em Ω , então V é dito ser uma função Liapunov.

Observe que uma função de Liapunov é monotonamente decrescente. Este fato é essencial para a identificação do tipo de estabilidade de um sistema dinâmico. Porém, antes de serem apresentados os teoremas de estabilidade do método direto de Liapunov, convém ressaltar que a vizinhança Ω da origem do espaço \mathfrak{R}^n de x é fielmente refletida da imagem de V aplicada em Ω . Isto é, para cada ponto $x \in \Omega$ existe um único ponto correspondente em $V(\Omega)$ e vice-versa. Além disto, esta correspondência é contínua em cada direção. O termo matemático que expressa esta afirmação é: a correspondência é homeomórfica.

Teorema 5.3: Estabilidade

Se houver em alguma vizinhança Ω da origem uma função de Liapunov $V(x)$, então a origem é estável.

Teorema 5.4: Estabilidade Assintótica

Se $-\dot{V}$ for definida positiva em Ω , então a estabilidade é assintótica.

As provas dos Teoremas 5.3 e 5.4 por interpretação geométrica permitem observar o quanto poderoso é o método direto Liapunov.

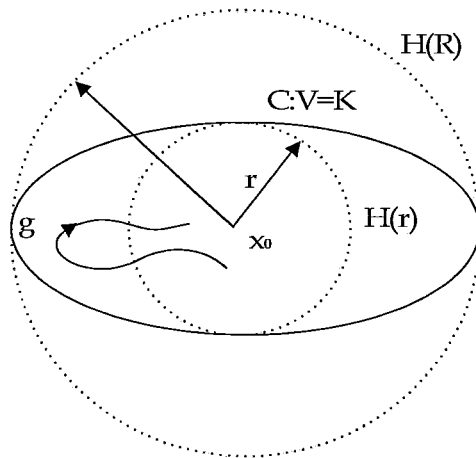


Figura 5.2: interpretação do Teorema 5.3.

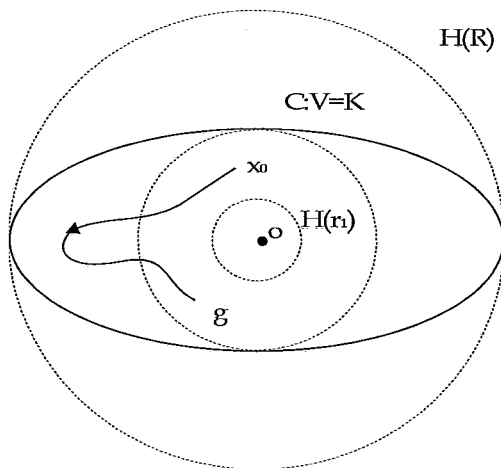


Figura 5.3: interpretação do Teorema 5.4.

Verificação da validade do Teorema 5.3:

As verificações dos Teoremas 5.3 e 5.4 serão realizadas na dimensão $n = 2$, não afetando suas características de generalização.

Na Figura 5.2, $V(x) = K$ é uma curva de nível C da superfície determinada pela função de Liapunov, representada pela linha fechada, e a fronteira $H(R)$ do círculo $S(R)$ é representada pela linha pontilhada. Então, dado que $R < A$, pode-se encontrar uma constante K tal que a curva C definida por $V(x) = K$ esteja completamente delimitada pela fronteira $H(R)$ e, por conseqüência, existe um $r > 0$ tal que $H(r)$ está delimitada por C . Seja g uma trajetória cujo ponto inicial x_0 está no interior de $S(r)$. Em x_0 tem-se que $V(x_0) < K$. Portanto, desde que V seja não-crescente ao longo da trajetória, ou seja $V(x_1) \leq V(x_2)$ se $x_2 < x_1$, então g nunca alcançará C e, desta forma, nunca alcançará $H(R)$. Logo, qualquer trajetória começando no interior de $S(r)$ deverá estar completamente no interior de $S(R)$, e isto significa estabilidade. Dado R , a existência de r é clara. Já que V é positiva e contínua no interior de $H(R)$, segue-se da compacticidade de $H(R)$ que V tem um mínimo positivo K no interior de $H(R)$; isto é, existe x tal que $V(x) \geq K$ no interior de $H(R)$. Como $V(x)$ é contínua e anula-se na origem, portanto para r suficientemente pequeno, $V(x) < K$ para x no interior de $S(r)$. Segue-se então a validade do Teorema 5.3.

Observações:

• Cobertura:

Seja $\mathcal{A} = \{G_i\}$ uma classe de subconjuntos de X tal que $A \subset \bigcup_i G_i$ para algum $A \subset X$.

\mathcal{A} é uma cobertura de A (uma cobertura aberta se cada G_i for aberto). Além disto, se uma subclasse finita de \mathcal{A} for também cobertura de A , isto é, se

$\exists G_{i_1}, \dots, G_{i_m} \in \mathcal{A}$ talvez $A \subset G_{i_1} \cup \dots \cup G_{i_m}$ então \mathcal{A} é redutível a uma cobertura finita, ou contém uma subcobertura finita.

- **Compaticidade:**

Um subconjunto A de um espaço topológico X é compacto se toda cobertura aberta de A for redutível a uma cobertura finita.

Verificação da validade do Teorema 5.4:

Considere a Figura 5.3 como ilustração. Pela hipótese do Teorema 5.4 ($-\dot{V}$ positiva definida), $V(x)$ decresce ao longo de g , ou seja se $x_1 < x_2$ então $V(x_1) > V(x_2)$. Será que V poderá estabilizar-se em algum $\ell > 0$? Certamente não, pois se $-\dot{V}$ tendesse a zero fora de uma região $H(r_i)$, que é marcada pelo fato de ser definida positiva, então teria um mínimo positivo m na região anular S_1^R . Porém, $V(x)$ tende a zero ao longo de g e isto acontece somente se g tender a origem. Logo ela é assintoticamente estável.

A determinação de uma função de Liapunov para um sistema dinâmico nem sempre é uma tarefa fácil. Na Definição 5.5, uma condição para uma função V ser de Liapunov é que ela seja definida positiva. Cohen e Grossberg [83] utilizaram outra definição de função de Liapunov, onde não há necessidade de provar-se que a função V seja definida positiva.

Definição 5.6: Função de Liapunov [Cohen & Grossberg]

Seja V uma função de \mathcal{R}^n em \mathcal{R} . Seja g uma trajetória sobre o sistema dinâmico $\dot{x} = D(x, t)$. Se $\dot{V} \leq 0$ sobre todas as trajetórias admissíveis g então V é uma função de Liapunov.

Definição 5.7: Conjunto Invariante

Um conjunto invariante G é caracterizado pela propriedade de que se um ponto x_0 estiver em G então sua trajetória completa (a frente e atrás) estará em G .

Os Teoremas padrões sobre funções de Liapunov implicam que cada trajetória converge para o maior conjunto invariante M contido em E onde $E = \left\{ x \in \mathfrak{R}^n / \dot{V}(x) = 0 \right\}$.

Teorema 5.5:

Seja V uma função de Liapunov no domínio Ω_ℓ onde $V(x) < \ell$. Seja Ω_ℓ limitado. Seja E o conjunto de todos os pontos de Ω_ℓ onde $\dot{V}(x) = 0$, e seja M o maior conjunto invariante em E . Então, toda solução $x(t)$ em Ω_ℓ tende a M quando $t \rightarrow \infty$.

Não será apresentada a prova deste teorema. Pode-se encontrá-la na maioria dos livros textos sobre o método direto de Liapunov, como por exemplo La Salle e Lefschetz [61].

Neste tópico foi apresentado o conhecimento mínimo necessário para o entendimento das provas de estabilidade e convergência do modelo proposto nesta tese.

5.4 - Análise da Estabilidade da Dinâmica e da Convergência do Modelo STM-CELR Temporal

Neste tópico serão estudadas as condições de estabilidade do modelo (5.26), proposto no Tópico 5.2, e introduzidas as restrições necessárias para garantir-se estabilidade e convergência sem descaracterizar a essência do modelo.

Considere o sistema dinâmico do modelo proposto no Tópico 5.2:

$$\dot{y}_i(t) = a_i(y_i) \left[b_i(y_i) - \sum_{j=1}^n \sum_{l=0}^{K-1} (c_{ijl} y_{jl}(t) + \hat{w}_{ijl} d_j(y_j(t))) \right]. \quad (5.34)$$

Seja a função V uma candidata a função de Liapunov para as trajetórias do sistema (5.34):

$$V(y) = - \sum_{i=1}^n \int_0^{y_i} b_i(\xi_i) d\xi_i + \sum_{j,p=1}^n \sum_{\ell=0}^{K-1} (c_{jpl} \int_0^{t} y_{p\ell}(s) ds + \hat{w}_{jpl} \int_0^{y_j} d_j(\xi_j) d\xi_j). \quad (5.35)$$

Considere as seguintes hipóteses sobre o sistema (5.34), onde y sempre estará na região \mathfrak{R}_+^n :

a) Coeficientes ARMA:

Os coeficientes c_{ijl} e \hat{w}_{ijl} são constantes e não-negativos;

b) Continuidade:

$a_i(\xi)$ é contínua para $\xi \geq 0$ e a função $b_i(\xi)$ é contínua para $\xi > 0$;

c) Positividade:

$a_i(\xi) > 0$ para $\xi > 0$ e $d_i(\xi) \geq 0$ para $\xi \in (-\infty, \infty)$;

d) Suavidade e Monotonicidade:

$d_i(\xi)$ é diferenciável e monotonamente não-decrescente para $\xi \geq 0$.

Observe que estas hipóteses estão de acordo com a proposta alcançada em (5.26).

De acordo com a Definição 5.6, para provar que $V(y)$ é uma função de Liapunov é necessário mostrar que os dados iniciais positivos geram trajetórias limitadas e positivas de (5.34) (trajetórias admissíveis) e que $\dot{V}(y) \leq 0$ sobre todas estas trajetórias.

Antes de provar que dados iniciais positivos levam a trajetórias admissíveis, torna-se necessária a introdução de mais hipóteses, pois b_i pode ser ilimitado quando $y_i \rightarrow 0^+$.

Lema 5.1: (Limitação e Positividade)

Limitação:

Para cada $i = 1, \dots, n$, suponha que

$$\limsup_{\xi \rightarrow \infty} [b_i(\xi) - \hat{w}_{iil} d_i(\xi)] < 0. \quad (5.36)$$

Positividade:

Para cada $i = 1, \dots, n$, suponha que

$$\lim_{\xi \rightarrow 0^+} b_i(\xi) = \infty \quad (5.37)$$

ou que

$$\lim_{\xi \rightarrow 0^+} b_i(\xi) < \infty \quad (5.38)$$

e

$$\int_0^\epsilon \frac{d\xi}{a_i(\xi)} = \infty \quad \text{para algum } \epsilon > 0. \quad (5.39)$$

então qualquer dado inicial positivo gera uma trajetória admissível.

O teorema a seguir, apresentado no Tópico 4.3.4, será utilizado durante a prova do Lema 5.1.

Teorema 5.6: (de Vries & Principe):

Soluções do sistema

$$\frac{d^K w}{dt^K}(t) = \sum_{l=0}^{K-1} a_l \frac{d^l w}{dt^l}(t) \quad (5.40)$$

podem ser escritas como uma combinação linear de funções $t^{r_i} e^{\mu_i t}$, onde

$$1 \leq i \leq m, \quad 0 \leq r_i < R_i \quad e \quad \sum_{i=1}^m R_i = R.$$

Prova do Lema 5.1:

Limitação:

A limitação é provada utilizando o Teorema 5.6 e a equação (5.36). A desigualdade

$$b_i(y_i) - \sum_{j=1}^n \sum_{l=0}^{K-1} (\hat{w}_{ijl} d_j(y_j)) \leq b_i(y_i) - \hat{w}_{iil} d_i(y_i) \quad (5.41)$$

é verdadeira, pois \hat{w}_{ijl} e d_j são não-negativos. Por (5.36), $\lim_{\xi \rightarrow \infty} b_i(\xi) \ll \infty$. Pela hipótese

(d), a função $d_i(\xi)$ é contínua para $\xi \geq 0$. Como $\lim_{\xi \rightarrow \infty} \sup [b_i(\xi) - \hat{w}_{iil} d_i(\xi)] < 0$ então,

por (5.41), $\sum_{j=1}^n \sum_{\ell=0}^{K-1} \hat{w}_{ijl} d_j(y_j)$ é limitado.

Como as soluções do sistema de equações diferentes do Teorema 5.6 são limitadas, pode-se concluir que as variáveis do sistema $w_{ijl}(t) = \frac{d^l w_{ij}}{dt^l}(t)$, $l = 0, \dots, K$, são limitadas. Logo $y_{jl}(t)$ é limitado.

Positividade:

Seja $\lim_{\xi \rightarrow 0^+} b_i(\xi) = \infty$. Na prova da limitação verificou-se que o termo

$\sum_{j=1}^n \sum_{\ell=0}^{K-1} (c_{ijl} y_{j\ell}(t) + \hat{w}_{ijl} d_j(y_j))$ é limitado. Portanto o termo

$b_i(y_i) - \sum_{j=1}^n \sum_{\ell=0}^{K-1} (c_{ijl} y_{j\ell}(t) + \hat{w}_{ijl} d_j(y_j))$ torna-se positivo quando $y_i \rightarrow 0^+$. Como

$a_i(y_i) > 0$ para $y_i > 0$, então $\dot{y}_i > 0$ antes que y_i alcance zero. Portanto y_i nunca alcançará zero.

Sejam $\lim_{\xi \rightarrow 0^+} b_i(\xi) < \infty$ e $\int_0^{\epsilon} \frac{d\xi}{a_i(\xi)} = \infty$ para algum $\epsilon > 0$. Então, para o

primeiro instante $t = T$ tal que $y_i(t) = 0$,

$$-\infty = \int_{y_i(T)}^0 \frac{d\xi}{w_i(\xi)} = \int_T^0 \frac{d}{dt} (y_i(t)) \frac{1}{a_i(y_i(t))}$$

$$= \int_T^0 \left[b_i(y_i(t)) - \sum_{j=1}^n \sum_{\ell=0}^{K-1} (c_{ij\ell} y_{j\ell}(t) + \hat{w}_{ij\ell} d_j(y_j(t))) \right] dt > -\infty. \quad (5.42)$$

Isto é uma contradição. Portanto, $y_{j\ell}(t)$ permanece positivo para qualquer $t \geq 0$. Em outras palavras, qualquer dado inicial positivo gera uma trajetória admissível.

Utilizando-se do fato que dados iniciais positivos geram trajetórias admissíveis, necessita-se apenas mostrar que $\dot{V} \leq 0$ sobre tais trajetórias para, de acordo com a Definição 5.6, verificar-se que V é uma função de Liapunov.

Proposição 5.1: (Função de Liapunov)

A função $V(y)$ satisfaz a

$$\frac{dV}{dt}(y(t)) \leq 0 \quad (5.43)$$

sobre as trajetórias admissíveis.

Prova:

$$V(y) = -\sum_{i=1}^n \int_0^{y_i} b_i(\xi_i) d\xi_i + \sum_{j,p=1}^n \sum_{\ell=0}^{K-1} (c_{jpl} \int_0^{y_{p\ell}} y_{p\ell}(s) ds + \hat{w}_{jpl} \int_0^{y_j} d_j(\xi_j) d\xi_j) \quad (5.44)$$

sabe-se que

$\dot{V}(y) = \langle \nabla V, \dot{y} \rangle$, onde ∇V é o gradiente da função V e a função \langle, \rangle é o produto escalar.

Logo,

$$\dot{V}(y) = \left(\begin{array}{c} -b_1(y_1) + \sum_{p=1}^n \sum_{\ell=0}^{K-1} [c_{1p\ell} y_{p\ell}(t) + \hat{w}_{1p\ell} d_1(y_1)] \\ \vdots \\ -b_n(y_n) + \sum_{p=1}^n \sum_{\ell=0}^{K-1} [c_{np\ell} y_{p\ell}(t) + \hat{w}_{np\ell} d_n(y_n)] \end{array} \right) \cdot \left(\begin{array}{c} \dot{y}_1 \\ \vdots \\ \dot{y}_n \end{array} \right),$$

$$= -\sum_{i=1}^n a_i(y_i) \left[b_i(y_i) - \sum_{j=1}^n \sum_{\ell=0}^{K-1} [c_{ij\ell} y_{j\ell}(t) + \hat{w}_{ij\ell} d_j(y_j)] \right]^2. \quad (5.45)$$

logo, a função $\dot{V}(y) \leq 0$, pois, pela hipótese (c), $a_i(y_i) > 0$. Portanto, V é uma de Liapunov.

Observe que o Teorema 5.5 trata da estabilidade e convergência das trajetórias admissíveis. Mas para aplicá-lo, é necessário mostrar que a função de Liapunov é limitada. O Teorema 5.5 é também conhecido como o Princípio Invariante de La Salle.

Proposição 5.2:

Se as hipóteses do Lema 5.1 ocorrerem, então $V(y)$ (ou uma redefinição dele) é limitado e contínuo nas trajetórias admissíveis.

Prova:

Se (5.38) ocorrer, $\lim_{\xi \rightarrow 0^+} b_i(\xi) < \infty$, então as integrais $\int_0^{y_i} b_i(\xi_i) d\xi_i$ em $V(y)$

limitadas, pois as trajetórias admissíveis y_i são limitadas e b_i é contínua.

$\int_0^t y_{p\ell}(s) ds$ é limitada em $V(y)$, porque $y_{p\ell}(t)$ é limitada, como verificado na

prova do Lema 5.1. O termo $\int_0^{y_j} d_j(\xi_j) d\xi_j$ é limitado, pois a função $d_j(y_j)$ é contínua

de variável limitada. Logo $V(y)$ é limitada.

Se (5.37) ocorrer, $\lim_{\xi \rightarrow 0^+} b_i(\xi) = \infty$, então a integral $\int_0^{y_j} b_i(\xi_i) d\xi_i$ em $V(y)$ poderá

ser substituída por uma integral $\int_{\lambda_i}^{y_j} b_i(\xi_i) d\xi_i$, onde λ_i é uma constante positiva que será

escolhida posteriormente. Tal escolha é possível devido a alguns fatos que ocorrem juntos. Cada d_j é uma função de y_j não-negativa e monotonamente não-decrescente e

$y_{j\ell}$ é limitado para a variação t , onde $0 \leq y_j(t) \leq L_j$ para t suficientemente grande, $j = 1, \dots, n$. Conseqüentemente, existe um L positivo e finito tal que

$$\sum_{j=1}^n \sum_{\ell=0}^{K-1} \left(c_{ij\ell} y_{j\ell}(t) + \hat{w}_{ij\ell} d_j(y_j) \right) \leq L \quad (5.46)$$

sobre todas as trajetórias admissíveis num instante de tempo suficientemente grande.

Como (5.37) ocorre, existe um intervalo $[0, 2\lambda_i]$ tal que

$$b_i(y_i) - \sum_{j=1}^n \sum_{l=0}^{K-1} (c_{ijl} y_{jl}(t) + \hat{w}_{ijl} d_j(y_j(t))) \geq L \quad (5.47)$$

e portanto

$$\dot{y}_i \geq L a_i(y_i) \quad (5.48)$$

em qualquer instante $0 < y_i \leq 2\lambda_i$ sobre qualquer trajetória admissível num instante de tempo suficientemente grande. Como a função a_i é positiva no intervalo $[y_i(T), 2\lambda_i]$ onde $y_i(T) > 0$, então a_i possui um limite inferior neste intervalo. Portanto, por (5.48), se T for escolhido tão grande que (5.46) ocorra para $t \geq T$, então $y_i(t)$ crescerá pelo menos numa taxa linear até exceder λ_i e tenderá a ficar maior após λ_i . Como este argumento

ocorre para qualquer trajetória admissível, a escolha de λ_i na integral $\int_{\lambda_i}^{y_i} b_i(\xi_i) d\xi_i$ é

justificada.

A continuidade segue por inspeção de cada termo de $V(y)$, substituindo a integral

$$\int_0^{y_i} \text{ por } \int_{\lambda_i}^{y_i} \text{ onde for necessário.}$$

Para provar-se a convergência e a estabilidade pelo princípio invariante de La Salle, considere o seguinte teorema.

Teorema 5.7: (Convergência das Trajetórias)

Em qualquer sistema

$$\dot{y}_i(t) = a_i(y_i) \left[b_i(y_i) - \sum_{j=1}^n \sum_{l=0}^{K-1} (c_{ijl} y_{jl}(t) + \hat{w}_{ijl} d_j(y_j(t))) \right] \quad (5.49)$$

Tal que

a) Coeficientes ARMA:

Os coeficientes c_{ijl} e \hat{w}_{ijl} são constantes e não-negativos;

b) Continuidade:

$a_i(\xi)$ é contínua para $\xi \geq 0$ e a função $b_i(\xi)$ é contínua para $\xi > 0$;

c) Positividade:

$a_i(\xi) > 0$ para $\xi > 0$ e $d_i(\xi) \geq 0$ para $\xi \in (-\infty, \infty)$;

d) Suavidade e Monotonicidade:

$d_i(\xi)$ é diferenciável e monotonamente não-decrescente para $\xi \geq 0$.

e) Limitação:

Para cada $i = 1, \dots, n$,

$$\lim_{\xi \rightarrow \infty} \sup [b_i(\xi) - \hat{w}_{iil} d_i(\xi)] < 0. \quad (5.50)$$

f) e ocorre que (Positividade):

Para cada $i = 1, \dots, n$,

$$\lim_{\xi \rightarrow 0^+} b_i(\xi) = \infty \quad (5.51)$$

ou que

$$\lim_{\xi \rightarrow 0^+} b_i(\xi) < \infty \quad (5.52)$$

e

$$\int_0^\epsilon \frac{d\xi}{a_i(\xi)} = \infty \quad \text{para algum } \epsilon > 0. \quad (5.53)$$

então, todas as trajetórias admissíveis aproximam-se para o maior conjunto invariante M contido no conjunto

$$E = \{y \in \mathfrak{R}^n / \dot{V}(y) = 0, y \geq 0\}, \quad (5.54)$$

onde

$$\dot{V}(y) = -\sum_{i=1}^n a_i(y_i) \left[b_i(y_i) - \sum_{j=1}^n \sum_{\ell=0}^{K-1} [c_{ij\ell} y_{j\ell}(t) + \hat{w}_{ij\ell} d_j(y_j)] \right]^2. \quad (5.55)$$

Prova:

Pelas hipóteses do teorema, já foi verificado que a função V é uma função de Liapunov limitada sobre todas as trajetórias admissíveis do sistema (5.49). Verificou-se também que as trajetórias admissíveis são limitadas, então pelo Teorema 5.5 prova-se o Teorema 5.7.

Desta forma verifica-se que o modelo proposto é consistente, estável e convergente.

5.5 - O Modelo STM-CELR Temporal com o Filtro Gamma

No Tópico 4.3.5 foi apresentado o modelo gamma. Verificou-se pelo Teorema 4.4 que o gamma pode ser uma aproximação do modelo de convolução. Verificou-se também que as soluções $w(t)$ do modelo gamma são casos particulares das soluções $w(t)$ do modelo ARMA. Logo, considerando as devidas restrições, o modelo gamma pode ser aplicada ao sistema descrito pela equação (5.6) e manter-se estável e convergente conforme análise realizada no Tópico 5.4.

Portanto, sejam:

$$h_\ell(t) \equiv \frac{\mu^\ell}{(\ell-1)!} t^{\ell-1} e^{-\mu t}, \quad \ell = 1, \dots, K, \quad (0 < \mu < 2); \quad (5.56)$$

$$x_{i\ell}(t) \equiv \int_0^t h_\ell(t-s) f_i(x_i(s)) ds; \quad (5.57)$$

$$x_{j\ell}(t) \equiv \int_0^t h_\ell(t-s) g_j(x_j(s)) ds ; \quad (5.58)$$

$$\frac{dx_{j\ell}}{dt}(t) = -\mu_j x_{j\ell}(t) + \mu_j x_{j,\ell-1}(t), \quad \ell = 1, \dots, K, \quad (5.59)$$

$$0 < \mu < 2;$$

$$net_i(t) = \sum_{\ell=0}^{K-1} w_{i\ell} x_{i\ell}(t); \quad (5.60)$$

e

$$net_j(t) = \sum_{\ell=0}^{K-1} w_{ij\ell} x_{j\ell}(t). \quad (5.61)$$

De (5.60) e (5.61), (5.6) torna-se então:

$$\dot{x}_i(t) = -A_i x_i(t) + (B_i - x_i) \left[I_i + \sum_{\ell=0}^{K-1} w_{i\ell} x_{i\ell}(t) \right] - (x_i + C_i) \left[J_i + \sum_{j=1}^n \sum_{\ell=0}^{K-1} w_{ij\ell} x_{j\ell}(t) \right] \quad (5.62)$$

Desta forma, a equação (5.62) representa o modelo da STM-CELR temporal com o filtro gamma embutido.

VI - A Arquitetura ART-Temporal com o Filtro Gamma

6.1 - Introdução ao Modelo ART

O modelo neurocomputacional ART (Adaptive Resonance Theory) foi desenvolvido por Grossberg [76] para comparar padrões estáticos de forma similar a utilizada pelos seres humanos. Outra característica importante deste modelo é a aprendizagem não-supervisionada, cujo mérito é a classificação de novos padrões desconhecidos sem a necessidade de qualquer monitoração externa quanto ao processo paralelo aprendizagem-classificação.

Em geral, os sistemas que utilizam a aprendizagem não-supervisionada possuem problemas quanto à estabilidade operativa, pois são adaptáveis continuamente em ambientes aleatórias e complexos. Neste tópico, o modelo será observado analisando-se as vantagens de seu uso no processo de classificação. Com o intuito de apenas abordar a filosofia do modelo, não será descrita sua formulação matemática. No Capítulo 3 desta tese há uma descrição detalhada sobre a teoria cuja evolução levou ao desenvolvimento dos modelos ART.

6.1.1 - O Modelo e a Arquitetura ART

A aprendizagem no ART respeita a Doutrina de Helmholtz: "uma informação pura (crua) modifica-se através de experiências anteriores, via um incremento imaginário de aprendizado, antes de tornar-se uma percepção verdadeira".

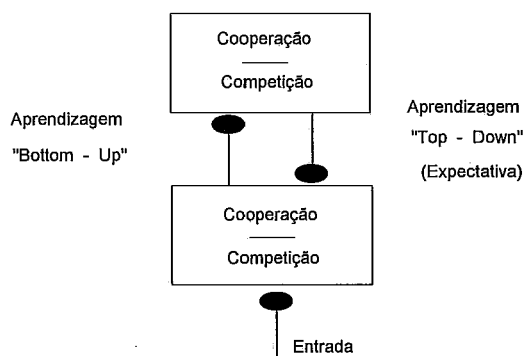


Figura 6.1: Doutrina de Helmholtz.

Há dois tipos de operações sobre a informação de entrada (Figura 6.1):

1 - aprendizagem no sentido "Bottom-Up" - o sinal é levado ao ambiente de aprendizagem para cooperar ou competir com os conhecimentos anteriores;

2 - aprendizagem no sentido "Top-Down" - após a primeira etapa de aprendizagem, cria-se uma expectativa sobre a imagem da entrada e desenvolve-se um ambiente de cooperação ou competição para que evoluam as divergências até que se encontre um consenso (neste ponto obtém-se a "percepção verdadeira").

A Figura 6.2 representa como a doutrina de Helmholtz é aplicada no Modelo de Grossberg (Capítulo 3).

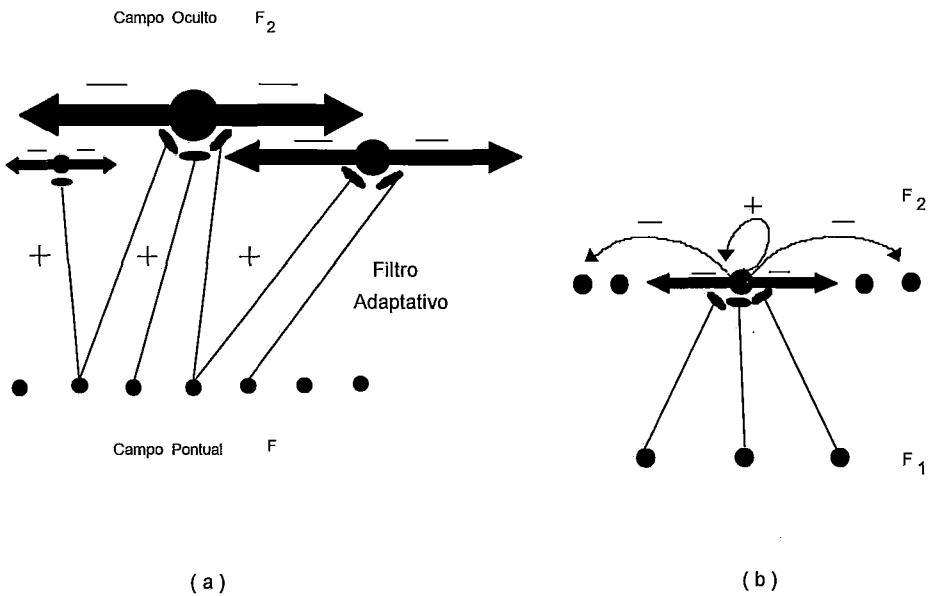


Figura 6.2: Aplicação da Doutrina de Helmholtz.

Pelo modelo descrito na Figura 6.2, observa-se que além da aplicação da doutrina de Helmholtz, também há a política de centralização de poder sobre um neurônio dentro de uma região cooperativa. Ou seja, define-se uma vizinhança na qual ela torna-se mais sensível na presença de uma determinada classe de padrões, responsabilizando-se por inibir lateralmente

o restante com potência proporcional à distância ao centro da vizinhança. Com isto, "particiona-se" a camada sensitiva de neurônios em regiões de aglomerados diferentes.

No item (a) da figura, as células do campo pontual F_1 crescem aleatoriamente em direção ao campo oculto F_2 respeitando as direções dos gradientes sensitivos. Os neurônios do campo oculto fortificam-se em proporção à quantidade de elos que cada qual possui com o campo pontual. Após atingir um tamanho ótimo, tais nós são capazes de ativar um vasto número de outros neurônios e produzir fortíssimas interações inibitórias. As interações entre o campo oculto exibem propriedades de auto-similaridade. Na figura (b), as sinapses nos finais das conexões $F_1 \rightarrow F_2$ adaptativamente sintonizam os filtros definidos por estes fluxos para amplificar os sinais correntes de F_1 para F_2 .

Uma visão mais sistêmica e completa de um modelo ART para padrões binários é exibida na Figura 6.3.

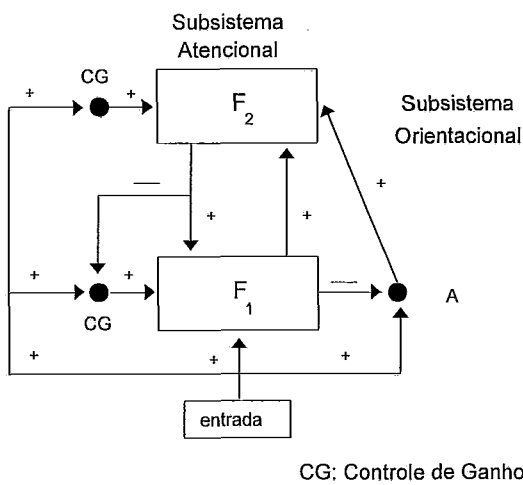


Figura 6.3: Um modelo ART para padrões binários.

O funcionamento básico do modelo ART pode ser dividido em três etapas:

1 - Filtragem Adaptativa "Bottom-Up" e Pré-Classificação: após uma entrada P_k no sistema, os neurônios de F_2 competem entre si e o que receber maior estímulo de entrada será pré-classificado;

2 - Comparação Com a Expectativa "Top-Down": o neurônio ativado em F_2 envia um sinal de expectativa de aprendizagem ("padrão estimado") para F_1 . Este "padrão estimado", (um representante da categoria pré-classificada), é comparado com o padrão P_k

para verificar se a pré-classificação foi satisfatória. Os neurônios de F_1 sofrem agora uma perturbação causada por P_k e pelo "padrão estimado", gerando um padrão \overline{P}_k . Analisando-se as características de P_k e \overline{P}_k , o modelo determinará no futuro se a pré-classificação procede ou se é necessário criar-se uma nova classe (aglomerado);

3 - Interações Entre os Subsistemas Atencional e Orientacional: o subsistema orientacional possui um mecanismo de cancelamento de seleção, **A**, que é acionado quando o "padrão estimado" for suficientemente diferente de P_k . O mecanismo **A** é um comparador entre o padrão sentido por F_1 , inicialmente P_k , e o \overline{P}_k , necessitando portanto de ligações com F_1 e com a entrada. O mecanismo **A** disparará quando receber uma excitação da entrada maior do que a inibição de F_1 . Nesse instante, as unidades ativadas de F_2 sofrerão uma forte excitação inibitória e duradoura de **A**, mantendo-as isoladas, isto é, cancelando a seleção do "padrão estimado". Com a eliminação do "padrão estimado", F_1 será capaz de restaurar P_k e esquecerá definitivamente \overline{P}_k . A entrada agora não poderá criar a mesma ativação sobre os neurônios de F_2 , pois os mesmos estão fortemente inibidos. Com isto, um novo "padrão estimado" deverá ser acionado em F_2 e o ciclo prosseguirá até classificar-se corretamente ou criar-se uma nova classe. Após este ponto, o sistema entrará em estado de ressonância: os sinais de $F_1 \rightarrow F_2$ e de $F_2 \rightarrow F_1$ se reforçarão e os pesos das sinapses se ajustarão aumentando a eficácia dos processos de filtragem adaptativa e classificação.

Após um sistema ART ser totalmente explorado e sua memória for totalmente ocupada, qualquer padrão novo que não pertencer a uma classe do sistema será automaticamente rejeitado, eliminando qualquer possibilidade de esquecimento e catástrofe.

No ciclo descrito na Figura 6.4 [Carpenter & Grossberg, 88], verifica-se que em (a) o padrão de entrada I gera a ativação do padrão X nos neurônios de F_1 e alimenta o mecanismo **A**. Em seguida, o padrão X inibe **A** e gera um sinal "Bottom-Up" através do padrão S . O padrão S é transformado pelo filtro adaptativo no padrão $T = Z(S)$ e este gera o padrão Y em F_2 pela ativação de um conjunto de neurônios de F_2 . Em (b), o padrão Y gera o sinal "Top-Down" através do padrão V que é transformado pelo filtro adaptativo "Top Down" em $V = Z(U)$, que é o "padrão estimado". Se V não for similar à I em F_1 , então um novo padrão X^* será gerado em F_1 pelo acionamento de um subconjunto de neurônios já ativados em F_1 . Esta redução de neurônios ativados em F_1 enfraquece a inibição que F_1

provoca ao mecanismo A. Em (c), o mecanismo A envia um sinal inibidor aos neurônios de F_2 que pré-classificaram Y, cancelando a seleção. Em d), X é recuperado em F_1 enviando o padrão T pelo filtro adaptativo para F_2 . Em F_2 um novo padrão é estimado e o ciclo continua.

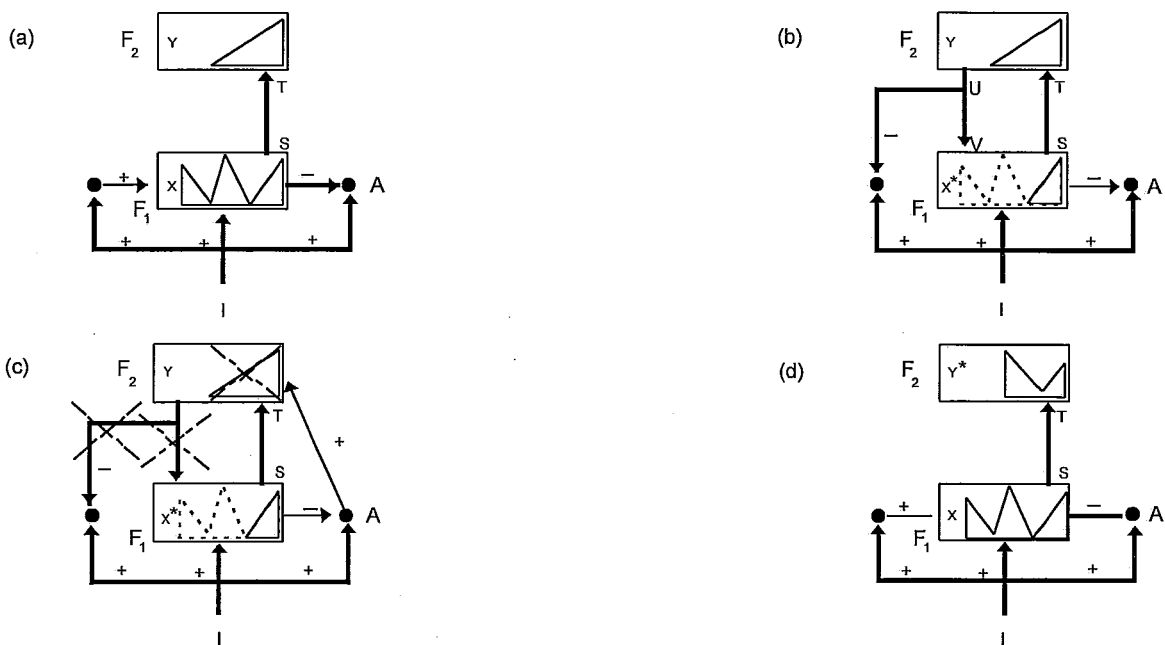


Figura 6.4: Ciclo de um teste de hipóteses num modelo ART para padrões binários.

A Tabela 4.1 [Carpenter & Grossberg, 88] compara as propriedades dos modelos ART e as dos outros modelos que se propõem a realizar a tarefa de classificação.

Tabela 4.1: Arquitetura ART Comparada Com Outros Esquemas de Aprendizagem.

ARQUITETURA ART	PROPRIEDADES ALTERNATIVAS DE APRENDIZAGEM
<p>Aprendizagem em Tempo Real</p> <p>Mundo Não-Estacionário</p> <p>Auto-Organização (não-supervisionamento)</p> <p>Auto-estabilização de memória em resposta à várias entradas arbitrárias</p> <p>Uso efetivo da capacidade total de memória</p> <p>Preservação da plasticidade num mundo desconhecido</p> <p>Auto-aprendizagem "TOP-DOWN" de expectativas</p> <p>Mecanismo ativo de alerta que supervisiona a aprendizagem</p> <p>Aprendizagem lenta ou rápida</p> <p>Aprendizagem na fase: aproximação - casamento de padrões</p> <p>Uso do teste de hipóteses do auto-ajuste para reorganizar a distribuição de energia</p> <p>Rápida pesquisa adaptativa em direção ao melhor casamento de padrões</p> <p>Rápido acesso direto a códigos de eventos similares</p> <p>Critério de erro variável em resposta a realimentação do meio</p> <p>Todas as propriedades são escaladas, independentes de qualquer capacidade do sistema</p>	<p>Aprendizagem "OFF-LINE"</p> <p>Mundo Estacionário</p> <p>Tutoriameto (supervisionamento)</p> <p>Capacidade de catástrofe em resposta à várias entradas arbitrárias</p> <p>Uso parcial da capacidade de memória</p> <p>Não flexibilidade à plasticidade para prevenir a capacidade de catástrofe.</p> <p>Imposição externa de aprendizagem</p> <p>Aprendizagem passiva</p> <p>Aprendizagem lenta ou oscilação catastrófica</p> <p>Aprendizagem na fase: não-casamento de padrões</p> <p>Uso de ruído para excitar o sistema num pré-determinado estado de energia fora dos mínimos locais</p> <p>Pesquisa em Árvore</p> <p>O tempo de reconhecimento aumenta em relação ao acréscimo da complexidade do código</p> <p>Critério de erro fixo, independente das respostas de realimentação do meio</p> <p>Propriedades importantes deterioram-se com o aumento da capacidade do sistema</p>

6.2 - O Modelo ART2

O ART2 é a versão da arquitetura ART para o processamento de padrões do espaço \mathfrak{R}^m que implementa o modelo neurocomputacional:

$$\dot{x}_i = -Ax_i + (B - x_i) \left[I_i + f(x_i) \right] - x_i \left[J_i + \sum_{k \neq i} f(x_k) \right]. \quad (6.1)$$

A Figura 6.5 esboça a arquitetura ART2.

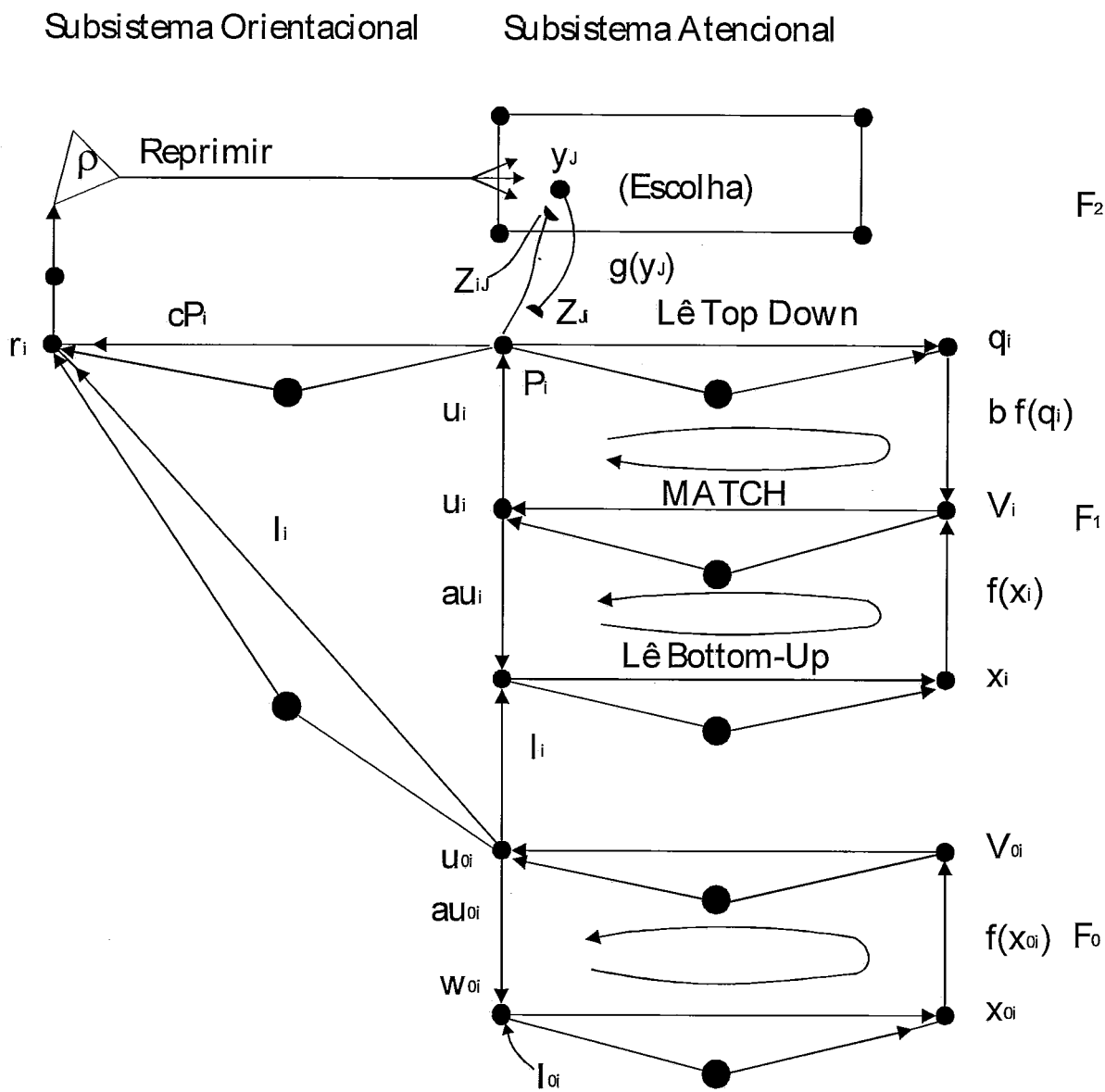


Figura 6.5: Arquitetura ART2. Os grandes círculos fechados representam operações de normalizações.

A camada F_o realiza um pré-processamento do padrão de entrada adicionando informações de entradas anteriores. O resultado do pré-processamento é a entrada I para a camada F_1 . A tarefa da camada F_o é resumida pelos seguintes passos:

- Considere o padrão de entrada $I^o \in \mathfrak{R}^m$, F_o gera o vetor w^o como sendo uma entrada para a camada F_o adicionada a uma realimentação interna au^o :

$$w^o = I^o + au^o. \quad (6.2)$$

- w^o sofre uma operação de normalização:

$$x^o = \frac{w^o}{\|w^o\|}. \quad (6.3)$$

- x^o é transformado em v^o através da aplicação de uma função não-linear:

$$v^o = f(x^o), \text{ onde} \quad (6.4)$$

$$f(x_i^o) = \begin{cases} x_i^o & \text{se } x_i^o > \theta \\ 0 & \text{caso contrario} \end{cases}, \quad (6.5)$$

e $0 < \theta \leq \frac{1}{\sqrt{m}}$.

- Os sinais acima do limiar são amplificados por uma nova operação de normalização:

$$u^o = \frac{v^o}{\|v^o\|}. \quad (6.6)$$

- u^o é enviado para a camada F_1 como padrão de entrada dessa camada:

$$I = u^o. \quad (6.7)$$

Observe que a normalização na equação (6.3) corresponde ao efeito de controle automático de ganho inibitório no sistema competitivo apresentado no Capítulo 3.

A camada F_1 engloba o processamento efetuado pela camada F_o . As operações nos dois níveis mais baixos de F_1 são as mesmas designadas para os dois níveis de F_o . No nível superior de F_1 , P soma u (sinal interno de F_1) e todos os sinais filtrados de $F_2 \rightarrow F_1$, isto é:

$$P_i = u_i + \sum_J g(y_J) Z_{ji}, \quad (6.8)$$

onde $g(y_j)$ é o sinal de saída do j -ésimo neurônio de F_2 e Z_{ji} é a LTM no sentido do j -ésimo neurônio de F_2 para o i -ésimo neurônio de F_1 .

Resumindo, F_1 executa as seguintes operações:

- F_1 gera o vetor w através da soma de I com uma realimentação interna au :

$$w = I + au. \quad (6.9)$$

- w sofre uma normalização:

$$x = \frac{w}{\|w\|}. \quad (6.10)$$

- v soma informações vindas de F_o (entrada I **bottom-up** $F_o \rightarrow F_1$) e de F_2 (entrada **top-down** $F_2 \rightarrow F_1$):

$$v = f(x) + bf(q). \quad (6.11)$$

- Os sinais acima do limiar são amplificados por uma nova operação de normalização:

$$u = \frac{v}{\|v\|}. \quad (6.12)$$

- F_1 gera o vetor P utilizado para medir a similaridade entre a entrada I e a expectativa da camada F_2 :

$$P_i = u_i + \sum_j g(y_j) Z_{ji}. \quad (6.13)$$

- F_1 executa um controle automático de ganho sobre P :

$$q = \frac{P}{\|P\|}. \quad (6.14)$$

A camada F_2 é um campo competitivo que estará ativa (ou capaz de realizar uma escolha) se e somente se um neurônio ($j = J$) tiver STM acima de um limiar. Este é o neurônio que recebe a maior entrada de F_1 . Neste caso, $g(y_J)$ é igual a uma constante d , $0 < d < 1$ e $g(y_j)$ para $j \neq J$ é nulo. Portanto, quando F_2 estiver ativo, a equação (6.8) torna-se:

$$P_i = u_i + dZ_{Ji} \quad (6.15)$$

e quando estiver inativo,

$$P = u. \quad (6.16)$$

Uma entrada qualquer para um neurônio j de F_2 , no sentido $F_1 \rightarrow F_2$, é uma soma ponderada dos sinais de cada neurônio i de F_1 que alcançar F_2 pela via representada pela sinapse Z_{ij} :

$$T_j = \sum_{i=1}^m P_i Z_{ij}, \quad (6.17)$$

onde $j = 1, \dots, n$.

Como verificado, a escolha do neurônio vencedor na camada F_2 é dada pelo neurônio que receber maior entrada de F_1 , ou seja, será o neurônio J de F_2 cuja entrada T_J é máxima em relação as outras entradas T_j , $j = 1, \dots, n$.

O modelo ART, em geral, após receber uma realimentação **top-down** de $F_2 \rightarrow F_1$ mede o grau de similaridade entre a entrada I para F_1 e P realimentado. Como observado, a realimentação **top-down** de $F_2 \rightarrow F_1$ para P é dada por dZ_J . O vetor r é responsável por monitorar este grau de similaridade pela seguinte definição:

$$r = \frac{I + cP}{\|I\| + \|cP\|}, \quad (6.18)$$

onde $c > 0$. Note que, se P for proporcional (paralelo) a I então $\|r\| = 1$, pois:

$$\|r\| = \frac{[\|I\|^2 + 2c\|I\|\|P\|\cos(I, P) + c^2\|P\|^2]^{1/2}}{\|I\| + c\|P\|} \quad (6.19)$$

e se P for proporcional a I então $\cos(I, P) = 1$ e $P = kI$, onde k é uma constante real. Desta forma:

$$\|r\| = \frac{[\|I\|^2 + 2c|k|\|I\|^2 + c^2k^2\|I\|^2]^{1/2}}{\|I\| + c|k|\|I\|}. \quad (6.20)$$

ou seja

$$\|r\| = \frac{[(1 + 2c|k| + c^2k^2)\|I\|^2]^{1/2}}{(1 + c|k|)\|I\|} = 1. \quad (6.21)$$

Verifica-se então que a similaridade ideal ocorre quando P e I são paralelos. Mas o modelo possui um parâmetro de vigilância que permite associar a entrada I com o neurônio J de F_2 , a partir de P , se e somente se

$$\|r\| \geq \rho, \quad (6.22)$$

onde ρ é o parâmetro de vigilância não dimensional entre 0 e 1.

Se a condição (6.22) for burlada, o mecanismo orientacional do modelo ART reprime por inibição o neurônio J de F_2 até as camadas F_1 e F_2 ressonarem. Isto é, o neurônio J de F_2 não participará da competição da sua camada até que a condição (6.22) ocorra para um novo vencedor L de F_2 (ressonância). Se nenhum neurônio da camada F_2 provocar a

ressonância, a camada F_2 criará uma nova categoria (novo neurônio) para representar a entrada I para F_1 .

Antes de F_2 tornar-se ativa, todos os seus neurônios são rotulados como não associados. Isto é, F_2 não representa qualquer categoria de padrões; não está associado a qualquer padrão. Quando forem calculadas as entradas T_j para F_2 , $j=1, \dots, n$, e for verificado que o neurônio vencedor tiver rótulo de não associado então ele irá aprender o padrão I , não necessitará realizar o teste de similaridade e se tornará associado deste instante em diante. Além disto, qualquer neurônio novo de F_2 criado a partir de uma nova categoria é rotulado como não associado.

Finalmente, após a ressonância entre F_1 e F_2 ocorre a aprendizagem do padrão P pelo neurônio J de F_2 :

$$\frac{dZ_{iJ}}{dt} = d[P_i - Z_{iJ}]. \quad (6.23)$$

Note que a aprendizagem é contínua. Mas, como verificado no Tópico 6.1, não ocorre esquecimento nem capacidade de catástrofe, garantidos pelo teste de similaridade.

A regra de aprendizagem determinada por (6.23) é chamada de regra **bottom-up**. Ela é equivalente ao Modelo Instar [Carpenter, 89]. Na Figura 6.5 verifica-se a existência da sinapse Z_{Ji} , utilizada na definição de P em (6.8), (6.13) e (6.15). Esta sinapse está no sentido **top-down** e sua regra, de mesmo nome, é dada por

$$\frac{dZ_{Ji}}{dt} = d[P_i - Z_{Ji}]. \quad (6.24)$$

A regra (6.24) é equivalente ao modelo outstar [Carpenter, 89].

Note que só há variação sináptica para Z_{Ji} e Z_{iJ} . Todos os outros sinapses Z_{ji} e Z_{ij} , $j \neq J$, possuem variações nulas.

6.3 - O Modelo ART2-Gamma

O modelo STM-CELR temporal com filtro gamma apresentado no Tópico 5.5 será implementado neste capítulo em uma arquitetura chamada agora de ART Temporal. Esta nova arquitetura propõe manter as capacidades e vantagens da teoria de ressonância adaptativa em um espaço com dinâmica temporal. Mais especificamente é uma extensão do ART2 para o espaço temporal através da aplicação do filtro gamma. Desta forma, o ART2-Gamma é definido pelas seguintes equações:

$$\dot{x}_i(t) = -Ax_i + (B - x_i) \left[I_i + \sum_{\ell=0}^K Z_{i\ell} x_{i\ell} \right] - x_i \left[J_i + \sum_{j \neq i} \sum_{\ell=0}^K Z_{ij\ell} x_{j\ell} \right], \quad (6.25)$$

onde

$$x_{f\ell} = \int_0^t h_\ell(t-s) g_f(x_f(s)) ds, \quad (6.26)$$

$$\frac{dx_{f\ell}}{dt} = -\mu x_{f,\ell-1}, \quad \ell = 1, \dots, K, \quad (6.27)$$

$$h_\ell(t) = \frac{(\mu)^\ell}{(\ell-1)!} t^{\ell-1} e^{-\mu t}, \quad \mu > 0, \quad (6.28)$$

e g_f é uma função de realimentação: excitatória se $i = f$ e inibitória se $i \neq f$.

O esboço do fluxo operacional de ART2-Gamma é apresentado na Figura 6.6.

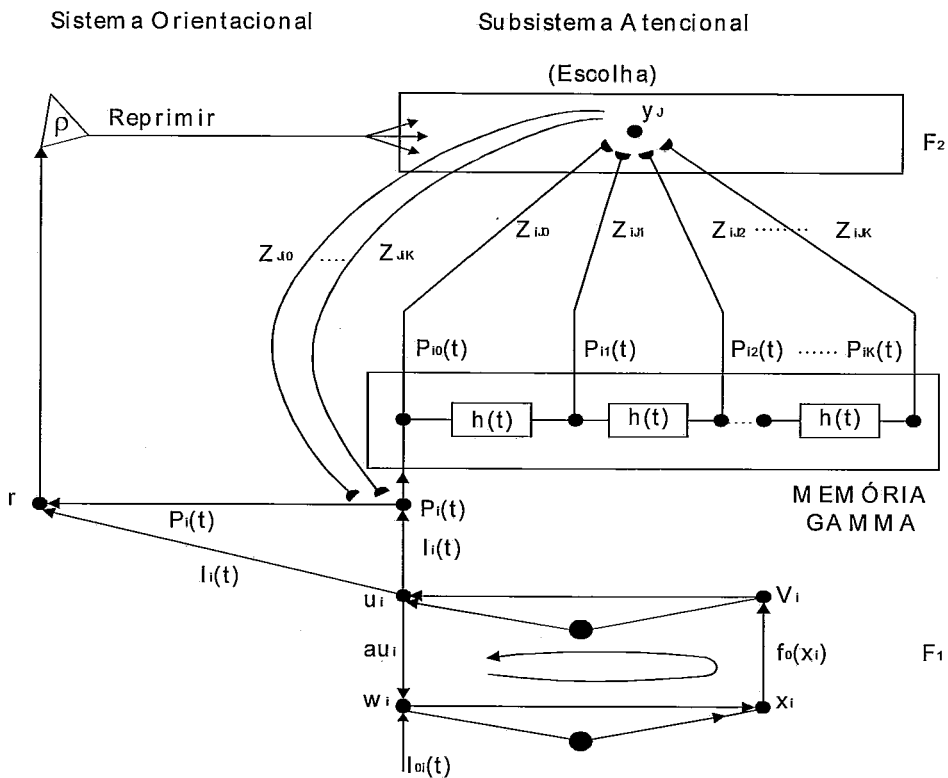


Figura 6.6: Arquitetura ART2-Gamma.

O mecanismo do ART2-Gamma é similar ao ART2. A extensão temporal introduzida pelo filtro gamma gera a memória gamma sobre o vetor $P(t)$. Portanto é

testada a similaridade entre a entrada temporal $I(t)$ e o vetor $P(t)$ realimentado pelo neurônio J de F_2 através das sinapses $Z_{Ji\ell}$, $i = 1, \dots, m$ e $\ell = 0, \dots, K$, de forma que o neurônio J tenha recebido a maior entrada referente a soma ponderada entre a memória gamma $P_{i\ell}(t)$ e as sinapses **bottom-up** $Z_{iJ\ell}$.

As operações ocorridas na camada F_1 do ART2-Gamma são um subconjunto das mesmas definidas pelo ART2. A escolha do neurônio vencedor na camada F_2 é dada pela maior entrada total T_J tal que:

$$T_J = \sum_{t=0}^{T_{total}} T_J(t), \text{ onde} \quad (6.29)$$

$$T_J(t) = \text{Max} \left\{ T_J(t) / T_J(t) = \sum_{i=1}^m \sum_{\ell=0}^K P_{i\ell}(t) z_{ij\ell} \right\}, \quad (6.30)$$

e o instante T_{total} é a última amostra temporal referente à apresentação completa do padrão de entrada.

O neurônio y_J vencedor na camada F_2 tem ativação regrada pela função sigmóide aplicada sobre a entrada total T_J .

A memória gamma $P_{i\ell}(t)$ é obtida pela relação de recorrência:

$$P_{i\ell}(t) = (1 - \mu) P_{i\ell}(t-1) + \mu P_{i(\ell-1)}(t-1), \quad (6.31)$$

onde as condições iniciais são:

$$P_{i0}(t) = P_i(t); \quad P_{i1}(0) = \mu P(0);$$

$$\text{e } P_{i\ell}(0) = 0 \quad \text{para } \ell = 2, \dots, K. \quad (6.32)$$

O vetor $P(t)$ recebe realimentação do neurônio vencedor y_J de F_2 , tornando-se:

$$P_i(t) = I_i(t) + \frac{R_i(t)}{\|R_i(t)\|}, \quad (6.33)$$

onde a realimentação $R_i(t)$ é

$$R_i(t) = \sum_{l=0}^K Z_{Ji\ell} g(y_J(t-l)), \quad (6.34)$$

sendo $g(y_J)$ a função de realimentação excitatória. A função $g(y_J)$ será igual a função f definida pela equação (6.5) se o neurônio y_J de F_2 não for reprimido pelo subsistema orientacional. Caso ocorra a repressão, $g(y_J)$ será nulo.

O coeficiente r da Figura 6.6 é o coeficiente de correlação entre todas as amostras temporais de $I(t)$ e de $P(t)$. O coeficiente de correlação é definido estatisticamente pela fórmula:

$$r(I, P) = \frac{COV(I, P)}{\sigma(I) \cdot \sigma(P)}, \quad (6.35)$$

onde a função $\sigma(\cdot)$ é o desvio padrão e

$$COV(I, P) = E[(I - \bar{I})(P - \bar{P})] \quad (6.36)$$

é a função estatística covariância, com $E[\cdot]$ sendo a função expectância e \bar{X} sendo a média da variável aleatória X .

O coeficiente de correlação mede o grau de similaridade linear entre duas variáveis aleatórias. Pela definição de r verifica-se que $-1 \leq r \leq 1$. Quando $r = 1$ diz-se que a correlação é perfeita. Quando $r = -1$ diz-se que as variáveis têm sinais opostos. E quando $r = 0$ diz-se que não há correlação linear entre as variáveis.

No problema específico desta tese, r é o coeficiente de correlação e ρ é o parâmetro de vigilância do modelo ART2-Gamma. A similaridade ocorrerá entre $I(t)$ e $P(t)$ se e somente se

$$r \geq \rho, \quad (6.37)$$

onde ρ está entre 0 e 1.

Os subsistemas orientacional e atencional do ART2-Gamma operam como os mesmos subsistemas do ART2.

Como nos modelos ART, após a ressonância entre F_1 e F_2 é executada a aprendizagem. A memória gamma $P_{i\ell}(t)$ é aprendida pelo neurônio y_J de F_2 , onde $t = T_{total}$ é o último instante da apresentação completa do padrão de entrada. Desta forma, a regra de aprendizagem **bottom-up** é obtida por:

$$\frac{dZ_{iJ\ell}}{dt} = d[P_{i\ell}(t) - Z_{iJ\ell}], \quad (6.38)$$

e a regra **top-down** por:

$$\frac{dZ_{Ji\ell}}{dt} = d[P_{i\ell}(t) - Z_{Ji\ell}], \quad (6.39)$$

onde $\ell = 0, \dots, K$ e $i = 1, \dots, m$.

Nota: na íntegra do modelo ART2-Gamma, o sistema dinâmico (6.25) traduz o comportamento de um neurônio da camada F_2 .

A implementação do modelo ART2-Gamma mostrada nesta tese é algorítmica. Os neurônios da camada F_2 não foram modelados pelo sistema (6.25). A solução algorítmica é

computacionalmente viável e simples. Esta solução não degenera a essência do modelo ART2-Gamma.

6.4 - Testes com o ART2-Gamma

6.4.1 - Seleção e Pré-Processamento dos Padrões de Entrada

Foram selecionados arquivos sonoros do tipo “wave”, padrão **Microsoft Windows 3.11**, do sistema de som de computadores pessoais **Compaq - Família Deskpro M**. Esses arquivos são didáticos e possuem alta relação sinal/ruído. Eles correspondem a sons de pássaro, grilo, trovão e de um tipo de macaco sul-africano. Os padrões de entrada para o ART2-Gamma foram espectrogramas destes sinais, pois a quantidade de informações relevantes neste espaço tri-dimensional (frequência x tempo x energia) é superior ao do espaço bi-dimensional temporal (tempo x amplitude).

A obtenção de padrões temporais para o treinamento da rede neuronal não é uma tarefa simples. Além do sistema de som Compaq, foram procuradas outras fontes de geração de sinais temporais para o uso em reconhecimento de padrões. Porém não havia número de categorias de padrões e quantidade de padrões por categoria suficientes para efetuar-se o treinamento. A solução encontrada foi gerar representantes distorcidos ou transformados dos arquivos acima citados.

O ambiente de processamento de sinais utilizado para distorcer e transformar as fontes sonoras originais foi a ferramenta **Cool versão 1.50** padrão **MS Windows 3.1**. Esta ferramenta possibilitou compatibilizar as taxas de amostragem dos quatro arquivos selecionados, fixando-as em 22050 amostras por segundo. Isto é, espectros padronizados na faixa de 0 a 11 KHz. Cada arquivo de som resultou em 688 espectros formadores de um padrão de entrada temporal (espectrograma) com quatro segundos de apresentação.

Para cada arquivo de som foram gerados quatro outros distorcidos e/ou transformados no ambiente Cool. As distorções e transformações utilizadas foram:

- modificação da rotação de apresentação do som original;
- reverberação em ambientes diferentes (auditório, sala de aula, reunião social, etc.);
- aplicação do som em outros ambientes, tais como: de baixo d'água e em um **canyon**;
- variação da frequência fundamental do som original;

- aplicações de envelopes do tipo: triangular; curva de extremidades suaves; e curva de decaimento suave;
- ecos variados;
- filtragens tipo passa-baixa e faixa audível;
- e distorções do nível do sinal de saída.

No Anexo II desta tese encontram-se os gráficos espectrais de cada arquivo original e de seus quatro derivantes. Os arquivos originais têm terminações “a”; os derivantes têm terminações “b” ou “c” ou “d” ou “e”.

6.4.2 - Resultados das Classificações com o ART2-Gamma

Foram realizados quatro experimentos importantes para o treinamento da rede neuronal a partir dos 20 padrões de entrada.

Como citado no tópico anterior, cada espectrograma tem a duração de 4 segundos de apresentação. Na avaliação dos espectrogramas, verificou-se que o filtro gamma que melhor se adapta ao problema é um filtro de frequências passa-baixa, pois permite diminuir os efeitos causados pelos ruídos de medição e reforça as características do sinal (padrão temporal com informações em frequências baixas). No Capítulo 4 há uma condição para que o filtro gamma seja estável e tipo passa-baixa. Esta condição é $0 < \mu < 1$. Um outro aspecto importante é que a profundidade da memória gamma D mantém uma relação com a resolução μ e a ordem K do filtro gamma. Para uma boa representação destes padrões espectrais, K deve ser no mínimo 3. Respeitando a profundidade de memória $D = 4$, tem-se que $\mu = 0.75$. Estes parâmetros foram fixados para os quatro experimentos a seguir descritos.

No Capítulo 5 foi apresentada a parte teórica desta tese. O modelo ART2-Gamma encaixa-se nas condições do sistema dinâmico proposto naquele capítulo. Provou-se que padrões de entrada temporais positivos aplicados ao sistema permitiriam a convergência e a estabilidade se algumas condições fossem respeitadas. O ART2-Gamma respeitou todas as condições, inclusive a dos coeficientes sinápticos serem todos não-negativos. Para os quatro experimentos foram rerepresentados os vinte padrões de entrada aleatoriamente até que a rede convergisse em cada experimento ou que fosse verificada a recodificação

contínua, invalidando tal experimento. Convém ressaltar que a recodificação não é sinônimo de instabilidade ou de não convergência da rede. Em todos os experimentos Z (matriz dos coeficientes sinápticos) sempre foi não-negativa.

As características que diferenciam os experimentos são o limiar do coeficiente de correlação ρ e a ordem aleatória de apresentação dos padrões de entrada. Nas tabelas 6.1 a 6.8 estão os resultados das classificações dos vinte padrões temporais submetidos a quatro redes ART2-Gamma. As oito tabelas correspondem a primeira apresentação a cada rede e a uma etapa posterior que relata a convergência ou a recodificação contínua. O item “Padrão Entrada” das tabelas codifica cada padrão temporal de entrada onde os códigos “pa” a “pe” representam as versões de “Pássaro Original” (pa) a última transformação deste padrão, “Pássaro Transformado - Versão E” (pe), resultando em um total de cinco padrões para a categoria “pássaro”. De forma análoga tem-se “ga” a “ge” para a categoria “Grilo”, “ta” a “te” para a categoria “Trovão” e “ma” a “me” para a categoria “Macaco”. O item “seqüência” das tabelas descreve em que ordem cada padrão foi apresentado à rede. E o item “neurônios” representa os neurônios vencedores (categorias) da camada $F2$ referente a cada entrada.

No primeiro experimento tentou-se obter categorias bem compactas, independentes das características do espaço dos padrões. Desta forma escolheu-se $\rho = 0.9$, ou seja, os padrões de cada categoria devem possuir uma grande correlação estatística. Esta situação é ideal para garantir-se uma boa margem de acertos. Para este experimento, 1ª etapa, os valores Z_{jil} ficaram entre 0 e 1.5936. A Figura 6.7 apresenta um gráfico dos pesos Z (memórias gamma para cada categoria). Neste gráfico, o eixo das abcissas contém as 4 memórias gamma para cada categoria. No eixo das ordenadas tem-se os 124 neurônios da camada $F1$. Cada neurônio da $F1$ representa uma **feature** do padrão de entrada. E os tons de cinza do gráfico correspondem as amplitudes de cada Z_{jil} .

No Anexo II, além dos gráficos dos padrões de entrada, existem os gráficos do padrão $I(t)$ (saída da camada $F1$) e do padrão $P(t)$ após a realimentação do neurônio vencedor da camada $F2$. Estes gráficos correspondem a apenas os resultados do quarto experimento, décima etapa (Tabela 6.8). Não foram colocados os gráficos das sete outras tabelas, pois a quantidade de informações é muito grande.

Tabela 6.1 : 1º Experimento - 1ª Etapa

Seqüência =====	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Neurônio																				
01			•		•											•		•		•
02		•															•			
03	•																			
04				•			•			•			•	•						
05						•														
06								•							•					
07									•											
08											•									
09												•								
10																			•	
Padrão de Entrada	pc	gb	te	m b	td	pb	m e	gc	pe	m d	pd	pa	m c	m a	ga	tc	ge	ta	gd	tb

Observe que o neurônio 01 da Tabela 6.1 criou categoria “Trovão”, classificando todos os cinco padrões de entrada. O neurônio 04 criou a categoria “Macaco”, classificando todos as 5 versões de “Macaco”. Porém, para classificar os padrões das cinco versões de “Grilo” foi necessária a criação de três neurônios: 02, 06 e 10. Para os padrões de “Pássaro”, cada versão gerou uma categoria e, portanto um neurônio correspondente.

Pesos Sinápticos Z - ART2-Gamma

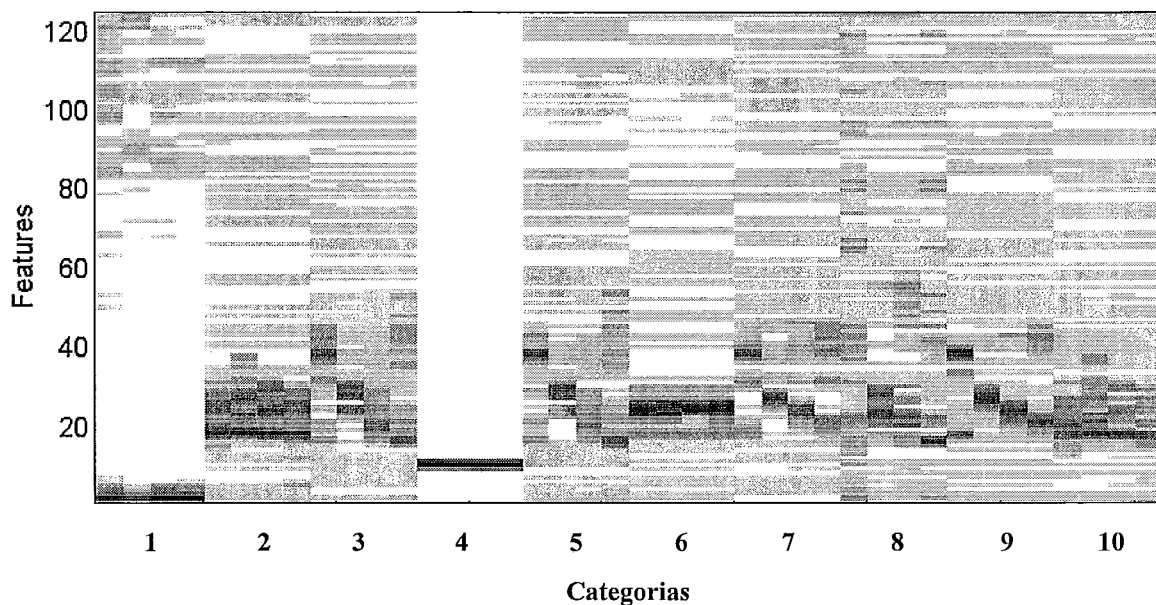


Figura 6.7: Pesos da rede. 1º Experimento - 1ª Etapa.

Tabela 6.2 : 1º Experimento - 2ª Etapa

Seqüência =====	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Neurônio																				
01				•			•					•				•				•
02						•														
04								•												
06																			•	
11	•																			
12		•																		
13			•						•	•			•							
14				•											•					
15											•									
16													•							
17																	•			
18																		•		
Padrão de Entrada	pd	pa	md	td	gc	ge	tb	ma	mc	mb	pb	tc	me	pe	gb	te	gd	pc	ga	ta

Os pesos Z_{ji} correspondentes aos resultados apresentados na Tabela 6.2 ficaram da seguinte forma: $0 < Z_{ji} \leq 1.6066$ (Figura 6.8). Analisando esta tabela, verifica-se que novas categorias foram criadas para representar os mesmos padrões já aprendidos (recodificação). Durante a geração destes resultados notou-se que a correlação entre $I(t)$ e $P(t)$ após a realimentação diminuiu um pouco, o suficiente para que ficasse abaixo do limiar ρ . Em uma primeira análise somos levados a acreditar que tenha ocorrido o efeito esquecimento. Mas na realidade houve adaptação e a criação da categoria 13 só ocorreu pois os coeficientes de correlação entre os padrões de “Macaco” estavam muito próximos do limiar na primeira etapa. Um problema similar ao da categoria “Macaco” levou a explosão da categoria “Grilo”. Figura 6.8 apresenta os pesos da rede após esta segunda etapa de apresentação. Note que os pesos da categoria 04 e da categoria 13 são muito similares.

Ao analisar-se esse cenário, conclui-se que poderá haver várias recodificações para as próximas etapas de apresentações até que a rede convirja. Ou seja, até estabilizar-se a rede terá várias categorias bem semelhantes para os mesmos tipos de padrões de entrada.

Pesos Sinápticos Z - ART2-Gamma

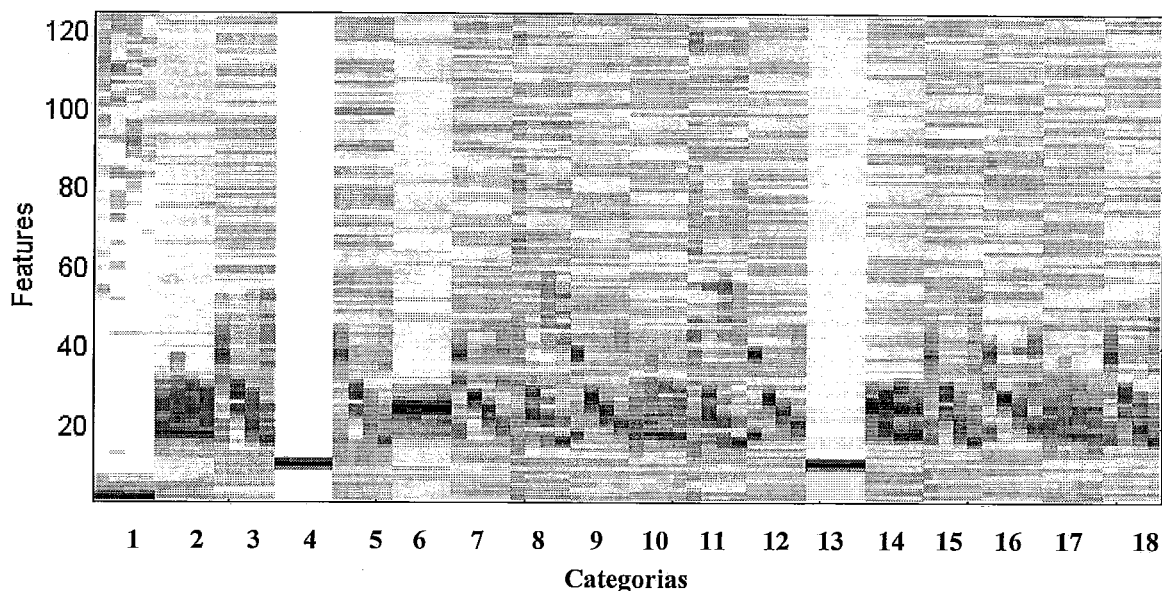


Figura 6.8 : Pesos da Rede : 1º Experimento - 2ª Etapa.

No segundo experimento baixou-se um pouco o limiar de correlação ρ , porém mantendo ainda uma grande correlação entre os padrões de uma categoria, $\rho = 0.8$. Na primeira etapa deste experimento, $0 < Z_{jil} \leq 1.6070$ (Figura 6.9).

Tabela 6.3 : 2º Experimento - 1ª Etapa

Sequência =====	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	
Neurônio																					
01	•					•				•		•			•						
02			•																		
03		•			•						•					•	•				
04				•																	
05							•	•	•				•								
06														•							
07																		•			
08																				•	
09																					•
Padrão de Entrada	md	ta	pa	pb	tc	mb	gd	ga	ge	mc	td	ma	gb	pc	me	te	tb	gc	pe	pd	

Com esta pequena redução do limiar ρ , os neurônios 01 e 03 criaram e representaram todos os padrões das categorias “Macaco” e “Trovão”, respectivamente.

Observe que o neurônio 05 possui uma melhor representação da categoria “Grilo” do que no experimento anterior. Porém, ainda não foi possível agrupar os padrões “Pássaro”.

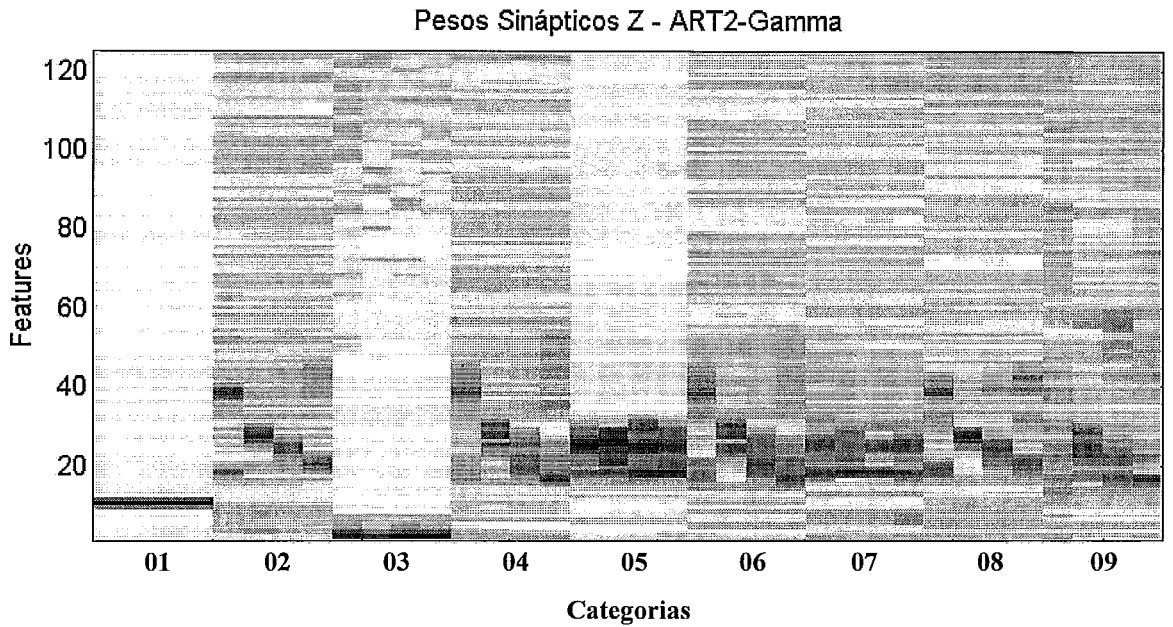


Figura 6.9 : Pesos da Rede : 2º Experimento - 1ª Etapa.

Tabela 6.4 : 2º Experimento - 2ª Etapa

Seqüência =====	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Neurônio																				
01								•	•				•	•						•
03				•		•	•			•							•			
07	•										•									
10		•													•				•	
11			•																	
12					•															
13												•								
14																•				
15																				•
Padrão de Entrada	ga	gd	pe	td	pb	tc	tb	me	mb	te	gb	pd	ma	mc	gc	pc	ta	ge	md	pa

Os valores dos pesos estiveram entre $0 < Z_{jil} \leq 1.6209$ (Figura 6.10) para a segunda etapa do segundo experimento. Note que a categoria 05 (“Grilo”) manteve-se desprezível, explodindo nas categorias 07 e 10. As novas categorias para “Pássaro” foram criadas

(recodificação), pois os pesos ao serem gerados possuem definições aleatórias e próximas de zero. Portanto, como cada “Pássaro” ficou em um único neurônio na primeira etapa, tal neurônio carrega muito ruído e o padrão $I(t)$ quando comparado com o padrão $P(t)$ realimentado possui uma correlação muito abaixo do limiar, gerando um novo neurônio representante. Notou-se novamente que o valor do limiar ρ estava muito próximo da correlação entre os padrões da categoria “Grilo”, como ocorrido na 2ª etapa do experimento anterior.

Novas reapresentações levarão a várias recodificações de “Pássaro” até a rede estabilizar.

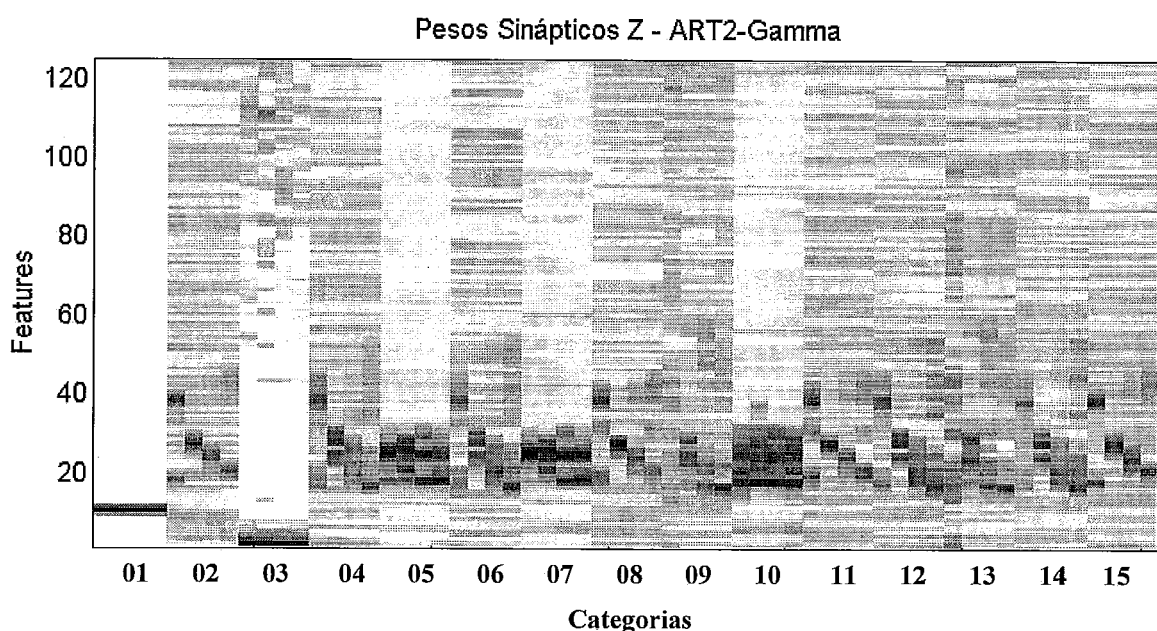


Figura 6.10 : Pesos da rede : 2ª Experimento - 2ª Etapa.

Surge neste instante uma questão: como resolver os problemas das explosões de categorias conhecidas e como agregar a categoria “Pássaro”? Este é um problema de reconhecimento de padrões e há dois caminhos que podem resolvê-lo:

- escolher um conjunto de **features** que melhor discrimine as categorias e agregue os padrões de uma mesma categoria;
- ou estudar um subconjunto do espaço de padrões e segmentá-lo em hiper-regiões agregando os padrões de uma hiper-região em uma única categoria.

Supõe-se que após esta segmentação, a rede tenha a capacidade de extrapolar corretamente para a classificação de novos padrões de entrada pertencentes às hiper-regiões.

Para os experimentos desta tese, é mais simples escolher-se a segunda opção. Isto não implica em dizer que seja mais eficiente. Na maioria dos casos, a primeira opção é a mais eficiente; e uma combinação delas é o ideal.

Uma forma de avaliar o espaço de padrões é calcular os coeficientes de correlação entre os padrões de uma categoria. Repete-se este processo para todas as categorias e calculam-se os coeficientes de correlação entre as categorias duas a duas. Os resultados desta avaliação foram (onde $\bar{\rho}$ e σ são a média e o desvio padrão dos coeficientes de correlação) :

- categoria “Pássaro”: $\bar{\rho} = 0.5393$ e $\sigma = 0.2171$;
- categoria “Grilo”: $\bar{\rho} = 0.6551$ e $\sigma = 0.1905$;
- categoria “Macaco”: $\bar{\rho} = 0.9335$ e $\sigma = 0.0553$;
- categoria “Trovão”: $\bar{\rho} = 0.7293$ e $\sigma = 0.1390$;
- “Pássaro” x “Grilo”: $\bar{\rho} = 0.0192$ e $\sigma = 0.015$;
- “Pássaro” x “Macaco”: $\bar{\rho} = -0.0135$ e $\sigma = 0.0031$;
- “Pássaro” x “Trovão”; $\bar{\rho} = -0.0030$ e $\sigma = 0.0099$;
- “Grilo” x “Macaco”: $\bar{\rho} = 0.0095$ e $\sigma = 0.0062$;
- “Grilo” x “Trovão”: $\bar{\rho} = 0.0082$ e $\sigma = 0.0118$;
- “Macaco” x “Trovão”: $\bar{\rho} = 0.0215$ e $\sigma = 0.0095$.

Pelos resultados acima, conclui-se que as categorias são linearmente muito bem separadas uma das outras, viabilizando a segunda proposta de solução do problema. Porém, as categorias “Pássaro” e “Grilo” não possuem bons coeficientes de correlação entre os seus representantes, podendo realmente haver mais de um neurônio para representar uma única categoria em subcategorias. Pois se forçar a representação de uma categoria, por exemplo “Pássaro”, em um único neurônio é provável que ocorram erros de classificação. Isto é , agrupar representantes de “Pássaro” e “Grilo” em um único neurônio.

Portanto, na escolha do limiar ρ considerou-se a análise acima realizada e o limiar do último experimento. Calculou-se a média entre 0.88 (experimento anterior) e $\bar{\rho} + \sigma$ da categoria “Pássaro”. Desta forma, o terceiro experimento empregou o limiar $\rho = 0.8315$.

Tabela 6.5 : 3º Experimento - 1ª Etapa

Seqüência ----- Neurônio	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
01	•								•			•		•			•			
02			•		•	•	•			•										
03		•		•												•			•	•
04								•			•		•							
05															•			•		
Padrão de Entrada	md	tc	pc	ta	ga	ge	gc	pd	ma	gb	pe	mc	gd	mb	pa	td	me	pb	te	tb

Para a primeira etapa do terceiro experimento $0 < Z_{ji} \leq 1.6139$ (Figura 6.11). Note que as categorias ficaram mais agrupadas, mas ocorreram dois erros de classificação. O padrão “Pássaro C” ficou agrupado na categoria “Grilo” (neurônio 02) e o padrão “Grilo D” agrupou-se em uma das subcategorias de “Pássaro” (neurônio 04).

Pesos Sinápticos Z - ART2-Gamma

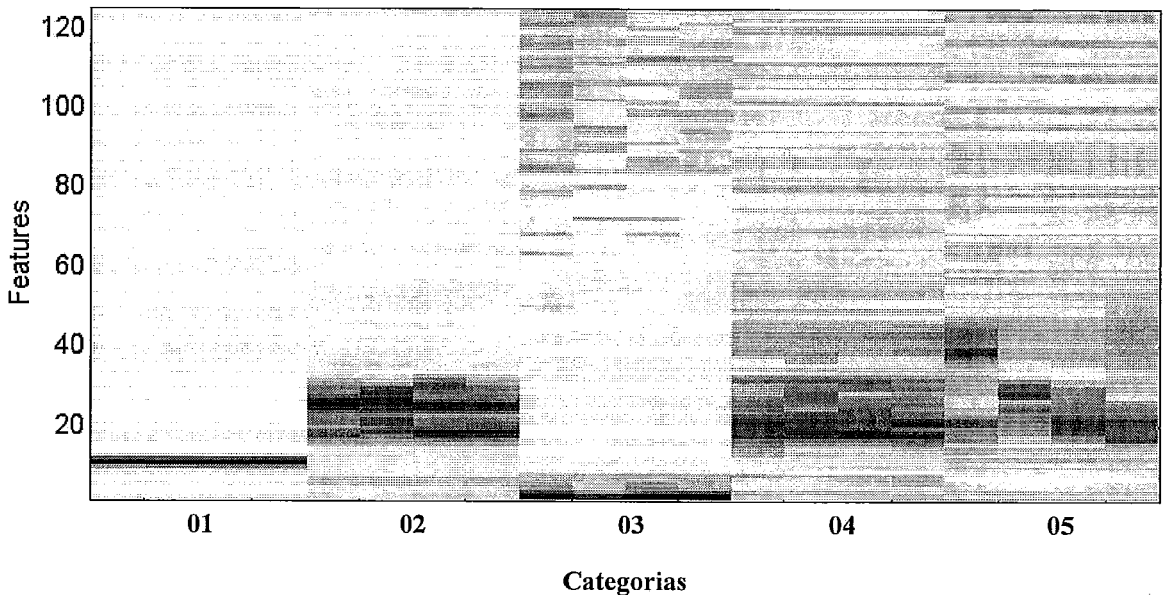


Figura 6.11 : Pesos da Rede : 3º Experimento - 1ª Etapa.

Tabela 6.6 : 3º Experimento - 2ª Etapa

Seqüência	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Neurônio																				
01					•						•						•	•		•
02								•	•											
03	•	•	•										•			•				
04										•				•					•	
05				•																
06						•	•													
07												•			•					
Padrão de Entrada	tc	ta	te	pa	ma	pd	pc	gc	ga	gd	mc	pe	tb	ge	pb	td	me	mb	gb	md

Na segunda etapa do terceiro experimento $0 < Z_{jil} \leq 1.6168$ (Figura 6.12). Nesta etapa as categorias e subcategorias se estabilizaram corretamente. O neurônio 01 criou a categoria “Macaco” agregando os cinco padrões. O neurônio 03 criou a categoria “Trovão”, agregando todos o cinco representantes. Os neurônios 02 e 04 dividiram a categoria “Grilo”. E os neurônios 05, 06 e 07 dividiram a categoria “Pássaro”. Ou seja, novamente ocorreu recodificação e novas apresentações poderão levar a novas recodificações até a estabilidade da rede.

Pesos Sinápticos Z - ART2-Gamma

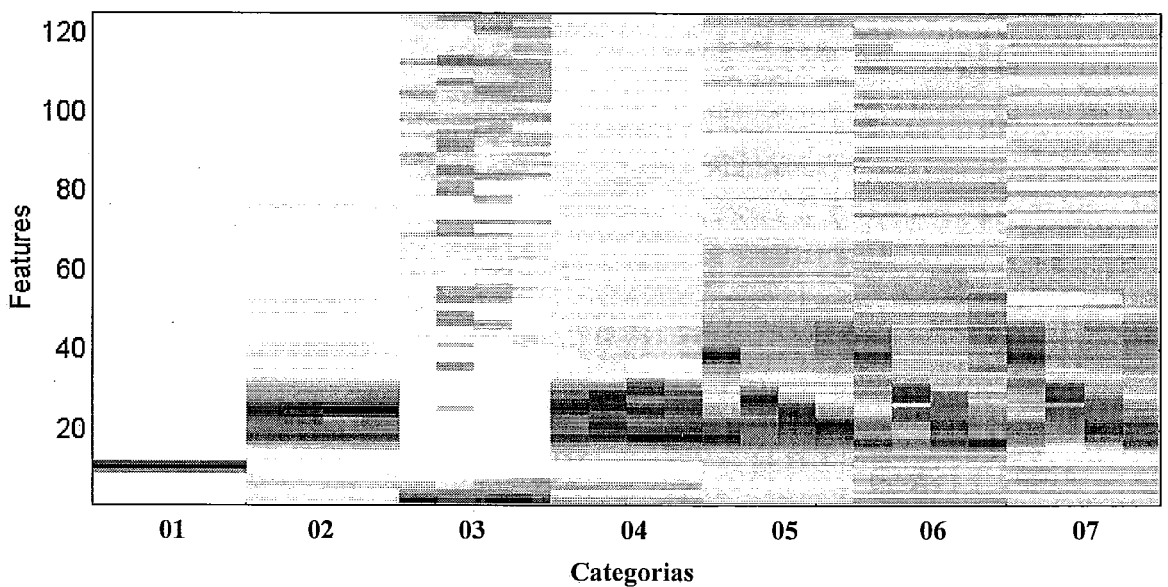


Figura 6.12 : Pesos da Rede : 3º Experimento - 2ª Etapa.

No quarto experimento escolheu-se o limiar de correlação $\rho = 0.7564$. Este valor corresponde a média de ρ somada com o desvio padrão σ referentes a categoria “Pássaro”. Na primeira etapa deste experimento, $0 < Z_{jil} \leq 1.6059$ (Figura 6.13).

Tabela 6.7 : 4º Experimento - 1ª Etapa

Seqüência =====	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Neurônio																				
01	•	•	•				•											•	•	
02																	•			•
03										•	•	•	•	•	•					
04				•	•	•		•	•	•										
Padrão de Entrada	pe	mc	mb	ge	pa	ga	ma	gb	gc	gd	te	tb	td	pd	tc	ta	pc	md	me	pb

As categorias “Macaco”, “Trovão” e “Grilo” agruparam todos os seus componentes e mais um representante de “Pássaro”. Com as repetições do experimento espera-se que a categoria “Pássaro” fique melhor representada no neurônio 02 e as outras categorias não aceitem mais os seus representantes.

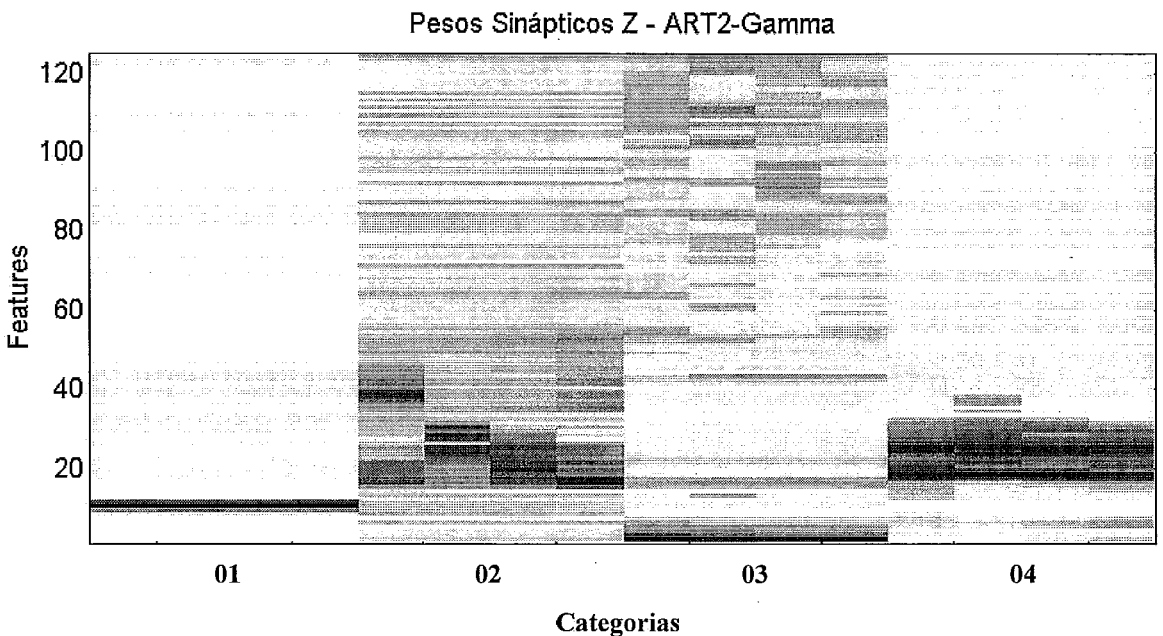


Figura 6.13: Pesos da Rede: 4º Experimento - 1ª Etapa.

Tabela 6.8 : 4º Experimento - Após a 3ª Apresentação

Seqüência	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Neurônio																				
01						•	•		•	•								•		
02	•	•						•				•								
03			•	•						•		•			•				•	
04					•								•	•		•				•
Padrão de Entrada	pb	pe	td	tc	gb	me	mc	pa	mb	ma	tb	pc	ta	gd	ga	pd	gc	md	te	ge

Os valores dos pesos estiveram entre $0 < Z_{ji} \leq 1.6204$ (Figura 6.14) para a décima etapa do quarto experimento.

Neste experimento a rede convergiu após a terceira apresentação e não recodificou mais, num total de dez apresentações. O único erro foi a caracterização do padrão “Pássaro Versão D” como pertencente a categoria “Trovão”. Acredita-se que mais apresentações podem corrigir este erro. Foram mantidas as quatro categorias (4 neurônios na camada F_2) criadas na primeira etapa.

Este experimento foi realizado mais duas vezes. As três realizações foram completamente independentes e obtiveram os mesmos resultados de convergência após a terceira apresentação (total de 10) e do mesmo erro de categorização.

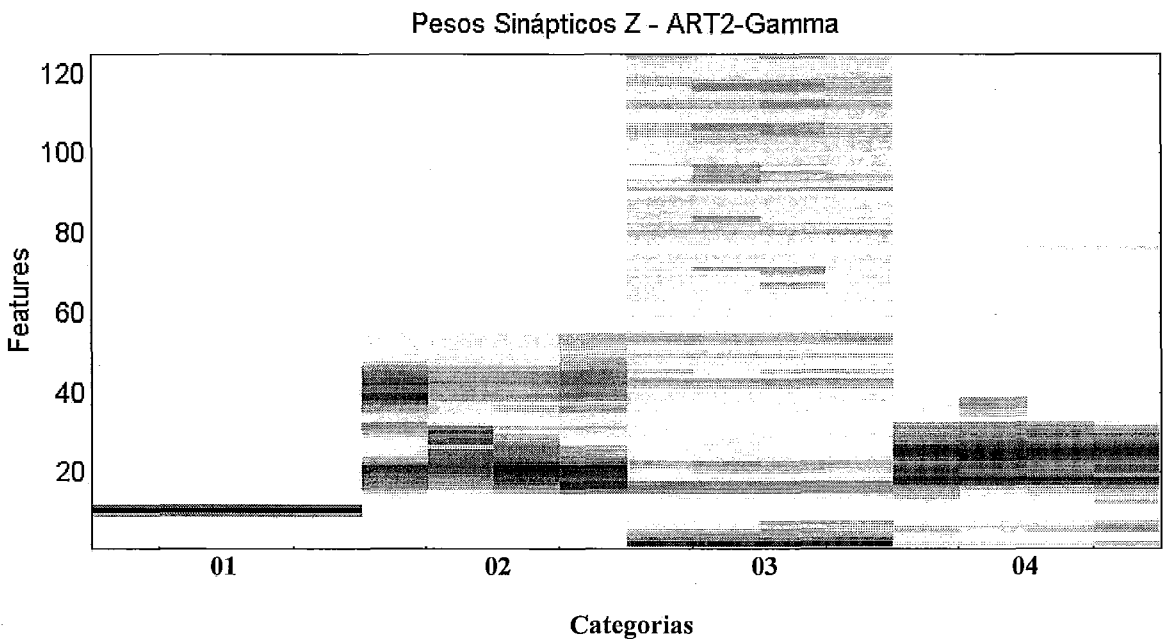


Figura 6.14: Pesos da Rede: 4º Experimento - 10ª Etapa.

Conclusões

Esta tese aborda especificamente uma estrutura neurocomputacional para reconhecimento de padrões temporais. Alguns pontos foram fundamentais para o desenvolvimento do modelo proposto. As pesquisas iniciais de Grossberg [68, 73, 78, 87] sobre os comportamentos e as estruturas dos neurônios e suas interações nas regiões do cortex e do hipocampus levaram-no a desenvolver modelos matemáticos que simulavam a fisiologia e o mecanismo biológico destas regiões. A aplicação imediata das suas pesquisas encaixou-se nos problemas de reconhecimento de padrões, cujas soluções dos seres vivos apresentam resultados mais eficientes do que os métodos estatísticos, matematicamente bem fundamentados. Um outro ponto importante dos estudos de Grossberg, juntamente com Cohen [83], foi a proposta e as provas de estabilidade e convergência de uma família de sistemas dinâmicos que modelam o comportamento dos neurônios daquelas pesquisas por ele desenvolvidas. Esta tese é uma extensão da família de sistemas dinâmicos proposta por Grossberg e Cohen. Nela introduziu-se a memória temporal, aspecto de consagrada importância para as aplicações em reconhecimento de padrões temporais não-estacionários. Esta extensão é uma contribuição de vasta aplicabilidade e, até o presente, única no âmbito de estrutura biológica. Além disto, foram provadas sua estabilidade e sua convergência, viabilizando sua implementação com segurança.

O fato desta tese estar direcionada a uma linha carente em comparações torna difícil compará-la com outras propostas. Entretanto, como um dos objetivos foi preservar todas as características de simulação das estruturas biológicas desenvolvidas por Grossberg no seu modelo, pode-se considerar que tais características foram herdadas pela extensão.

O filtro gamma introduzido no modelo teórico e na arquitetura ART Temporal (ART2-Gamma) adaptou-se perfeitamente ao problema apresentado. Contudo, há situações que sejam necessários estudos, representações e definições de comportamento de padrões temporais dos quais o gamma não se mostre a melhor solução. Tanto a proposta teórica quanto a aplicada são flexíveis à introdução de novos filtros. A única restrição a princípio é atender as condições de estabilidade e convergência da família de sistemas dinâmicos proposta neste trabalho.

Como citado, a maioria de modelos neurocomputacionais direcionados ao espaço temporal dedicam-se a outros campos, tais como: predição, filtragem adaptativa, identificação de sistemas, controle, etc. [Bianchini, Gori & Maggini, 94], [Nerrand, Roussel-Ragot, Urbani, Personnaz & Dreyfus, 94], [Olurotimi, 94], [Piché, 94], [Srinivasan, Prasad & Rao, 94], [Connor, Martin & Atlas, 94], [Parlos, Chong & Atiya, 94], [Kechriotis, Zervas & Manolakos, 94], [Puskorius & Feldkamp, 94] e [Sastry, Santharam & Unnikrishnan, 94]. Além disto, as arquiteturas utilizadas são modelos puramente de engenharia que utilizam algumas definições neurocomputacionais. Seus métodos de aprendizagem também são aplicações de teorias das quais têm-se padrões de comportamento conhecidos (métodos supervisionados) e, portanto, utilizam técnicas de filtragem adaptativa para ajustarem os pesos sinápticos [Bengio, Simard & Frasconi, 94], [Tsoi & Back, 94], [Robinson, 94], [Zeng, Goodman & Smyth, 94], [Principe, Kuo & Celebi, 94], [Lawrence, Tsoi, Back, 96].

O método de aprendizagem utilizado no ART2-Gamma foi não supervisionado: sem o conhecimento prévio do padrão a ser aprendido e sem qualquer controle sobre a etapa de aprendizagem.

Há quatro pontos importantes a serem sugeridos como continuidade de pesquisa desta tese:

- implementação da camada F_2 do ART2-Gamma segundo o sistema dinâmico (6.25);
- procura de outros métodos de ajuste adaptativo dos pesos sinápticos;
- criação da arquitetura ARTMAP Temporal;
- e a pesquisa de outras funções de concentração de informações.

A arquitetura ARTMAP [Carpenter, Grossberg & Reynolds, 91] é uma evolução da ART. São duas ART que são supervisionadas por um sistema de controle e comparação dos padrões de aprendizagem. Ela possui um aprendizado supervisionado onde são apresentados dois padrões de entrada. O primeiro é o destino de aprendizagem e o segundo é o padrão a ser aprendido. Cada um destes padrões é disponibilizado simultaneamente à ART entrada e à ART destino. O mecanismo de supervisão do ARTMAP responsabiliza-se em comparar os resultados dos dois ART e efetuar a aprendizagem.

BIBLIOGRAFIA

- [01] - Amari, S. (1977): "Neural Theory of Association and Concept-Formation"; Biological Cybernetics 26, 175-185.
- [02] - Attneave, F. (1954): "Some Informational Aspects of Visual Perception"; Psychological Review 61, 183-193.
- [03] - Bendat, J. S. & Piersol, A. G. (1980): "Engineering Applications of Correlation and Spectral Analysis"; John Wiley & Sons, Inc..
- [04] - Bengio, Y.; Simard, P. & Frasconi, P. (1994): "Learning Long-Term Dependencies with Gradient Descent is Dificult", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [05] - Bianchini, M.; Gori, M. & Maggini, M. (1994): "On The Problem of Local Minima in Recurent Neural Networks", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [06] - Bock, H. H. (Ed.) (1988): "Classification and Related Methods of Data Analysis"; North-Holland.
- [07] - Boyce, W. E. & DiPrima, R. C. (1977): "Equações Diferenciais Elementares e Problemas de Valores de Contorno", Editora Guanabara Dois, 3ª edição.
- [08] - Braun, N. (1983): "Ordinary Differential Equations and Their Applications"; (3 rd ed.). Berlin: Springer-Verlag.
- [09] - Buhmann, J. & Schulten, K. (1987): "Noise-Driven Temporal Association Neural Networks", Europhysics Letters, 4 (10), 1205-1209.
- [10] - Carpenter, G. A. & Grossberg, S. (1987a): "ART 2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns", Vol. 26, No. 23, Applied Optics.
- [11] - Carpenter, G. A. & Grossberg, S. (1987b): "A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine"; Computer Vision, Graphics and Image Processing, Vol. 37, 54-115.
- [12] - Carpenter, G. A. & Grossberg, S. (1988): "The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network"; Computer, Vol. 21, No. 3, march.

- [13] - Carpenter, G. A. & Grossberg, S. (1990): "ART3 Hierarchical Search Using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures"; Neural Networks, Vol. 3, 129-152.
- [14] - Carpenter, G. A. (1989): "Neural Network Models for Pattern Recognition and Associative Memory"; Neural Networks, Vol. 2, 243-257.
- [15] - Carpenter, G. A.; Grossberg, S. & Meharian, C. (1989): "Invariant Recognition of Cluttered Scenes by Self-Organizing ART Architecture: CORT-X Boundary Segmentation"; Neural Networks, Vol. 2, 169-181.
- [16] - Carpenter, G. A.; Grossberg, S. & Rosen, D. B. (1991a): "ART2-A: An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition"; Neural Networks, Vol. 4, 493-504.
- [17] - Carpenter, G. A.; Grossberg, S. & Rosen, D. B. (1991b): "Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by Adaptive Resonance System"; Neural Networks, Vol. 4, 759-771.
- [18] - Carpenter, G. A.; Grossberg, S. & Reynolds, J. H. (1991): "ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network", Neural Networks, Vol. 4, 565-588.
- [19] - Carpenter, G. A.; Grossberg, S.; Markuzon, N.; Reynolds, J. H. & Rosen, D. B. (1992): "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps", IEEE Transactions on Neural Networks, Vol. 3, No. 5.
- [20] - Cesari, L. (1963): "Asymptotic Behavior and Stability Problems in Ordinary Differential Equations", 2nd edition, Springer-Verlag.
- [21] - Chauke Nehme, C. (1989): "Reconhecimento de Sinais de Voz Contínua", Tese de Mestrado, Instituto Militar de Engenharia.
- [22] - Chauke Nehme, C. & Maia, M. A. G. M. (1992): "Uma Abordagem do Problema de Reconhecimento de Padrões", Publicação Técnica, COPPE/Sistemas, ES-267/92.
- [23] - Chauke Nehme, C. & de Carvalho, L. A. V. (1993): "O Enfoque Neurocomputacional para o Problema de Classificação"; Publicação Técnica, COPPE/Sistemas, ES-275/93.

- [24] - Chauke Nehme, C. & Mendes, S. B. T. (1994): "Modelos Neurocomputacionais Aplicados ao Reconhecimento de Padrões Temporais", Publicação Técnica, COPPE/Sistemas, ES-332/94.
- [25] - Cohen, M. A. & Grossberg (1983): "Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks", IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-13, september/october.
- [26] - Connor, J. T.; Martin, R. D. & Atlas, L. E. (1994): "Recurrent Neural Networks and Robust Time Series Prediction", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [27] - Crochiere, R. E. & Rabiner, L. R. (1983): "Multirate Digital Signal Processing", Prentice-Hall Signal Processing Series.
- [28] - Day, S. P. & Davenport, M. R. (1993): "Continuous-Time Temporal Back-Propagation with Adaptable Time Delays", IEEE Transactions on Neural Networks, Vol. 4, No. 2, march.
- [29] - Dehaene, T.; Changeux, J. P. & Nadal, J. P. (1987): "Neural Networks that Learn Temporal Sequences by Selection"; Proc. Natl. Acad. Sci. USA, Vol. 84, 2727-2731.
- [30] - Devijver, P. A. & Kittler, J. (1982): "Pattern Recognition: A Statistical Approach"; Prentice Hall.
- [31] - Elman, J. L. (1990): "Finding Structure in Time", Cognitive Science, 14, 179-211.
- [32] - Farley, B. G. & Clark, W. A. (1954): "Simulation of Self-Organizing Systems by Digital Computer"; IRE Transactions on Information Theory IT-14, 76-84.
- [33] - Fu, K. S. (Ed.) (1980): "Digital Pattern Recognition"; 2nd Edition, Communication and Cybernetics 10, Springer-Verlag.
- [34] - Fukunaga, K. (1972): "Introduction to Statistical Pattern Recognition"; Electrical Science Series, Academic Press.
- [34] - Gjerdingen, R. O. (1992): "Learning Syntactically Significant Temporal Patterns of Chords: A Masking Field Embedded in an ART3 Architecture"; Neural Networks, Vol. 5, 551-564.
- [36] - Gonzalez, R. C. & Thomason, M. G. (1978): "Syntactic Pattern Recognition: An Introduction"; Addison-Wesley Publishing Company.

- [37] - Grossberg, S. (1968): "Some Physiological and Biochemical Consequences of Psychological Postulates", *Applied Mathematics*, Vol. 60.
- [38] - Grossberg, S. (1969): "Some Networks that can Learn, Remember, and Reproduce any Number of Complicated Space-Time Patterns, I"; *Journal of Mathematics and Mechanics*, Vol. 19, No. 1.
- [39] - Grossberg, S. (1970): "Neural Pattern Discrimination", *J. Theor. Biol.*, 27, 291-337.
- [40] - Grossberg, S. (1973): "Contour Enhancement, Short Term Memory, and Constancies in Reverberating Neural Networks", *Studies in Applied Mathematics*, Vol. LII, No. 3, september, MIT.
- [41] - Grossberg, S. (1976): "Adaptive Pattern Recognition and Universal Recoding II: Feedback, Expectation, Olfaction, and Illusions", *Biological Cybernetics*, Vol. 23, 187-202.
- [42] - Grossberg, S. (1978): "Competition, Decision, and Consensus"; *Journal of Mathematical Analysis and Applications*, Vol. 66, 470-493.
- [43] - Grossberg, S. (1987): "Neural Dynamics of Attentionally Modulated Pavlovian Conditioning: Blocking, Interstimulus Interval, and Secondary Reinforcement", *Applied Optics*, Vol. 26, No. 23, december.
- [44] - Grossberg, S. (1988): "Nonlinear Neural Networks: Principles, Mechanisms, and Architectures", *Neural Networks*, Vol. 1, 17-61.
- [45] - Gutfreund, H. & Mezard M. (1988): "Processing of Temporal Sequences in Neural Networks", *Physical Review Letters*, Vol. 61, No. 2, july.
- [46] - Guyon, I.; Personnaz, L.; Nadal, J. P. & Dreyfus, G. (1988): "Storage and retrieval of Complex Sequences in Neural Networks", *Physical Review A*, Vol. 38, No. 12, december.
- [47] - Haken, H. (1987): "Information Compression in Biological Systems", *Biological Cybernetics*, 56, 11-17.
- [48] - Hale, J. K. (1963): "Oscillations in Nonlinear Systems", McGraw-Hill.
- [49] - Hall, D. W. & Spencer II, G. L. (1955): "Elementary Topology", John Wiley & sons.
- [50] - Hartigan, J. A. (1937): "Clustering Algorithms", Wiley Series in Probability and Mathematical Statistics, John Wiley & sons.
- [51] - Haykin, S. (1991): "Adaptive Filter Theory", second ed., Englewood Cliffs, NJ: Prentice-Hall.

- [52] - Haykin, S. (1994): "Neural Networks: A Comprehensive Foundation", Macmillan College Publishing Company, Inc.
- [53] - Hebb, D. O. (1949): "The Organization of Behavior: A New Psychological Theory"; New York: Wiley.
- [54] - Hecht-Nielsen, R. (1987): "Nearest Matched Filter Classification of Spatiotemporal Patterns", *Applied Optics*, Vol. 26, No. 10, 1892 - 1899.
- [55] - Hecht-Nielsen, R. (1990): "Neurocomputing"; Addison-Wesley Publishing Company, Inc..
- [56] - Hush, D. R. & Horne, B. G.(1993): "Progress in Supervised Neural Networks: What's New Since Lippmann?", *IEEE Signal Processing Magazine*, January.
- [57] - Kechriotis, G.; Zervas, E. & Manolakos, E. S. (1994): "Using Recurrent Neural Networks for Adaptive Communication Channel Equalization", *IEEE Transactions on Neural Networks*, Vol. 5, No. 2, March.
- [58] - Kosko, B. (1988): "Bidirectional Associative Memory"; *IEEE Trans. Sys. Man Cybern.*, Vol. 18, 49-60.
- [59] - La Salle, J. & Lefschetz, S. (1961): "Stability by Liapunov's Direct Method With Applications", *Mathematics in Science and Engineering*, Vol. 4, Academic Press.
- [60] - La Salle, J. & Lefschetz, S. (1963): "International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics", edited, Academic Press.
- [61] - Lawrence, S.; Tsoi, A. C. & Back, A. D. (1996): "The Gamma MLP for Speech Phoneme Recognition", Advances in Neural Information Processing Systems 8, edited by Touretzky, D.; Mozer, M. & Hasselmo, M., MIT Press.
- [62] - Lefschetz, S. (1949): "Introduction to Topology", Princeton University Press.
- [63] - Ljung, L. (1987): "System Identification: Theory for User", Prentice-Hall.
- [64] - von der Malsburg, C. (1973): "Self-Organization of Orientation Sensitive Cells in the Striate Cortex", *Kybernetik* 14, 85-100.
- [65] - McCulloch, W. S. & Pitts, W. (1943): "A Logical Calculus of Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics* 5, 115-133.
- [66] - Miller, R. (1987): "Representation of Brief Temporal Patterns, Hebbian Synapses, and the Left-Hemisphere Dominance for Phoneme Recognition"; *Psychobiology*, Vol. 15, No. 3, 241-247.

- [67] - Nerrand, O.; Roussel-Ragot, P.; Urbani, D.; Personnaz, L. & Dreyfus, G. (1994): "Training Recurrent Neural Networks: Why and How? An Illustration in Dynamical Process Modeling", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [68] - Olurotimi, O. (1994): "Recurrent Neural Network Training with Feedforward Complexity", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [69] - Oppenheim, A. & Schaffer, R. (1975): "Digital Signal Processing", Englewood Cliffs, NJ: Prentice-Hall.
- [70] - Pao, Y.-H. (1989): "Adaptive Pattern Recognition and Neural Networks"; Addison-Wesley Publishing Company, Inc..
- [71] - Parlos, A. G.; Chong, K. T. & Atiya, A. F. (1994): "Application of The Recurrent Multilayer Perceptron in Modeling Complex Process Dynamics", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [72] - Piché, S. W. (1994): "Steepest Descent Algorithms for Neural Network Controllers and Filters", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [73] - Principe, J. C., de Vries, B. & de Oliveira, P. G. (1993): "The Gamma Filter - A New Class of Adaptive IIR Filters with Restricted Feedback"; IEEE Transactions on Signal Processing, Vol. 41, No. 2.
- [74] - Principe, J. C.; Kuo, J. M. & Celebi, S. (1994): "An Analysis of Gamma Memory in Dynamic Neural Networks", IEEE Transactions on Neural Networks, Vol. 5, No. 2.
- [75] - Puskorius, G. V. & Feldkamp, L. A. (1994): "Neurocontrol of Nonlinear Dynamical Systems with Kalman Filter-Trained Recurrent Networks"; IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [76] - Ratcliff, F., (1965): "Mach Bands: Quantitative Studies on Neural Networks in the Retina", San Francisco: Holden Day.
- [77] - Robinson, T. (1994): "An Application of Recurrent Nets to Phone Probability Estimation", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [78] - Rosenblatt, F. (1958): "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", Psychological Review, 65, 386-408.
- [79] - Rudin, W. (1964): "Principles of Mathematical Analysis", McGraw-Hill.

- [80] - Sastry, P. S.; Santharam, G. & Unnikrishanan, K. P. (1994): "Memory Neuron Networks for Identification and Control and Control of Dynamical Systems", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [81] - Sejnowski, T. J. (1981): "Skeleton Filters in the Brain", In Hilton & Anderson (eds.), Parallel Model of Associative Memory, 189-212, LEA Inc., Hillsdale, NJ.
- [82] - Shen, L. (1989): "Neural Integration by Short Term Potentiation"; Biological Cybernetics, Vol. 61, 319-325.
- [83] - Srinivasan, B.; Prasad, U. R. & Rao, N. J. (1994): "Back Propagation Through Adjoints for The Identification of Nonlinear Dynamic Systems Using Recurrent Neural Models", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [84] - Stanley, C. & Kilmer, W. L. (1975): "A Wave Model of Temporal Sequence Learning"; Int. J. Man-Machine Stud.; Vol. 7, 397-412.
- [85] - Szego, G. (1939): "Orthogonal Polynomials"; Providence, RI: American Mathematical Society.
- [86] - Tank, D. W. & Hopfield, J. J. (1987a): "Concentrating Information in Time: Analog Neural Networks with Applications to Speech Recognition Problems", IEEE 1st International Conference on Neural Networks, Vol. IV, 455-468.
- [87] - Tank, D. W. & Hopfield, J. J. (1987b): "Neural Computation by Concentrating Information in Time", Proc. Natl. Acad. Sci., Biophysics, USA, Vol. 84, 1896-1900, april.
- [88] - Tsoi, A. C. & Back A. (1994): "Locally Recurrent Globally Feedforward Networks, A Critical Review of Architectures", IEEE Transactions on Neural Networks, Vol. 5, No. 2, march.
- [89] - de Vries, B. & Principe, J. C. (1992): "The Gamma Model - A New Neural Model for Temporal Processing"; Neural Networks, Vol. 5, 565-576.
- [90] - Wang, D. & Arbib, M. A. (1990): "Complex Temporal Sequence Learning Based on Short-Term Memory"; Proceedings of The IEEE, Vol. 78, No. 9.
- [91] - Widrow, B. & Stearns, S. (1985): "Adaptive Signal Processing", Englewood Cliffs, NJ: Prentice-Hall.
- [92] - Widrow, B. & Winter, R. (1988): "Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition"; Computer, Volume 21, Número 3, march.

[93] - Wilson, H. R. & Cowan, J. D. (1972): "Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons"; *Journal of Biophysics* 12, 1-24.

[94] - Zeng, Z.; Goodman, R. M. & Smyth, P. (1994): "Discrete Recurrent Neural Networks for Gramatical Inference", *IEEE Transactions on Neural Networks*, Vol. 5, No. 2, march.

ANEXO I

Regra de Leibniz (Derivação sob o sinal de Integral)

Dado $U \subset \mathfrak{R}^n$, aberto, seja $f: U \times [a,b] \rightarrow \mathfrak{R}$ uma função com as seguintes propriedades:

1 - Para todo $x \in U$, a função $t \rightarrow f(x,t)$ é integrável em $a \leq t \leq b$.

2 - A i -ésima derivada parcial $\frac{\partial f}{\partial x_i}(x,t)$ existe para cada $(x,t) \in U \times [a,b]$ e a função

$\frac{\partial f}{\partial x_i} : U \times [a,b] \rightarrow \mathfrak{R}$, assim definida, é contínua.

Então a função $\varphi : U \rightarrow \mathfrak{R}$, dada por $\varphi(x) = \int_a^b f(x,t) dt$, possui i -ésima derivada parcial em cada ponto $X \in U$, sendo

$$\frac{\partial \varphi}{\partial x_i}(x) = \int_a^b \frac{\partial f}{\partial x_i}(x,t) dt. \quad (\text{A.I.1})$$

Em suma: pode-se derivar sob o sinal de integral, desde que o integrando resultante seja uma função contínua.

Um resultado da Regra de Leibniz:

Seja $f: U \times [a,b] \rightarrow \mathfrak{R}$ contínua, com derivadas parciais contínuas

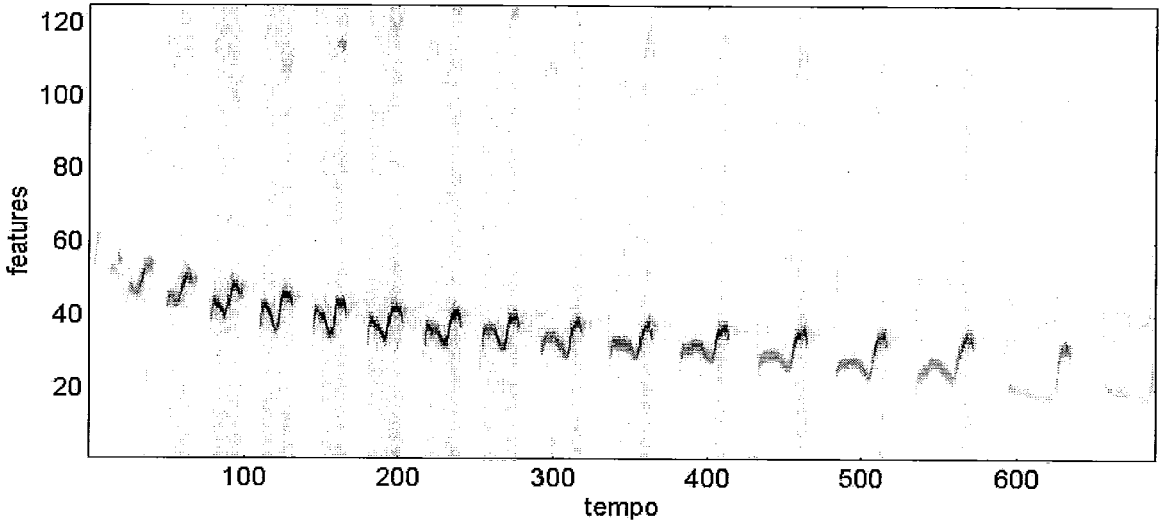
$\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} : U \times [a,b] \rightarrow \mathfrak{R}$. Seja $g : U \rightarrow [a,b]$ de classe C^1 . Então, a função $j : U \rightarrow$

\mathfrak{R} , definida por $\varphi(x) = \int_a^{g(x)} f(x,t) dt$ é da classe C^1 , e suas derivadas parciais são expressas pela fórmula:

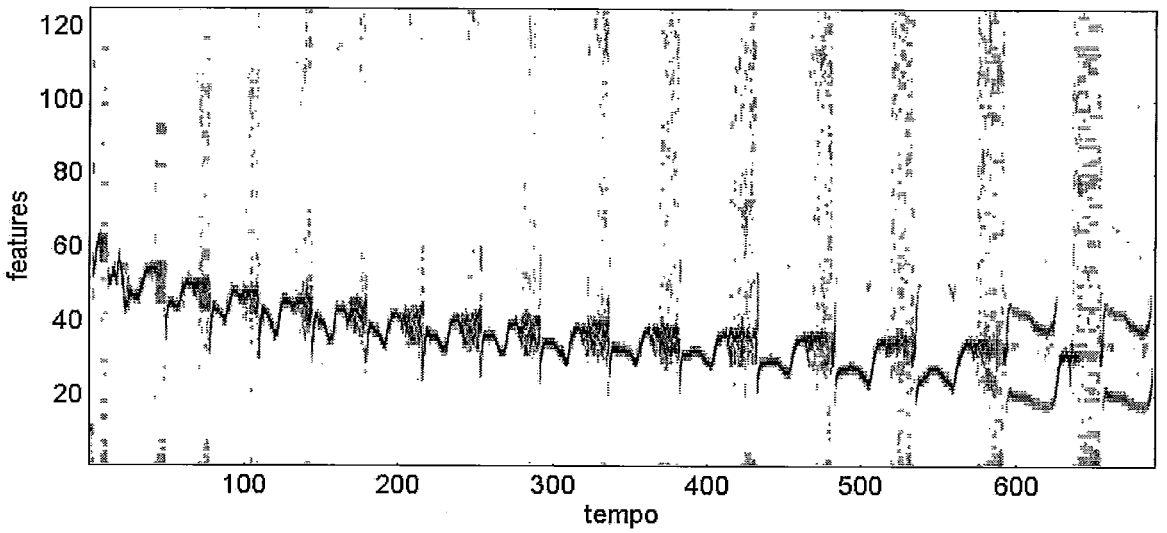
$$\frac{\partial \varphi}{\partial x_i}(x) = \int_a^{g(x)} \frac{\partial f}{\partial x_i}(x,t) dt + \frac{\partial g}{\partial x_i}(x) \cdot f(x, g(x)) \quad (\text{A.I.2})$$

ANEXO II

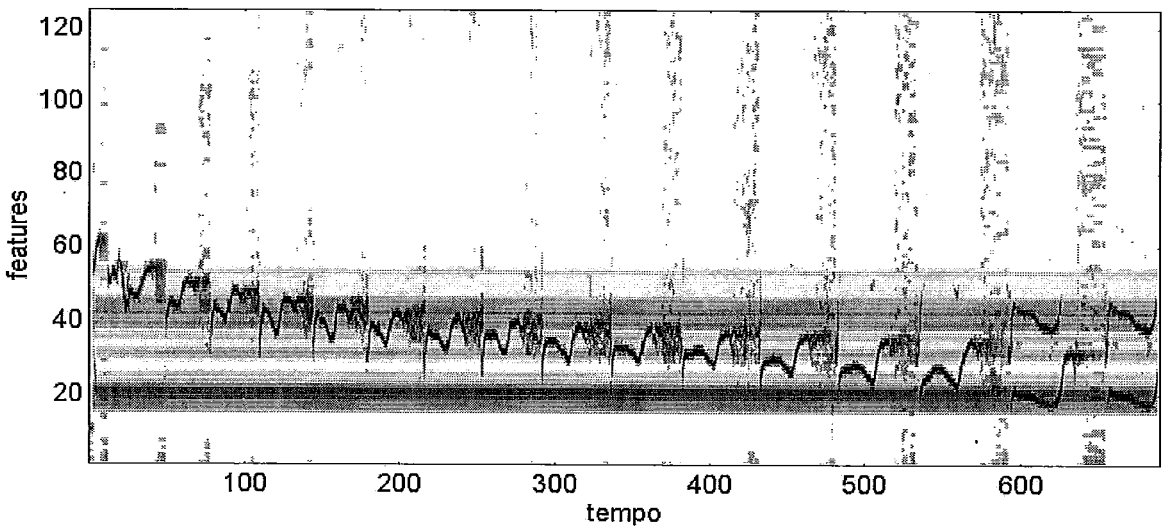
Padrão de Entrada - Pássaro Original



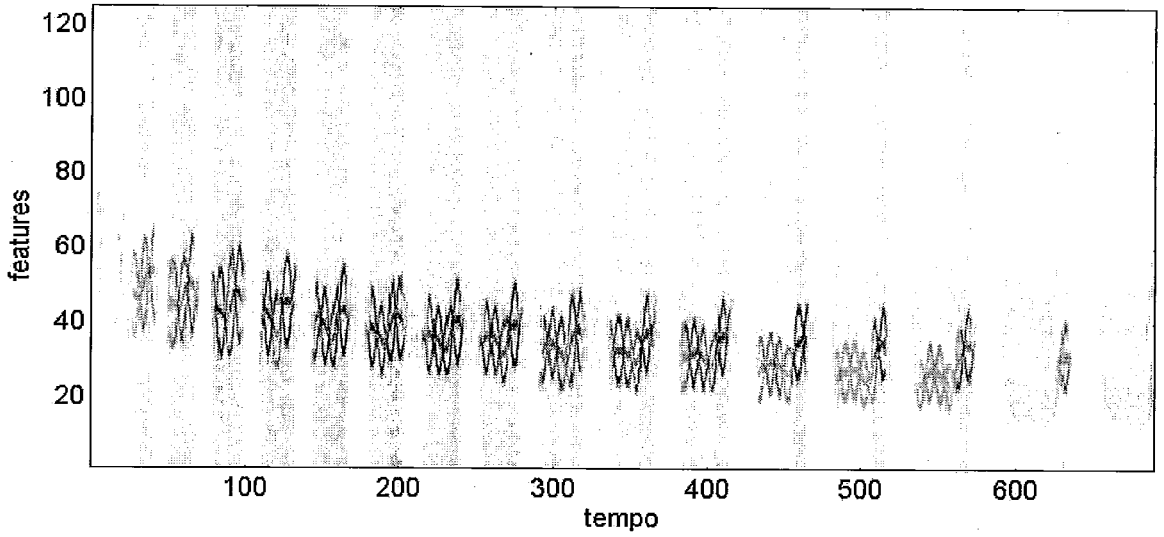
$I(t)$ - Padrão de Entrada Para $P(t)$



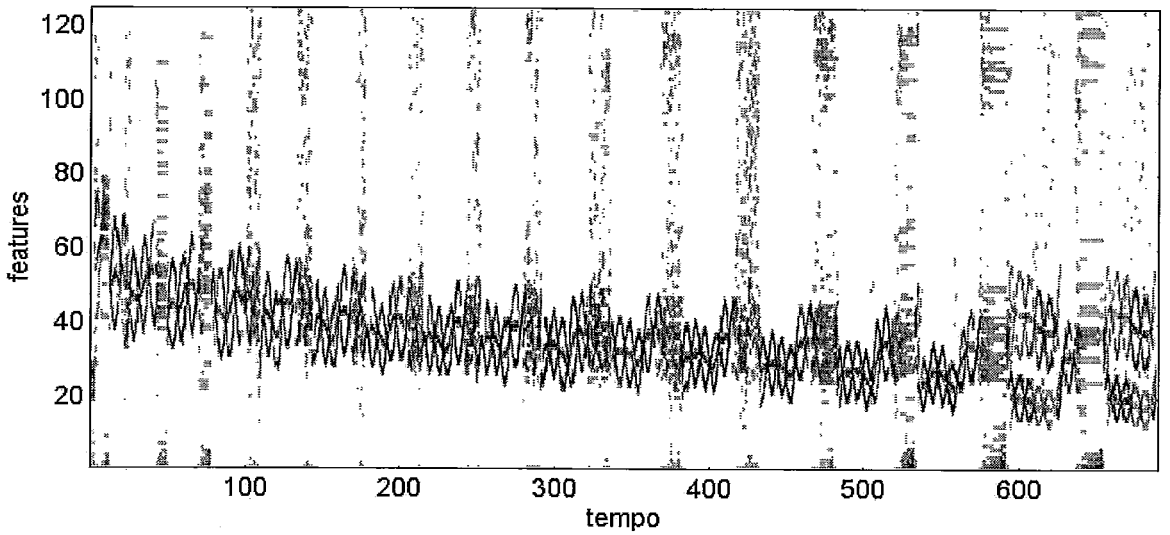
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



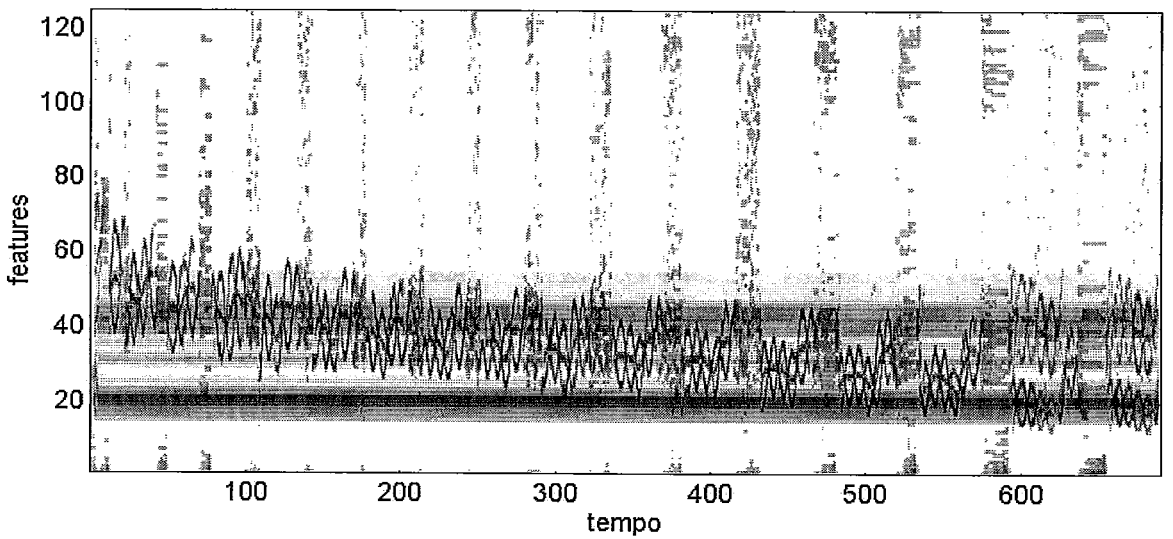
Padrão de Entrada - Pássaro Versão B



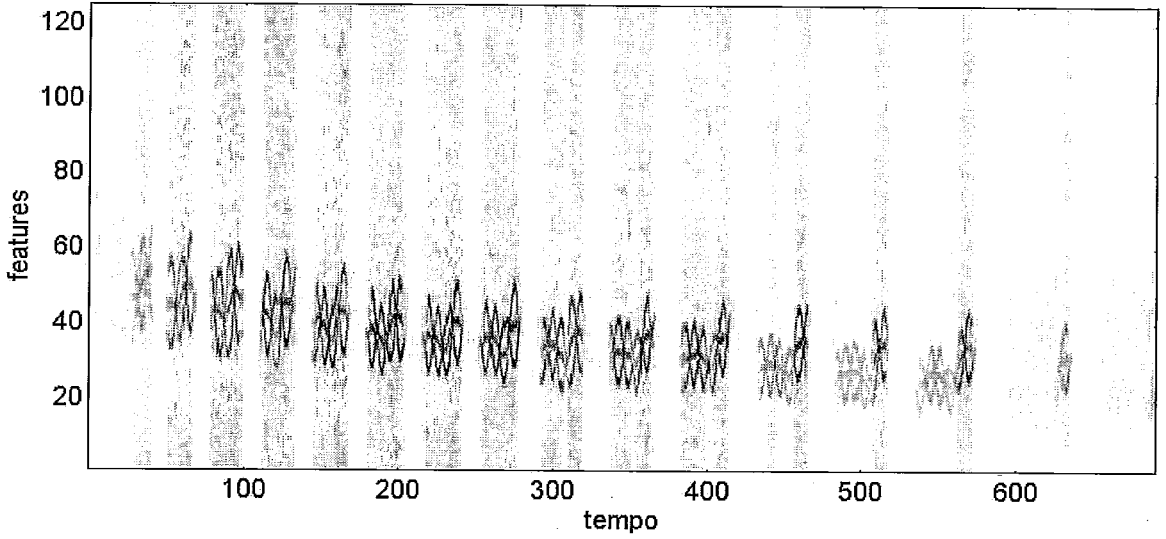
$I(t)$ - Padrão de Entrada Para $P(t)$



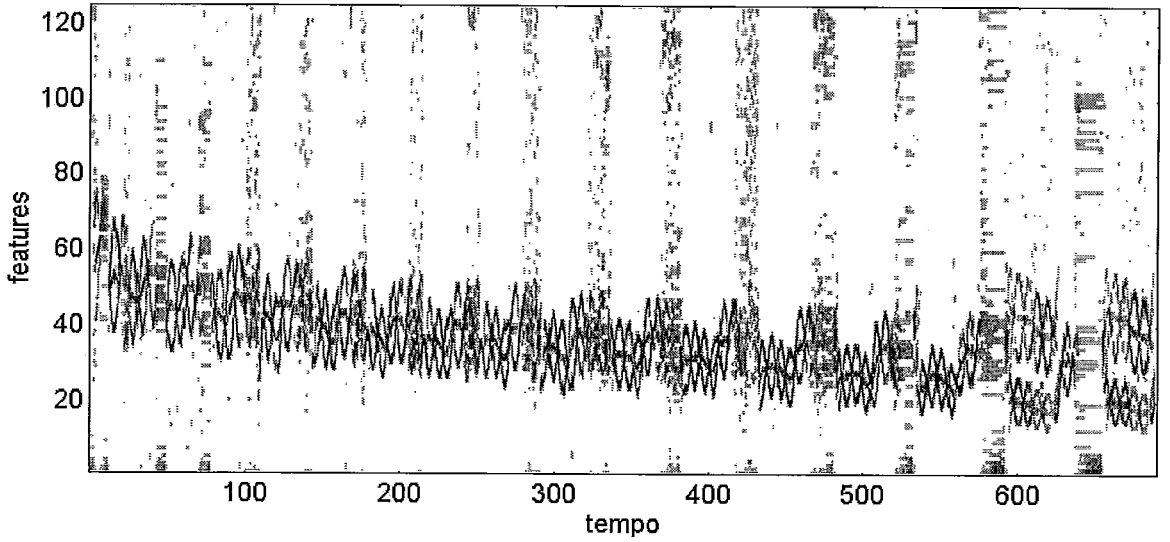
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



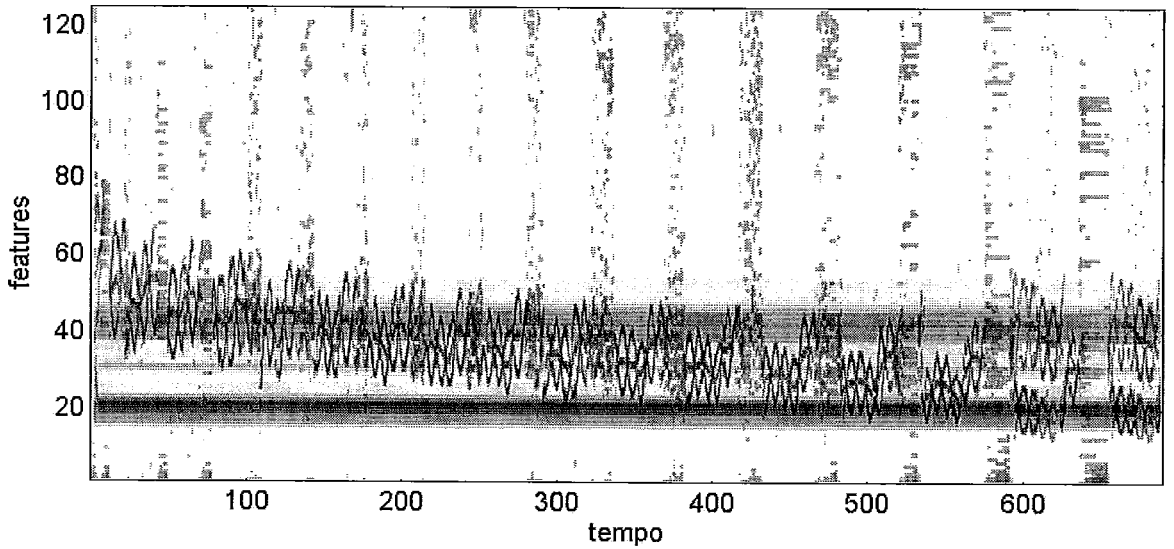
Padrão de Entrada - Pássaro Versão C



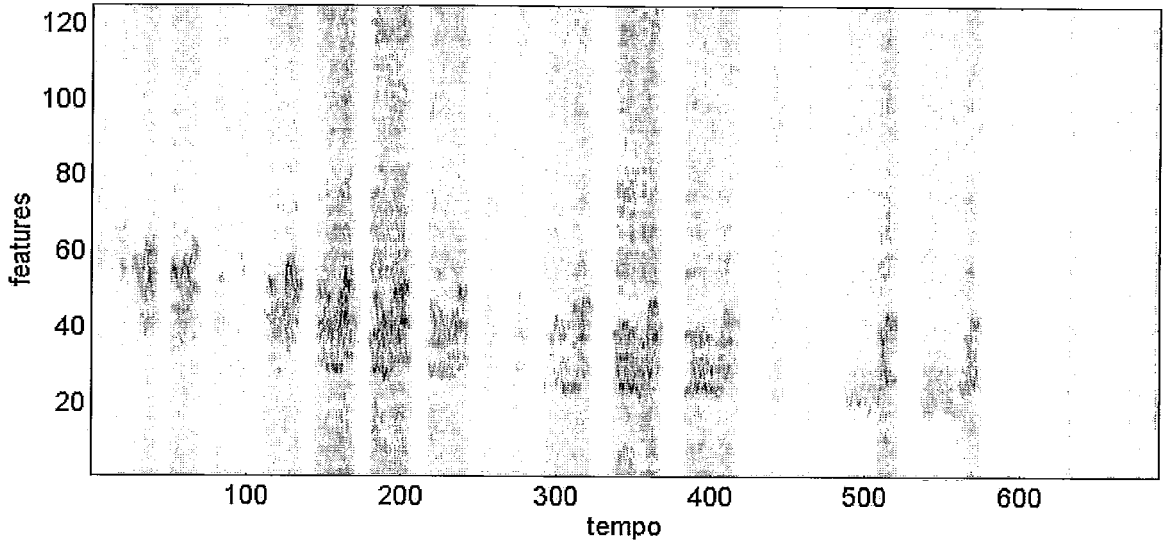
$I(t)$ - Padrão de Entrada Para $P(t)$



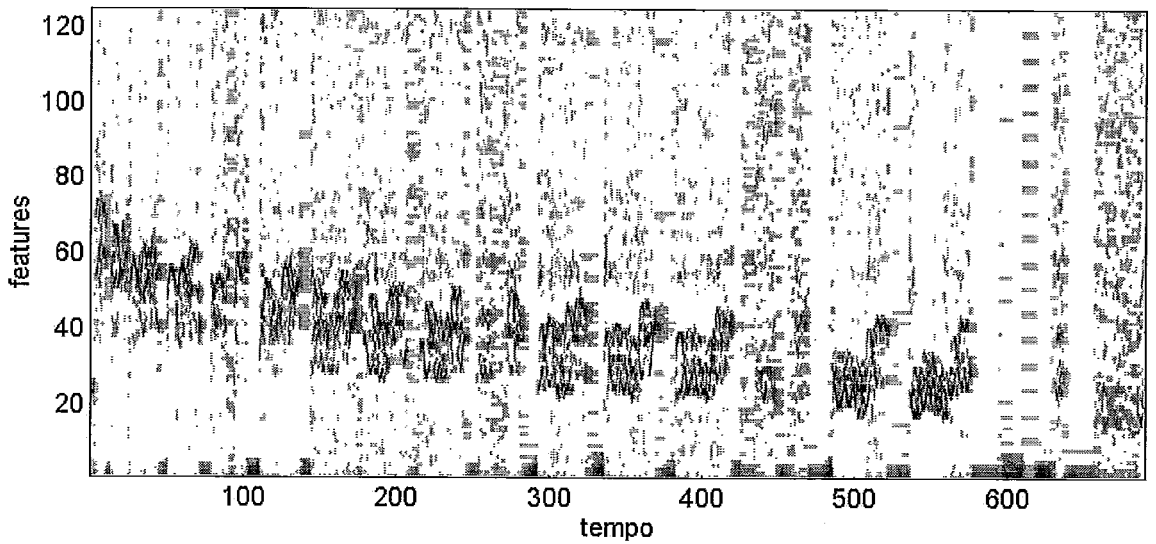
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



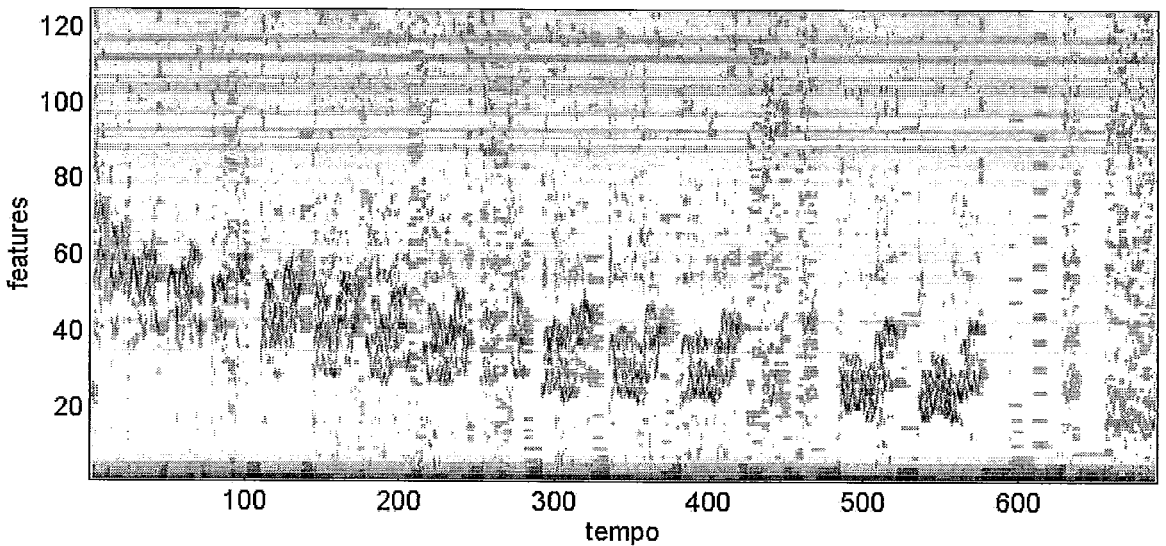
Padrão de Entrada - Pássaro Versão D



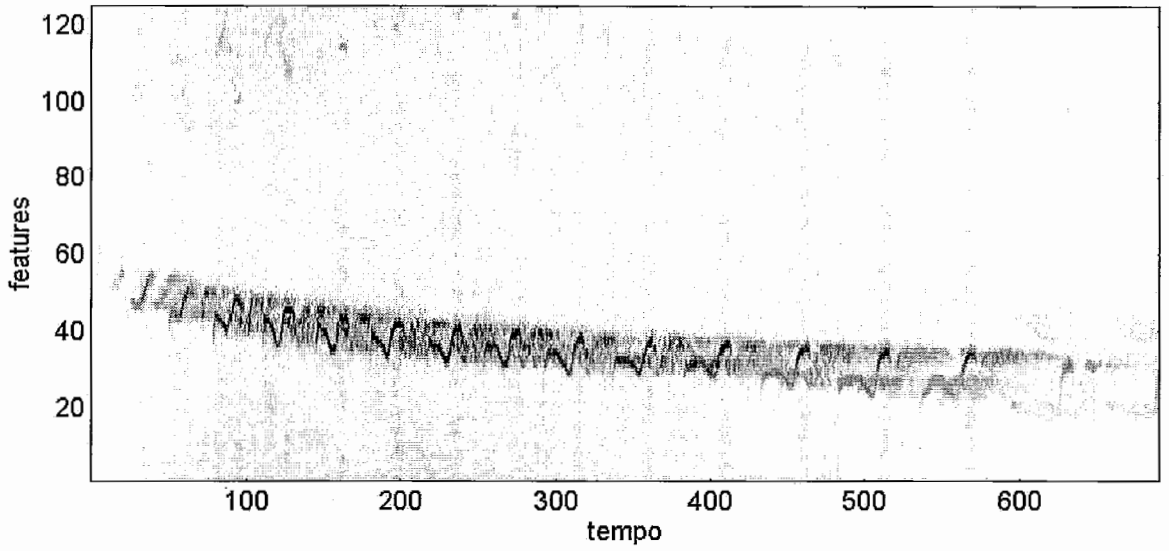
I(t) - Padrão de Entrada Para P(t)



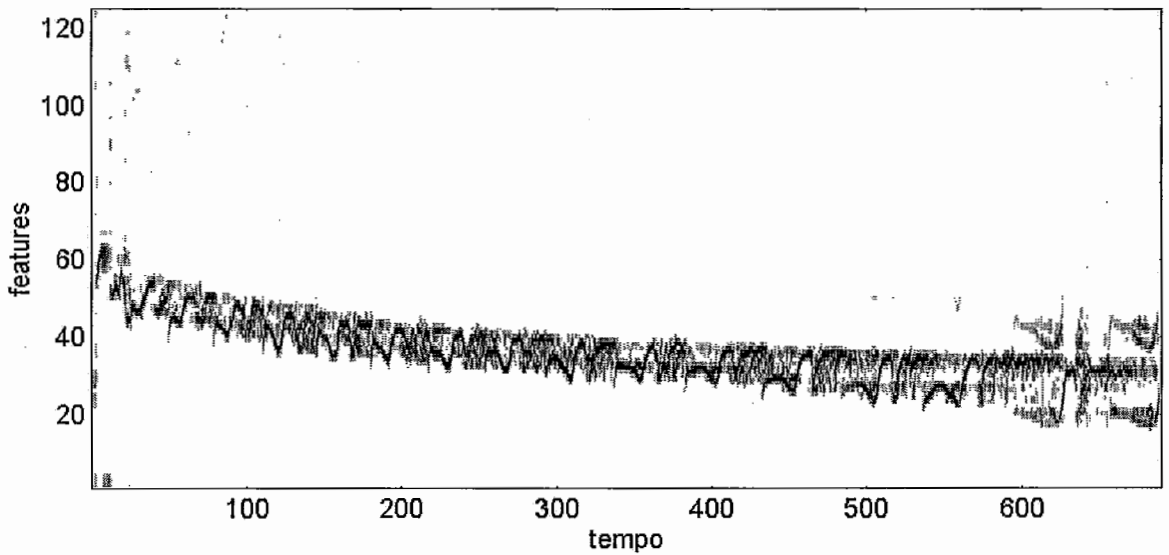
P(t) Após Realimentação do Neurônio J Vencedor na Camada F2



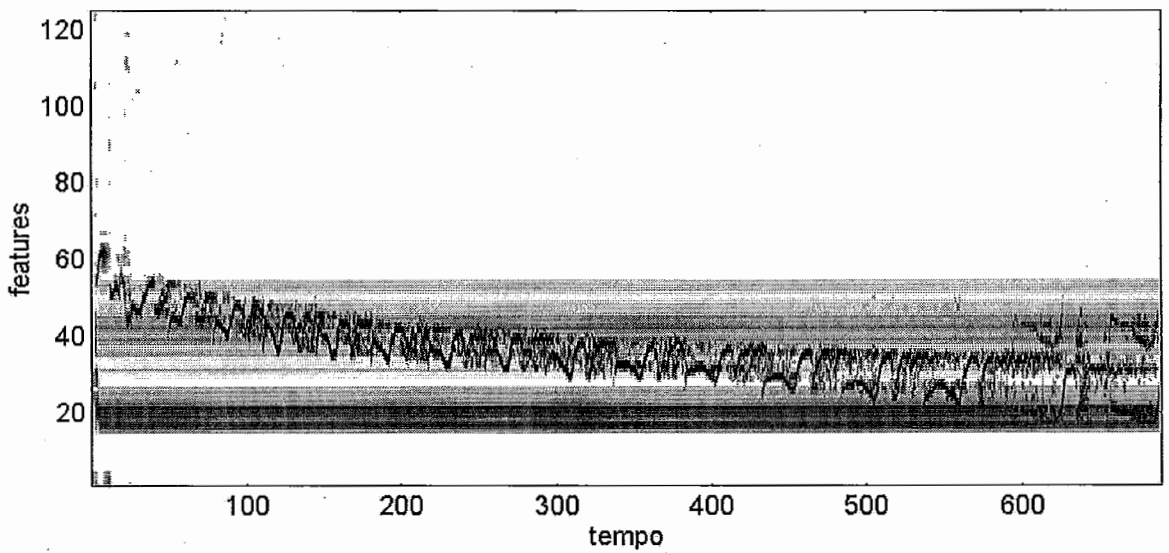
Padrão de Entrada - Pássaro Versão E



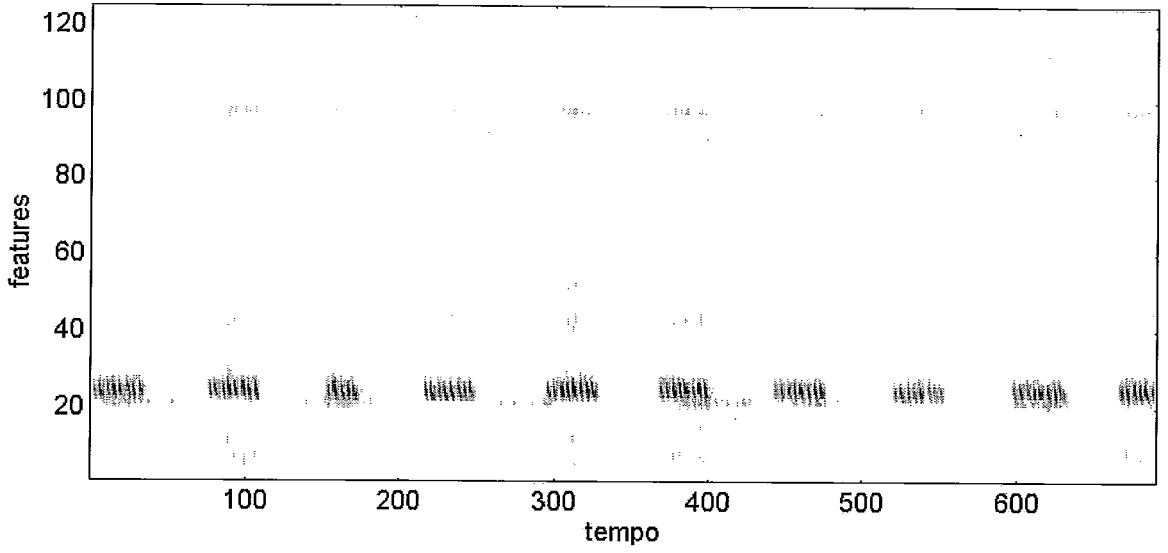
$I(t)$ - Padrão de Entrada Para $P(t)$



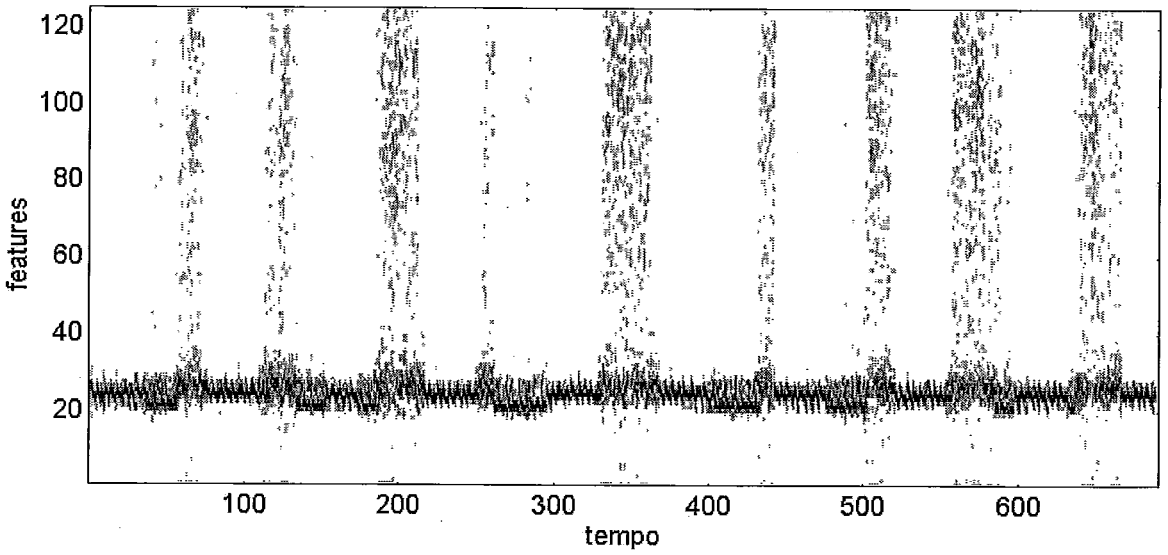
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



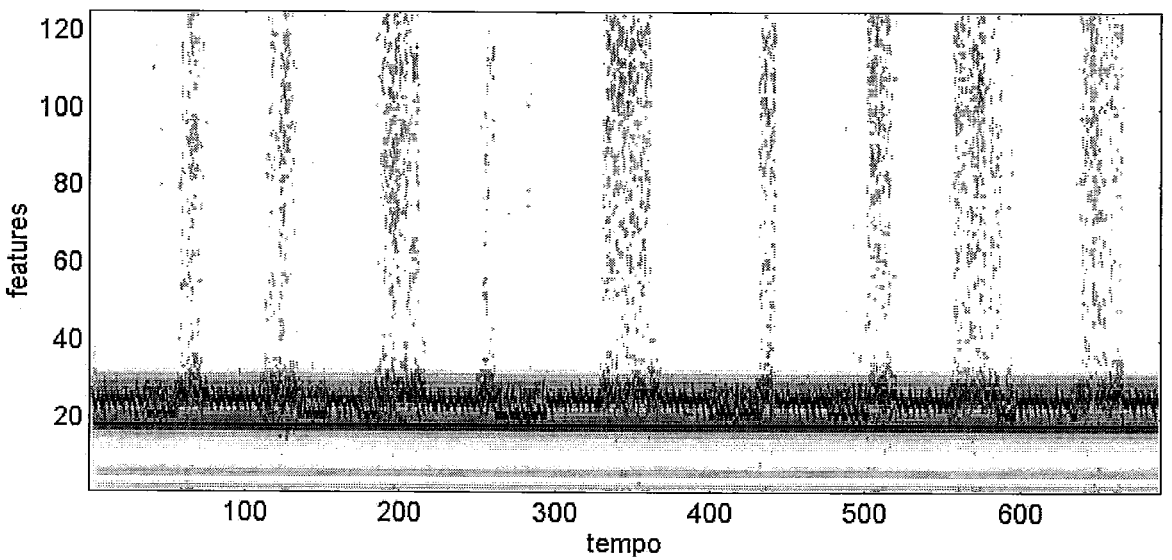
Padrão de Entrada - Grilo Original



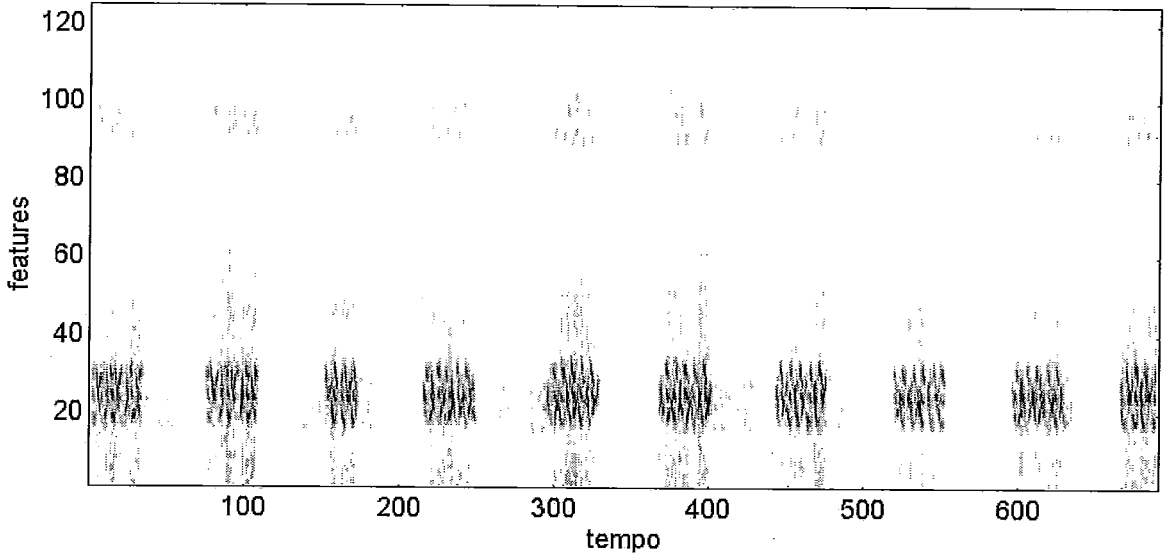
$I(t)$ - Padrão de Entrada Para P(t)



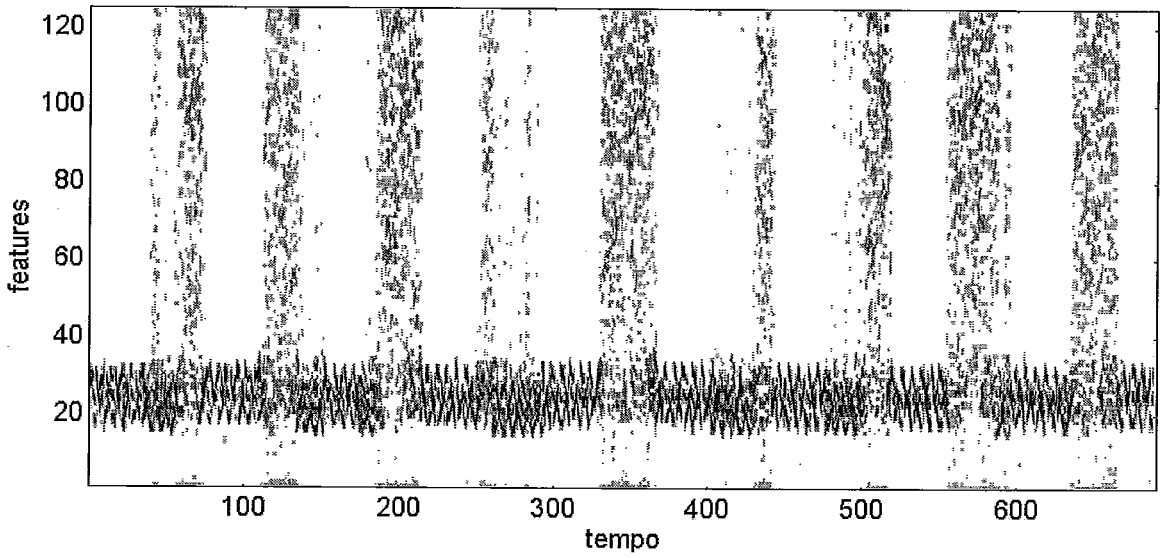
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



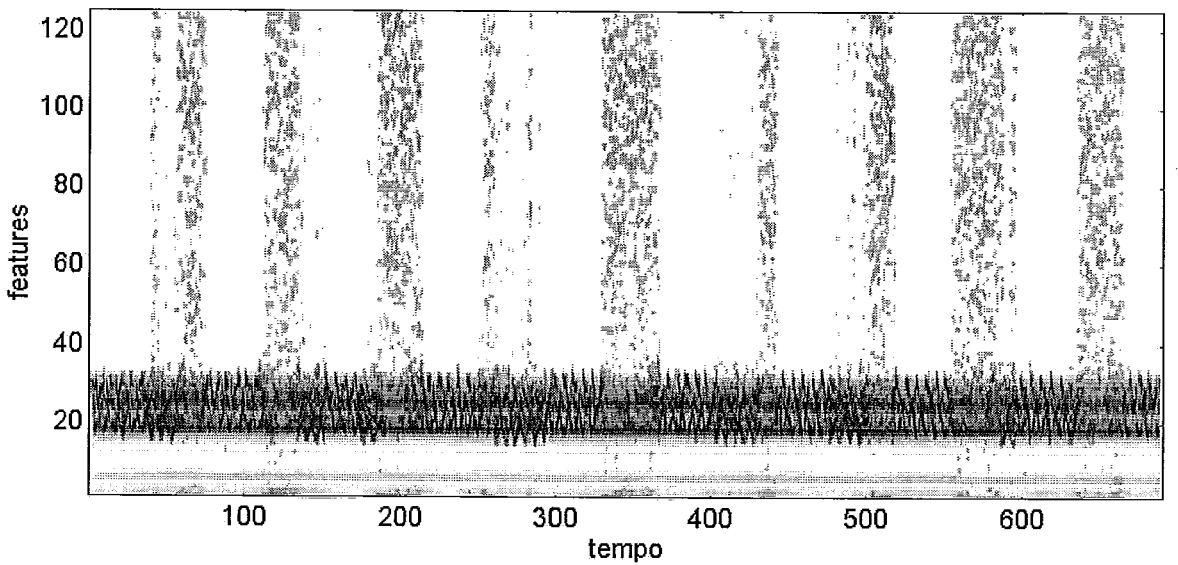
Padrão de Entrada - Grilo Versão B



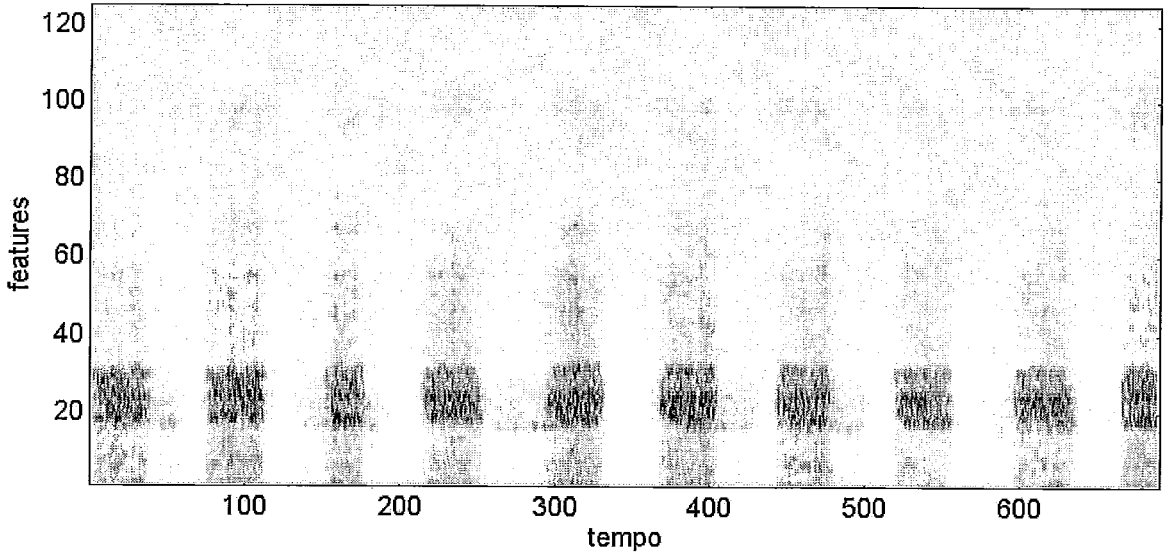
$I(t)$ - Padrão de Entrada Para $P(t)$



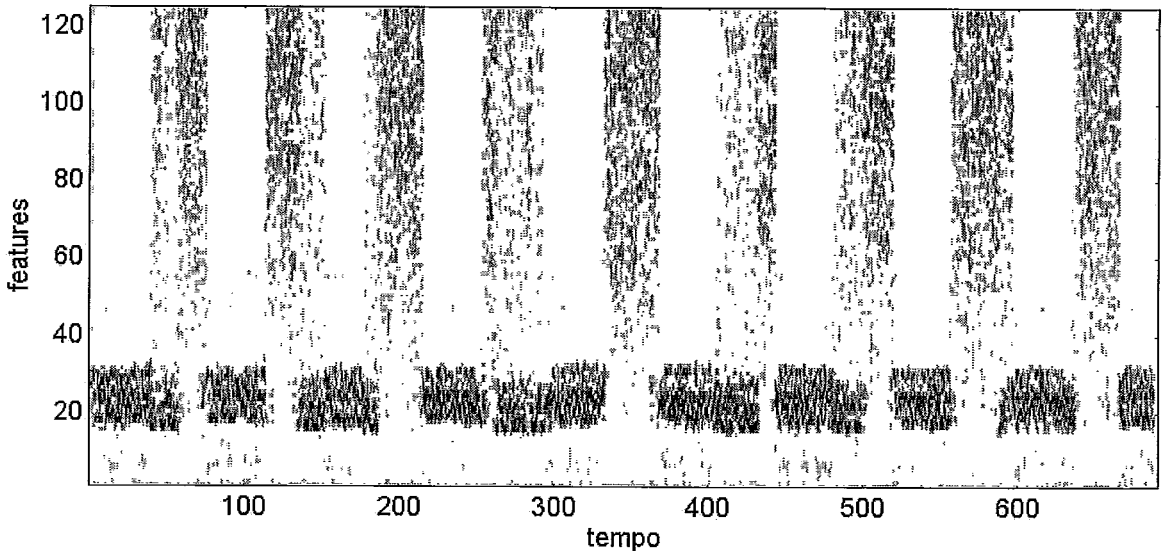
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



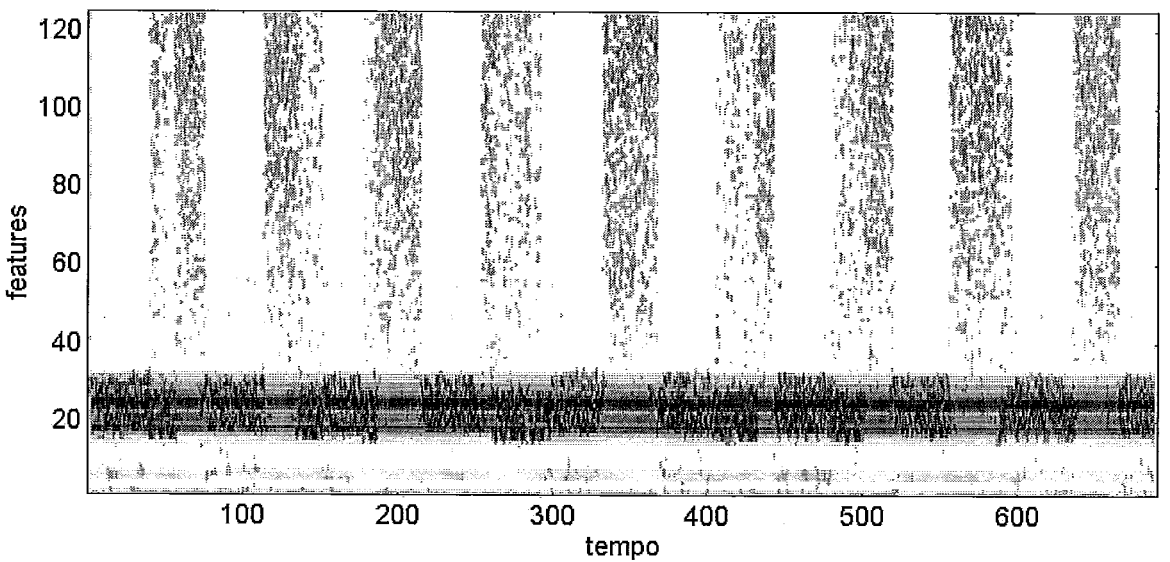
Padrão de Entrada - Grilo Versão C



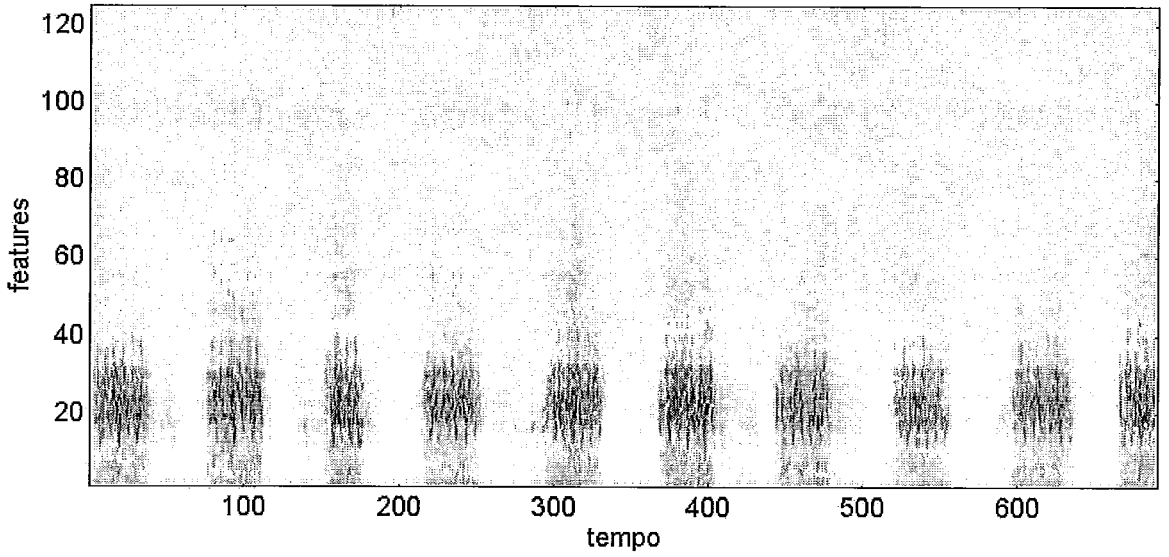
$I(t)$ - Padrão de Entrada Para $P(t)$



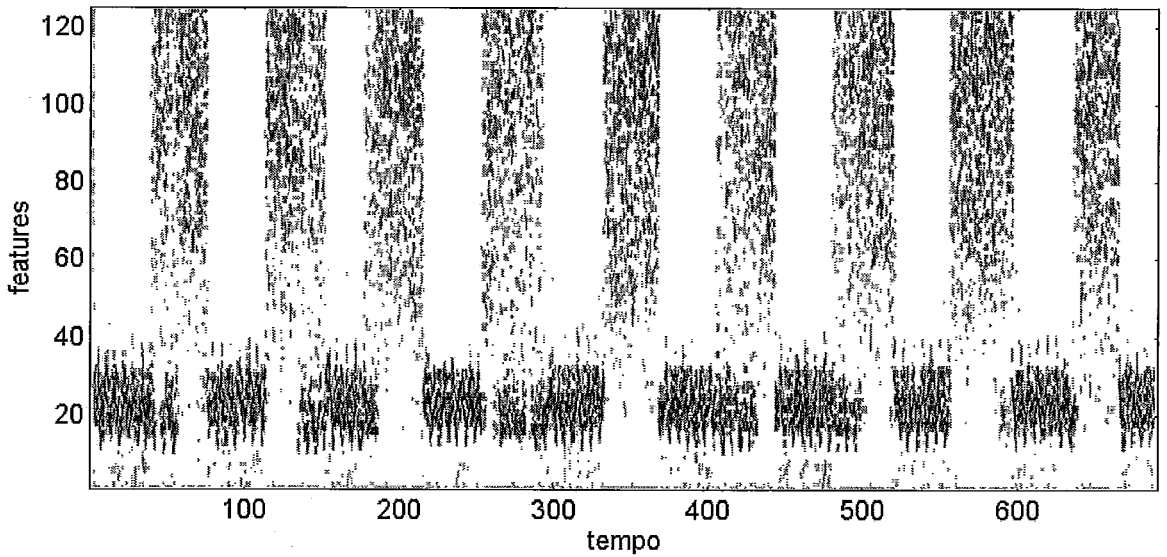
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



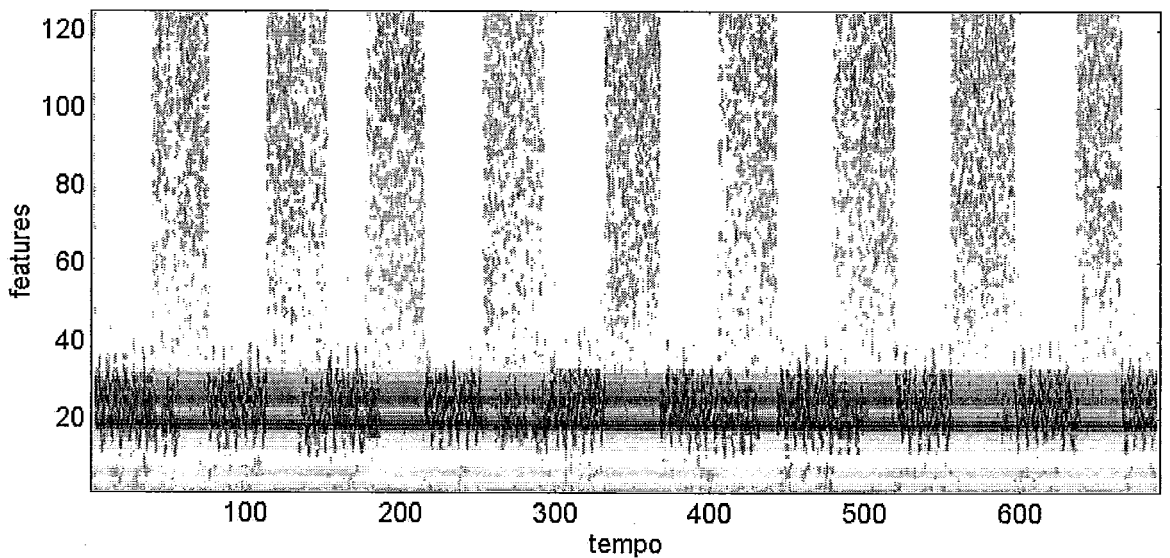
Padrão de Entrada - Grilo Versão D



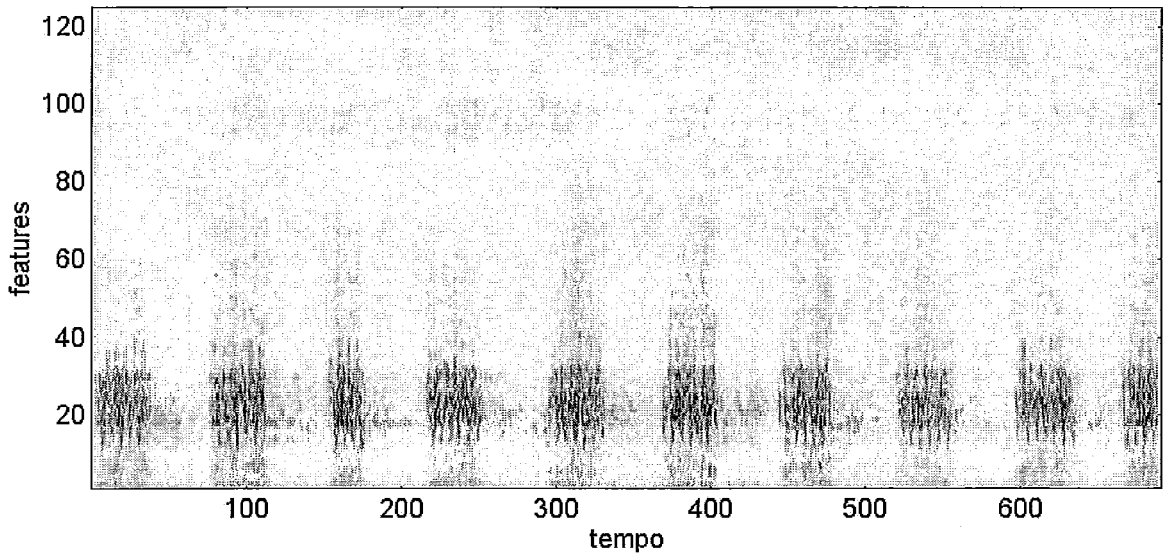
$I(t)$ - Padrão de Entrada Para $P(t)$



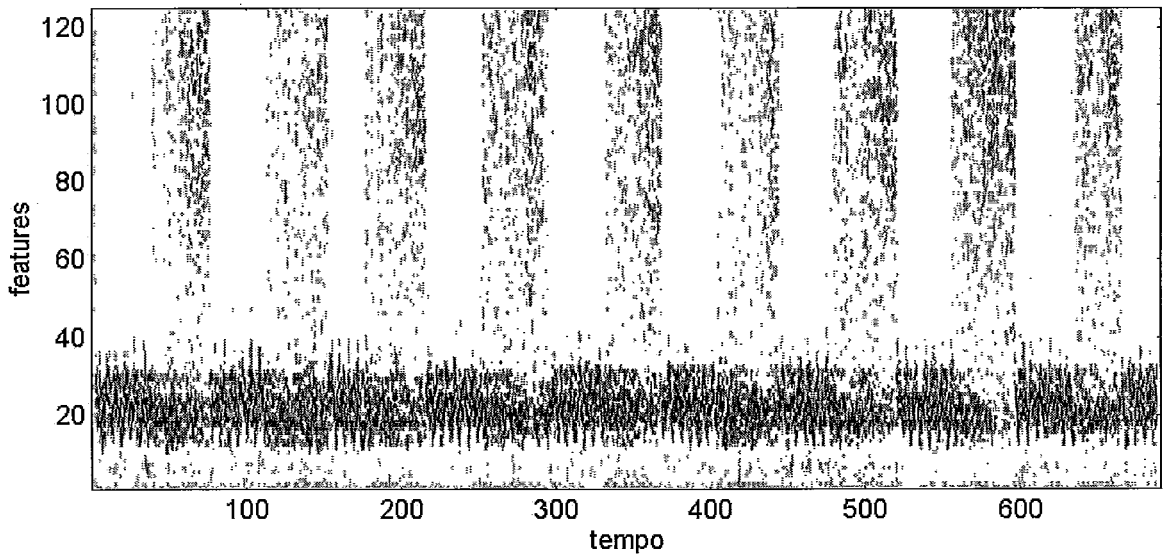
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



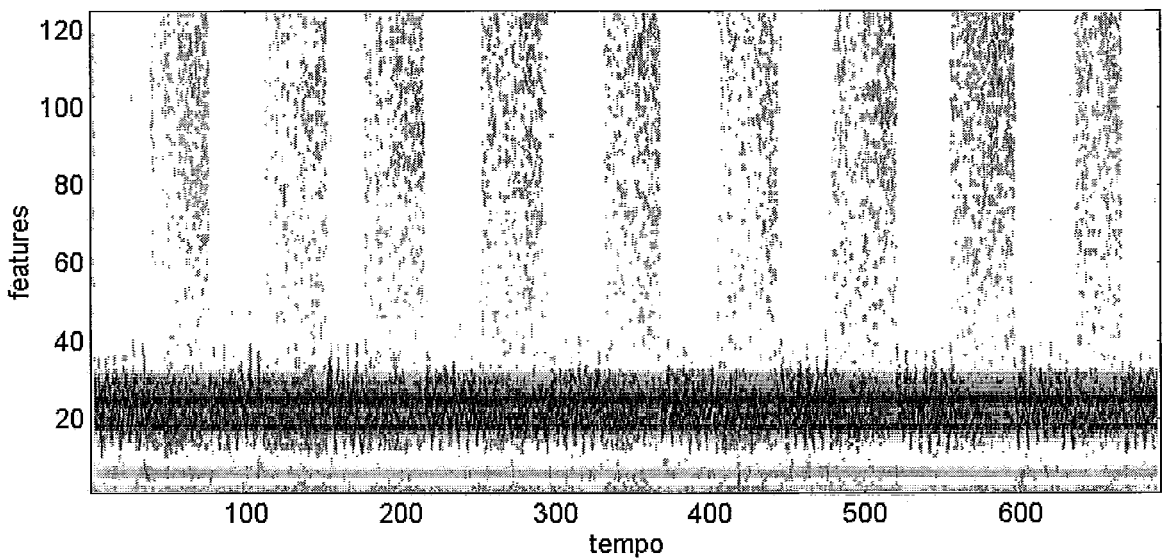
Padrão de Entrada - Grilo Versão E



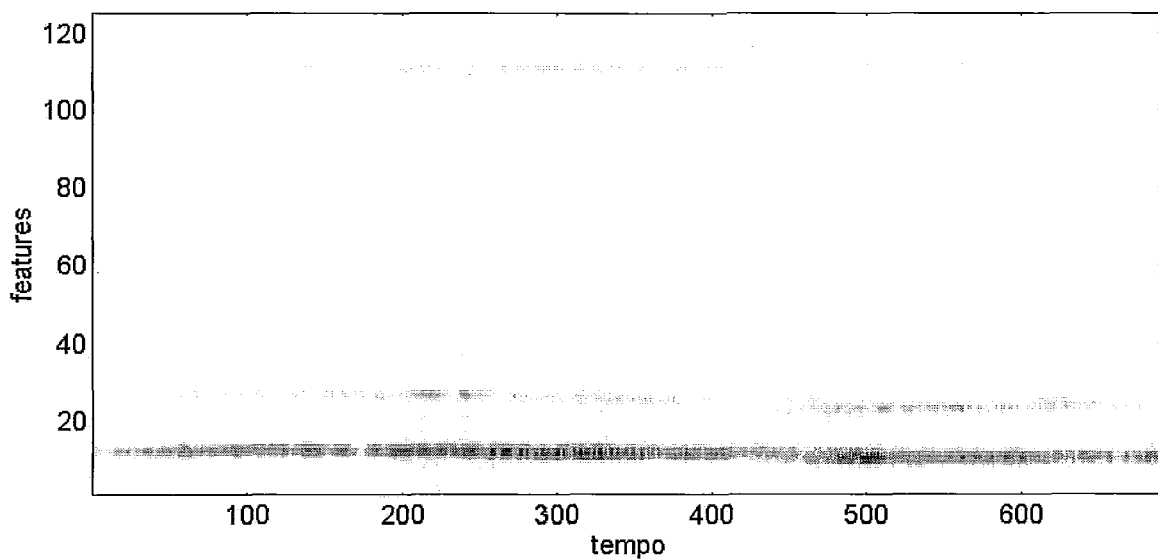
$I(t)$ - Padrão de Entrada Para $P(t)$



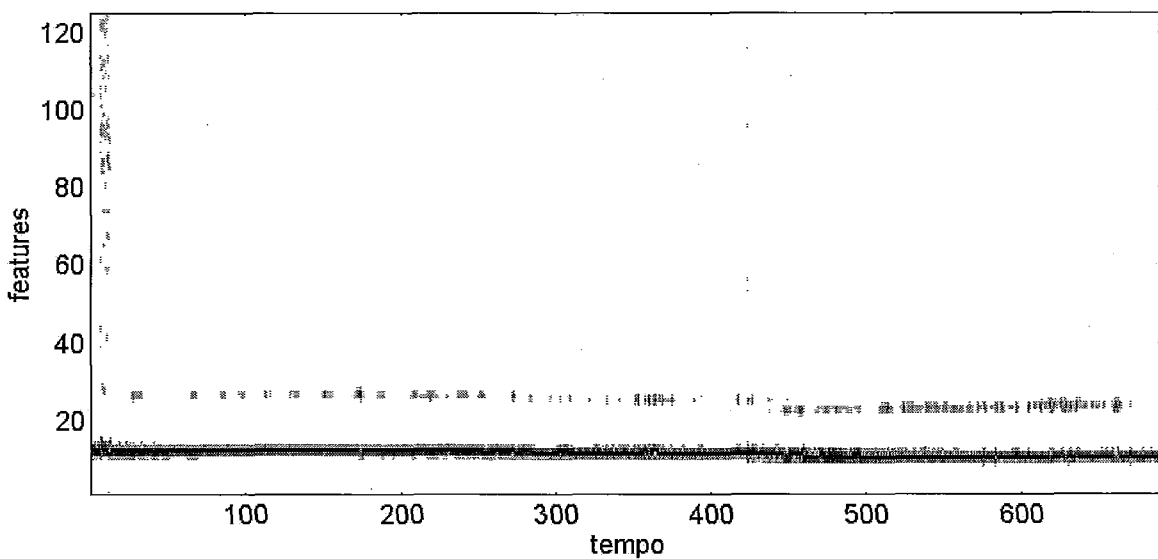
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



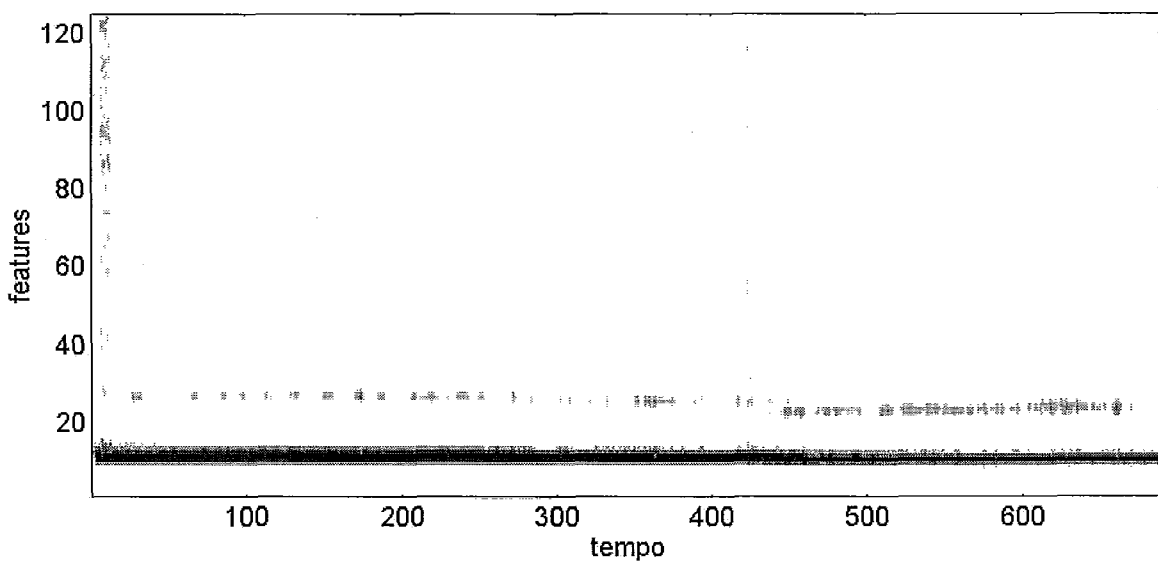
Padrão de Entrada - Macaco Original



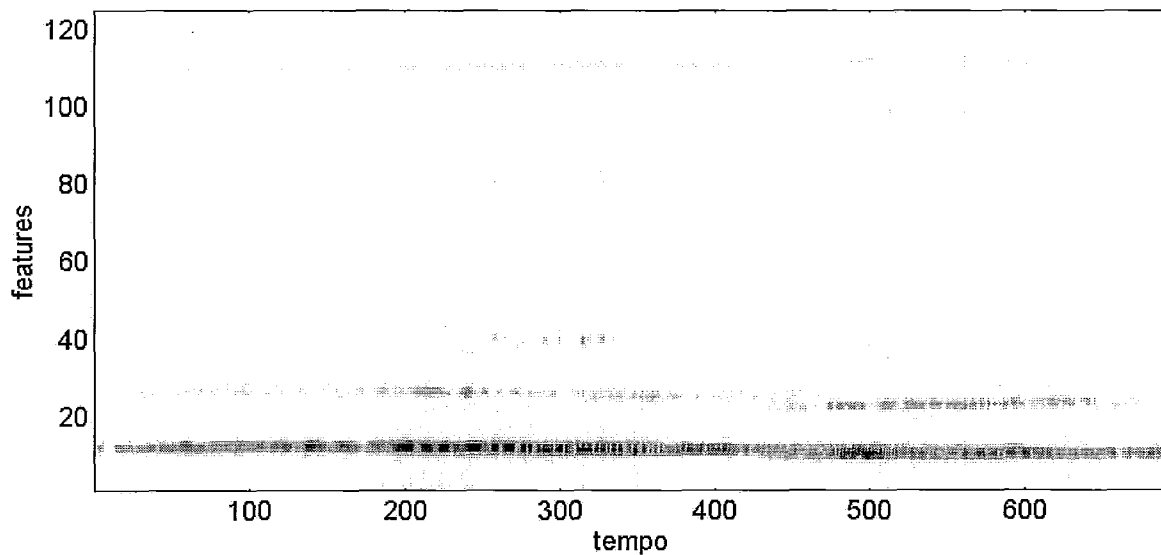
$I(t)$ - Padrão de Entrada Para $P(t)$



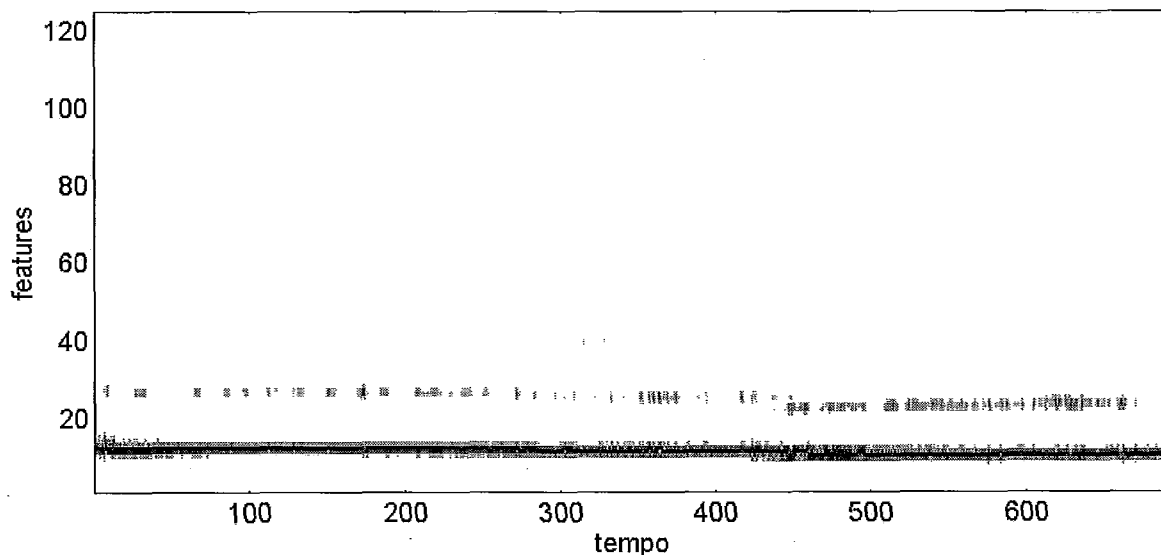
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



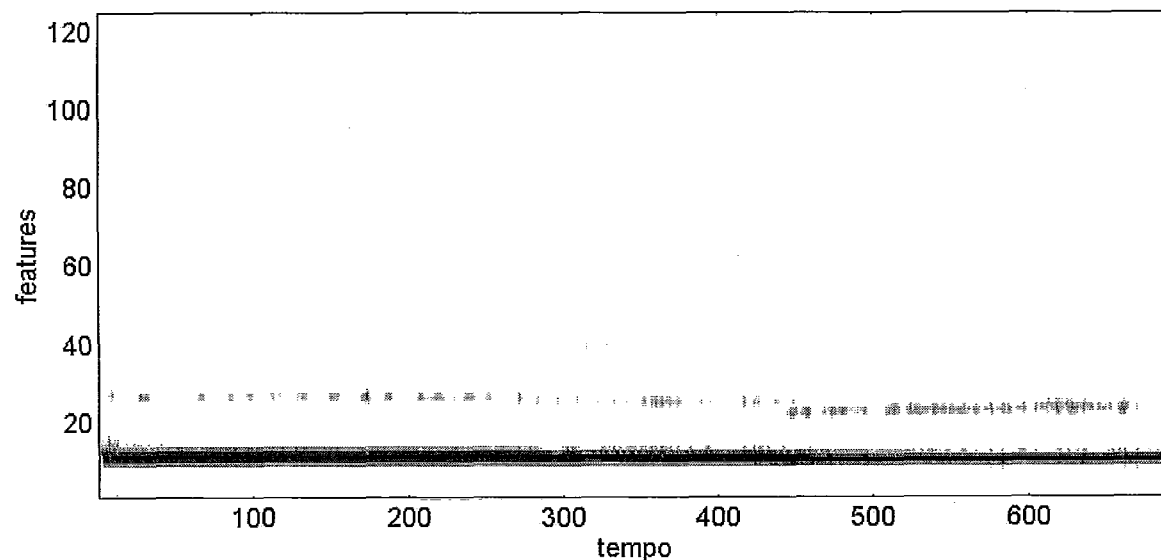
Padrão de Entrada - Macaco Versão B



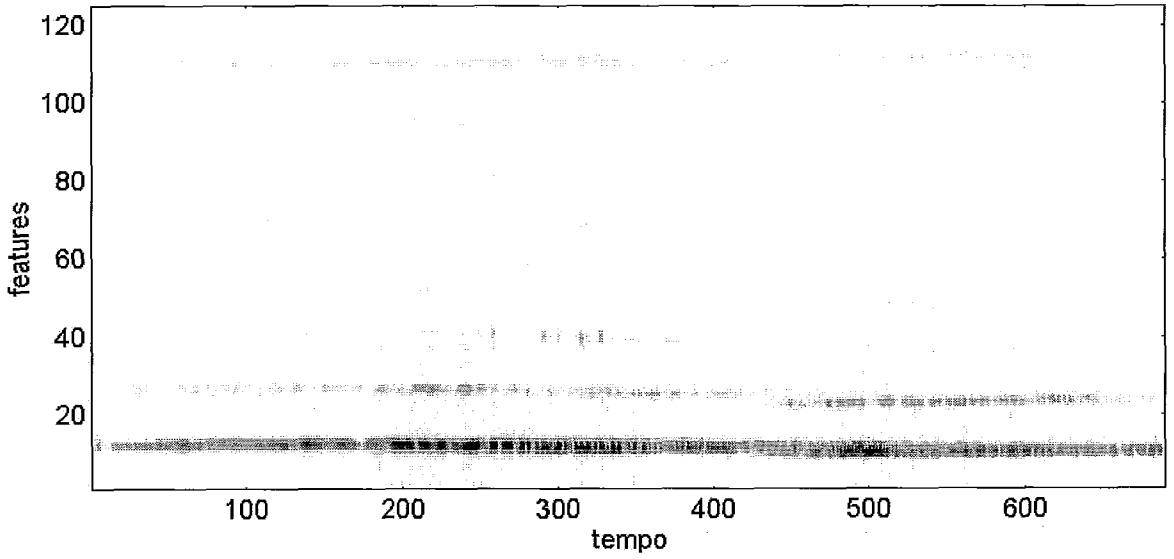
I(t) - Padrão de Entrada Para P(t)



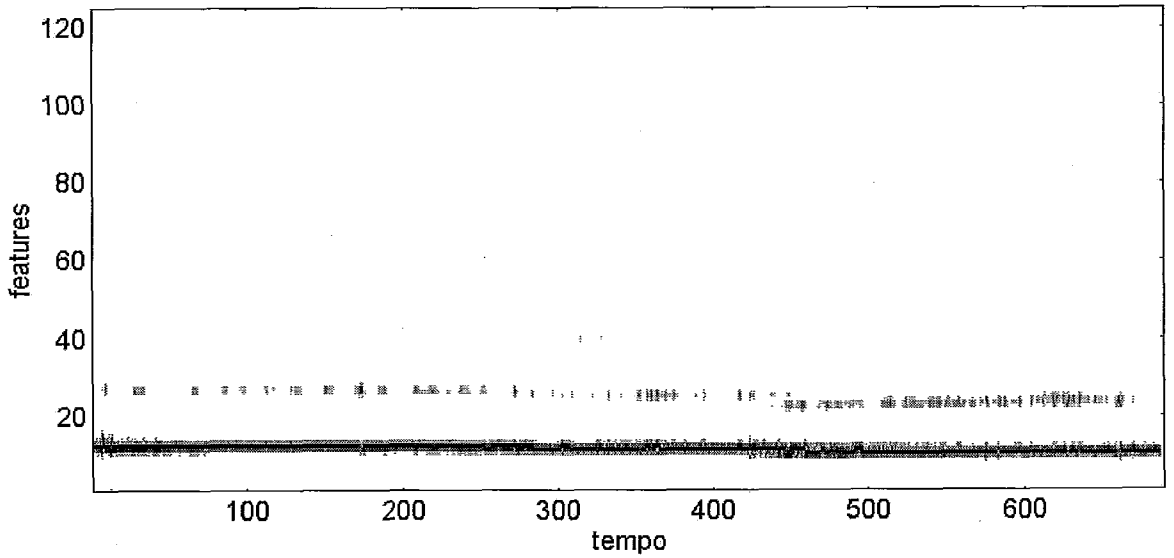
P(t) Após Realimentação do Neurônio J Vencedor na Camada F2



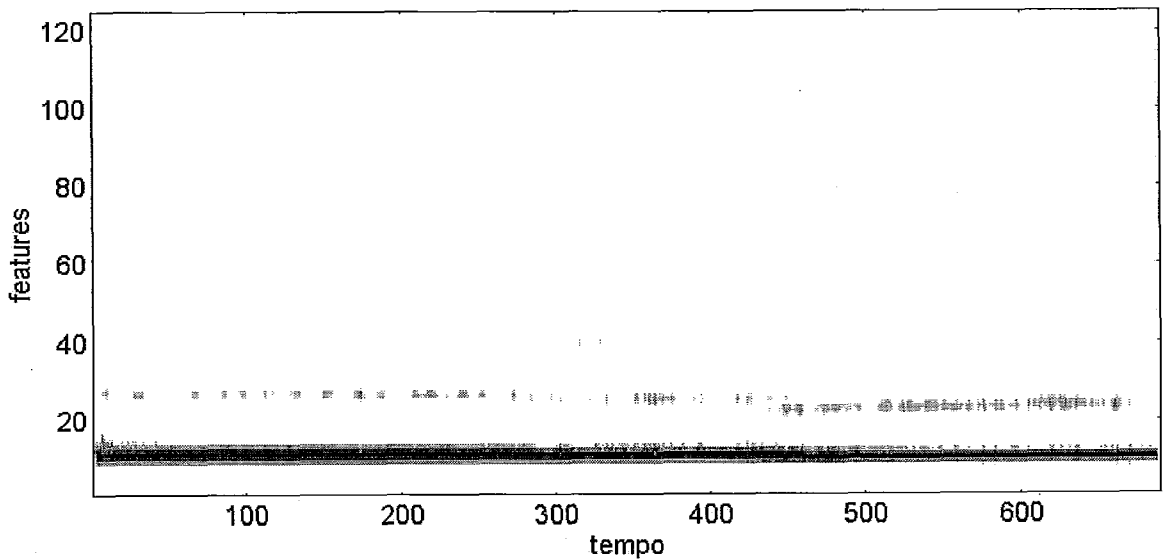
Padrão de Entrada - Macaco Versão C



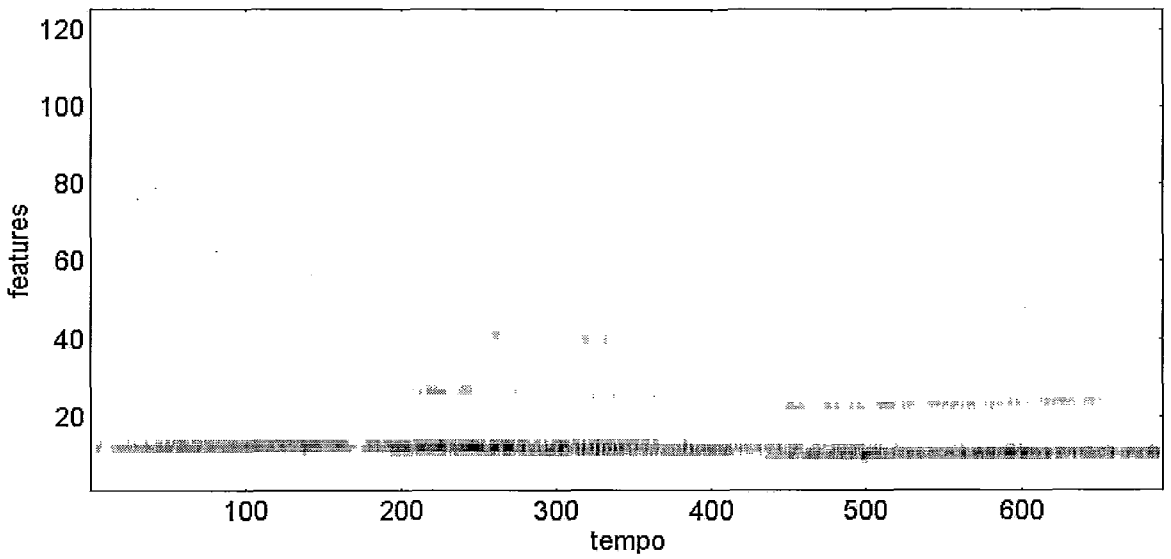
$I(t)$ - Padrão de Entrada Para $P(t)$



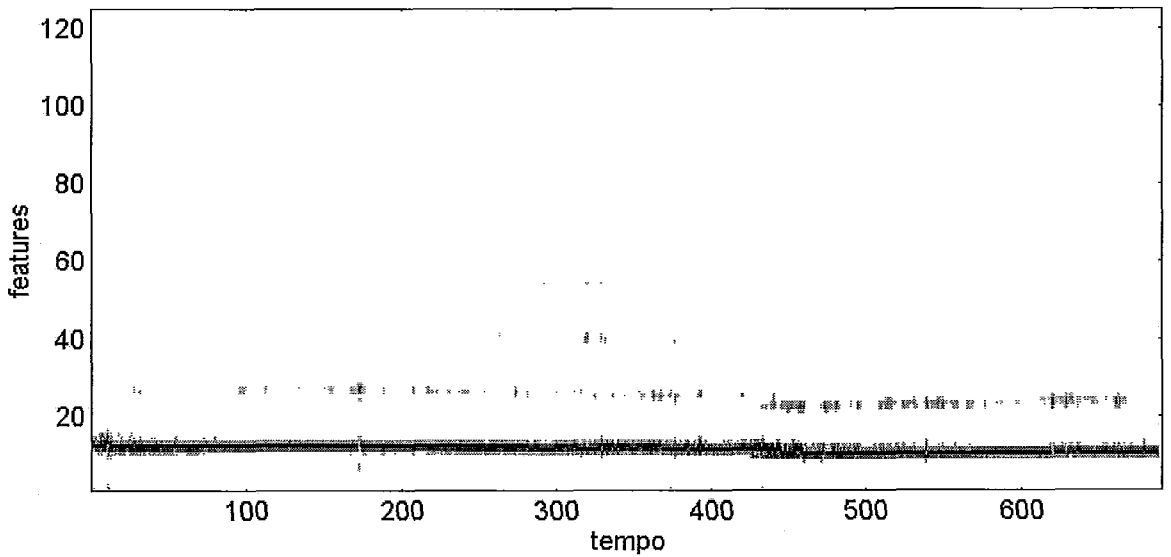
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



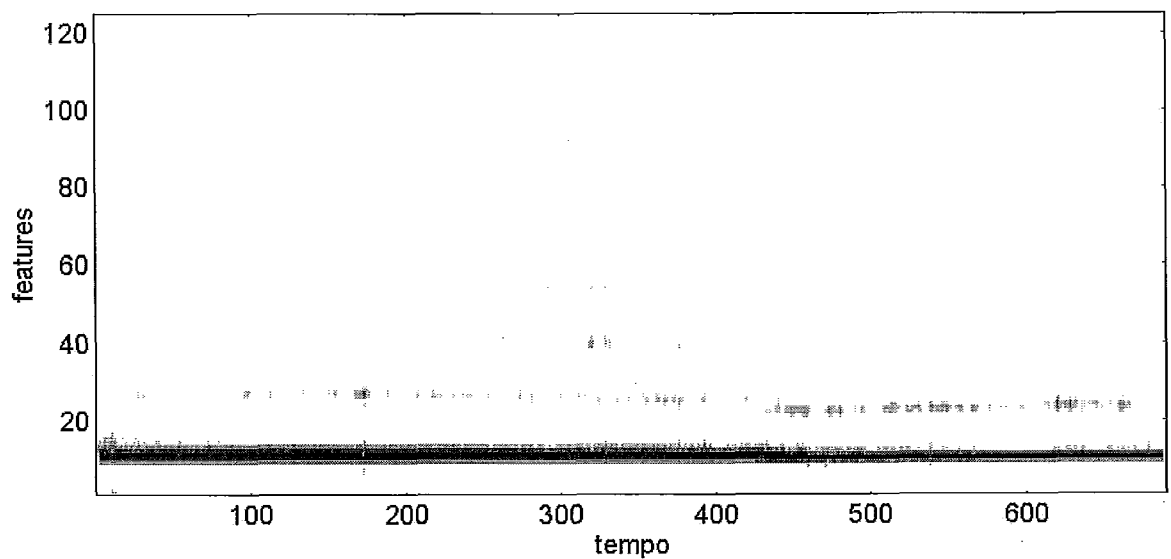
Padrão de Entrada - Macaco Versão D



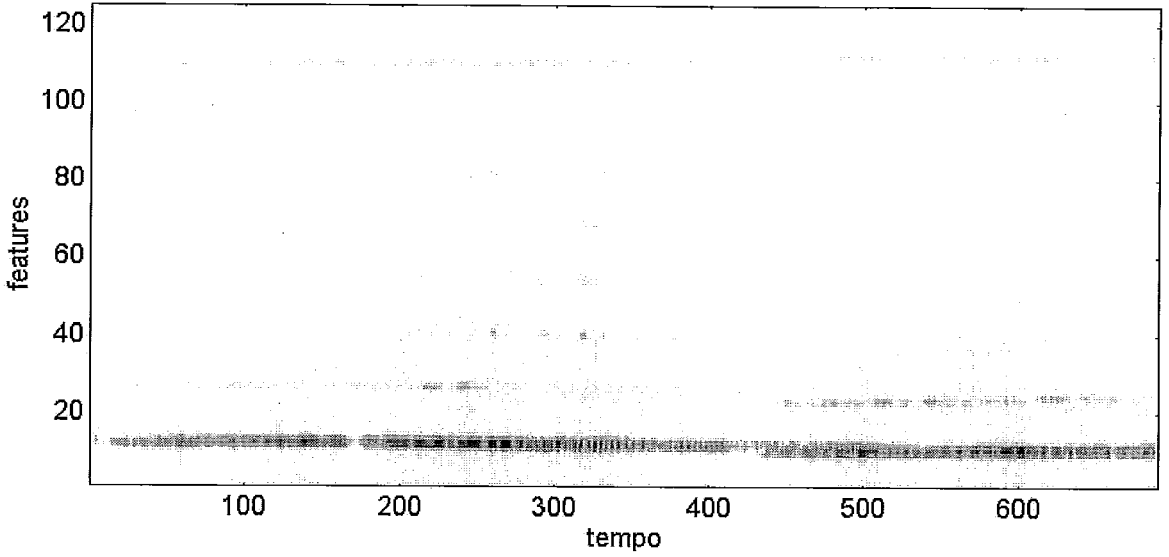
$I(t)$ - Padrão de Entrada Para $P(t)$



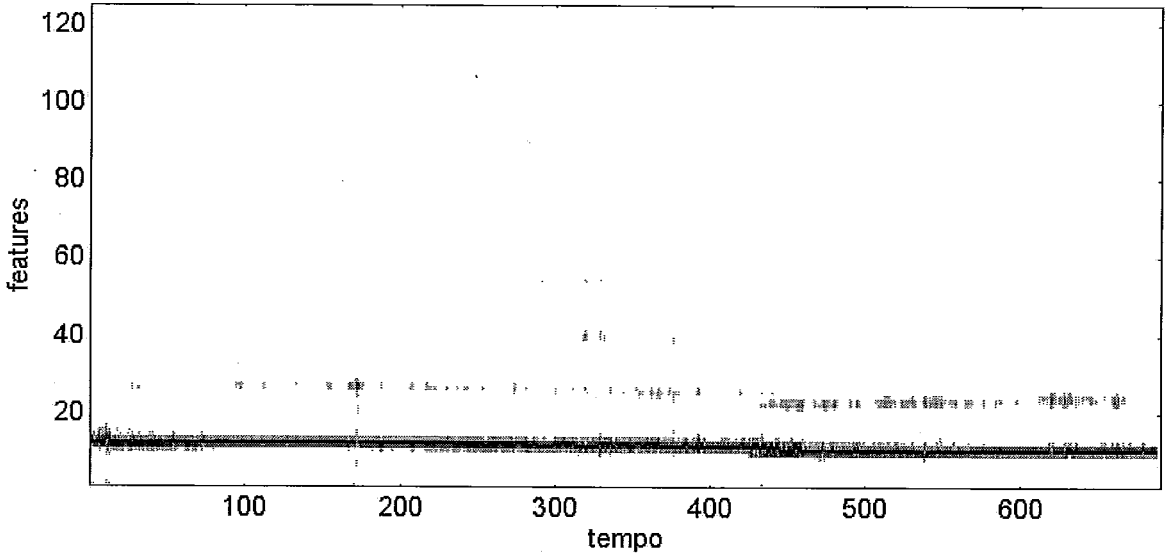
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



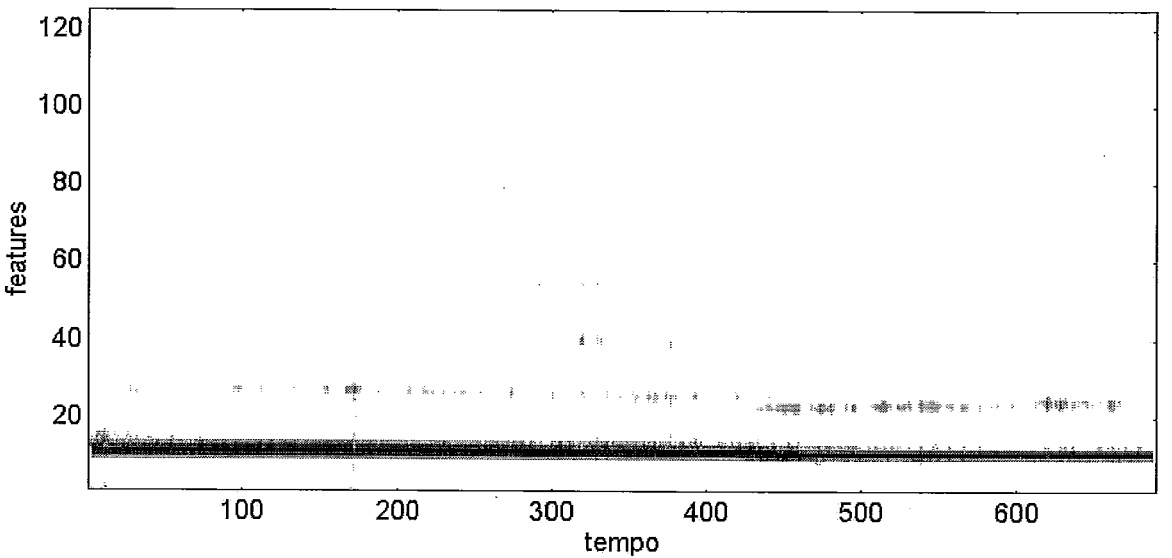
Padrão de Entrada - Macaco Versão E



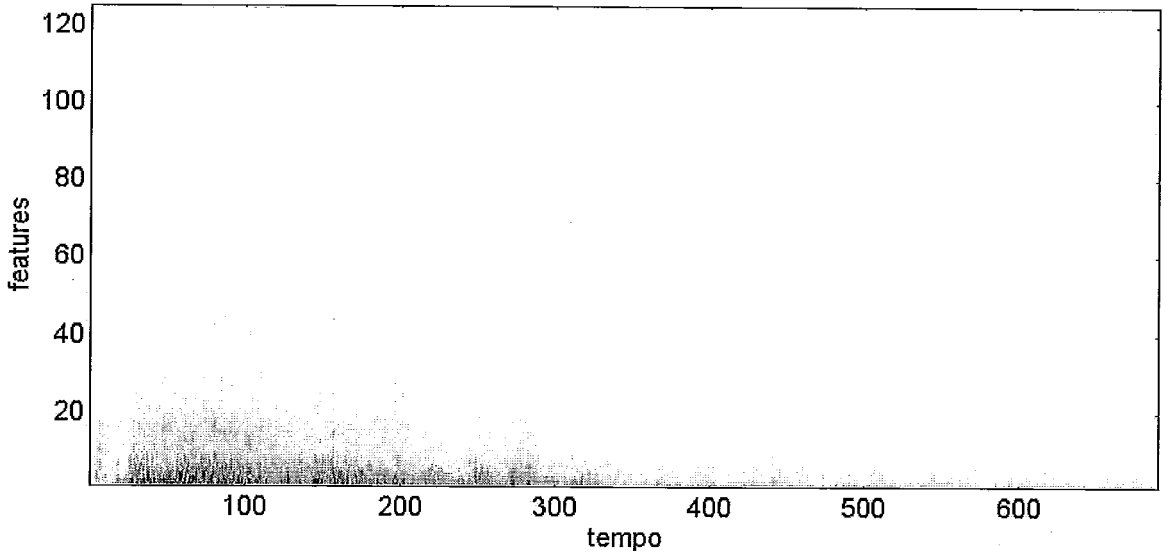
I(t) - Padrão de Entrada Para P(t)



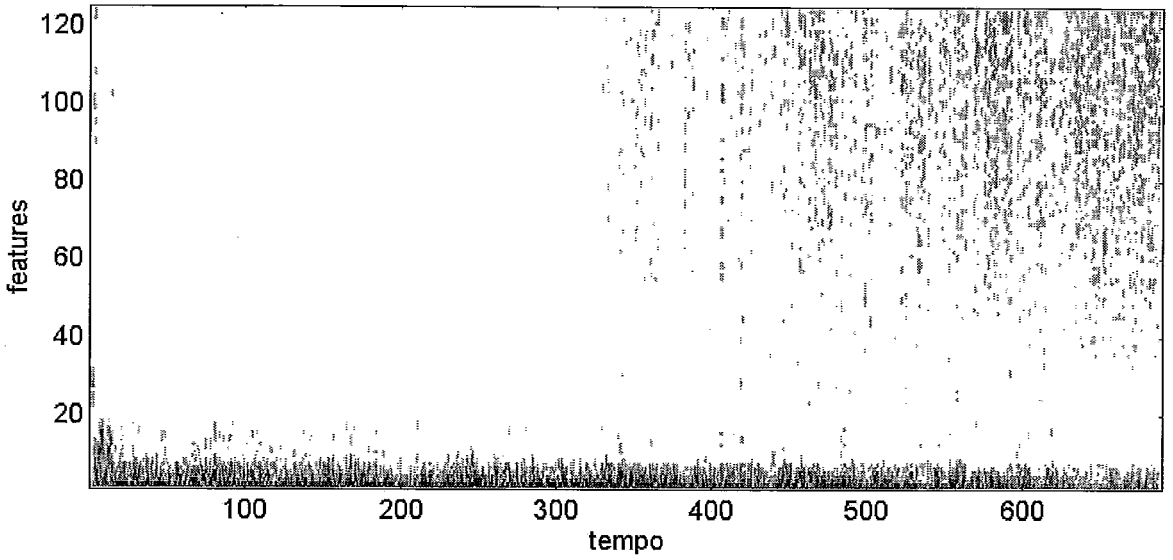
P(t) Após Realimentação do Neurônio J Vencedor na Camada F2



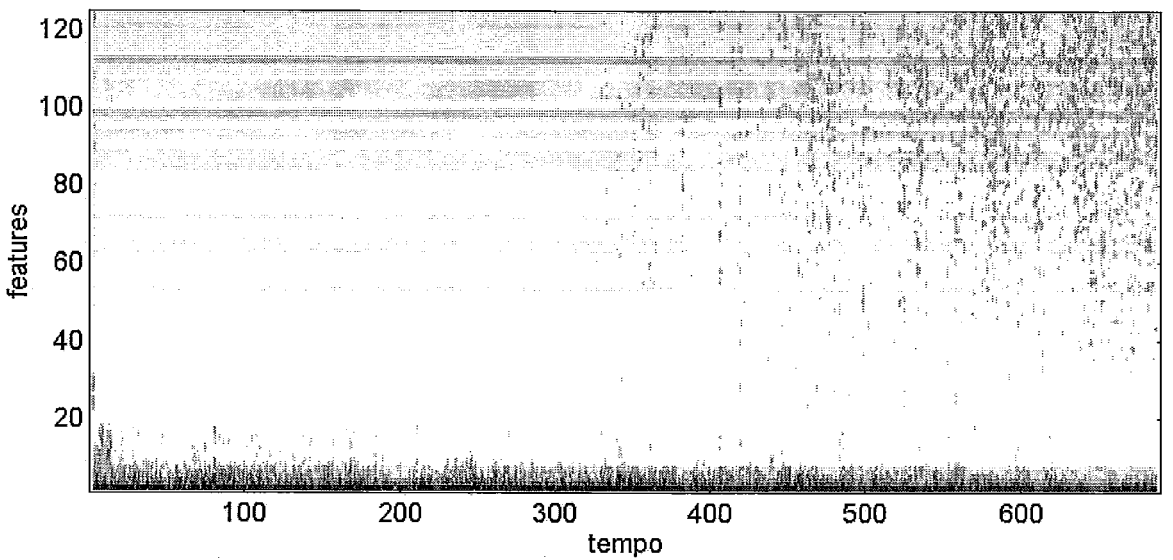
Padrão de Entrada - Trovão Original



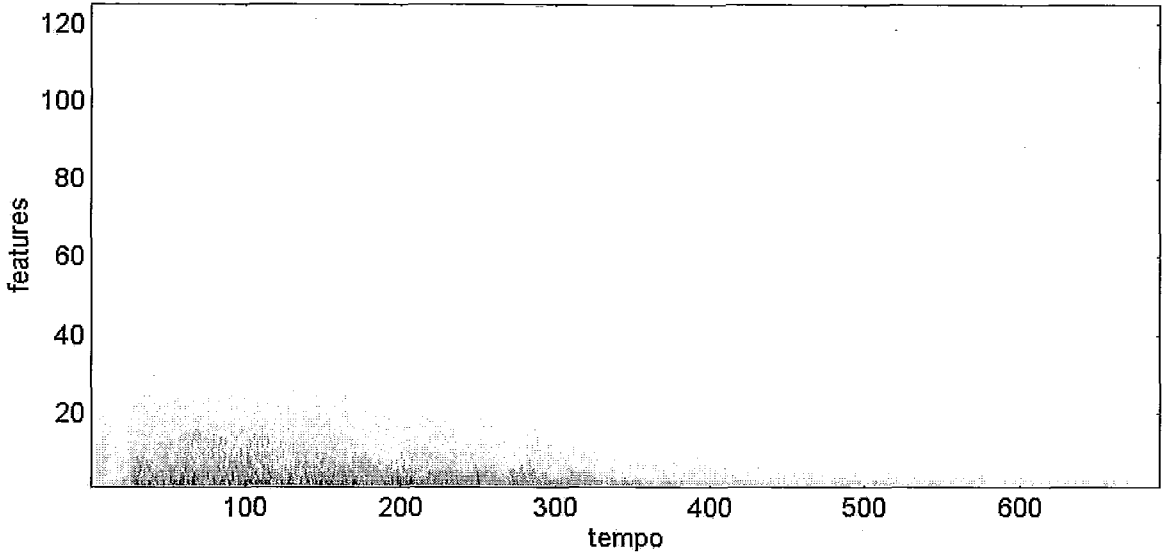
$I(t)$ - Padrão de Entrada Para $P(t)$



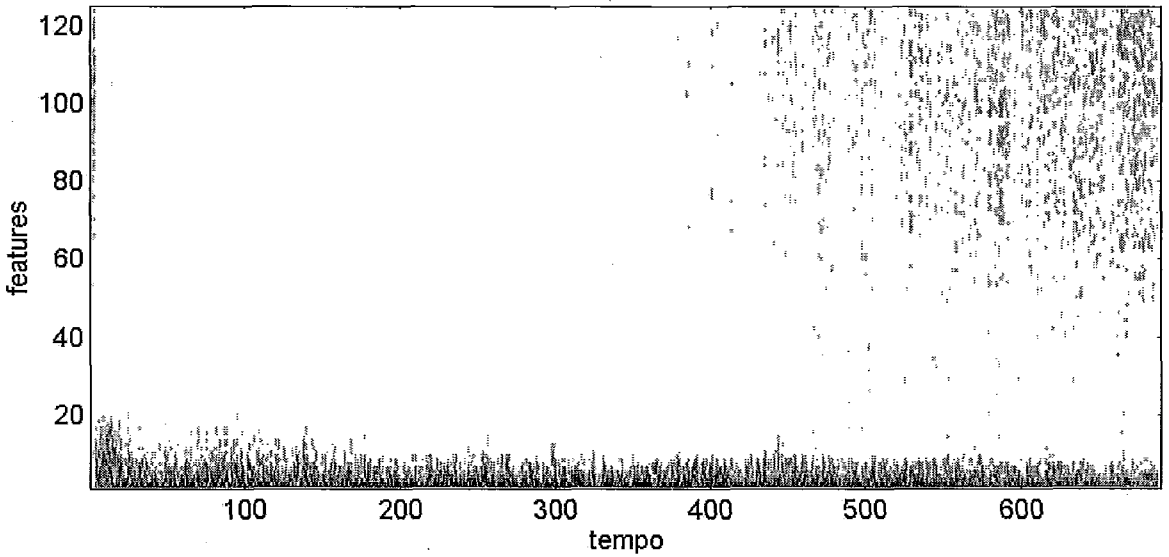
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



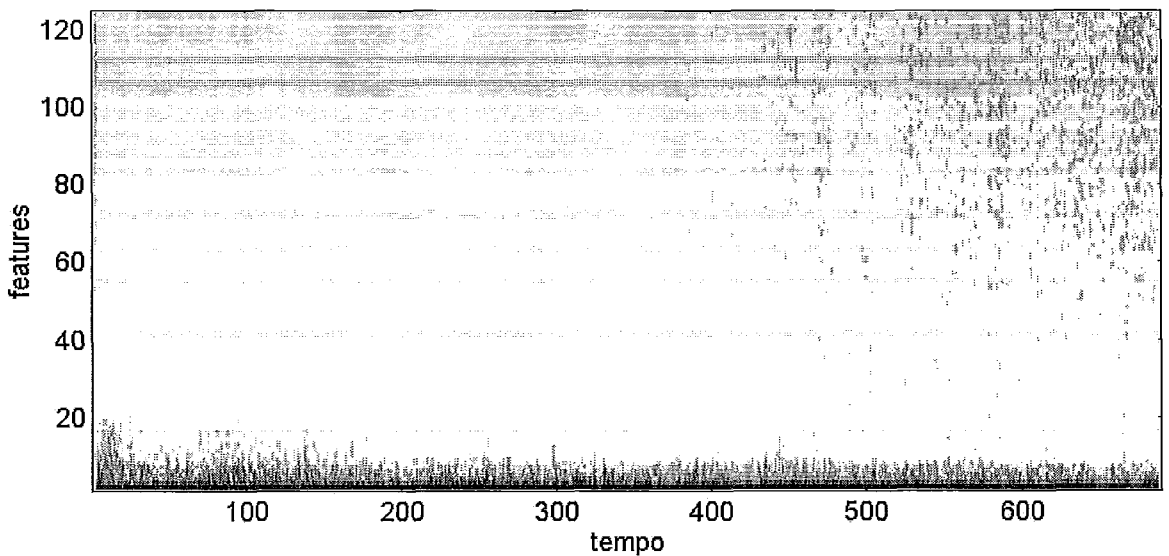
Padrão de Entrada - Trovão Versão B



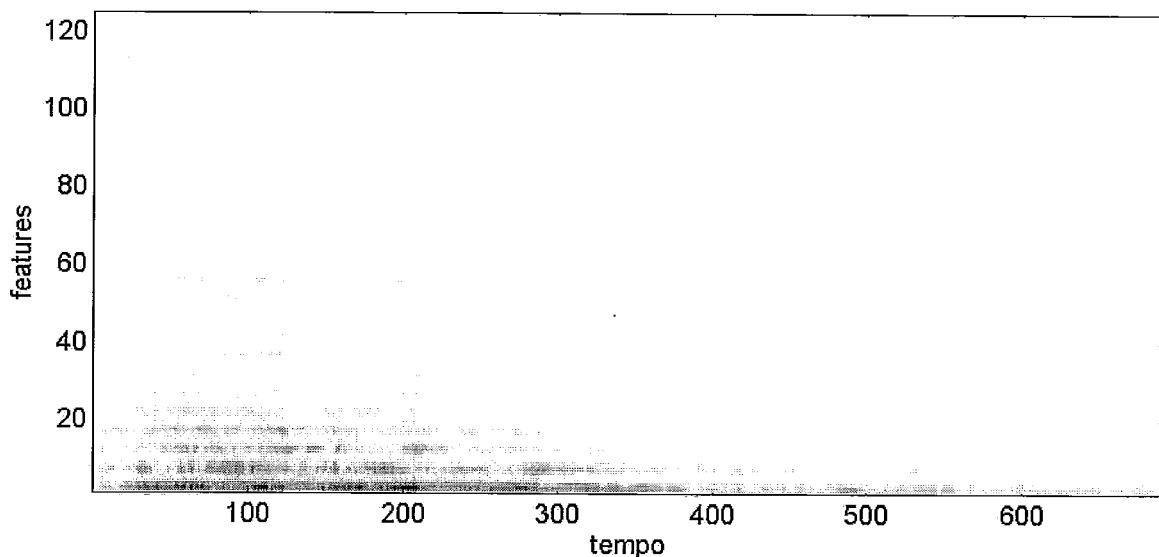
$I(t)$ - Padrão de Entrada Para $P(t)$



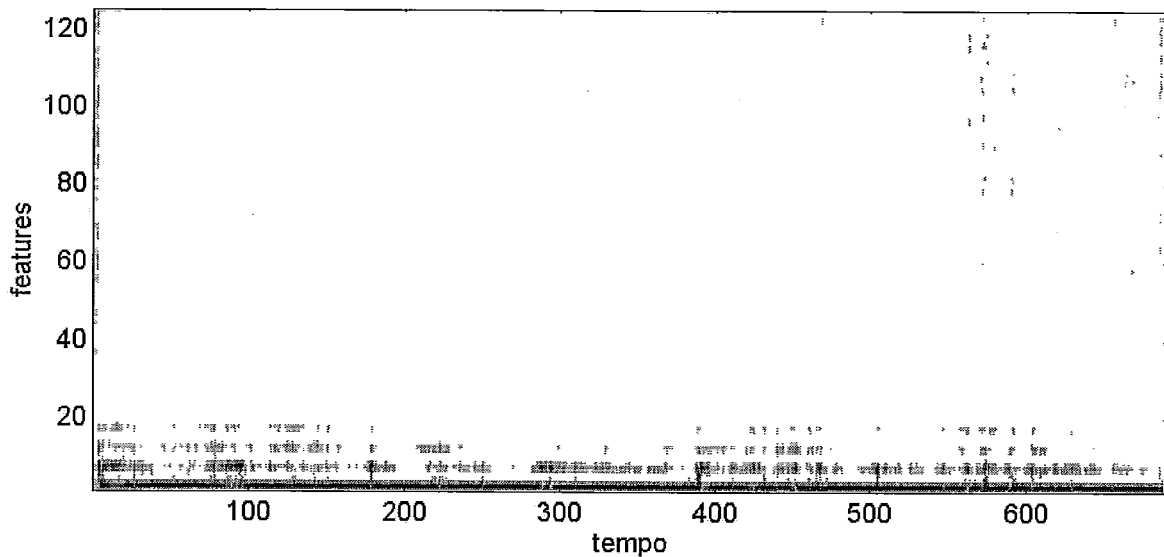
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



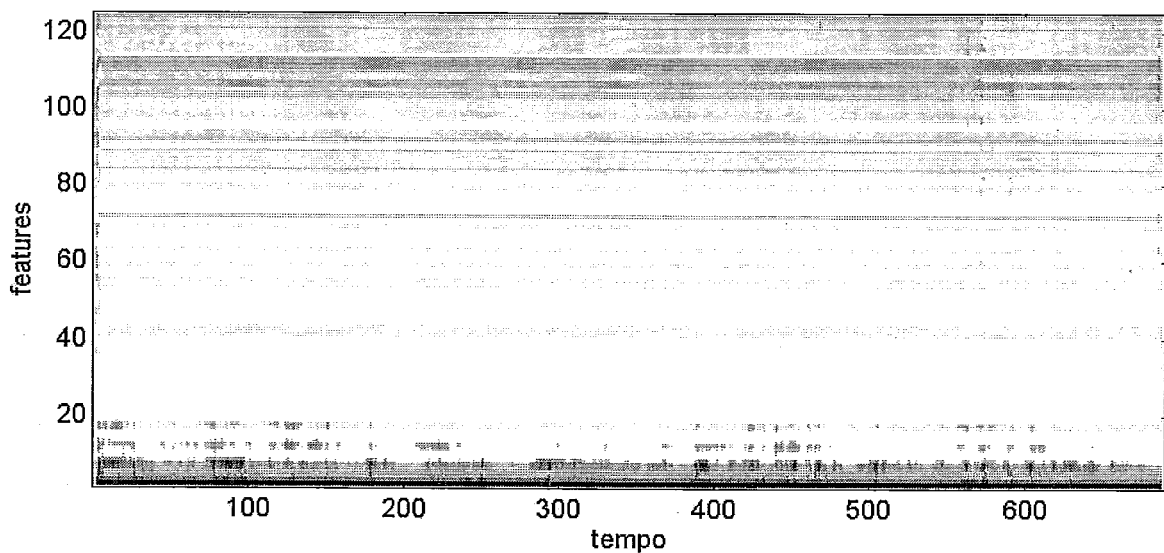
Padrão de Entrada - Trovão Versão C



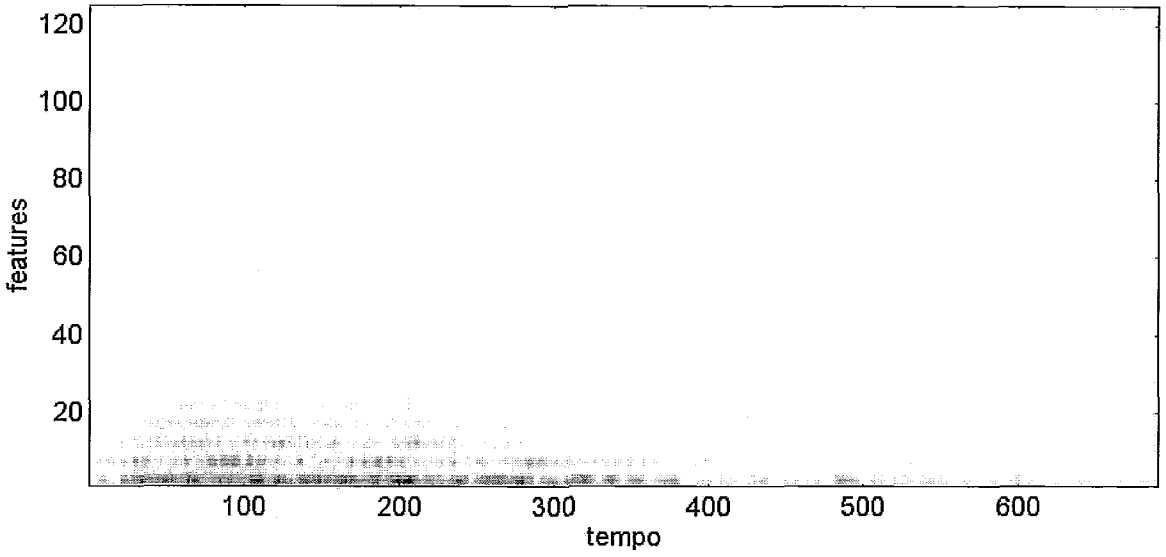
$I(t)$ - Padrão de Entrada Para $P(t)$



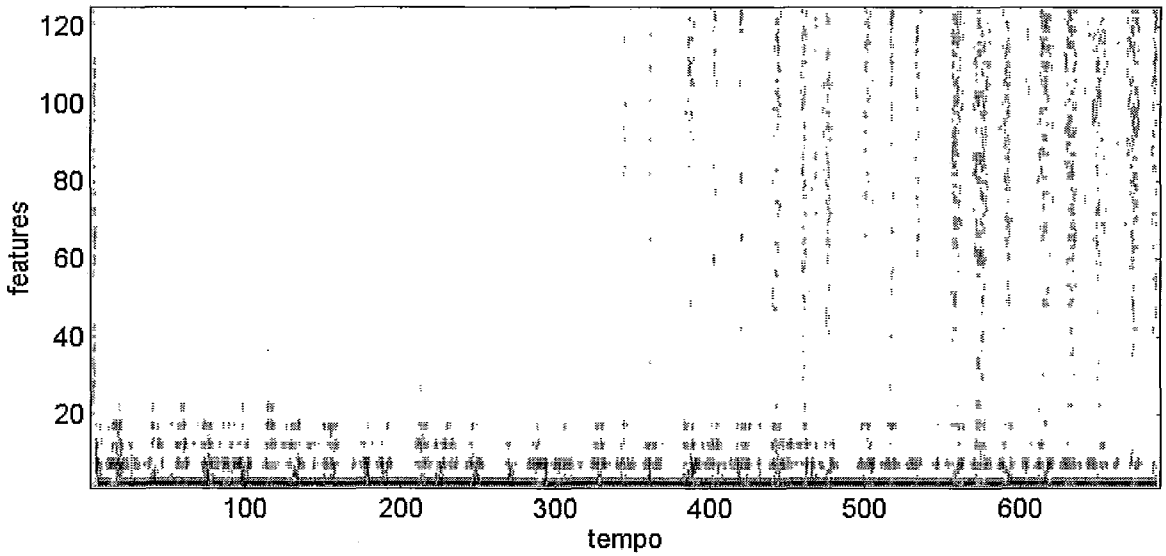
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



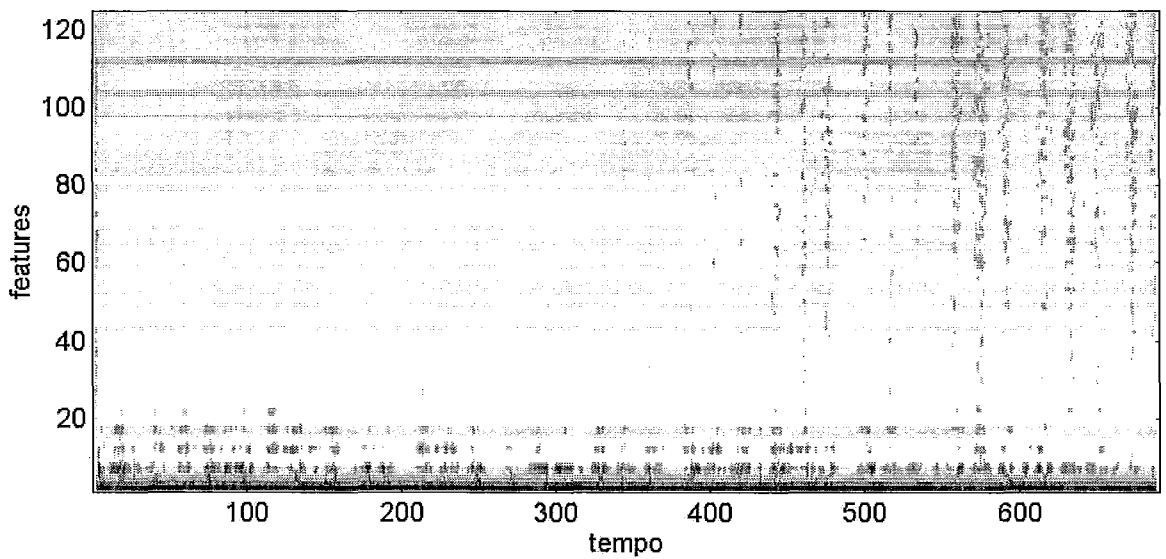
Padrão de Entrada - Trovão Versão D



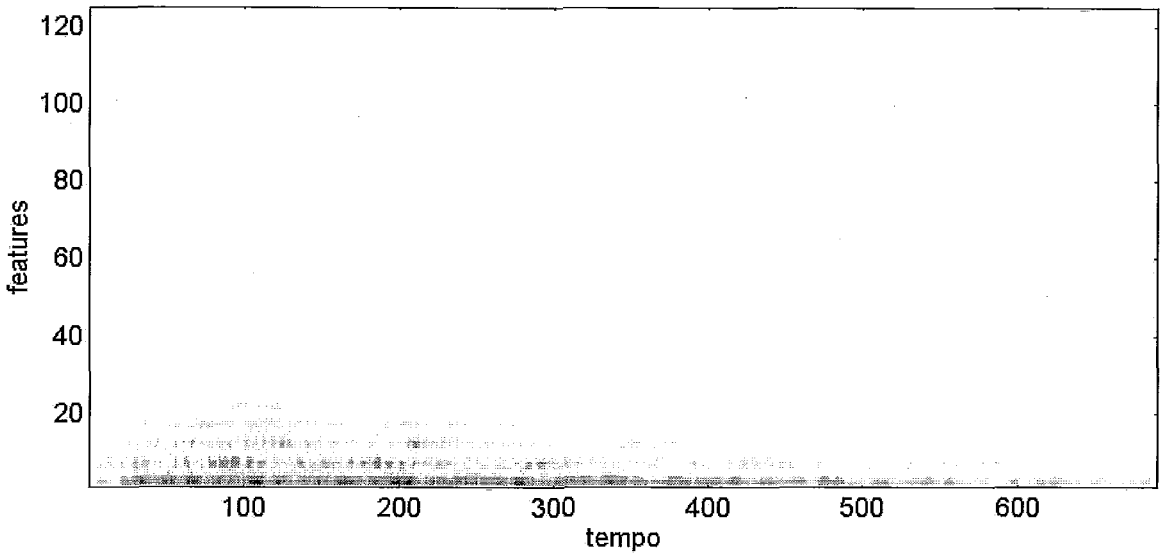
$I(t)$ - Padrão de Entrada Para $P(t)$



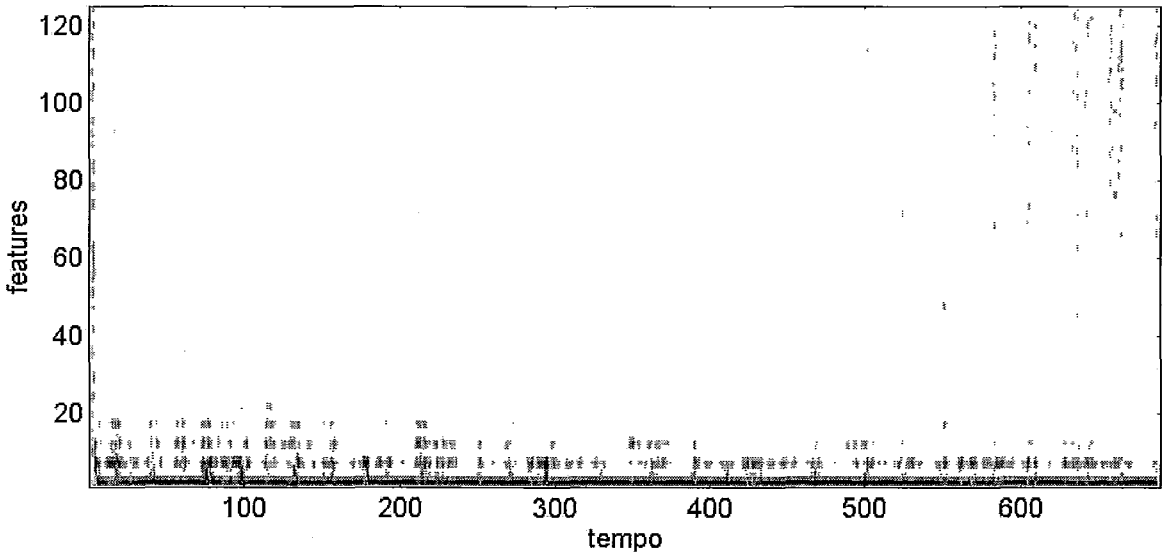
$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2



Padrão de Entrada - Trovão Versão E



$I(t)$ - Padrão de Entrada Para $P(t)$



$P(t)$ Após Realimentação do Neurônio J Vencedor na Camada F2

