

UNIVERSIDADE FEDERAL DE PELOTAS
INSTITUTO DE FÍSICA E MATEMÁTICA
CURSO DE BACHARELADO EM INFORMÁTICA

Reconhecimento de palavras faladas utilizando Redes Neurais Artificiais

por

MAURÍCIO BRAGA DE PAULA
INFORMÁTICA – UFPEL

Trabalho Individual

Prof. Anderson Priebe Ferrugem
Orientador

Pelotas, dezembro de 2000.

Maurício Braga de Paula

Reconhecimento de palavras faladas utilizando Redes Neurais Artificiais

Monografia apresentada ao Curso de Bacharelado em Informática do Instituto de Física e Matemática da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Bacharel em Informática.

Ênfase: Sistemas de Computação

Orientador: Prof. Anderson Priebe Ferrugem
Co-orientador: Prof. Dr. João Artur de Souza

Pelotas, dezembro de 2000.

“Nunca se achou que o degrau da escada se destinasse a alguém permanecer em cima dele, mas sim que se destina a sustentar o pé de um homem pelo tempo suficiente para que coloque o outro um pouco mais alto”

Huxley

Agradecimentos

Ao Prof. Orientador Anderson Priebe Ferrugem pela confiança, incentivo e por sempre ter acreditado na realização deste trabalho.

Aos amigos João Artur de Souza e Gertrudes Aparecida Dandolini, pelos ensinamentos, apoio e contribuição para o enriquecimento deste trabalho.

Ao Prof. Luiz Fernando Minello, pelo fornecimento de material, revisão e sugestões.

Aos meus pais Ivanor e Neusa, que me ensinaram os verdadeiros caminhos a serem trilhados, revestindo minha existência de amor, carinho e dedicação.

À minha irmã Jaqueline, que sempre me apoiou.

À minha avó Diná, pelo incentivo e acompanhamento da minha vida acadêmica.

À minha namorada Angelisa, pela paciência e compreensão.

A todos aqueles que de alguma forma contribuíram para a realização deste trabalho.

Sumário

SUMÁRIO	V
LISTA DE FIGURAS.....	VII
LISTA DE TABELAS	IX
RESUMO	X
1 INTRODUÇÃO.....	1
1.1 Apresentação do tema	1
1.2 O Reconhecimento da Fala	2
1.3 Objetivos.....	4
1.3.1 Objetivos Gerais	4
1.3.2 Objetivos Específicos	5
1.4 Estrutura do Trabalho	5
2 AS ONDAS	7
2.1 Direção de Vibração	7
2.2 Ondas Simples e Periódicas	8
2.3 Dimensões das Ondas	8
2.4 Frente de Onda	8
2.5 Elementos de uma Onda	9
2.6 As Ondas Sonoras.....	9
3 A VOZ HUMANA	11
3.1 Introdução	11
3.2 O mecanismo de fonação.....	12
4 O SISTEMA NERVOSO HUMANO	16
4.1 Introdução	16
4.2 Cérebro e Conduta	17
4.3 Relação entre cérebro e conduta	17
4.4 Regiões do Encéfalo	20
4.5 A Linguagem e outras funções cognitivas estão localizadas no córtex cerebral	21
4.6 O Neurônio	29
5 AS REDES NEURAIS ARTIFICIAIS	45
5.1 Introdução	45
5.2 Histórico	46
5.3 Aplicações.....	49
5.4 O Neurônio Artificial.....	50
5.5 Arquiteturas.....	54
5.6 Aprendizado	57
5.6.1 Supervisionado.....	58
5.6.2 Não supervisionado.....	59
5.6.3 Velocidade de aprendizado	60
5.6.4 Algoritmos de aprendizado	61
5.7 Redes Perceptron.....	62
5.8 A lei de aprendizado do perceptron.....	65
5.9 Limitações: O problema do OU-EXCLUSIVO	67
5.10 Redes Multilayer Perceptron.....	68
5.11 Algoritmo de treinamento das redes MLP	70
6 O SISTEMA MAPA FAN	75
6.1 Introdução	75
6.2 Arquitetura	76

6.3 Pré-Processamento: Ordenação	78
6.3.1 Teoria da Informação.....	78
6.3.2 Ordenação	79
6.4 O classificador Mapa FAN.....	81
6.4.1 FAN	81
6.4.2 O Mapa FAN	83
7 PROTÓTIPO DESENVOLVIDO	88
7.1 A captura do sinal de voz	88
7.1.1 Componentes e características de uma placa de som [TOR 99]	90
7.2 Testes Realizados	91
8 CONCLUSÕES E RECOMENDAÇÕES FINAIS	93
9 REFERÊNCIAS BIBLIOGRÁFICAS.....	94

Lista de Figuras

Figura 2.1 - Visão bidimensional de uma frente de onda circular.....	8
Figura 2.2 - Gráfico de uma onda longitudinal.....	9
Figura 3.1 - Funcionamento básico do aparelho fonador [BAR 98]	14
Figura 3.2 - Esquema do ouvido humano [BAR 98]	15
Figura 4.1 - Principais divisões do encéfalo [MINb 00]	20
Figura 4.2 - Sistema Nervoso Central	23
Figura 4.3 - O córtex cerebral (Modificado de Geschwind, [GES 74]).....	24
Figura 4.4 - O Arco reflexo [MINb 00].....	30
Figura 4.5 - Estrutura do Neurônio [KAN 98]	32
Figura 4.6 - Classificação dos neurônios quanto ao número e forma dos prolongamentos citoplasmáticos. [KAN 98]	34
Figura 4.7 - Estágios do processamento da informação na organização cerebral [KAN 98]	37
Figura 4.8 - Regiões funcionais dos neurônios [KAN 98]	37
Figura 4.9 - Eletrofisiologia neuronal [KAN 98]	39
Figura 4.10 - Tipos e terminologia das sinapses de distintas partes do neurônio [FAW 87]	42
Figura 4.11 - Esquema de uma sinapse neuromuscular (Placa Motora). A união neuromuscular comumente recebe o nome de “placa terminal” devido ao aspecto da terminação axônica em algumas espécies [ALB 86]	42
Figura 4.12 - Seqüência de sinais produzindo uma ação reflexa [KAN 98]	43
Figura 5.1 - O neurônio de McCulloch e implementações de algumas funções booleanas [KOV 96]	51
Figura 5.2 - O Neurônio artificial [TAFb 96].....	52
Figura 5.3 - Funções de transferência [KOV 96]	54
Figura 5.4 - Rede neural artificial.....	54
Figura 5.5 - RNA de uma única camada	55
Figura 5.6 - RNA multicamada	56
Figura 5.7 - RNA feedforward ou acíclica	56
Figura 5.8 - RNA feedback ou cíclica	57
Figura 5.9 - O perceptron elementar de Roseblatt [BIS 95]	63
Figura 5.10 - A unidade de processamento do perceptron	64
Figura 5.11 - Perceptron de duas entradas.....	66
Figura 5.12 - Plano que representa as combinações possíveis do XOR.....	68
Figura 5.13 - Uma rede MLP.....	70
Figura 5.14 - Rede MLP com os acoplamentos retrógrados para os ajustes sinápticos [PAT 95]	72
Figura 5.15 - Fluxo de treinamento de uma MLP com <i>backpropagation</i>	73
Figura 6.1 - Sinal de voz da palavra cálculo.....	76

Figura 6.2 - Trechos do sinal (amostras)	77
Figura 6.3 - Arquitetura do Mapa FAN	77
Figura 6.4 - Amostra do sinal de voz.....	80
Figura 6.5 - Ordenação do sinal de voz	80
Figura 6.6 - Aprendizagem no Mapa _c indicada pela matriz do padrão de entrada.....	84
Figura 6.7 - Intersecção entre um padrão de entrada (quadrados pequenos com bordas) e um mapa treinado (onde o nível de cinza caracteriza o grau de reforço de uma célula)	86
Figura 6.8 – Visualização tridimensional de duas palavras treinadas (dois mapas).....	87
Figura 6.9 - Visão bidimensional de duas palavras treinadas.....	87
Figura 7.1 - Regiões de silêncio e voz da palavra capturada.....	89
Figura 7.2 - Aplicativo para a captura do sinal de som	90
Figura 7.3 – Conjunto de palavras de teste com um percentual de 100% de acerto.....	91
Figura 7.4 - Conjunto de palavras de teste com um percentual de 75% de acerto	92

Lista de Tabelas

Tabela 4.1 e Figura 4.2 - Subdivisão Anatômica do Sistema Nervoso Central.....	22
Tabela 4.2 - Tomografia por Emissão de Pósitrons do Córtex Cerebral – vista lateral do hemisfério esquerdo em distintos momentos de conduta relacionados a linguagem.....	28
Tabela 5.1 - Tabela verdade do ou-exclusivo.....	67
Tabela 7.1 - Resultados para um conjunto de 3 palavras	91
Tabela 7.2 - Resultado para um conjunto de 4 palavras	92

Resumo

O advento da tecnologia do reconhecimento da fala, tem tornado realidade a visão futurista apresentada na maioria dos filmes de ficção científica. O processamento computacional da fala é uma área em ascensão, devido a sua vasta gama de aplicações. Muitas pesquisas nesta área revelam que ainda não é possível manter um domínio pleno sobre esta arte e, a possibilidade de conversar com uma máquina em uma linguagem natural ainda vai levar algum tempo. O fator motivador para a investigação nesta área é decorrente da necessidade de sistemas/produtos que sejam flexíveis quanto a usabilidade, independentemente do nível gradual do usuário. Portanto, este poderia usufruir desta tecnologia de maneira tão simples quanto ao uso de um eletrodoméstico (reservado as devidas proporções de cada caso e sua aplicabilidade). O principal objetivo deste trabalho foi empregar o uso de um modelo de Redes Neurais Artificiais (Mapa FAN) para o reconhecimento de palavras faladas da língua portuguesa. O reconhecimento de palavras faladas não é um processo de fácil manuseio, devido a complexidade da voz humana. O modelo utilizado obteve bons resultados, visto que, nenhum pré-processamento de sinais mais elaborado, a não ser a ordenação, foi utilizado. O reconhecedor proposto é aplicado ao reconhecimento de palavras isoladas da fala. Os resultados obtidos são dependentes de vários fatores, como por exemplo, o tamanho da amostra devido a taxa de amostragem da palavra capturada. A principal meta deste trabalho foi alcançada, face a constatação de que é possível realizar o reconhecimento da fala através do modelo acima citado. E, ainda, mostrando que o Sistema Mapa FAN, que nunca havia sido utilizado para este tipo de estudo, torna-se viável para futuras aplicações.

Abstract

The advent of the technology of the speech recognition, has become reality the futurist vision presented in most of the sci-fi movies. The computational processing of the speech is an area in ascension, due to your vast range of applications. Many researches in this area reveal that is not possible still to maintain a full domain about this art and, the possibility to talk with a machine in a natural language will still take some time. The factor motivador for the investigation in this area is due to the systems need that are flexible as the usabilidade, independently of the user's level. Therefore, it's able to use this technology in a such simple way with relationship to the use of an appliance (reserved the due proportions of each case and your aplicabilidade). The principal objective of this work was to use the use of a model of Artificial Neural Nets (FAN Map) for the recognition of spoken words of the Portuguese language. The recognition of spoken words is not a process of easy handling, due to complexity of the human voice. The used model obtained good results, because, any pre-processing of signs more elaborated, not being the ordination, it was used. The proposed prototype is applied to the recognition of isolated words of the speech. The obtained results are dependents of several factors, for example, the size of the sample due the sample rate of the captured word. The main goal of this work was reached, in the face of the verification that is possible to accomplish the recognition of the speech above through the model mentioned. And, still, showing that the FAN Map System, that had been used never for this study type, becomes viable for future applications.

1 INTRODUÇÃO

1.1 Apresentação do tema

A constante evolução tecnológica tem proporcionado à humanidade desfrutar de novas ferramentas que ultimamente vêm auxiliando o homem de forma significativa na realização de inúmeras tarefas. Uma destas ferramentas é o computador, que foi proposto em meados do século XVIII pelo matemático Charles Babbage. O computador era composto inicialmente por elementos básicos como: unidade aritmética, unidade de controle, memória, entrada e saída. No princípio da computação, os dados e instruções eram introduzidos em seqüência e eram feitas operações girando-se mecanismos contadores ou através de cartões perfurados, chaves e fios que se conectavam em soquetes.

O conceito da máquina mecânica de Babbage foi efetivamente evoluindo até a construção do famoso ENIAC (Eletronic Numerical Interpreter And Calculator), o primeiro computador digital eletrônico da história.

A revolução surgiu quando o matemático húngaro Von Neumann, propôs que os dados, os resultados intermediários e as instruções fossem armazenados em um único lugar denominados memória principal do computador. Surge, então, o conceito de processador com programa armazenado. Assim, programas diferentes podiam ser executados pela máquina sem a necessidade de alterar a configuração da mesma. A partir de 1947, os computadores passaram a ser projetados e construídos baseados nas idéias de Von Neumann; e além disso, o aperfeiçoamento tecnológico foi marcado pelo surgimento do transistor, seguido dos circuitos integrados e, mais recentemente pelos microprocessadores (larga escala de integração dos circuitos integrados).

A forma de alimentação e comunicação com estas máquinas variaram muito desde os seus primórdios. A substituição dos cartões perfurados, fios e chaves pelos chamados periféricos de entrada (responsáveis pela alimentação do computador com variáveis do mundo externo) como o mouse, teclado, scanner, drives, proporcionaram uma maior flexibilidade na relação humano x computador.

O progresso computacional geralmente busca o sentido de aumentar o desempenho e exatidão de processamento, criando novas tecnologias para a otimização

destes processos. Um dos meios ainda pouco explorados neste progresso são os dispositivos de entrada. Trata-se de uma evolução não nestes dispositivos, mas sim de um novo conceito de comunicação com o computador. O teclado, por exemplo, evoluiu possuindo um maior número de teclas, novos símbolos e formatos mais ergonômicos, garantindo e proporcionando ao usuário maior praticidade e conforto. Falamos de um novo paradigma, que tem como principal objetivo transferir algumas funções para um microfone que receba os comandos falados pelo usuário. Não se trata da eliminação do mouse ou do teclado, mas sim conjugar o uso destes com um novo conceito de interação com os computadores. O mouse poderia continuar sendo usado para operações auxiliares e o teclado para funções mais específicas.

A tecnologia do reconhecimento da fala não parece ser tão futura assim, pois grandes empresas como a IBM, a Intel, dentre outras já estão distribuindo no mercado softwares que viabilizam a criação de cartas e relatórios através de comandos de voz. Este advento está dominando o mercado mundial, sendo citado como uma das tecnologias mais importantes e emergentes para os próximos tempos.

1.2 O Reconhecimento da Fala

Desde os primórdios da revolução industrial, o homem sonhava em desenvolver autômatos, posteriormente chamados de robôs, dotados de certa inteligência e capacidade de compreensão, para poder desempenhar tarefas de forma autônoma a fim de substituir o trabalho humano, como por exemplo, através de comandos da fala. A tecnologia existente na época não era suficiente para que tais avanços pudessem ser alcançados. As primeiras conquistas foram obtidas através da descoberta de novos conhecimentos relacionados ao processamento de informações, que proporcionaram o surgimento de computadores com um maior poder de processamento e armazenamento, e o surgimento da Inteligência Artificial.

A Inteligência Artificial, proposta em 1956 por MacCarthy, é composta por um agregado de métodos que visa solucionar determinados tipos de problemas na área da engenharia e computação. Estes métodos, mais conhecidos como algoritmos computacionais ou programas, tentam se espelhar na maneira de resolução de problemas adotados pelo homem.

As redes neurais artificiais (RNA), um dos nichos da IA, propõe uma metodologia para solucionar alguns problemas relacionados a área de inteligência artificial, modelando sistemas através de conexões que possam simular o sistema nervoso humano, abrangendo a capacidade que o mesmo possui de aprender e agir perante as mais adversas situações apresentadas.

Uma das aplicações das RNA's é o reconhecimento de palavras faladas, que por sua vez auxilia e principalmente agiliza algumas tarefas desempenhadas manualmente, tais como a digitação de um texto, controle de eletroeletrônicos, comandos de voz para robôs para a realização de algumas funções que podem ou não colocar em risco a vida humana. Cabe também ressaltar o considerável auxílio que esta tecnologia pode proporcionar a pessoas portadoras de deficiência física.

O reconhecimento da fala consiste em identificar fonemas, sílabas, palavras para formar a mensagem original [HUG apud LUN 00], ou uma informação onde existe uma seleção mais direta da resposta, sem interpretação dela. Assim, uma ação pode ser executada diretamente quando um padrão falado é reconhecido.

O reconhecimento automático da voz consiste no processo de extrair automaticamente a informação lingüística do sinal da fala, a qual está codificada. Este processo normalmente acontece em três etapas [BAR 99]:

- 1ª. Aquisição do Sinal de Voz
- 2ª. Extração de Parâmetros
- 3ª. Reconhecimento do padrão

A primeira etapa consiste em realizar a captação do sinal de voz através de uma interface analógica digital (uma placa de som por exemplo), que consiste basicamente em entrarmos com a informação (voz) de forma analógica e recolhermos na saída essa mesma informação de forma digital. A segunda etapa (pré-processamento) procura extrair do sinal capturado as características que descrevem adequadamente o sinal de voz. A terceira e última etapa consiste em identificar os dados selecionados na fase de pré-processamento (treinamento do sistema).

A segunda etapa do processo de reconhecimento de voz pode ser auxiliada pela área da ciência que tem como objeto de estudo os sinais. Os DSP (Processamento Digital de Sinais) são as técnicas matemáticas e os algoritmos usados para manipular os

sinais após eles terem sido convertidos na forma digital [DAN 00]. Isso inclui uma ampla variedade de objetivos, tais como: intensificação de imagens visuais, reconhecimento e geração de voz, compressão de dados para armazenagem e transmissão.

Como o Processamento Digital de Sinais trata da representação matemática discreta de sinais contínuos que variam com o tempo e com a frequência, a sua representação é realizada através de amostras do domínio do tempo, ou com amostras do domínio de frequência. Com estas características extraídas, os algoritmos mais usados para estas amostragens são [RIOS 99]:

- Codificação Preditiva Linear (LPC) – Baseado na diferença entre os tipos de sons (vogais, consoantes) emitidos pelo aparelho fonador, que são tratados separadamente;
- Modelo de Mistura Gaussiano – Baseia-se nos componentes individuais vocais com suas classes acústicas amplas e na densidade empregada em amostras subjacentes de termos longos;
- Transformada Rápida de Fourier (FFT) – Baseia-se na modelagem do sinal de palavras isoladas, realizando, assim, a transferência de abordagem do sinal em função do tempo para sinal em função da frequência.

O reconhecimento propriamente dito pode ser feito utilizando-se várias técnicas, desde o alinhamento temporal não linear até as redes neurais neuronais, modelos escondidos de Markov e lógica difusa [ESP 99].

1.3 Objetivos

1.3.1 Objetivos Gerais

A principal meta deste trabalho é desenvolver um protótipo que reconheça um pequeno conjunto de palavras isoladas da língua portuguesa, usufruindo a tecnologia das redes neurais artificiais.

1.3.2 Objetivos Específicos

- Realizar uma pequena abordagem estudo sobre o sistema nervoso humano e a fala humana.
- Estudar um método de pré-processamento (ordenação) e um método de classificação (Mapa FAN) para trabalhar com padrões de alta dimensão, proposto por Dandolini, 2000.
- Realizar a aquisição do som através de uma interface analógica digital (A/D).
- Melhorar a representatividade do sinal digital, oriundo da interface A/D, em relação ao sinal original.
- Realizar um estudo sobre as ferramentas utilizadas no desenvolvimento do protótipo tais como o *Borland Delphi* e o *MatLab*.

1.4 Estrutura do Trabalho

Este trabalho está disposto em 8 capítulos. O capítulo 1 aborda os aspectos introdutórios do trabalho, como a apresentação do tema e uma breve introdução sobre o reconhecimento da fala. O capítulo 2 trata da fundamentação teórica sobre as ondas mecânicas, apresentando as características das mesmas, a fim de proporcionar um maior entendimento na formação dos sons da fala humana.

No capítulo 3 e 4 é feita uma abordagem geral referente a fisiologia da fala e do sistema nervoso, demonstrando a estrutura e composição do aparelho fonador e do cérebro. O funcionamento básico desde a formação do som até a fala propriamente dita, os centros de controle específicos e os mecanismos de fonação são mostrados no capítulo 3. Os conceitos referentes às bases biológicas da consciência do cérebro, a sua conduta e relação, e a descrição e funcionamento dos neurônios, fazem parte do capítulo 4.

No capítulo 5 é feita uma introdução sobre as redes neurais artificiais, como o seu surgimento, a arquitetura e disposição dos elementos formadores destas redes (os

neurônios artificiais) e os conceitos referentes aos métodos e paradigmas de treinamento.

Ainda no capítulo 5 é apresentado um dos primeiros modelos matemáticos de redes neurais artificiais, chamado de Perceptron. Também é abordada a evolução desta rede, devido as limitações que este modelo apresenta.

O modelo utilizado para o reconhecimento de palavras faladas, é apresentado no capítulo 6. Trata-se do Mapa FAN, que é um modelo que trata de padrões que possuem uma quantidade muito grande de características, como por exemplo o som.

Finalmente, no capítulo 7, é descrito os testes realizados com o protótipo desenvolvido. No capítulo 8 são feitas as considerações finais e conclusões do trabalho.

2 AS ONDAS

Defini-se onda como o movimento causado por uma perturbação que se propaga. Os fenômenos de propagação aparecem em inúmeros nichos da física. Fazem parte do mesmo conjunto das ondas, as ondas produzidas em cordas, ondas sonoras, ondas na superfície da água, ondas em molas, ondas de luz, ondas de rádio [SAN 72]. Um exemplo simples de ser observado é a perturbação produzida por uma pedra, quando jogada na água. Ao jogarmos a pedra, produz-se uma perturbação no ponto em que ela atinge a água, propagando-se em todas as direções sob a forma de círculos concêntricos.

Há várias maneiras de se classificar as ondas, dependendo do ponto de vista a ser considerado. Pode-se classificar as ondas quanto à necessidade ou não de algum meio de propagação. As ondas mecânicas, objeto de nosso estudo, são aquelas ondas que necessitam de um meio material para se propagar, como por exemplo, as ondas sonoras. Já as ondas eletromagnéticas não necessitam de um meio material para se propagar, pois são constituídas pela vibração de campos eletromagnéticos, como por exemplo as ondas de luz.

As ondas não propagam matéria, sendo que cada partícula do meio oscila apenas em torno de sua posição de equilíbrio. A onda propaga somente energia e qualquer propagação de energia pode ser considerada como uma onda [GON 72].

2.1 Direção de Vibração

Em relação à direção de propagação, as ondas podem ser transversais e longitudinais. Uma onda é dita transversal quando as partículas do meio vibram numa direção perpendicular àquela em que a onda se propaga. As ondas eletromagnéticas são ditas transversais, pois não se considera a vibração de partículas, mas sim de um campo elétrico.

Uma onda mecânica é dita longitudinal quando as partículas do meio vibram na mesma direção em que a onda se propaga. As ondas sonoras são ditas longitudinais [HAL 96].

2.2 Ondas Simples e Periódicas

Um meio é percorrido por uma onda simples (ou por um pulso) quando se produz numa porção qualquer deste meio um único abalo. Se este abalo é repetido por inúmeras vezes, o meio é percorrido por um conjunto de ondas simples que se sucedem, chamado de trem de ondas [GON 72]. Quando o abalo ocorre de forma periódica, obtém-se um trem de ondas periódicas.

2.3 Dimensões das Ondas

As ondas podem ser classificadas quanto ao número de direções que ela se propaga. As unidimensionais se propagam numa única direção, tais como: as ondas nas cordas ou nas molas. As bidimensionais se propagam em duas direções (distribuições superficiais), como as ondas na superfície da água. As tridimensionais se propagam em três direções (distribuições volumétricas - espaço), como as ondas sonoras.

2.4 Frente de Onda

A frente de onda é o lugar geométrico dos pontos do meio, que no instante considerado, são primeiramente atingidos pela perturbação que a onda propaga. A frente de onda divide a região perturbada do meio, da que ainda não foi perturbada.

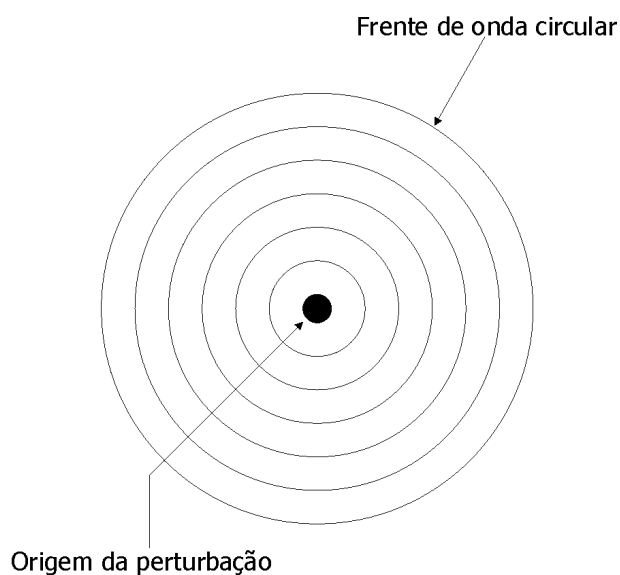


Figura 2.1 - Visão bidimensional de uma frente de onda circular

2.5 Elementos de uma Onda

Os elementos que compõem as ondas são: o comprimento de onda, a amplitude, frequência, período e velocidade.

O comprimento de onda é a distância que a onda avança num intervalo de tempo igual a um período. Normalmente é representado pela letra grega λ (lambda). A Figura 2.2 mostra o gráfico de uma onda longitudinal. A amplitude (A) é a distância que vai de uma crista ao eixo de propagação da onda; pode ser também a distância do ponto máximo de depressão (vale) ao eixo de propagação. A frequência de uma onda é o número de vibrações ou ciclos por segundo que a mesma efetua. O período é, portanto, o tempo decorrido de uma oscilação.

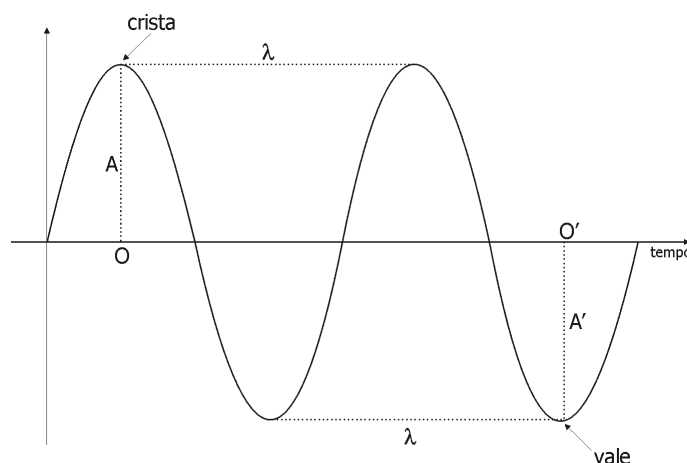


Figura 2.2 - Gráfico de uma onda longitudinal

Quando uma onda passa de um meio para outro, sua frequência e seu período não se alteram, pois ambos dependem somente da fonte. Já a velocidade de propagação é alterada em função do módulo de elasticidade e da massa específica do meio, o que resultará na modificação do seu comprimento de onda.

2.6 As Ondas Sonoras

O som é o efeito produzido por ondas mecânicas longitudinais no aparelho auditivo desde que sua frequência e amplitude sejam suficientes para impressionar o órgão da audição. Em virtude de o som ser uma onda mecânica, o mesmo só poderá se propagar através de meios materiais como: sólidos, líquidos ou gasosos [GON 72].

Estas ondas sonoras podem ser produzidas por um elemento vibrador que pode ser um cristal, uma corda de algum instrumento musical ou as próprias cordas vocais. Esses elementos vibradores causam variações na densidade ou pressão do meio. Os sons distinguem-se por três características [GON 72]:

- a altura: qualidade de um som ser mais ou menos agudo ou grave, ligada a frequência das vibrações; os sons graves são baixos e os sons agudos são altos;
- a intensidade: qualidade de um som mais ou menos forte ou fraco, ligada à amplitude das vibrações;
- o timbre: permite distinguir os sons emitidos de mesma altura e de mesma frequência; está ligado à complexidade das vibrações, i.e, à presença de harmônicos superpostos ao som natural.

Os sons perceptíveis ao homem têm uma frequência compreendida entre 20 e 20.000 Hz; os infra-sons (terremotos) têm uma frequência inferior e os ultra-sons uma frequência superior. Para a produção de uma sensação auditiva, a frequência da onda deve estar compreendida nestes limites. Estes, porém não são iguais para todos os seres humanos, variando de indivíduo para indivíduo.

Qualquer corpo que oscile com uma frequência compreendida entre 20 e 20.000Hz, pertencendo a um meio elástico, pode trabalhar como uma fonte sonora. As fontes sonoras podem ser divididas em 4 grupos:

- fontes que utilizam cordas vibrantes – cordas vocais, violão, piano, violino, etc;
- fontes que utilizam colunas de ar vibrantes – tubo de órgão, flauta, saxofone, clarinete, etc;
- fontes que utilizam membranas ou placas vibrantes – tambor, pratos, xilofone, etc;
- fontes que utilizam hastes vibrantes – diapasão;

3 A VOZ HUMANA

3.1 Introdução

Não existe certeza sobre o momento em que o homem começou a falar, mas se supõe que a linguagem tenha começado a desenvolver-se há cerca de 1,5 milhão de anos, com o *Australopithecus*. Um dos últimos passos na evolução da linguagem ocorreu com o *Homo sapiens spp* a mais de 40000 anos [BAR 98]. O desenvolvimento da linguagem permitiu que os conhecimentos adquiridos pelos homens pudessem ser compartilhados e possivelmente passados de uma geração para outra, abrindo novos horizontes a espécie.

A fala é uma das capacidades ou aptidões que os seres humanos possuem de comunicação, manifestando seus pensamentos, opiniões e sentimentos através de vocábulos que se transladam textualmente quando necessários. Consiste no principal sinal entre os distintos sinais abordados pela linguagem natural, como por exemplo, ideogramas, gestos, gritos, trejeitos e outros tipos de linguagem corporal.

O processo pelo qual os seres humanos produzem palavras e orações audíveis para se comunicar, possibilita a obtenção de informações a respeito do ambiente no qual o indivíduo está inserido. Grande parte das espécies animais possuem algum grau de comunicação, porém o homem, em virtude da complexidade social em que está inserido, adquiriu o mais alto grau de comunicação conhecido, dentre as quais, a fala tem uma extrema relevância.

Os sons da fala não devem ser confundidos com os fonemas da Língua Portuguesa, uma vez que, o som é entendido como uma complexa realidade físico acústica de cada unidade sonora da fala, enquanto que os fonemas correspondem a percepção eclética e interpretativa realizada pelo falante e ouvinte, respectivamente. O fonema é a menor unidade sonora ou articulatória da palavra [LUF 95]. Os fonemas são produzidos pelo aparelho fonador, que é constituído por um conjunto de órgãos que auxiliam no ato da fala. A fala representa a emissão de sons articulados na linguagem oral enquanto que a produção da voz é a fonação.

3.2 O mecanismo de fonação

A fonação envolve os centros de controle específicos da fala no córtex cerebral, funções mecânicas da produção da voz e o controle do som para produzir um fonema definido [OKU 82]. O aparelho fonador está constituído por regiões definidas dos aparelhos digestivo e respiratório, sendo auxiliado pelo mecanismo de audição. Sua principal função é produzir os sons, que podem ser a voz cantada ou a voz falada. O aparelho fonador pode ser dividido em seis partes [MINa 00]:

1. Produtor
2. Condutor
3. Vibrador
4. Ressonador
5. Articulador
6. Sensor/Coordenador

O **produtor** é formado pelos pulmões, músculos abdominais, diafragma, músculos intercostais e músculos extensores da coluna. Os pulmões se apresentam como uma massa esponjosa e elástica, que ocupa toda a cavidade torácica e estão protegidos pela caixa de mesmo nome, que é formada por diversos músculos e ossos. Além de serem responsável pelos fenômenos respiratórios (inspiração e expiração) e pelas troca gasosa, os pulmões juntamente com as vias respiratórias superiores, servem de reservatório de ar que possibilitam a vibração das cordas vocais.

O diafragma é um músculo que separa o tórax do abdome e sua contração provoca um aumento de volume da caixa torácica, e portanto, a inspiração. Quando o ar sai dos pulmões, o diafragma sobe, diminuindo o volume pulmonar.

O **condutor** é caracterizado pelo conjunto de tubos e túbulos que compõe o aparelho fonador sendo responsáveis pela produção da coluna de ar que pressiona a laringe fazendo vibrar as cordas vocais e produzindo o som. Os órgãos desta parte são brônquios, traquéia, laringe e faringe. A traquéia é um órgão anelado que possui a forma de um tubo, medindo aproximadamente doze centímetros, e é vista como uma via de passagem do ar que proporciona um suporte para a vibração das cordas vocais. Os brônquios principais consistem em duas bifurcações principais da traquéia que se

dividem em esquerdo e direito e, a sua vez, se subdividem gradativamente em unidades menores que constituirão os bronquíolos, bronquíolos transitórios e por final bronquíolos respiratórios, que atingem os alvéolos pulmonares. Os bronquíolos respiratórios, juntamente com os alvéolos, realizam as trocas gasosas, enquanto que as demais ordens de bronquíolos, brônquios, traquéia, laringe e faringe, consistem na porção condutora de gases do aparelho respiratório.

A laringe é um dos principais órgãos da fala e se apresenta como um tubo cartilaginoso situado na parte anterior do pescoço que se comunica com a faringe através de um canal, a glote. Nesta região há uma válvula responsável pela abertura e fechamento da glote, a epiglote, que auxilia durante a deglutição e impede que substâncias alimentícias adentrem na porção respiratória do aparelho respiratório. Como elemento condutor, leva a corrente de ar da faringe para a traquéia. Consiste em uma cavidade alongada que possui a forma de um funil e está localizada logo após a faringe, cerca da cavidade oral, que possui comunicação com a boca, as fossas nasais, a laringe e com o esôfago. A faringe como elemento de fonação amplia os sons e serve como uma caixa de ressonância.

O **vibrador** é composto pelas cordas vocais e a laringe. A laringe, como mencionada anteriormente, é considerada o órgão da fala, pois nela está localizada as cordas vocais. As cordas são formadas por duas membranas que se localizam nas bordas da glote e que vibram à passagem do ar. Logo, quando uma pessoa está respirando normalmente, as cordas vocais estão separadas uma da outra, formando uma abertura em na forma de um V. A produção do som se dá quando os músculos da laringe tornam rijas as cordas vocais, produzindo o som quando o ar as atravessa.

O **ressoador** formado pela cavidade nasal, a faringe e a boca, tem como função ampliar o som. As fossas nasais são duas cavidades existentes no nariz, onde o ar entra para o organismo. Elas são separadas por uma estrutura chamada de septo nasal que se comunicam com o exterior através das narinas, com a boca e a faringe pelas coanas. Sua função biológica é de filtrar, aquecer e umidificar o ar. Como contribuição fonatória, visa a vibração e amortização do som, resultando na ressonância nasal.

O **articulador** constitui-se pelos componentes responsáveis pela articulação dos sons, transformando-os em orais e nasais. O sentido dado ao som é outra função desempenhada por este componente. O articulador compreende os órgãos localizados na boca como lábios, língua, palato, maxilar, dentes, úvula e véu palatino. O principal

órgão articulador é a língua, que participa de todos os sons produzidos e interfere diretamente na formação das vogais e consoantes. Os lábios, são dobras membranosas localizadas na parte externa da boca e que se dividem em superior e inferior. Têm como função fonatória produzir os sons bilabiais da Língua Portuguesa. Os dentes são órgãos que guarnecem as maxilas e não só auxiliam na mastigação, mas também fazem parte do sistema fonador, contribuindo para o escoamento do som produzido.

O palato se divide em duas partes: o palato duro, céu da boca, e o palato mole onde se encontra a úvula. O palato duro como articulador é responsável pela projeção da voz. Já o palato mole (úvula), permite ou não a passagem do som produzido pelo ressoador. O véu palatino, conjunto de músculos ligado ao palato, caracteriza a distinção entre os fonemas produzidos. Se a coluna de ar encontrar o véu levantado, fechando a passagem pelas fossas nasais, haverá sons/fonemas orais, caso contrário, nasais [LUF 95]. A Figura 3.1 mostra o funcionamento básico do aparelho fonador, onde os pulmões forçam o ar em direção a glote, que atingindo as cordas vocais faz com que elas vibrem. A boca e o nariz contribuem para a modelagem da voz.

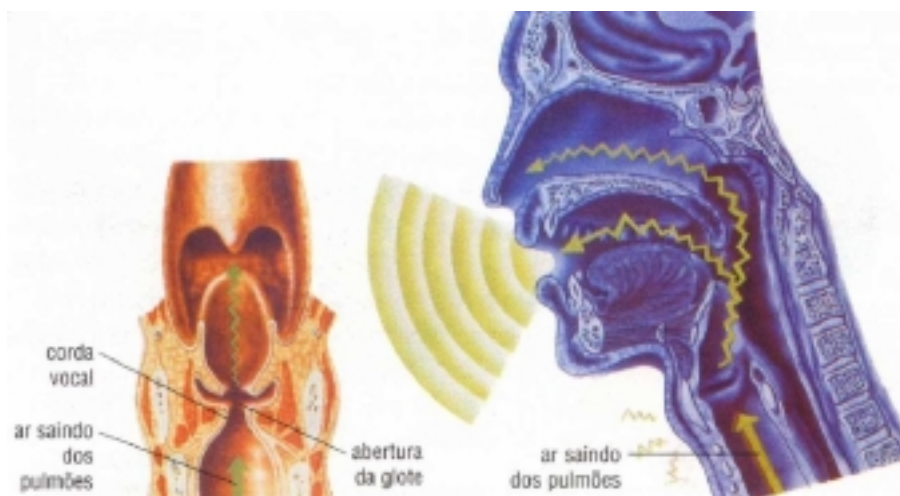


Figura 3.1 - Funcionamento básico do aparelho fonador [BAR 98]

A última parte que integra o mecanismo de fonação é o ouvido (mecanismo de audição como *feedback*) que tem como função captar, selecionar e interpretar o som. O ouvido humano é um órgão sensível que possibilita captar, perceber e interpretar ondas sonoras numa faixa de frequência que está compreendida entre 20 e 20khz [OKU 82]. Uma série de fatores e transformações de energia são necessários para a conclusão deste processo. Os ouvidos (esquerdo e direito) estão encaixados nos ossos temporais e cada

um possui três partes: ouvido externo, ouvido médio e ouvido interno. O ouvido externo é formado pelo pavilhão auditivo, chamada de orelha, e sua função é captar os sons, direcionando-os para o interior do conduto auditivo. Também compõe o ouvido externo o canal auditivo externo, que é um canal que conduz o som para o interior do ouvido. O ouvido médio é também conhecido como a caixa do tímpano. Nele, se encontra o tímpano, que é uma fina membrana responsável por transformar as vibrações sonoras em mecânicas. No ouvido médio se encontram três ossos, chamados de bigorna, martelo e estribo, que se articulam uns com os outros e recebem as vibrações oriundas do tímpano. Do ouvido médio sai um canal (trompa de Eustáquio) que vai até a faringe e sua função é manter a pressão da caixa do tímpano igual a pressão atmosférica. O ouvido interno recebe o nome de labirinto e está dividido em três partes: vestíbulo – cavidade separada do ouvido médio pela janela oval; canais semicirculares – são três tubos em forma de semicírculo (Figura 3.2); cóclea ou caracol – canal de aproximadamente 2,5 centímetros com forma de espiral.

Os ossos que compõe o ouvido, transmitem mecanicamente as vibrações da membrana timpânica até a membrana que cobre o vestíbulo. A seguir, as vibrações mecânicas se transformam em pressão hidráulica, que se propagam no fluído que preenche o caracol. Receptores sensoriais captam estas ondas e enviam estes sinais, agora elétricos, até o cérebro, que serão posteriormente interpretados como som [TAF 96].

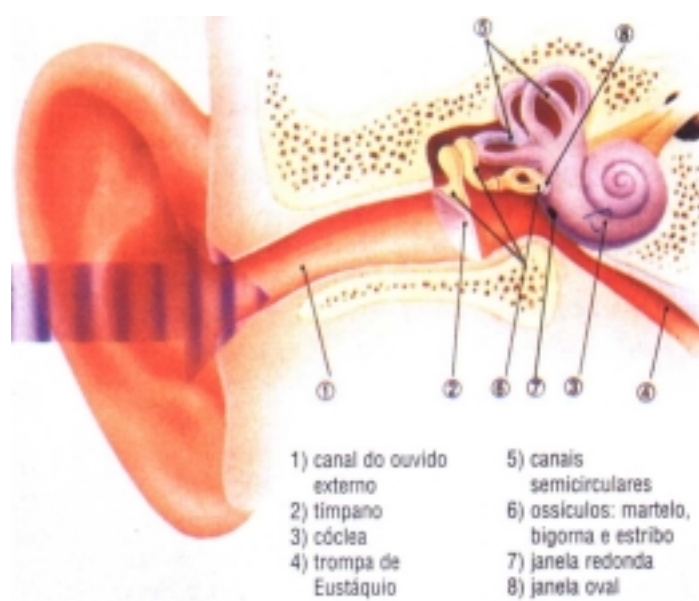


Figura 3.2 - Esquema do ouvido humano [BAR 98]

4 O SISTEMA NERVOSO HUMANO

4.1 Introdução

O propósito principal da Neurociência é compreender como o encéfalo produz a acentuada individualidade da ação humana. A compreensão da conduta humana necessita de um estudo neurobiológico que parte da mente e atravessa o mundo molecular, ou seja, como se relacionam as moléculas responsáveis pelas atividades das células nervosas para resultar nos complexos processos mentais.

O encéfalo é uma rede de mais de 100.000 milhões (100 bilhões) de células nervosas delimitadas, que se interconectam em sistemas que produzem nossa percepção do mundo exterior, concentram nossa atenção e controlam o mecanismo da ação. Por tanto, o primeiro passo para conhecer a mente é entender como os neurônios se organizam em vias de comunicação e como as células nervosas individualizadas do encéfalo se comunicam mediante transmissão sináptica. Por fim, seria necessário estudar como alterações em genes individuais afetam a comunicação entre células nervosas e como alterações na comunicação alteram a conduta [KAN 98].

Esses estudos, desde a embriogênese até a neurofisiologia são objeto de estudo da Neurociência que utilizou, inicialmente, a Anatomia macroscópica via disseções anatômicas de órgãos e sistemas (clássicos gregos, Idade Média), depois a Anatomia microscópica (Histologia e Citologia, onde se encontram os célebres estudos de Ramón y Cajal - Espanha), a Biofísica e, recentemente, a Biologia Molecular como ferramenta de estudo. Assim, a Neurociência surgiu no último século como resultado de estudos do sistema nervoso realizados por várias disciplinas clássicas. Atualmente, novas técnicas aportam os meios para vincular diretamente a dinâmica molecular de células nervosas individuais com representações de atos perceptivos e motores do encéfalo e para relacionar estes mecanismos com a conduta observável. As novas técnicas de neuroimagem, por exemplo, permitem observar o encéfalo humano em ação (identificar as regiões específicas que se associam com o pensamento e o sentimento). A neurociência, com sua capacidade de interligar a Biologia Molecular e os estudos cognitivos possibilitou que se comece a explorar a Biologia do potencial humano, de modo que, possamos entender o que determina o que somos (por exemplo, estudos recentes sobre neurotransmissores, associam o comportamento mais agitado, a busca

por esportes de riscos, a pessoas com maior quantidade de receptores adrenérgicos em suas terminações nervosas; os problemas de depressão ligados diretamente a quantidade de mediadores serotoninérgicos na corrente sanguínea, e assim por diante.

4.2 Cérebro e Conduta

Talvez a última fronteira da ciência seja entender as bases biológicas da consciência e dos processos mentais pelos que percebemos, atuamos, aprendemos e recordamos.

A tarefa da Neurociência é aportar explicações da conduta em termos de atividades do encéfalo, explicar como atuam os milhões de células nervosas individuais no encéfalo para produzir a conduta e como, por sua vez, estas células estão influenciadas pelo meio ambiente, incluindo a conduta de outros indivíduos. Para responder perguntas como – “Estão localizados os processos mentais em regiões específicas do encéfalo ou representam uma propriedade coletiva e emergente do encéfalo em sua totalidade?” “Vários processos mentais podem localizar-se em diferentes regiões cefálicas?” “Que regras relacionam a Anatomia e a fisiologia de uma região com sua função específica na percepção, no pensamento ou no movimento?” é necessário entender como está estruturado o Sistema Nervoso Central em todos aspectos desde sua embriogênese até sua histofisiologia, biologia molecular e também incluir possíveis alterações teratogênicas.

O estudo integral destes aspectos não pode ser abordado em uma síntese introdutória do Sistema Nervoso, mas que, no entanto é necessária para a construção da estrutura do presente estudo, ou seja, a construção de redes neuronais. Ressalta-se que estes estudos devem ser realizados pois o simples modelo de organização da estrutura das redes neuronais durante a embriogênese [HAR 87], por exemplo, pode ser útil na estruturação de uma rede neuronal artificial [CHU 92].

4.3 Relação entre cérebro e conduta

O conceito atual das células nervosas, o encéfalo e a conduta surgiu ao longo do último século a partir de cinco disciplinas experimentais principais: Anatomia, Embriologia, Fisiologia, Farmacologia e Psicologia.

Antes que se inventasse o microscópio óptico (século XVIII) a idéia vigente era de que o sistema nervoso tinha função glandular (idéia atribuída a Antigüidade e a proposta de Galeno de que os nervos são canais que conduzem um fluído secretado pelo encéfalo e medula espinhal até a periferia do corpo). No entanto, somente no final do século XIX através dos estudos de Camilo Golgi (Italiano - Médico) e Ramon y Cajal (Espanhol - Histólogo) que o tecido nervoso passou a ser reconhecido como tal. De seus estudos realizados com colorações de prata, estes individualizaram a estrutura do neurônio e Ramon y Cajal estruturou alguns conceitos básicos que se denominaram “Doutrina do Neurônio”, ou seja, o princípio básico que os neurônios individuais são os elementos básicos da sinalização do sistema nervoso.

Esta Teoria foi reforçada pelos estudos de Ross Harrison (EUA-Embriologista) que com seus estudos de cultivos celulares observou que os dendritos e axônio são prolongações do corpo celular neuronal e crescem a partir deste. Demonstrou que o extremo do axônio da origem a um cone que determina a direção de crescimento deste em direção a outros tecidos [HAR 35].

No final do século XVIII Luigi Galvani (Italiano – Médico e Físico) descobriu com seus estudos eletrofisiológicos que o músculo vivo excitável e os neurônios produzem eletricidade. Já no século XIX Emil Du Bois-Reymond, Johannes Müller e Hermann von Helmholtz (Alemanha – Fisiologista) estabeleceram as bases da eletrofisiologia ao descobrir que a atividade elétrica de uma célula nervosa afeta a atividade de outra célula de modo previsível.

Claude Bernard (Francia – Farmacologista), Paul Ehrlich (Alemanha – Farmacologista) e John Landley (Inglaterra – Farmacologista) no final do século XIX demonstraram que as drogas interatuam com receptores específicos das células. Tal descobrimento é a base para o estudo da natureza química das comunicações entre células nervosas.

A investigação psicológica da conduta tem suas raízes nos princípios da ciência ocidental com a filosofia grega clássica. Muitos dos temas principais da investigação moderna da conduta, particularmente na área da percepção, foram planteados nos escritos de René Descartes (morto em 1650), John Locke (morto em 1704) e David Hume (morto em 1776). A princípios do século XIX as investigações de Charles Darwin sobre a evolução foram o ponto de partida da observação sistemática dos atos e

da conduta. Daí resultaram a Psicologia Experimental (comportamento humano) e a Etologia (Comportamento animal).

Os intentos de relacionar os conceitos biológicos com os psicológicos no estudo da conduta ocorreram muito cedo, a finais do século XVIII, Franz Joseph Gall (Viena – Médico e Neuroanatomista) propôs que regiões delimitadas do córtex cerebral controlam funções específicas. Gall defendeu que o encéfalo não funciona como um órgão unitário, mas sim composto por 35 órgãos (mais tarde foram acrescentados outros), cada um deles correspondentes a uma faculdade mental específica. Gall também propôs que o centro de cada função aumenta de acordo com o seu grau de uso, a exemplo do que ocorre com o tecido muscular. Dai surge a Teoria de que com o crescimento de uma dada área esta se sobrepunha as demais e com isto propôs a teoria da Frenologia (descrição do caráter em bases meramente anatômicas) [COO 84].

Em finais de 1820 Pierre Flourens (França) experimentou em animais a Teoria de Gall extirpando-lhes partes do encéfalo e concluiu que as condutas específicas não dependem exclusivamente de regiões específicas do encéfalo, se não que todas as regiões do encéfalo (especialmente dos hemisférios cerebrais e pró-encéfalo) participam em cada função mental. Com seus escritos propôs a Teoria do Campo Agregado que deixava um pouco de lado a visão estritamente materialista da conduta.

Na metade do século XIX J. Hughlings Jackson (Inglês – Neurologista) realizando estudo clínico de epilepsia focal, demonstrou que diferentes processos sensoriais e motores se localizam em diferentes partes do córtex cerebral. Posteriormente Karl Wernicke (Alemanha – Neurologista), Charles Sherrington (Inglaterra – Fisiologista) e Ramón y Cajal (Espanha – Histólogo) realizaram estudo elaborados sistematicamente com um enfoque oposto da função cerebral, denominado “Conexionismo celular”. Segundo está hipótese, os neurônios individuais são as unidades de sinalização do encéfalo; organizam-se pelo geral em grupos funcionais e se conectam um com outro de modo preciso. Wernicke demonstrou em particular que determinadas condutas estão mediadas por diferentes regiões do encéfalo, as quais se interconectam mediante vias neurais particulares.

A história do enfrentamento entre partidários da Teoria do Campo Agregado e do Conexionismo Celular pode ser exemplificada analisando como o encéfalo controla a linguagem, função mental humana culminante e certamente mais característica. Para

entender este processo é necessário realizar previamente uma revisão anatômica breve da estrutura do encéfalo.

4.4 Regiões do Encéfalo

O sistema nervoso é bilateral e essencialmente simétrico, estando constituído de sete partes principais: medula espinhal, bulbo raquídeo, protuberância, cerebelo, cérebro médio, diencéfalo e hemisférios cerebrais (Figura 4.1).

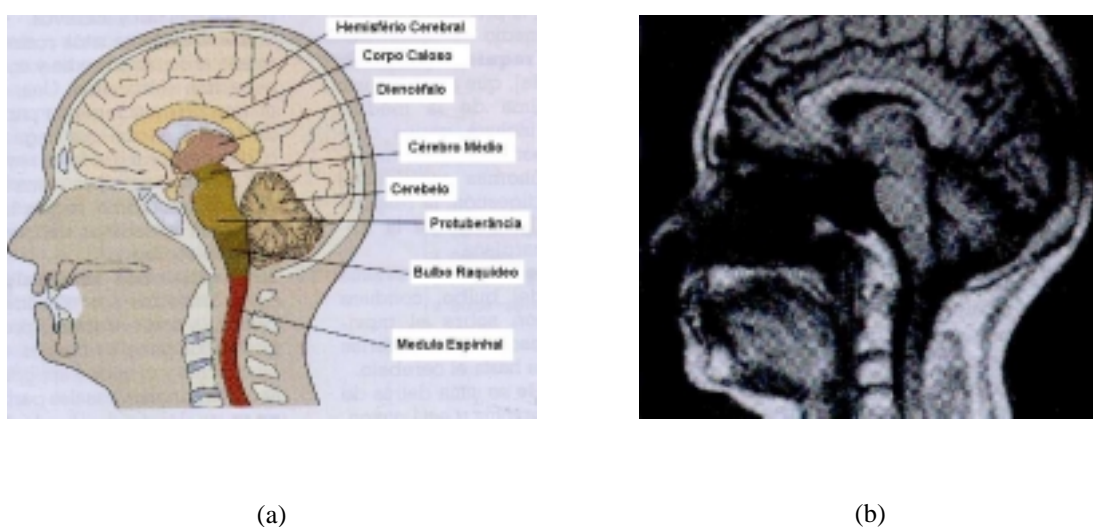


Figura 4.1 - Principais divisões do encéfalo [MINb 00]

As principais divisões do encéfalo se distinguem claramente quando se secciona o encéfalo seguindo a linha média entre os dois hemisférios [MINb 00].

- (a) Este esquema mostra a posição das principais estruturas do encéfalo referentes às delimitações externas. Os estudantes de anatomia cerebral aprendem logo a reconhecer as delimitações internas tais como o corpo caloso, um extenso feixe de fibras nervosas que conecta o hemisfério esquerdo e direito.
- (b) A mesma seção do esquema (a) vista em uma imagem de ressonância magnética de um encéfalo *in vivo*.

Com as técnicas de neuroimagem se pode visualizar *in vivo* estas estruturas do encéfalo humano. Através destes estudos realizados por diversos métodos experimentais se demonstrou que distintas regiões do encéfalo têm funções específicas. Como

consequência, a idéia de que diferentes regiões estão especializadas em diferentes funções, se considera atualmente um dos pilares da ciência do cérebro.

Com o conhecimento do “Processamento Paralelo” (onde em geral, cada uma das principais funções sensoriais, motoras e outras de integração se servem de mais de uma via neural), quando se lesiona uma via, *a priori* outras podem compensar parcialmente a perda, obscurecendo assim a evidência de sua localização. Sem dúvida, a precisão com a qual estão localizadas certas funções superiores é evidente no momento em que se examina a linguagem, tema que será analisado agora.

4.5 A Linguagem e outras funções cognitivas estão localizadas no córtex cerebral

As funções cerebrais relacionadas com a linguagem estão localizadas basicamente no córtex cerebral que envolve e recobre os hemisférios cerebrais como a casca de uma árvore. Em cada um dos hemisférios do encéfalo, o córtex que os rodeia se divide em quatro lobos anatomicamente distintos: frontal, parietal, occipital e temporal (Tabela 4.1). Os lobos denominados assim originalmente pelos ossos específicos do crânio que os cobrem, têm funções especializadas. O lobo frontal está consideravelmente implicado na planificação da ação futura e no controle do movimento; o lobo parietal na sensação tátil e imagem corporal; o lobo occipital na visão e o lobo temporal na audição e em certos aspectos da aprendizagem, memória e emoção. Cada lobo tem várias circunvoluções ou pregas características; isto é uma estratégia evolutiva para aumentar a área de superfície útil em um espaço restringido. As proeminências das circunvoluções se denominam “giros”. Os canais intercalados se denominam “sulcos”. As circunvoluções e sulcos mais proeminentes são os mesmos em todos indivíduos e têm nomes específicos [KAN 98].

Tabela 4.1 e Figura 4.2 - Subdivisão Anatômica do Sistema Nervoso Central

Parte do SNC	Localização / Função
Medula Espinhal	A Medula Espinhal, a parte mais caudal do sistema nervoso central, recebe e processa informação sensorial da pele, articulações e músculos das extremidades e tronco, controla o movimento das extremidades e do tronco. Se subdivide nas regiões cervical, torácica, lombar e sacra. A medula espinhal se une rostralmente com o tronco cerebral, que conduz informação nos dois sentidos entre a medula espinhal e o encéfalo. O tronco cerebral contém vários grupos distintos de corpos celulares, os núcleos dos nervos craniais. Alguns destes núcleos recebem informação da pele e músculos da cabeça; outros controlam o <i>output</i> motor dos músculos da face, pescoço e olhos. Outros, por sua vez, estão especializados em informação de sentidos específicos: o ouvido, o equilíbrio e o paladar. O tronco cerebral também regula os níveis de alerta e de consciência através da formação reticular, que tem projeções difusas. O tronco cerebral consta de três partes: o bulbo raquideo, a protuberância e o cérebro médio.
Bulbo Raquideo	O Bulbo Raquideo (ou medula oblonga), que se situa justo acima da medula espinhal, inclui vários centros responsáveis de funções autônomas como a digestão, a respiração e o controle da frequência cardíaca.
Protuberância	A Protuberância, que se situa acima do bulbo, conduz a informação sobre o movimento desde os hemisférios cerebrais até o cerebelo.
Cerebelo	O Cerebelo está situada atrás da protuberância e está conectado com o tronco cerebral por vários tratos destacados de fibras chamadas pedúnculos. O cerebelo modula a força e a disposição do movimento e está implicado na aprendizagem das habilidades motoras.
Cérebro Médio	O Cérebro Médio (mesencéfalo) que se situa rostralmente a protuberância, controla muitas funções sensoriais e motoras, incluindo os movimentos oculares e a coordenação dos reflexos visuais auditivos.
Diencefalo	O Diencefalo situado rostralmente ao cérebro médio contém duas estruturas. Uma, o tálamo, processa a maior parte da informação que chega ao córtex cerebral desde o restante do Sistema Nervoso Central. Outra, o hipotálamo, regula as funções autônomas e vísceras.
Hemisférios Cerebrais	Os Hemisférios Cerebrais estão formados pelo córtex cerebral e três estruturas profundas: os gânglios basais, o hipocampo e o núcleo amigdalóideo. Os gânglios basais participam da regulação da conduta motora, o hipocampo está implicado em processos de armazenamento da informação e o núcleo amigdalóideo coordena as respostas autônomas e endócrinas com os estados emocionais. Recobrindo ambos hemisférios está a capa do córtex densamente pregueada, que se divide em quatro lóbulos: frontal, parietal, temporal e occipital. O encéfalo também se divide em três regiões mais amplas: o cérebro posterior (o bulbo, a protuberância e o cerebelo), o cérebro médio e o cérebro anterior (o diencefalo e os hemisférios cerebrais). O cérebro médio e o cérebro posterior(excluindo o cerebelo) constituem o tronco cerebral.

Fonte: [MINb 00]

O Sistema Nervoso Central tem sete partes principais, representadas na Figura 4.2.



Figura 4.2 - Sistema Nervoso Central

Grande parte do que sabemos sobre a localização da linguagem procede do estudo da afasia, uma categoria de transtorno de linguagem devido a problemas médicos. A afasia é mais freqüente em pacientes que tenham sofrido um acidente (obstrução ou rompimento de um vaso sanguíneo que irriga uma parte do hemisfério cerebral) [MINb 00].

Inúmeros estudos desta patologia se desenvolveram desde a segunda parte do século XIX, sendo que o primeiro avanço ocorreu em 1861 quando Pierre Paul Broca (França – Neurologista) descreveu o caso de um paciente que podia compreender a linguagem mas não podia falar. O paciente não tinha problemas funcionais na língua, boca ou cordas vocais que poderiam evitar que falasse. Podia pronunciar palavras soltas e cantar uma melodia sem dificuldade, mas não podia falar gramaticalmente ou com frases completas, nem podia expressar as idéias por escrito. O exame post-mortem (*post mortem*) do encéfalo do paciente revelou uma lesão na região posterior do lóbulo frontal (área hoje denominada “área de Broca”) (Figura 4.3). De seus estudos com outros oito pacientes com a mesma patologia Broca enunciou: “Falamos com o hemisfério esquerdo” (*Nous parlons avec l’hémisphère gauche*).

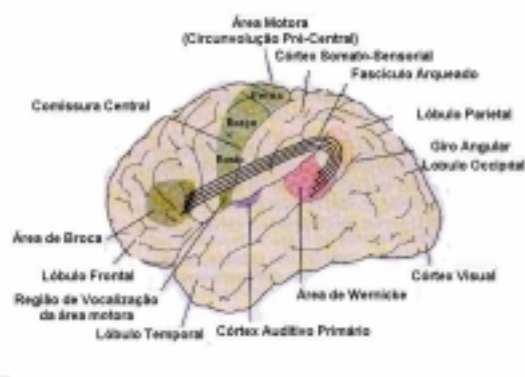


Figura 4.3 - O córtex cerebral (Modificado de Geschwind, [GES 74])

Na vista lateral do hemisfério esquerdo se apresentam as principais áreas do córtex cerebral implicadas na linguagem. A área de Wernicke processa o *input* auditivo para a linguagem e é importante para a compreensão da fala. Está localizada perto do córtex auditivo primário e do giro angular, que integra o *input* auditivo com a informação procedente de outros sentidos. A área de Broca controla a produção da fala. Está localizada perto da região da área motora que controla os movimentos da boca e da língua que produzem as palavras. A área de Wernicke se comunica com a área de Broca através de um trato de fibras, o fascículo arqueado.

A partir deste trabalho se iniciou a busca da sede cortical de outras funções comportamentais específicas. Em 1870 Gustav Fritsch (Alemanha - Fisiologista) e Eduard Hitzig (Alemanha – Psiquiatra) descobriram que a estimulação elétrica de certas regiões do encéfalo de cão produzia movimentos característicos nas extremidades. Descobriram que cada movimento particular está controlado por uma pequena região do córtex, bastante delimitada; além disso, estes movimentos de uma extremidade se produziam ao estimular a circunvolução pré-central do córtex motor contralateral. Assim, em humanos, a mão direita, utilizada habitualmente para escrever e realizar movimentos de precisão está controlada pelo hemisfério esquerdo, o mesmo que controla a fala. Por tanto, se considera que o hemisfério esquerdo é dominante na maioria das pessoas [KAN 98].

A seguinte etapa iniciou com Carl Wernicke (Alemanha – Neurologista) quando em 1876, descreveu um segundo tipo de afasia (caracterizada por uma disfunção receptiva por contraposição a uma expressiva) publicando um trabalho intitulado “O Complexo de sintomas da afasia: Um estudo psicológico sobre uma base anatômica”. Enquanto os pacientes de Broca podiam entender mas não falar, o paciente de Wernicke

podia falar mas não entender a linguagem (inclusive suas próprias palavras). Esta afasia era determinada por lesões na parte posterior do lóbulo temporal, onde se une com os lobulos parietal e occipital (Figura 4.3).

Baseando-se em seus estudos, nos de Broca, Fritsch e Hitzig, Wernicke propôs que somente as funções mentais mais básicas, as concernentes as atividades perceptivas e motoras, estão localizadas em áreas particulares do córtex; enquanto que, as funções intelectuais mais complexas são o resultado de interconexões entre várias zonas funcionais. Situando o princípio de função localizada em um contexto connexionista, Wernicke considerou que os diferentes componentes de uma conduta determinada se processam em diferentes regiões do encéfalo. Assim, avançou a primeira prova da idéia do “processamento distribuído”, que na atualidade é um conceito central do conhecimento da função cerebral.

Wernicke postulou que a linguagem envolve programas sensoriais e motores específicos, distintos; estando cada um deles sob o controle de regiões corticais específicas. Propôs que o programa motor, que rege os movimentos da boca implicados na fala, se localiza na área de Broca; que, adequadamente, situa-se justo diante da área motora que controla a boca, a língua, o palato e as cordas vocais (Figura 4.3). O programa sensorial que regula percepção da palavra foi atribuído a área do lóbulo temporal (atual área de Wernicke). Esta área também se localiza adequadamente, estando rodeada pelo córtex auditivo como por áreas que integram as sensações auditivas, visuais e somáticas para formar percepções complexas (áreas denominadas Córtex de Associação).

Este trabalho de Wernicke serviu de base estrutural do modelo de organização da linguagem, atualmente modificado por novos achados, mas mantido em sua essência. Inspirada neste trabalho no início do século XX na Alemanha surgiu uma nova escola de localização cortical coordenada por Korbinian Brodmann (Anatomista). Foram descritas 52 áreas funcionais do córtex cerebral humano pelo método “citoarquitetônico”.

Todos estes estudos deixaram uma evidente base anatômica e funcional de que no córtex existiam muitas áreas delimitadas, a algumas das quais podiam atribuir-se um papel específico em certas condutas. Não obstante, a teoria do Campo Agregado foi dominante e não a Connexionista no início do século XX. Isto foi devido a argumentos de neurocientistas destacados como Henry Head (Inglaterra-Neurologista), Kurt

Goldstein (Alemanha – Neuropsicólogo), Ivan Pavlov (Rússia – Fisiologista Comportamental), Jacques Loeb e Karl Lashley (EUA) em defesa da Teoria do Campo Agregado.

Lashley foi o maior defensor desta teoria e deixou a importância da célula neuronal e suas conexões em segundo plano, dando importância ao conjunto da “Massa cerebral”, utilizando experimentos com ratos em labirintos seguindo a tradição do trabalho de Fluorens (busca de uma sede específica da aprendizagem). Este lesionava regiões do cérebro dos ratos e observava seus comportamentos, avaliando assim, somente a extensão do dano.

Os estudos de Lashley, Head e Goldstein foram reinterpretados e reavaliados, concluindo-se que a tarefa empregada era inapropriada para estudar a localização de função já que implica muitas capacidades sensoriais e motoras complexas. Quando se priva o animal de uma capacidade sensorial (por exemplo, a visão), pode aprender com outras (por exemplo, seguindo sinais táteis e olfativos).

Em finais dos anos 30, Edgar Adrian (Inglaterra), Wade Marshall [MAR 88a] [MAR 88b] e Phillip Bard (EUA) descobriram que estímulos táteis promovem a atividade elétrica em regiões definidas do córtex cerebral. A seguir Jerzy Rose e Clinton Woolsey reexaminaram o conceito de “rede arquitetônica” e muitos outros pesquisadores. Segundo estes estudos as áreas corticais podem ser definidas com precisão segundo vários critérios independentes, incluindo o tipo de célula e a estratificação celular, as conexões aferentes e eferentes e o mais importante, a função (fisiologia).

Baseado nestes estudos, durante cirurgias de epilepsia, Wilder Penfield estimulava áreas de pacientes submetidos a anestesia local para não lesar áreas da fala descritas por Broca e Wernicke, em estudos no encéfalo *in vivo*.

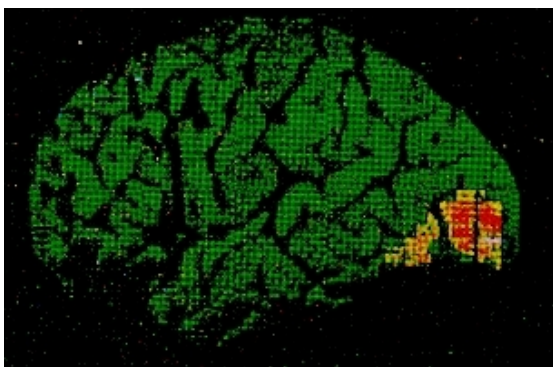
Atualmente Michael Raichle e colaboradores passaram a estudar indivíduos saudáveis utilizando a exploração mediante Tomografia por Emissão de Pósitrons (TEP ou PET). Esta é uma técnica de neuroimagem não invasiva para visualizar mudanças locais no fluxo sanguíneo cerebral e o metabolismo que se associa com atividades mentais, tais como ler, falar e pensar. Posner [POS 94] e colaboradores descobriram que *inputs* neurais para a produção e compreensão da linguagem são processados por mais de uma via. Estes concluíram que se utilizam diferentes vias encefálicas e códigos sensoriais

para perceber palavras apresentadas oralmente ou visualmente. Propuseram que estas vias têm acesso independente às regiões de nível superior, que se encarregam de dar o significado e da expressão da linguagem.

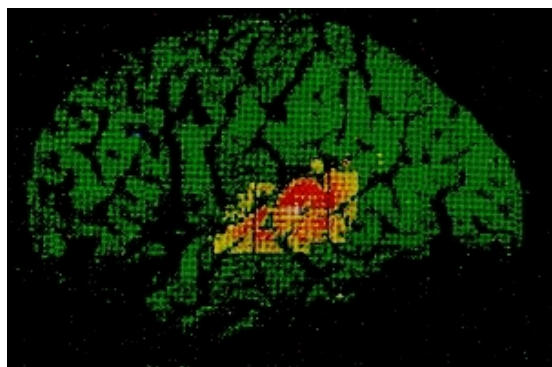
Não só a leitura e a escuta ativas se processam por separado, se não também quando um indivíduo simplesmente pensa o significado de uma palavra, sem recorrer a outros *inputs*, se ativa uma área diferente no córtex frontal esquerdo. Assim, o processamento da linguagem ocorre tanto em série como em paralelo. Estes estudos demonstraram que o processamento da informação requer que áreas corticais particulares estejam interconectadas apropriadamente e que respondam a, e portanto codifiquem, somente certos aspectos de estímulos sensoriais específicos ou de movimentos motores e não outros [KAN 98].

Tabela 4.2 - Tomografia por Emissão de Pósitrons do Córtex Cerebral – vista lateral do hemisfério esquerdo em distintos momentos de conduta relacionados a linguagem

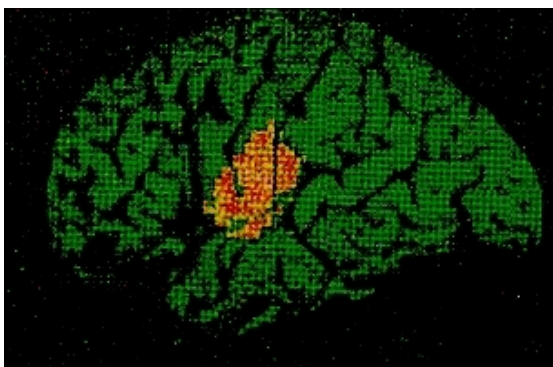
A. Lendo as palavras



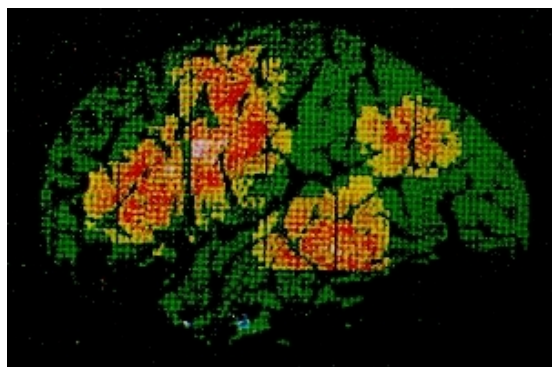
B. Escutando as palavras



C. Pronunciando as palavras



D. Pensando as palavras



Fonte: Kandel *et al.*, 1998

O córtex cerebral tem duas características de organização importantes. A primeira, cada hemisfério se ocupa basicamente dos processos sensoriais e motores do lado oposto ou contralateral do corpo. A informação sensorial que chega a medula espinhal da parte esquerda do corpo cruza ao lado direito do sistema nervoso antes de ser conduzida ao córtex cerebral. De modo similar, as áreas motoras de um hemisfério do encéfalo exercem o controle dos movimentos da metade oposta do corpo. A segunda é que, mesmo que os hemisférios cerebrais parecem ser semelhantes em humanos, não apresentam uma estrutura completamente simétrica (tão pouco tem uma função equivalente) [MINb 00].

Com uma TEP (Tomografia por Emissão de Pósitrons - Tabela 4.2) é possível identificar regiões específicas do córtex implicadas no reconhecimento de uma palavra falada ou escrita. Cada uma das quatro imagens do encéfalo humano mostradas (vistas laterais do hemisfério esquerdo) representa a média de atividade cerebral que está ocorrendo em vários indivíduos normais. Nas imagens de TEP, o branco representa as áreas de maior atividade, o amarelo as áreas de atividade elevada e o azul as áreas de atividade mínima. O componente de *input* da linguagem (ler ou escutar uma palavra) ativa as regiões do encéfalo mostradas em **A** e **B**. O componente de *output* (fala ou pensamento) ativa as regiões mostradas em **C** e **D**.

- A.** A leitura de uma só palavra produz uma resposta no córtex visual primário e no córtex visual de associação.
- B.** Escutar uma palavra ativa uma série completamente diferente de áreas no córtex temporal e na convergência do córtex temporal e parietal. Isto demonstra que o encéfalo utiliza vias auditivas e visuais separadas para processar a linguagem e não uma única via comum. (Para controlar diferenças que poderiam interferir foram utilizadas as mesmas palavras nas provas de leitura e escuta).
- C.** Pronunciar uma palavra ativa a área motora suplementária na região medial do córtex frontal. Foi solicitado aos indivíduos que repetissem uma palavra apresentada mediante auriculares ou em uma tela. A área de Broca se ativa tanto se a palavra é apresentada oralmente ou visualmente. Assim, a via visual e a via auditiva convergem na área de Broca, a região comum para a produção motora da fala.
- D.** O pensamento, tal como analisar o significado de uma palavra, ativa o córtex frontal. Para identificar as regiões ativas durante o pensamento foi solicitado aos indivíduos que respondessem a palavra “encéfalo” com um verbo apropriado (por exemplo, “pensar”). As áreas de Broca e de Wernicke também estão implicadas.

4.6 O Neurônio

Todos os animais, inclusive o homem, obtêm informação sobre o seu entorno através de vários receptores sensoriais. A informação conseguida pelos receptores se transforma no encéfalo em percepções ou ordens para o movimento. Respostas tão notáveis são conseguidas somente com a utilização de células nervosas e as conexões estabelecidas entre elas. O comentário realizado neste parágrafo refere-se ao que em

fisiologia se denomina de “Arco Reflexo”. Neste caso, como é uma resposta elaborada e interpretada pelos centros nervosos superiores, diz-se que se trata de um Arco Reflexo Central. Já um reflexo de sobrevivência (saltar durante um susto, retirar a mão de uma superfície quente, reflexo patelar, e outras coisas do gênero) são respostas imediatas, sem interpretação detalhada e coordenadas pela medula espinhal, sendo denominadas de Arco Reflexo Periférico. O esquema a seguir ilustra um arco-reflexo de forma simplificada.

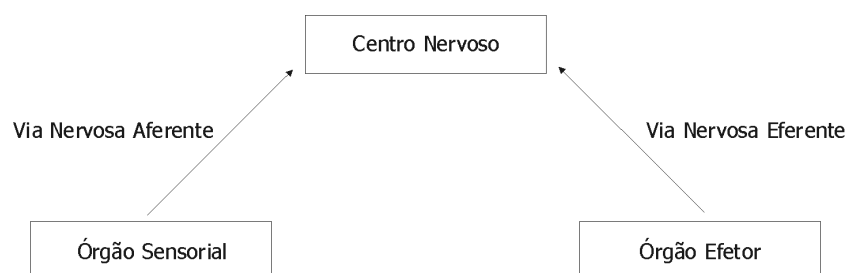


Figura 4.4 - O Arco reflexo [MINb 00]

As unidades básicas do encéfalo, as células nervosas, são muito simples. O encéfalo é capaz de gerar comportamentos tremendamente complexos porque tem uma grande quantidade de células nervosas que se comunicam entre si mediante interconexões específicas. As células nervosas apesar de sua grande quantidade compartilham muitas características. Um dos descobrimentos mais importantes para a compreensão do encéfalo foi que o potencial de ação para produzir condutas complexas não depende, em grande parte, da variedade das células nervosas, mas sim de seu número e de suas conexões específicas entre si e com os receptores sensoriais e os músculos.

A diversidade de células do sistema nervoso é maior do que a de qualquer outro sistema do organismo. Os neurônios diferem entre si em muitos aspectos, mas compartilham atributos que os tornam distintos dos hepatócitos (fígado), fibroblastos (conjuntivo) e de outras células presentes em outros tecidos. Por exemplo, os neurônios possuem, caracteristicamente, diferentes regiões: exibem polarização regional, que é a base citológica do princípio de polarização funcional ou dinâmica, enunciado por Ramón y Cajal. O corpo celular contém o núcleo e os organóides para sintetizar ácido ribonucléico (RNA) e proteínas, é só uma das distintas regiões importantes do neurônio e em na maioria deles somente supõe uma décima parte do volume total celular. O resto

do volume se distribui nas prolongações para a comunicação celular, os dendritos e o axônio, que têm origem no corpo celular. Estas regiões, que exercem funções diferentes, podem estar separadas por distâncias relativamente grandes. Por exemplo, alguns axônios excedem um metro de comprimento [KAN 98].

Os neurônios também se diferenciam da maioria das outras células por serem excitáveis, quer dizer, podem experimentar mudanças rápidas no potencial elétrico através de sua membrana, devido ao fluxo de íons desde, ou em direção ao interior da célula. Esta excitabilidade é devida a ação de determinadas proteínas existentes na membrana celular (os canais e as bombas iônicas).

A grande diversidade neuronal está bem ilustrada no cerebelo, região do encéfalo importante para a conduta motora. O cerebelo contém cinco tipos de células nervosas, cada uma delas com uma morfologia distinta e desempenhando uma função característica. As células de Purkinje do cerebelo são umas das maiores células do sistema nervoso dos vertebrados. Seus somas(corpos celulares) têm oitenta micrômetros de diâmetro e seus dendritos se ramificam profusamente ao longo de distâncias consideráveis para receber distintos tipos de mensagens aferentes [KAN 98]. Ao contrário, os corpos celulares das células granulares do cerebelo têm somente seis a oito micrômetros de diâmetro, compondo-se de um núcleo rodeado por uma envoltura citoplasmática mínima. Os dendritos destas células não se estendem mais distante do que a própria distância que apresenta o corpo celular.

A diversidade celular é o resultado da diferenciação que acontece durante a ontogênese do sistema nervoso e que não será tratada neste trabalho direcionado ao aspecto citológico da célula nervosa.

De um modo geral se pode ilustrar a morfologia dos neurônios através da descrição de neurônios sensoriais e neurônios motores. A estrutura geral de um neurônio consiste em um corpo celular (soma) e seus prolongamentos citoplasmáticos, os dendritos e axônios (Figura 4.5). Nas extremidades dos dendritos se encontram botões ou vesículas que contêm mediadores químicos utilizados na comunicação celular. O mesmo ocorre nas extremidades dos axônios. As fibras nervosas (axônios) podem ou não estar envoltas por uma bainha lipídica, denominada de bainha de mielina. Esta funciona com um isolante elétrico e determina a condução mais rápida do impulso elétrico. A disposição da bainha de mielina é dada em função do arranjo das células de

Schwann, responsáveis por sua secreção no Sistema Nervoso Periférico e pelos Oligodendrócitos, no Sistema Nervoso Central.

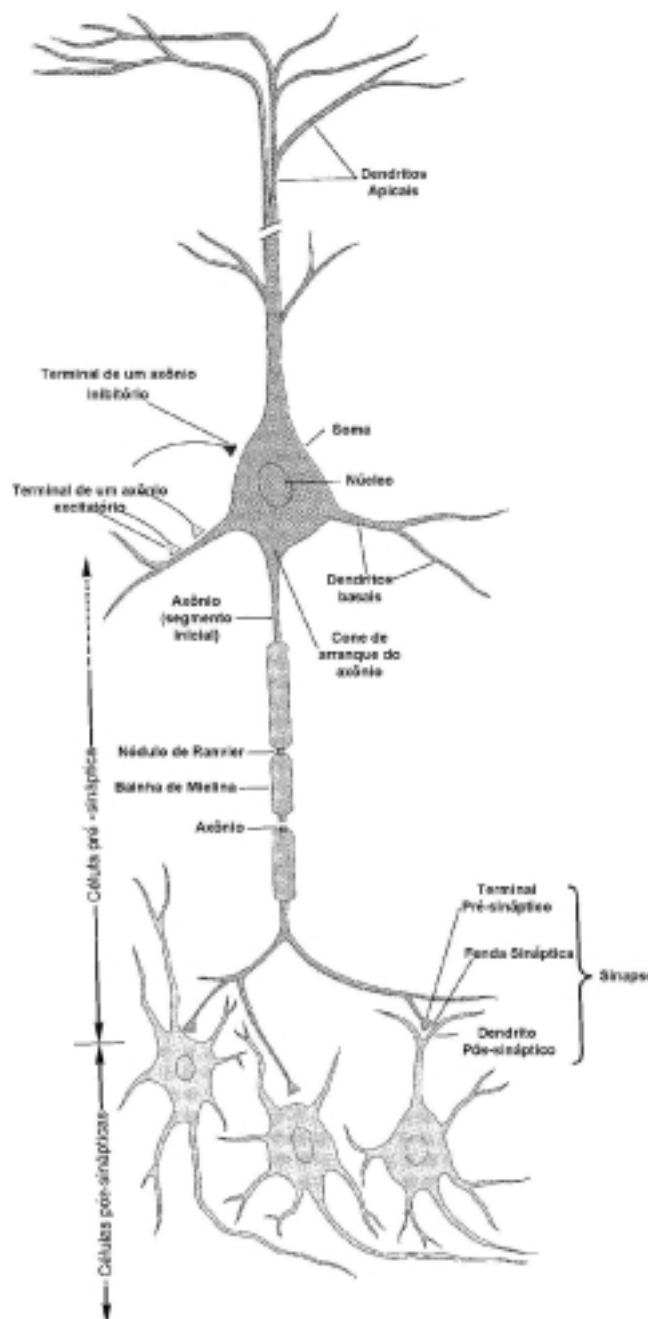


Figura 4.5 - Estrutura do Neurônio [KAN 98]

No sistema nervoso dos vertebrados a maior parte dos neurônios possui características principais comuns. O corpo celular contém o núcleo, o armazém da informação genética. O corpo celular origina dois tipos de prolongações os dendritos e o axônio. Os axônios, os elementos de transmissão dos neurônios, podem variar muito em

comprimento; alguns se estendem por mais de um metro além do corpo celular. A maioria dos axônios do sistema nervoso central são muito finos (0,2 a 20 micrômetros de diâmetro) se comparados com o diâmetro do corpo celular (50 micrômetros). O cone de arranque do axônio é o lugar onde inicia o potencial de ação, o sinal de comunicação celular. Muitos dos axônios se isolam graças a bainha de mielina (gordura) que se interrompe a intervalos regulares nas regiões conhecidas como nódulos de Ranvier. As ramificações do axônio de um neurônio (o neurônio pré-sináptico) transmitem sinais a outro neurônio (a célula pós-sináptica) em um lugar denominado sinapse. Os ramos de um só axônio podem estabelecer sinapses com outros mil neurônios. Os dendritos (apicais e basais) constituem a maior parte da superfície receptora de um neurônio e, junto com o corpo celular, recebem as mensagens aferentes procedentes das células pré-sinápticas.

A bainha de mielina deixa algumas incisuras de trecho em trecho, denominadas “incisuras de Schmidt-Lantermann” e de trechos em trechos sofre algumas constrictões, chamadas de “nódulos de Ranvier”. As fibras mielinizadas apresentam um processo mais rápido de condução dos estímulos devido ao fato dos impulsos se deslocarem ao longo de sua extensão de forma saltatória (de nódulo de Ranvier a nódulo de Ranvier) e não de forma retilínea como ocorre nas fibras amielinizadas.

É importante ressaltar que os neurônios são classificados de acordo com a sua morfologia celular (número de prolongamentos celulares) em unipolares, bipolares e multipolares (Figura 4.6). As células unipolares são os neurônios mais simples. Geralmente têm uma prolongação primária única, que habitualmente se divide em muitos ramos. Uma delas serve como axônio, outras funcionam como estruturas dendríticas de recepção. As células unipolares carecem de dendritos que emergem do soma. Este tipo de célula predomina no sistema nervoso de invertebrados e estão presentes nos vertebrados em certos gânglios do sistema nervoso autônomo.

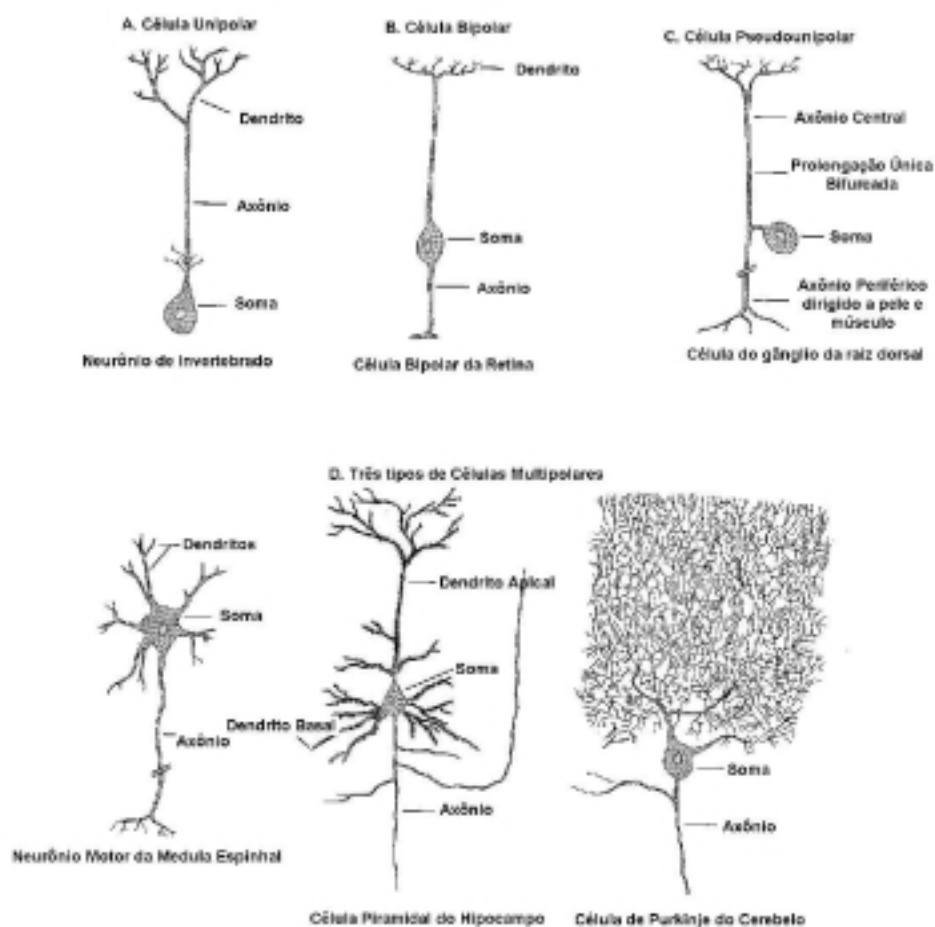


Figura 4.6 - Classificação dos neurônios quanto ao número e forma dos prolongamentos citoplasmáticos. [KAN 98]

Os neurônios podem classificar-se como unipolares, bipolares ou multipolares em função do número de prolongamentos que se originam desde o corpo celular (Figura 4.6).

- A.** As células unipolares têm uma prolongação única, com distintos segmentos que podem servir de superfícies receptoras, ou como terminais de liberação. As células unipolares são características do sistema nervoso dos invertebrados.
- B.** As células bipolares têm duas prolongações especializadas funcionalmente: o dendrito aporta a informação até a célula, e o axônio transmite a informação até outras células.
- C.** Certos tipos de neurônios que enviam informação sensorial a medula espinhal pertencem a uma subclasse de células bipolares denominadas de células pseudounipolares. À medida que tais células se desenvolvem, as duas prolongações da célula bipolar embrionária se fusionam e emergem do corpo celular como um

único processo. Este se divide a sua vez em dois ramos, os quais funcionam como axônios, dirigindo-se um a periferia até a pele ou músculo e o outro em direção a medula espinhal.

- D.** As células multipolares têm um axônio e muitos dendritos. São os tipos de neurônios mais comuns no sistema nervoso dos mamíferos. Três exemplos ilustram a grande diversidade na forma e organização das células multipolares. Os neurônios motores espinhais que inervam fibras musculares esqueléticas. As células piramidais têm um corpo celular triangular irregular; os dendritos emergem tanto do ápice(dendrito apical) como da base (dendrito basal). As células piramidais se encontram no hipocampo e ao longo do córtex cerebral. As células de Purkinje do cerebelo se caracterizam por sua árvore dendrítica extensa e profusa. Tal estrutura permite uma enorme quantidade de aferências sinápticas.

Os neurônios bipolares têm um soma de forma ovóide que dá lugar a dois prolongamentos – um dendrito que transporta a informação da periferia ao soma e um axônio que conduz a informação do soma ao Sistema Nervoso Central (SNC). Muitos neurônios bipolares são sensoriais, como as células bipolares da retina ou as do epitélio olfatório. As células sensoriais que portam a informação sobre o tato, pressão e dor são modelos especiais de células bipolares. Inicialmente se desenvolvem como células bipolares, mas a partir de um dado momento, as prolongações se fusionam para formar um axônio único que emerge do soma e se divide em dois. Um segmento se dirige para a superfície (órgãos sensoriais da pele, articulações e músculo) enquanto o outro se dirige a medula espinhal. Estes neurônios são denominados por estas características pseudounipolares [KAN 98].

Os neurônios multipolares são os predominantes no sistema nervoso de vertebrados. Estas células possuem um axônio único e de um a muitos dendritos que emergem tipicamente de qualquer zona do soma celular. O tamanho e a morfologia destas células varia enormemente. Em particular, as células multipolares diferem em número e comprimento de seus dendritos e também no comprimento do axônio. Na maioria destas células o número e o comprimento dos dendritos está relacionado com o número de contatos sinápticos que outros neurônios estabelecem com esta célula. Um neurônio motor espinhal, cujos dendritos tem uma extensão e um número moderados, recebe cerca de dez mil contatos, dois mil no corpo celular e oito mil nos dendritos. As

células de Purkinje do cerebelo recebem aproximadamente cento e cinquenta mil contatos.

Segundo sua função os neurônios se classificam como motores, sensoriais e interneurônios. Os neurônios sensoriais (ou aferentes) transmitem ao sistema nervoso a informação tanto perceptiva como a necessária para a coordenação motora. Os neurônios motores transmitem ordens aos músculos e glândulas. Já os interneurônios são a classe mais numerosa de neurônios e correspondem a neurônios que não são especificamente sensoriais ou motores. Os interneurônios de relevo ou de projeção têm axônios longos e portam a informação a grandes distâncias. Os interneurônios locais têm axônios curtos e processam a informação no interior de circuitos locais.

O Sistema Nervoso além dos neurônios possui outras células denominadas em conjunto de células gliais, que são: astrócitos, oligodendrócitos, microglia e células de Schwann. Estas células não serão abordadas no presente estudo por não apresentar enfoque direto ao tema enfocado.

As células nervosas são as unidades que codificam as respostas comportamentais através de suas comunicações divergentes ou convergentes (Figura 4.7) sendo que a comunicação é organizada do mesmo modo em todas as células nervosas. Para produzir uma conduta, cada célula nervosa e motora participante gera, seqüencialmente, quatro tipos de sinais: um de entrada *input*, um de integração (ativação), um sinal portador e um sinal de saída – *output*. Independente do tamanho, morfologia, bioquímica do transmissor, ou função comportamental, quase todas as células nervosas podem descrever-se mediante um modelo geral de neurônio que tem quatro regiões funcionais: um componente de entrada local (receptor), um elemento integrador (ativador), um componente condutor (sinalizador) e um elemento de saída (secretor) (Figura 4.8).

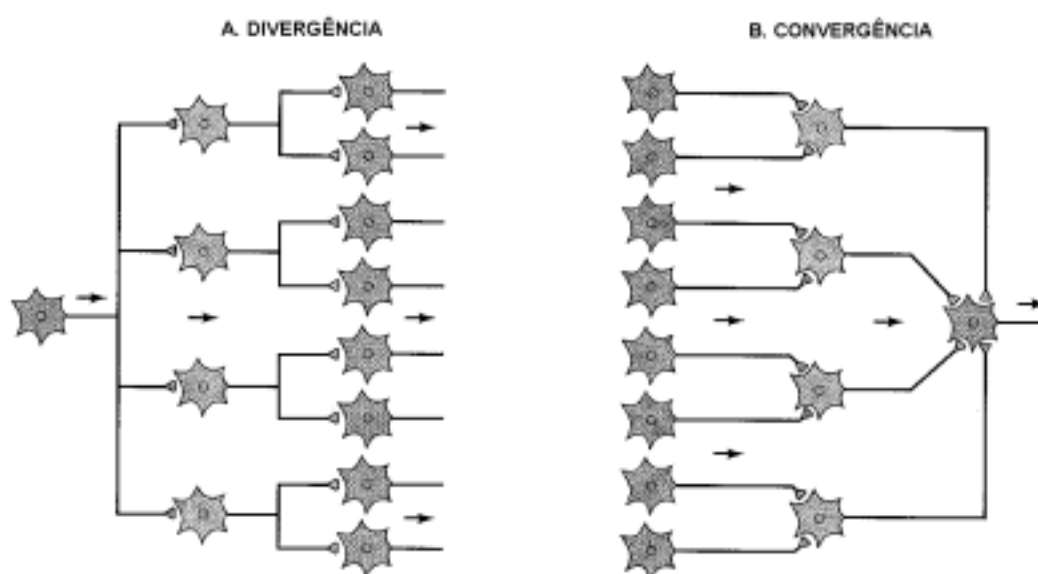


Figura 4.7 - Estágios do processamento da informação na organização cerebral [KAN 98]

A divergência e a convergência das conexões neuronais representam um princípio chave na organização do cérebro. Nos sistemas sensoriais os neurônios receptores se ramificam ao enviar suas mensagens aferentes, estabelecendo múltiplas conexões com neurônios que representam um segundo estágio no processamento da informação. (A) As conexões subsequentes divergem cada vez mais. Ao contrário, os neurônios motores são modelos de conexões progressivamente convergentes (B). Graças a convergência, as células alvo recebem a soma da informação de muitas células pré-sinápticas.

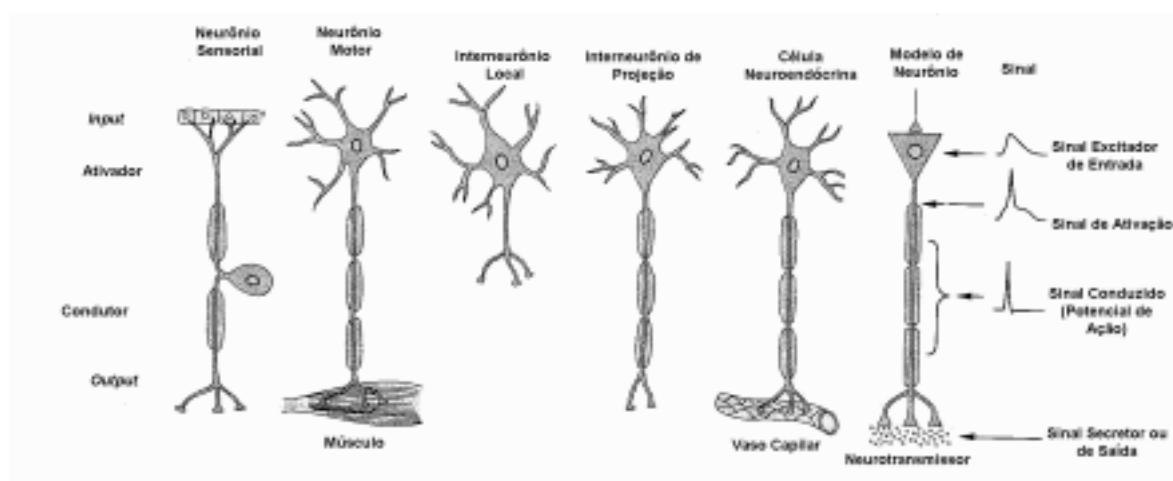


Figura 4.8 - Regiões funcionais dos neurônios [KAN 98]

A maioria dos neurônios, independentemente de seu tipo, possuem em comum quatro regiões funcionais: um elemento de recepção ou entrada, um elemento de ativação, um componente condutor e um elemento emissor. A organização funcional dos neurônios, portanto, pode ser representada por um neurônio modelo. Cada componente gera um sinal característico: a entrada, a ativação e os sinais propagáveis são elétricos; enquanto que, o sinal emitido é a liberação de um transmissor químico na fenda sináptica. Nem todos os neurônios compartilham todas estas características, por exemplo, os interneurônios locais carecem frequentemente de elemento condutor.

A interpretação destes quatro tipos de sinais requer o conhecimento das propriedades elétricas da membrana celular do neurônio. Os neurônios mantêm uma diferença de carga elétrica de 65 mV através de sua membrana plasmática externa. Esta diferença é chamada de **potencial de membrana de repouso** que é resultado de uma distribuição assimétrica de íons sódio, potássio e sódio, assim como de ânions orgânicos, através da membrana celular e da permeabilidade seletiva da membrana ao potássio, estando o interior da célula carregado negativamente em relação ao seu exterior. Considerando o exterior da membrana arbitrariamente como zero, diz-se que, o potencial de membrana é de - 65 mV. Este potencial de repouso pode oscilar entre - 40 e - 80 mV em diferentes células nervosas, sendo que nas musculares é maior (- 90 mV) [KAN 98].

Esta distribuição é mantida pela bomba de sódio-potássio que transporta sódio ao exterior e potássio ao interior da célula (sódio interior é dez vezes menor que sua concentração extracelular e potássio intracelular apresenta-se vinte vezes mais concentrado que o potássio extracelular).

Uma célula estimulada muda este perfil tornando-se mais positiva internamente, em decorrência da mudança de permeabilidade da membrana em relação ao sódio (torna-se, subitamente, permeável ao mesmo). A célula se despolariza e gera um potencial de ação. Cessado o estímulo ou diminuída sua intensidade este quadro volta a reverter-se, ou seja, através da bomba de sódio/potássio e com gasto de energia, o sódio é devolvido ao meio extracelular e o potássio reingressado ao meio intracelular. O estímulo que desencadeia o potencial de ação deve ser suficientemente forte para ser capaz de superar o **limiar de excitabilidade** da célula. Esta mudança ocasionada pelo ingresso de sódio no interior da célula é denominada de **despolarização de membrana**, que consiste numa diminuição do potencial de membrana, enquanto o aumento se

denomina **hiperpolarização**. Quando hiperpolarizada uma célula nervosa apresenta maior dificuldade em produzir um potencial de ação, sendo pouco provável que gere um sinal transmissível. Assim a hiperpolarização é **inibitória** enquanto que a despolarização **excitatória**.

A amplitude e duração de um sinal são sempre as mesmas ainda que a intensidade e variedade dos estímulos possam variar. O potencial de ação funciona num efeito tudo ou nada, ou seja, se o limiar de excitabilidade for superado temos um efeito “tudo”, caso contrário, temos um “nada”. A resposta a um potencial de ação é a geração de um estímulo que se propaga de forma saltatória (fibras mielinizadas), sem ruído ou perda de sinal através da fibra nervosa até sua extremidade distal, onde determina a liberação de mediadores químicos excitatórios (adrenalina) ou inibitórios (GABA) pelas vesículas encontradas nos botões terminais dos axônios, em geral, e dendritos. Distintos momentos de transformação de um estímulo (fenômeno físico) em atividade elétrica celular se observa na Figura 4.9.

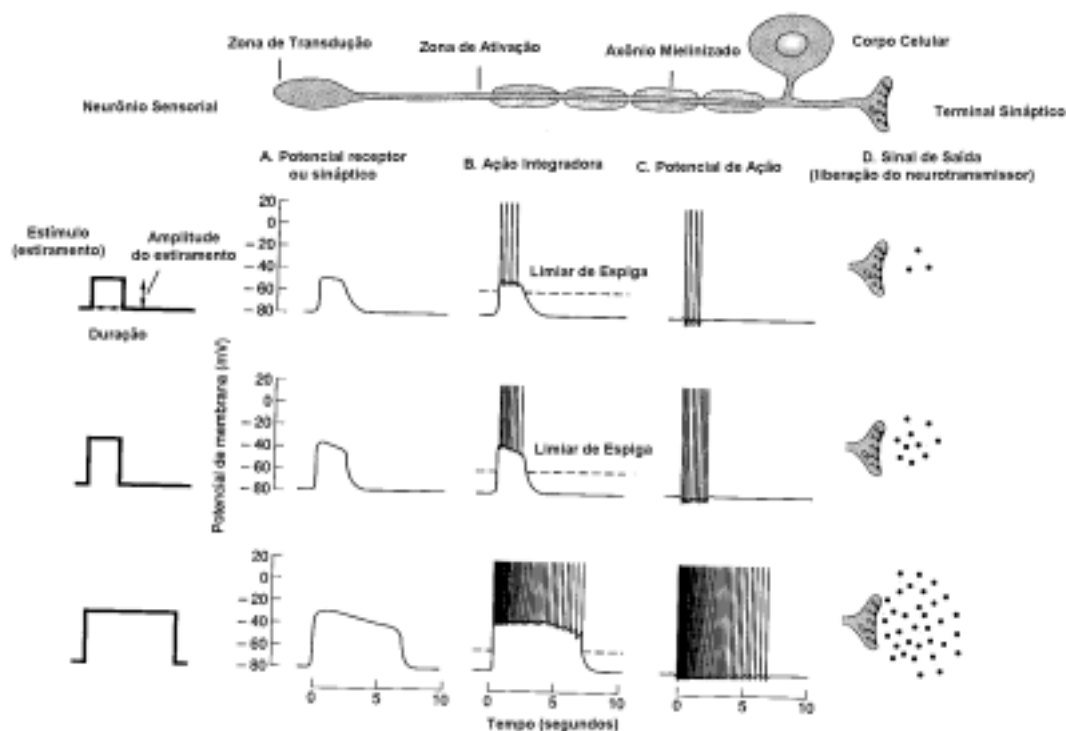


Figura 4.9 - Eletrofisiologia neuronal [KAN 98]

Um neurônio sensorial transforma um estímulo físico (neste exemplo um estiramento) em atividade elétrica celular. Cada um dos quatro elementos do neurônio gera um sinal característico.

- A. O sinal de *input* se gradua tanto em amplitude como em duração, proporcionalmente amplitude e duração do estímulo.
- B. O sinal de ativação transforma o sinal de entrada em potenciais de ação que se propagarão ao longo do axônio. Um potencial de ação será gerado unicamente se o potencial receptor (nos neurônios sensoriais) ou o potencial sináptico (nos neurônios motores) for superior ao limiar de excitabilidade – disparo – (limiar de excitabilidade ou “umbral de espiga”) determinado. Uma vez que o sinal de *input* sobrepassa este limiar, qualquer incremento adicional na amplitude do sinal de entrada aumentará a frequência de geração de potenciais de ação, sem variar a amplitude. Portanto, a natureza gradual do sinal de entrada se traduz a um código de frequência de potenciais de ação na zona de ativação. A duração do sinal de *input* determina o número de potenciais de ação gerados.
- C. Os potenciais de ação seguem a Lei do **tudo** ou **nada**. Cada potencial de ação tem a mesma amplitude e duração e portanto, a mesma morfologia de onda (as espigas) quando se registram num osciloscópio. Como potenciais de ação são conduzidos sem perdas ao longo de todo comprimento do axônio, a informação portada pelo sinal está representada somente pela frequência e número de espigas (pulsos) e não por sua amplitude. Quanto maior for a amplitude do estímulo maior será a frequência de pulsos. E, por lógica, quanto maior for a duração do estímulo, mais tempo se prolongará a geração de potenciais e portanto maior será o número de pulsos (espigas).
- D. Quando o potencial de ação chega ao terminal sináptico, a célula libera um neurotransmissor químico que serve como sinal emitido. O número total de potenciais de ação por unidade de tempo determina exatamente a quantidade de neurotransmissor que será liberado por célula.

Embora seja possível aportar um maior número de dados biofísicos e moleculares no presente estudo não será seguido este enfoque.

Para concluir o conhecimento da complexa estruturação da célula nervosa e sua resposta a estímulos há que entender os mecanismos de comunicação entre os neurônios

propriamente ditos e destes com outras células, ou seja, estudar as sinapses e seus principais modelos.

SINAPSE é a comunicação estabelecida entre um neurônio com outro(s) neurônio(s) ou com outros tecidos. A condução de um estímulo elétrico pela membrana celular de um neurônio é unidirecional, sendo assim, em uma comunicação entre um neurônio e outra célula qualquer, sempre se observa a distinção entre o neurônio que faz a sinapse e seu elemento subsequente. Assim, o elemento que fica antes da comunicação (sinapse) celular é denominado pré-sináptico e o que fica depois de pós-sináptico. O espaço entre o elemento pré-sináptico e o pós-sináptico é denominado de fenda sináptica e é onde são liberados os mediadores químicos inibidores ou excitadores de membrana.

O modelo sináptico mais freqüente é o que ocorre entre um axônio de um neurônio pré-sináptico e o dendrito de um neurônio pós-sináptico, que se denomina sinapse axo-dendrítica (Figura 4.10). No entanto esta comunicação poderia ocorrer entre um axônio de um neurônio pré-sináptico com o corpo celular (axo-somática) ou axônio (axo-axônica) de um neurônio pós-sináptico ou ainda através de um dendrito pré-sináptico com um axônio (dendro-axônica) ou o soma (dendro-somática) de um neurônio pós-sináptico e por final, entre um soma de um neurônio pré-sináptico com o soma (somática) de um neurônio pós-sináptico. Estes outros modelos são menos freqüentes e de pouco interesse ao presente estudo [KAN 98].

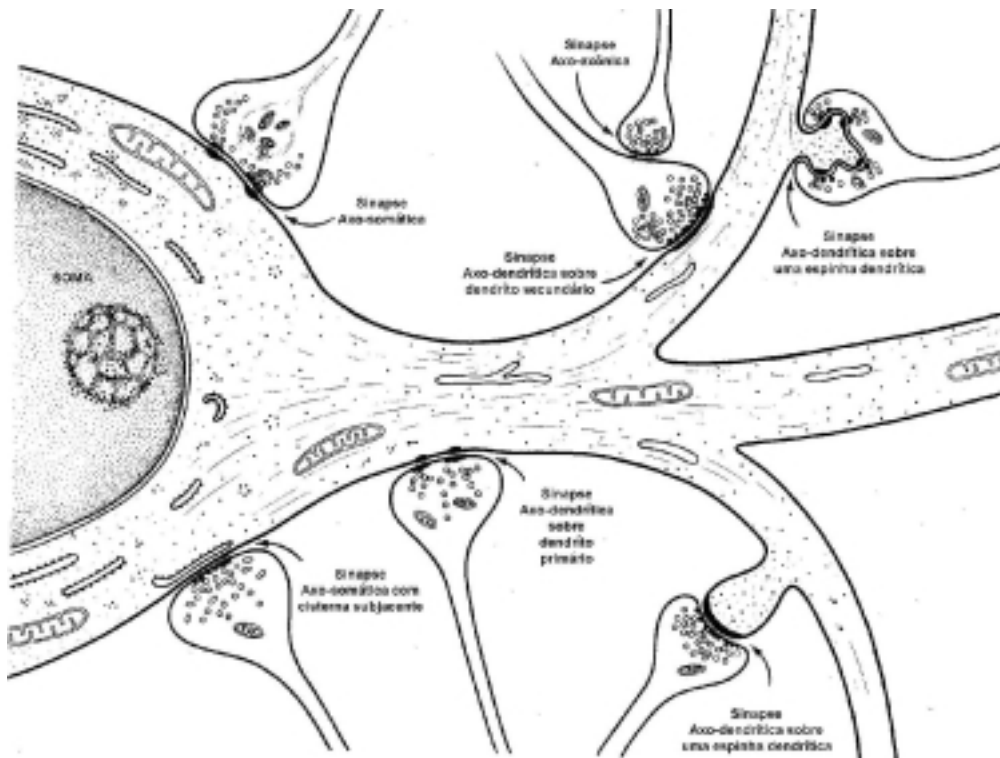


Figura 4.10 - Tipos e terminologia das sinapses de distintas partes do neurônio [FAW 87]

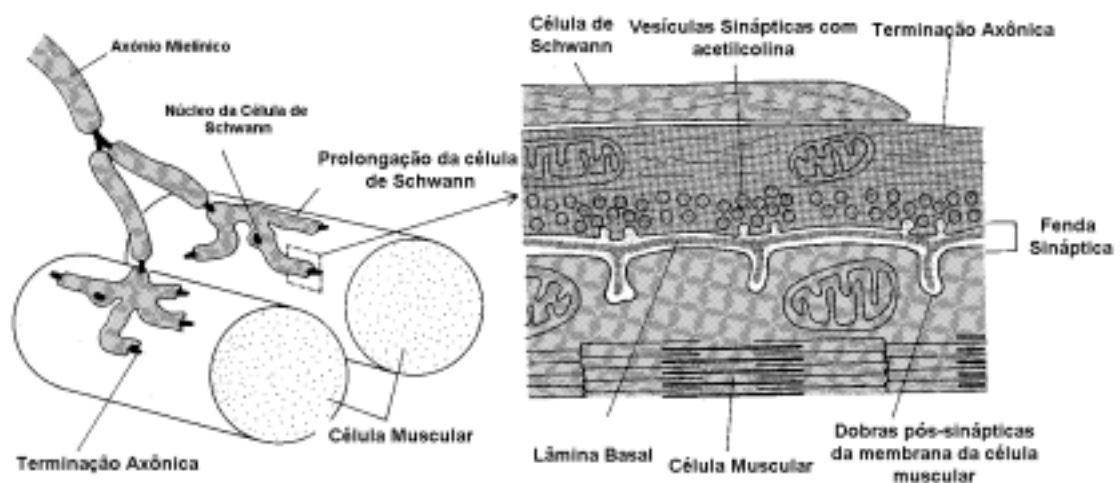


Figura 4.11 - Esquema de uma sinapse neuromuscular (Placa Motora). A união neuromuscular comumente recebe o nome de “placa terminal” devido ao aspecto da terminação axônica em algumas espécies [ALB 86]

O estiramento de um músculo produz um potencial receptor nas fibras terminais do neurônio sensorial (a célula do gânglio da raiz dorsal). A amplitude do potencial receptor é proporcional a intensidade do estiramento. Este potencial se propaga passivamente até a zona de ativação, no primeiro nó de Ranvier. Se o potencial receptor é suficientemente amplo, disparará um potencial de ação na zona de ativação, que se propagará ativamente e sem mudanças ao longo do axônio até a sua região terminal. Nesta região o potencial de ação produzirá um sinal de emissão: a liberação de um transmissor químico. O transmissor se difunde através da fenda sináptica e interage com moléculas receptoras, localizadas na face externa da membrana do neurônio motor que inerva o músculo estirado. Esta interação inicia um potencial sináptico na célula motora, que se propaga passivamente até a zona de ativação do axônio de tal neurônio motor, onde pode gerar um potencial de ação que se propagará ativamente ao terminal do neurônio motor. Aí, o potencial de ação ocasionará a liberação de um transmissor que ativa um potencial sináptico no músculo. Este sinal produz um potencial de ação no músculo que causa a contração da fibra muscular.

A Figura 4.12 apresenta o modelo clássico de comunicação entre duas células nervosas de forma simplificada, acompanhada de uma série de sinais com sua ação reflexa.

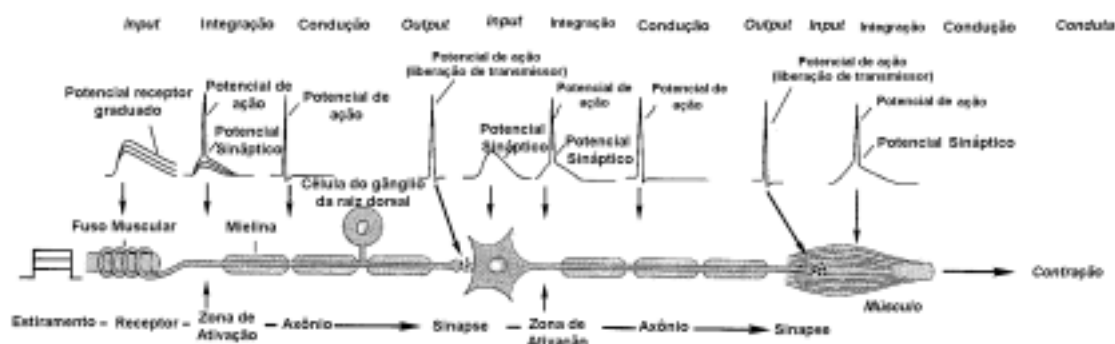


Figura 4.12 - Sequência de sinais produzindo uma ação reflexa [KAN 98]

Para finalizar cabe lembrar que a comunicação entre os neurônios e os demais tecidos, recebem nomes específicos, ou seja: a comunicação entre um neurônio e uma glândula se denomina neuroglandular; já a estabelecida entre o neurônio e um músculo – placa motora; entre um vaso sanguíneo e um neurônio – neuro-vascular e assim por diante.

Os temas discutidos até o presente momento de forma simplificada servem de base para a compreensão do estudo apresentado a seguir sobre **redes neurais artificiais**.

5 AS REDES NEURAIS ARTIFICIAIS

5.1 Introdução

A tecnologia das Redes Neurais Artificiais (RNA's) visa solucionar problemas de reconhecimento de padrões que geralmente são baseados em um conjunto de informações previamente conhecido. Geralmente os conjuntos de dados são divididos em conjunto de treinamento e conjunto de teste. Atualmente, pesquisadores em RNA's estão buscando uma compreensão das capacidades da natureza humana, as quais possibilitam que as pessoas construam soluções para problemas que não sejam resolvidos através de métodos tradicionais.

As redes neurais artificiais visam na sua maioria solucionar problemas de inteligência artificial, modelando sistemas através de circuitos (conexões) que possam simular o sistema nervoso humano, abrangendo a capacidade que o mesmo possui de aprender e agir perante as mais adversas situações apresentadas, bem como adquirir conhecimento através da experiência e da observação.

Segundo o pesquisador da Universidade de Helsinki Teuvo Kohonen, uma rede neural artificial tem a seguinte definição: "uma rede massivamente paralela de elementos interconectados e suas organizações hierárquicas que estão preparadas para iterar com objetos do mundo real do mesmo modo que um sistema nervoso biológico faz".

A complexidade das estruturas elementares das Redes Neurais Biológicas é muito maior do que a dos modelos matemáticos usados nas Redes Neurais Artificiais, demonstrando as dificuldades encontradas para se tentar imitar o funcionamento do sistema nervoso humano. O sistema nervoso é formado por bilhões de células nervosas, enquanto que uma rede neural artificial possui de dezenas a no máximo milhares de unidades de processamento (neurônios).

Uma rede neural artificial pode ser vista como um conjunto de várias unidades interconectadas (similar à estrutura do cérebro), denominadas de neurônios artificiais, cada qual contendo uma pequena porção local de memória. Estes conceitos foram baseados e fundamentados nos estudos realizados nas células nervosas naturais. Portanto, busca-se aproximar ao máximo o funcionamento das redes neurais artificiais

das redes neurais biológicas, na tentativa de buscar a desenvoltura com que o cérebro humano desempenha suas funções.

Alguns modelos de redes neurais artificiais possuem muitos neurônios conectados numa estrutura de pesos de conexão e com facilidade de adaptação, proporcionando uma estrutura paralela. A estrutura paralela é desejável pois se algum(s) neurônio(s) falhar (em), os efeitos na rede como um todo não será significativo para o desempenho do sistema se outro caminho de conexão entre os neurônios puder burlar a falha, surgindo então a tolerância à falha.

A princípio, as redes neurais podem calcular qualquer função computável que é realizada em um computador digital, ou seja, possuem a capacidade de modelar relações lineares e não lineares.

Principais características das RNA's [BAR99]:

- capacidade de "aprender" através de exemplos e de generalizar este aprendizado de forma a reconhecer elementos similares, que não foram apresentados no conjunto de exemplos (treinamento);
- bom desempenho em tarefas pouco ou mal definidas, onde falta o conhecimento explícito de como resolvê-las, o aprendizado se dá através de exemplos;
- robustez à presença de informações falsas ou ausentes, escolha dos elementos no próprio conjunto de treinamento (integridade do conjunto de treinamento);
- no contexto de classificação de padrões, uma rede neural pode fornecer informações sobre quais padrões selecionar em função do grau de confiança apresentado (confiabilidade do conjunto de treinamento);
- tolerância à falha.

5.2 Histórico

As Redes Neurais Artificiais surgiram na década de 40, mais precisamente em 1943, quando o neurofisiologista Warren McCulloch e o matemático Walter Pitts, da Universidade de Illinois, fizeram uma analogia entre as células nervosas e o processo

eletrônico num artigo publicado no *Bulletin of Mathematical Biophysics* com o título: *A Logical Calculus of the Ideas Immanent in Nervous Activity*.

Em 1949, o biólogo e psicólogo Donald Hebb, que estudava o comportamento dos animais, escreveu um livro chamado *The Organization of Behavior*, que reforçava as teorias de que o condicionamento psicológico estava presente em qualquer parte dos animais, pelo fato de que esta é uma propriedade de neurônios individuais. As idéias de Hebb não eram pioneiras, mas ele propôs um princípio de aprendizado em sistemas nervosos complexos, ou seja, uma lei que descreve o funcionamento quantitativo da sinapse e do processo de treinamento humano.

Desde, então, vários outros pesquisadores, entusiasmados com as novas descobertas, voltaram-se para esta linha de pesquisa.

Em 1951, Marvin Minsky, co-fundador do Laboratório de Inteligência Artificial do MIT, construiu o SNARC, o primeiro simulador de cadeia neural. O SNARC trabalhava com êxito e podia ajustar seus pesos sinápticos automaticamente. Ele nunca chegou a executar alguma função de processamento de informação interessante, servindo somente de fator motivador para idéias que surgiram posteriormente.

Em 1956, na Primeira Conferência Internacional de Inteligência Artificial, foi apresentado um modelo de rede neural artificial pelo pesquisador da IBM Nathaniel Rochester. Seu modelo consistia numa simulação de centenas de neurônios interconectados através de um sistema que verificaria como a rede responderia aos estímulos ambientais.

Já em 1959, Frank Rosenblatt na Universidade de Cornell, criou uma rede de múltiplos neurônios do tipo discriminadores lineares e a batizou de rede perceptron. Rosenblatt baseou-se nas linhas de pensamento de McCulloch para desenvolver o seu modelo matemático de sinapse humana. Devido as suas complexas pesquisas e inúmeras contribuições técnicas, muitos o consideram como fundador da neurocomputação.

No final da década de 50, Minsky e Seymour Papert lançaram em uma obra chamada Perceptron, a qual demonstrava que o modelo apresentado por Rosenblatt não era muito promissor, devido ao uso de técnicas empíricas, das grandes dificuldades da matemática envolvida e dos poucos recursos computacionais disponíveis na época. A publicação de Minsky e Papert acabou esfriando as pesquisas e praticamente todo o investimento financeiro nesta área foi cancelado.

Enquanto Rosenblatt trabalhava no perceptron, Bernard Widrow da Universidade de Stanford com a ajuda de alguns estudantes desenvolveu um novo modelo de processamento de redes neurais chamado de Adaline (ADaptive LINear Element), a qual se destacava pela sua poderosa lei de aprendizado. O princípio de treinamento para as redes Adalines ficou conhecido como a Regra Delta, que foi mais tarde generalizada para redes com modelos neurais mais sofisticados. Mais tarde, Widrow criou a Madaline, que era uma generalização multidimensional do adaline.

Nos anos seguintes, muitos artigos foram publicados, e várias previsões exageradas e pouco confiáveis para a época foram anunciadas [TAFb 96]. A maioria destas suposições falava de computadores com um poder de raciocínio e/ou processamento igual ou superior ao do cérebro humano. Desta forma, a credibilidade de futuros estudos das RNA's foram fortemente comprometidos.

No início da década de 80, muitos pesquisadores publicaram inúmeras propostas para a exploração de desenvolvimento e pesquisa em redes neurais. Foi quando o administrador de programas da DARPA (Defense Advanced Research Projects Agency) Ira Skurnick resolveu dar atenção às proposições da neurocomputação, contrariando todos os preceitos, fundando em 1983 as pesquisas em neurocomputação da DARPA. Este fato acabou abrindo novos horizontes para a neurocomputação.

O físico e biólogo de reputação mundial John Hopfield também se interessou pela neurocomputação e escreveu vários artigos em 1982 que levaram vários cientistas a se unirem nesta nova área emergente. Hopfield reascendeu as pesquisas em neurocomputação, criticando fortemente as teorias apresentadas por Minsky e Papert na década de 50.

Este campo de pesquisa explodiu mesmo em 1986, quando o professor de psicologia da Universidade de Stanford, David E. Rumelhart, e seu colega James L. McClelland, professor de psicologia da Universidade de CarnegieMellon, publicaram o livro *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (vol.1: *Foundations*, vol.2: *Psychological and Biological Models*). Nesse livro, eles apresentam um modelo matemático e computacional que propicia o treinamento supervisionado dos neurônios artificiais. Surgia, então, o algoritmo *backpropagation*, um algoritmo de otimização global sem restrições.

Em 1987 ocorreu a Primeira Conferência de Redes Neurais. Também foi formada a Sociedade Internacional de Redes Neurais (*International Neural Networks Society* - INNS) juntamente com o *INNS Journal* em 1989, do *Neural Computation* e do *IEEE Transactions on Neural Networks* em 1990.

A partir destes acontecimentos, muitas instituições formaram institutos de pesquisa e programas de educação em neurocomputação.

5.3 Aplicações

Um dos principais objetivos da pesquisa sobre redes neurais artificiais na computação é desenvolver modelos matemáticos das estruturas neurais, não necessariamente baseadas na biologia, que podem efetuar diversas funções. Na maior parte dos casos, os modelos neurais são compostos por conjuntos de elementos não lineares que operam em paralelo e que são classificados de acordo com modelos/padrões relacionados à biologia. Quando um método é criado visando utilizar aspectos de redes neurais artificiais, começam com o desenvolvimento de um neurônio artificial ou computacional baseado no entendimento de estruturas neurais biológicas, seguidas do aprendizado de mecanismos voltados para um determinado conjunto de aplicações e o treinamento do suposto sistema. Segue-se mais detalhadamente as seguintes fases:

- estudo do problema;
- desenvolvimento de modelos neurais motivados por neurônios biológicos;
- modelos de estruturas e conexões sinápticas;
- escolha de um algoritmo de aprendizado (um método de ajuste de pesos ou forças de conexões internodais);
- construção de um conjunto de treinamento;
- o treinamento propriamente dito;
- fase de testes;
- utilização da rede.

As diferenças entre as aplicações, os algoritmos de aprendizagem e as estruturas de interconexões entre os neurônios levam os pesquisadores a desenvolver diferentes

modelos (arquiteturas) de redes neurais. Do ponto de vista estrutural, a arquitetura de redes neurais pode ser classificada como estática, dinâmica ou fuzzy, podendo ter uma ou múltiplas camadas. Além disso, diferenças computacionais surgem devido a forma como são feitas as conexões entre os neurônios. Estas conexões podem ser *feed forward*, *backward*, lateralmente conectadas, topologicamente ordenadas ou híbridas. As aplicações de redes neurais podem ser classificadas em diversas classes como:

- reconhecimento e classificação de padrões;
- processamento de imagem;
- visão computacional;
- identificação e controle de sistemas;
- processamento de sinais;
- robótica;
- filtros contra ruídos eletrônicos;
- análise do mercado financeiro;
- controle de processos.

Cabe ressaltar que em uma determinada aplicação de um sistema, que faz o uso das redes neurais artificiais, não precisa necessariamente ser classificada em apenas uma das citadas acima.

5.4 O Neurônio Artificial

O primeiro modelo matemático para uma rede neural, proposto por McCulloch e Pitts, era simples diante das informações disponíveis naquela época sobre o funcionamento elétrico de uma célula nervosa (Figura 5.1). Era um dispositivo binário, sendo que a saída do neurônio poderia ser pulso ou não pulso (ativo ou não), e as várias entradas tinham um ganho arbitrário, podendo ser excitatórias ou inibitórias. Para se determinar a saída do neurônio, calculava-se a soma ponderada das entradas com os respectivos ganhos como fatores de ponderação, excitatórios ou inibitórios. Se o resultado atingisse um certo limiar, a saída do neurônio era pulso (ativo), caso contrário, não pulso (não ativo).

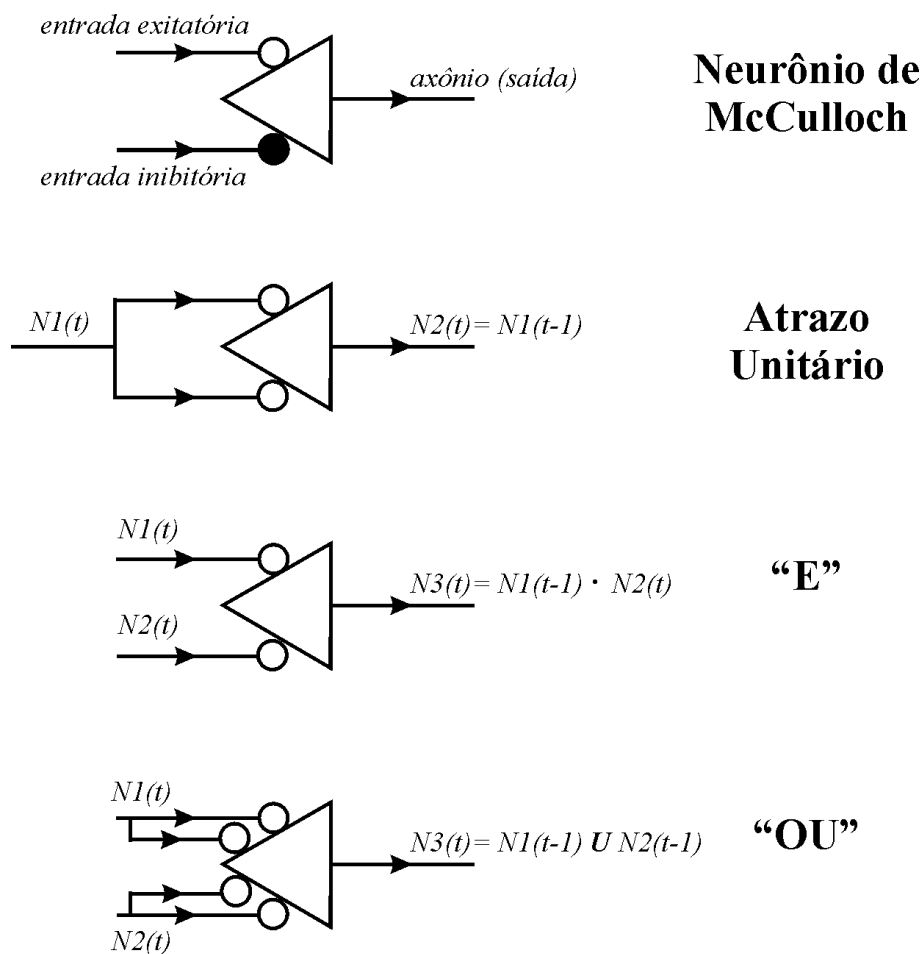


Figura 5.1 - O neurônio de McCulloch e implementações de algumas funções booleanas [KOV 96]

Assim como o neurônio biológico, o neurônio artificial possui um ou mais sinais de entrada e apenas um sinal de saída. As informações podem ser recebidas através de sensores ou de outros neurônios artificiais que fazem parte da Rede Neural Artificial (RNA). Estes sinais são processados e enviados para a saída. Os sinais de entrada (estímulos) devem chegar até o neurônio simultaneamente, isto é, todas as informações devem chegar ao núcleo do neurônio artificial ao mesmo tempo.

O processamento paralelo em computadores sequenciais (por exemplo, os microcomputadores atuais) pode ser paradoxal, mas não o é, ocorre de fato. A simulação de um ambiente paralelo é possível, e é desta forma que ocorre esse tipo de processamento para as redes neurais. O modelo matemático simula o paralelismo da rede neural através de um algoritmo [TAF 96].

Um dos atributos de grande importância do neurônio artificial é o peso. Os pesos, também conhecidos por pesos sinápticos, são representados pela letra **w** (*weight*)

e representam o grau de importância que determinada entrada possui em relação àquele determinado neurônio.

O valor do peso é alterado em função da intensidade do sinal de entrada, e dessa forma, o peso muda o seu valor representativo para a rede (processo de aprendizagem). Deduz-se que, quanto mais estimulada for uma entrada, mais estimulado será o peso correspondente, e quanto mais for estimulado um peso, mais significativo e influente o mesmo será para o resultado do sinal de saída do respectivo neurônio.

Matematicamente, os pesos são vistos como um vetor de valores $[w_1, w_2, \dots, w_n]$ para um neurônio, ou uma matriz de pesos, coleção de vetores, para um conjunto de neurônios.

O sinal de excitação do neurônio é resultante do somatório do produto dos sinais de entrada, representados por um vetor $[x_1, x_2, \dots, x_n]$, pelo vetor de pesos do neurônio ($\sum_{i=0}^n x_i w_i$ - o valor correspondente a $x_0 w_0$ será explicado adiante e corresponde ao viés, representando um estímulo inicial a rede). Após esta operação, os sinais de entrada passam a ser chamados de entradas ponderadas.

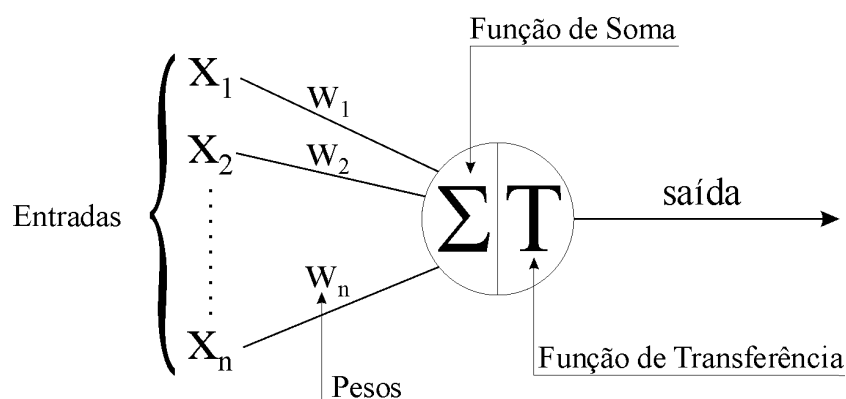


Figura 5.2 - O Neurônio artificial [TAFb 96]

A próxima tarefa a ser tomada pelo neurônio, é a de verificar se o valor resultante da soma entre o produto dos sinais de entrada pelos respectivos pesos atingiu ou não um valor predeterminado, chamado de limiar. Se o valor calculado atingiu o limiar, o mesmo é repassado adiante através da saída. Caso contrário, se o valor não atingiu o limiar, o sinal não será transferido. Esse processo de verificação é chamado de função de transferência, que também é conhecido como limiar lógico.

A resposta final da rede ou das camadas subjacentes está diretamente ligada com o resultado obtido pela função de transferência. Por isso, deve-se dar a devida atenção a este processo. A lógica neural expõe, que a intensidade dos sinais de entrada, dispara, ou não, o sinal do neurônio, fazendo com que este estimule o neurônio seguinte [TAFb 96].

Além da função de transferência, há a função de ativação, a qual antecede a mesma e tem como função, suceder um nível de ativação dentro do próprio neurônio, ou seja, o neurônio, através desta função, decidirá o que fazer com o resultado da soma ponderada das entradas (ativar ou não). Essa decisão tem efeito somente ao respectivo neurônio artificial.

Em alguns modelos simples de redes neurais artificiais, a função de ativação pode ser a própria função de soma das entradas ponderadas do neurônio. Já em modelos mais elaborados, a função de ativação pode possuir um processamento atribuído, o qual pode ser, por exemplo, o uso de um valor prévio de saída como uma entrada para o próprio neurônio, servindo de auto-excitação para o mesmo [TAFb 96].

O valor de saída do neurônio será produzido após a chamada da função de ativação, seguido pela função de transferência.

Em alguns casos, o neurônio artificial pode não ter efeito no neurônio seguinte se o valor de ativação não ultrapassar um certo valor mínimo. Este fator é resultante das características sigma ou ríspidas que a função de transferência tem como propriedade. Devido a esse fator, há vários tipos de funções de transferência (Figura 5.3).

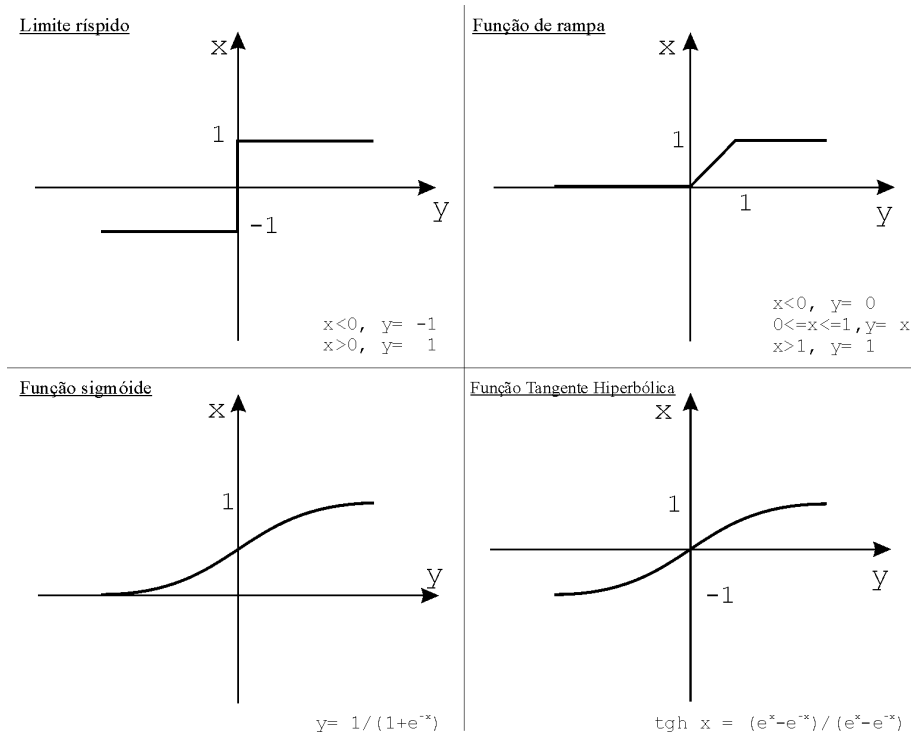


Figura 5.3 - Funções de transferência [KOV 96]

Assim como nas redes neurais biológicas, o conjunto de vários neurônios artificiais interconectados, formam as redes neurais artificiais.

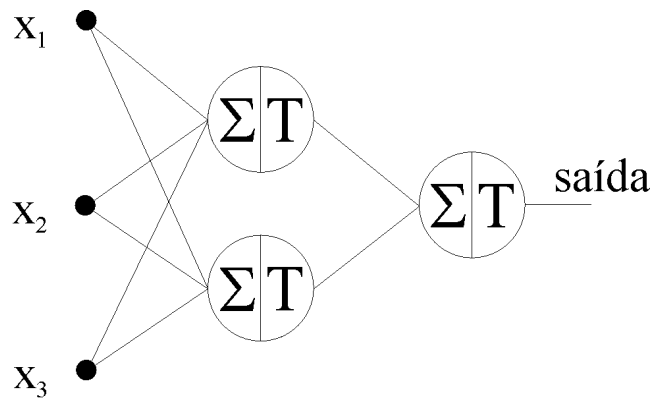


Figura 5.4 - Rede neural artificial

5.5 Arquiteturas

Um outro detalhe importante a ser considerado é a maneira como os neurônios artificiais podem ser agrupados. Este agrupamento se sucede no cérebro humano de maneira que as informações possam ser processadas de forma dinâmica ou interativa. Biologicamente, as redes neurais são organizadas e construídas de forma tridimensional

por componentes microscópicos. Há uma forte restrição no número de camadas que a rede pode conter, limitando consideravelmente o tipo e o escopo da implementação da mesma em silício.

Uma rede neural pode ter uma ou várias camadas. As redes que possuem uma única camada são as redes que possuem um nó entre uma entrada e uma saída da rede (Figura 5.5). Esse tipo de rede é indicado para a solução de problemas linearmente separáveis. Já as redes multicamadas possuem mais de uma camada entre as já existentes camadas de entrada e saída (Figura 5.6).

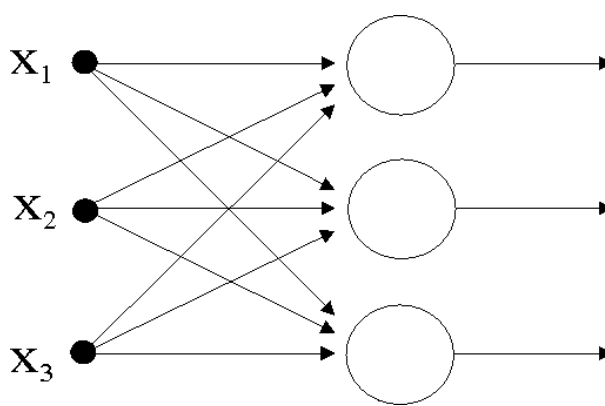


Figura 5.5 - RNA de uma única camada

As redes neurais artificiais multicamadas possuem as chamadas camadas escondidas (hidden), que também são chamadas de intermediárias ou ocultas. Esse número de camadas pode ser indeterminado, e estão situadas entre a camada de entrada e a camada de saída da rede neural [CAR 98].

As camadas ocultas são constituídas por neurônios artificiais, da mesma forma com que as camadas externas (entrada e saída) são compostas, e tendo como característica diferenciada o não contato com o mundo externo (Figura 5.6). Os sinais são passados para os outros neurônios obedecendo às funções de transferência que cada neurônio possui [NAS 94].

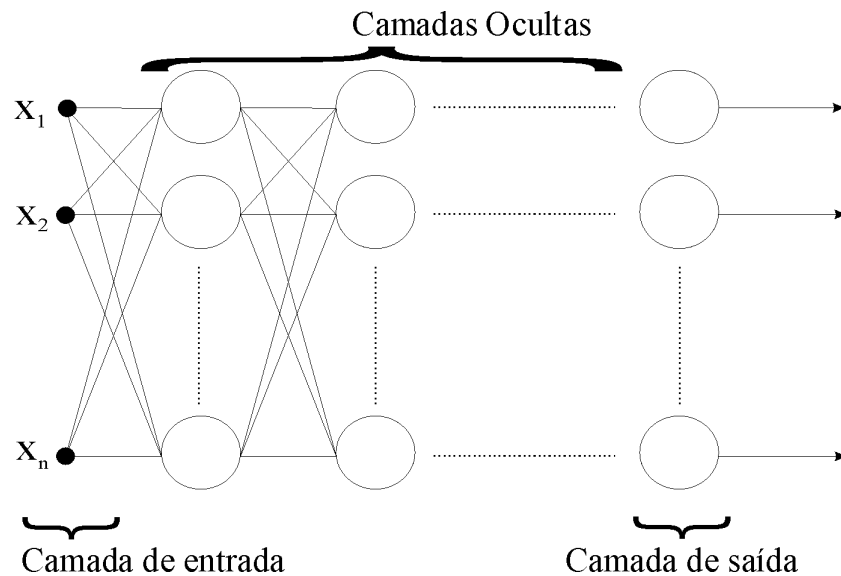


Figura 5.6 - RNA multicamada

Os nodos que compõe a rede neural artificial podem ter conexões do tipo:

- *feedforward* ou acíclicas (Figura 5.7) – a saída de um neurônio na i -ésima camada da rede não pode ser usada como entrada de nodos em camadas de índice menor ou igual a i [CAR 98]. Uma aplicação típica para as redes neurais artificiais *feedforward* é de desenvolver modelos não-lineares que também são usados para o reconhecimento e classificação de padrões. Uma rede *feedforward* pode ser vista como uma ferramenta que realiza a análise de regressão não linear [NAS 94].

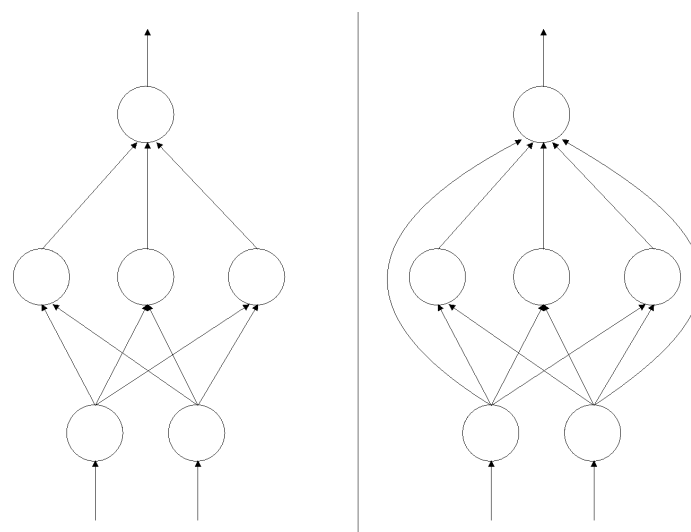


Figura 5.7 - RNA feedforward ou acíclica

- conexões feedback ou cíclica (Figura 5.8) – a saída de algum neurônio na i -ésima camada da rede é usada como entrada de nodos em camadas de índice menor ou igual a i . Se todas as ligações entre os neurônios forem cíclicas, a rede é chamada autoassociativa; estas redes associam um padrão de entrada com ele mesmo, e são particularmente úteis para a recuperação ou regeneração de um padrão de entrada [CAR 98].

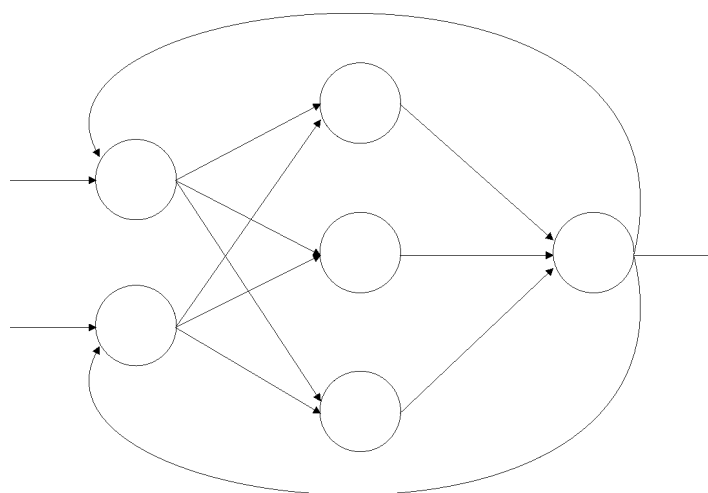


Figura 5.8 - RNA feedback ou cíclica

5.6 Aprendizado

Para o aprendizado das redes neurais, foram propostos diversos métodos de treinamento, sendo estes subdivididos em dois paradigmas principais: o aprendizado supervisionado e o não supervisionado. Para estes modelos existem vantagens e desvantagens que serão expostas a seguir. As RNA's possuem a capacidade de aprender por exemplos, determinando a intensidade de conexões entre os neurônios pertencentes à rede. Logo, um conjunto de procedimentos definidos para ajustar os parâmetros de uma RNA, a fim que a mesma possa aprender uma determinada função, é chamado de algoritmo de aprendizado. A designação de uma RNA, na resolução de um determinado problema, passa inicialmente por um processo de aprendizagem, onde a rede procura extrair informações relevantes de padrões de informação apresentados a ela, modelando uma representação própria.

5.6.1 Supervisionado

A vasta maioria das redes neurais artificiais tem utilizado o treinamento supervisionado. Deste modo, a saída atual da rede neural é comparada com a saída desejada. Os pesos terão os seus valores iniciais setados aleatoriamente, e serão ajustados, através do algoritmo de aprendizagem, pela rede na próxima iteração ou ciclo.

O ajuste sináptico é dependente do valor esperado e do sinal atual de saída. Desta maneira, o método de aprendizado tenta minimizar o fluxo corrente de erros de todos os elementos em processamento. Esta redução global de erros trabalha modificando continuamente os pesos até que a rede alcance uma certa precisão.

Com o aprendizado supervisionado, as redes neurais artificiais devem ser treinadas antes de serem usadas. O treinamento consiste da apresentação dos sinais de entrada e saída à rede. Estes dados são freqüentemente referenciados ao conjunto de treinamento. A fase de treinamento pode consumir uma grande fatia de tempo. Em alguns sistemas protótipos, com um inadequado poder de processamento, o aprendizado pode levar semanas. O treinamento é considerado completo quando a rede neural alcança um certo nível de performance. Este nível significa que a rede alcançou uma precisão estatística conforme as produções de saída necessárias para uma dada seqüência de entradas. Quando não há mais a necessidade de aprendizado, os pesos são praticamente “congelados” para a aplicação. Alguns tipos de redes neurais permitem um treinamento contínuo, com uma taxa muito baixa de aprendizado, enquanto a mesma está em operação. Este processo ajuda a rede a adaptar-se gradualmente as condições de mudança.

O conjunto de treinamento precisa ser suficientemente grande para conter as informações necessárias para que a rede aprenda os moldes e as relações importantes. Se a rede é treinada somente com um exemplo em um determinado tempo, todos os pesos serão setados meticulosamente para este fato, os quais poderiam sofrer alterações drásticas no aprendizado de um próximo fato. Conforme um resultado, o sistema precisa aprender com todos os fatos em conjunto, provendo posteriormente o melhor ajuste dos pesos para todo o conjunto de fatos.

A maneira com que os sinais de entrada são representados, ou codificados, determina o maior componente constituinte para o sucesso de instrução da rede.

Normalmente, as redes neurais artificiais somente manipulam, ou trabalham, com dados numéricos como entrada. Por este motivo, os dados do mundo exterior, devem ser tratados e convertidos para que se possa alimentar a rede. Esta captura de estímulos do mundo real pode ser feita através de vários tipos de dispositivos, tais como: câmeras de vídeo, diversos tipos de sensores, microfones, etc.

Várias técnicas de condicionamento já estão disponíveis para serem aplicadas a implementações de redes neurais artificiais, viabilizando e principalmente facilitando para que o desenvolvedor da rede encontre o melhor formato para os dados, e uma arquitetura adequada para a rede objetivando uma determinada aplicação.

Após o treinamento supervisionado, é importante analisar o que a rede pode realizar com os dados que ainda não foram apresentados à mesma. Se o resultado de saída do sistema não for razoável para este novo conjunto de dados (chamado conjunto de teste), presume-se que o treinamento da rede ainda não foi suficiente.

Esta avaliação é crítica para assegurar que a rede simplesmente não memorizou um dado conjunto de dados, mas sim aprendeu os modelos/padrões gerais envolvidos na aplicação (generalização). É importante ressaltar que às vezes o problema da generalização é devido à má qualidade dos dados usados para o treinamento e não um problema da rede.

5.6.2 Não supervisionado

O aprendizado não supervisionado é a grande promessa para o futuro, visto que implica que a rede aprenda se a necessidade de um conjunto de treinamento.

Estas redes não suportam influências externas para ajustar os seus pesos sinápticos, pois há um monitoramento de desempenho interno da mesma, analisando as regularidades e/ou tendências dos sinais de entrada, e conseqüentemente adaptando-se automaticamente as necessidades da rede.

Possuindo características de autonivelção, sem um suposto auxílio para determinar se o aprendizado converge ou não para o caminho certo, a rede possui mecanismos, mais precisamente, informações, de como se organizar. Esta propriedade e percepção da rede devem-se a topologia e as regras de aprendizado adotados pela rede neural artificial.

Uma rede com o algoritmo de aprendizado não supervisionado deve ter enfatizado a cooperação entre as camadas de unidades de processamento. A competição entre estas unidades é a base de aprendizado da rede. Normalmente, quando a competição pelo aprendizado ocorre de fato, somente os pesos pertencentes à unidade de processamento vencedora são ajustados.

5.6.3 Velocidade de aprendizado

A velocidade com que uma rede neural artificial aprende, depende de vários fatores. A baixa taxa de aprendizado resulta num tempo relativamente grande para a realização do aprendizado da rede, a fim de produzir um resultado adequado para o sistema em questão. Já com uma alta taxa de aprendizado, a rede pode não ser capaz de fazer uma possível discriminação fina em relação aos sistemas que aprendem de forma mais lenta.

Geralmente, vários fatores, além do tempo de aprendizado, precisam ser considerados quando se aborda a tarefa de aprendizado e treinamento da rede.

Alguns fatores que estão fortemente acoplados ao processo de aprendizado de uma RNA [DAC 92]:

- a complexidade da rede;
- o número de camadas (tamanho);
- o paradigma de seleção;
- a arquitetura adotada;
- algoritmo de aprendizado;
- as regras empregadas;
- a precisão desejada;

Todos estes fatores contribuem e alteram o tempo de treinamento da rede. A mudança de qualquer um destes fatores pode estender o tempo de treinamento para uma razão não muito significativa, ou resultando em uma precisão não satisfatória.

A maioria dos algoritmos de aprendizado possuem alguma provisão para a taxa de aprendizado ou em muitos casos, uma constante de tempo. Normalmente, este valor

está compreendido num intervalo entre zero e um $[0, 1]$. Se a taxa de aprendizado exceder o valor máximo, o algoritmo de aprendizado irá corrigir os pesos da rede.

Pequenos valores da taxa de aprendizado não retificarão os erros tão rapidamente, mas se pequenos passos são tomados na correção de erros, há uma grande possibilidade de se alcançar uma boa convergência de aprendizado da rede.

5.6.4 Algoritmos de aprendizado

Muitas das leis de aprendizado estão em uso, e a maioria delas são apenas variações da mais difundida lei, que é a lei de Hebb. As pesquisas em torno das funções de aprendizado continuam, e busca-se aproximar cada vez mais estes modelos dos tão sonhados e perfeitos padrões biológicos.

Embora o homem esteja muito além de descobrir realmente como ocorre o processamento biológico, e o aprendizado seja algo extremamente complexo, simplificações e morfologias matemáticas continuam em desenvolvimento.

Abaixo, são apresentadas algumas das mais conhecidas e difundidas leis de aprendizado [DAC 92]:

- **Princípio de aprendizado de Hebb** – a primeira e indiscutivelmente a mais conhecida regra de aprendizado, foi apresentada pelo biólogo e psicólogo Donald Hebb. A descrição da mesma foi exposta em seu famoso livro *The Organization of Behavior* em 1949. A regra parte do seguinte pressuposto. Se um neurônio recebe uma entrada, proveniente de outro neurônio, e ambos estão ativos, isto é, possuem o mesmo sinal, os pesos entre os neurônios precisam ser excitados;
- **A Lei de Hopfield** – é praticamente similar ao princípio de aprendizado de Hebb com apenas uma exceção: a lei de Hopfield especifica a magnitude da excitação ou inibição. Se a saída desejada e o sinal de entrada estão ambos ativos ou inativos, os pesos são incrementados pela taxa de aprendizado, caso contrário, decrementados;
- **A Regra Delta de Widrow** – esta regra é uma variação um pouco além do princípio de aprendizado proposto por Hebb. A regra delta apresentada por Widrow é a mais comumente usada nos dias atuais. Esta regra se baseia na simples idéia da contínua modificação da intensidade e a importância das

conexões de entrada; reduzindo consideravelmente a diferença entre o valor de saída desejado e o atual valor de saída da unidade de processamento, no caso, o neurônio artificial. A regra delta altera os pesos sinápticos de modo que minimize o erro quadrático da rede, trabalhando da seguinte forma: o erro calculado na saída é transformado pela derivação da função de transferência e conseqüentemente usado para ajustar os pesos de entrada da camada prévia da rede, ou seja, o erro é retro-propagado às camadas anteriores, sendo uma de cada vez. O processo de retro-propagação dos erros da rede continua até que a primeira camada da rede seja alcançada. Os tipos de redes chamadas de *feedforward* e *backpropagation* derivam seus nomes dos métodos adotados pelo processamento do erro. Quando se usa a regra delta, é importante assegurar que o conjunto de dados de entrada está disposto de forma aleatória ou gerado de forma randômica. Uma vez mal organizado, este conjunto de treinamento pode conduzir a rede a não convergência da precisão desejada, impossibilitando o aprendizado do problema em questão.

- **A Lei de aprendizado de Teuvo Kohonen** – desenvolvida por Teuvo Kohonen, a mesma foi inspirada nos sistemas biológicos, onde os elementos competem entre si por uma oportunidade de aprender, ou atualizar/ajustar seus respectivos pesos. A unidade de processamento que possuir o melhor sinal de saída será considerada o mais apto, e conseqüentemente passarão a ter a capacidade e privilégio de inibir os ajustes sinápticos de seus concorrentes e excitar seus vizinhos. Somente a unidade apta e seus respectivos vizinhos terão permissão para ajustar seus pesos. A abrangência e a possível área que uma unidade vizinha pertence está relacionada ao período de treinamento da rede. O paradigma atual usa o seguinte procedimento: é formada uma grande área de vizinhança e a medida com que ocorre o processo de treinamento, há uma seleção e conseqüentemente um estreitamento da mesma.

5.7 Redes Perceptron

As redes neurais artificiais com função de ativação foram inicialmente estudadas por Rosenblatt em meados de 1958, as quais foram chamadas por ele de Perceptrons. O

entusiasmo de Rosenblatt levou-o a construir suas redes em hardware, inclusive usando um algoritmo de aprendizado.

Estas redes foram aplicadas para a classificação de problemas que geralmente possuíam como fonte de alimentação imagens binárias de caracteres ou simplesmente moldes de informações [BIS 95]. O perceptron em sua origem era uma simulação computacional da retina, a qual demonstrou como o sistema nervoso visual reconhece padrões [TAF 96].

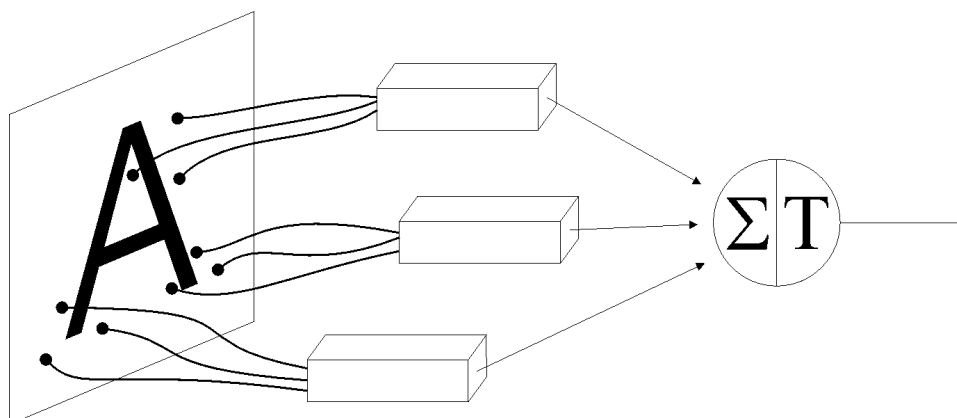


Figura 5.9 - O perceptron elementar de Rosenblatt [BIS 95]

Enquanto Rosenblatt estava desenvolvendo o perceptron, Widrow e seus colegas, estavam trabalhando em uma linha de pesquisa similar a de Rosenblatt; mais conhecida como ADALINE. Como já exposto, o termo **AD**aptive **LI**near **E**lement refere-se a uma única unidade de processamento com um limiar não linear.

Como as redes neurais artificiais de uma única camada possuem uma certa limitação, Rosenblatt resolveu então usar um número fixo de neurônios para transformar e tratar os dados provindos do mundo exterior. Estas unidades de processamento podem ser chamadas de função base de um discriminador limiar [BIS 95].

Rosenblatt propunha resolver problemas como a implementação das funções booleanas **E** e **OU** de duas variáveis, sendo que a escolha dos ganhos para este caso parecia ser trivial. Entretanto, para a implementação de uma função discriminatória arbitrária, a escolha não é tão simples e muito menos trivial, e dependendo do número de variáveis envolvidas, sem a existência de algum método, beira o impossível [KOV 96].

Inspirado também pelas idéias de McCulloch, Rosenblatt compôs a rede perceptron por uma camada de entrada, onde cada elemento pertencente à camada de entrada fazia a distribuição do sinal que ele recebia para todas as unidades de processamento. Os neurônios eram essencialmente compostos por unidades sigma e de funções de transferência, sendo que estas, eram responsáveis pela soma ponderada dos sinais oriundos das conexões com os dados de entrada. Foi adicionada a camada de entrada um elemento especial chamado viés, o qual possui um sinal de valor sempre um. A conexão entre o viés e a unidade sigma tem peso w_0 , que por sua vez é ajustado da mesma maneira com que os demais pesos o são.

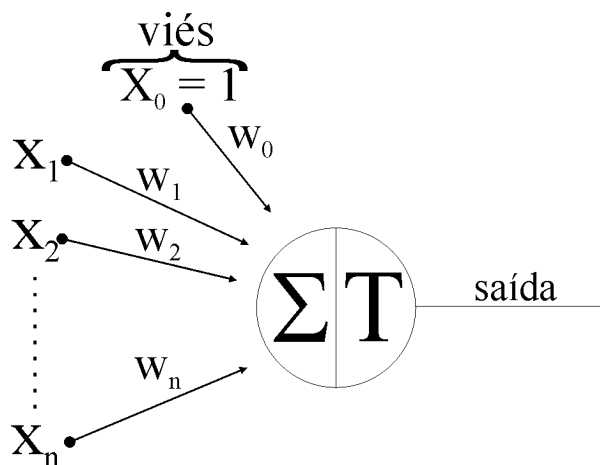


Figura 5.10 - A unidade de processamento do perceptron

O nível de ativação de uma rede perceptron é dado pela soma ponderada dos pesos sinápticos com os valores de entrada, $\sum x_i \cdot w_i$.

Estas redes usam uma função de transferência do tipo *hard-limiter* (limite ríspido), onde a ativação do limiar resulta num valor de saída 1, ou, -1 caso contrário. Dados os valores de entrada x_i , os pesos w_i , e um limiar t , o perceptron computa os valores de saída da seguinte maneira:

$$\begin{cases} \mathbf{1} & \text{se } \sum x_i w_i \geq t \\ -\mathbf{1} & \text{se } \sum x_i w_i < t \end{cases} \quad (5.1)$$

As redes perceptron usam como configuração, o treinamento supervisionado. O perceptron altera os seus pesos, visando reduzir o erro.

5.8 A lei de aprendizado do perceptron

Sendo o parâmetro c uma referência a taxa de aprendizado na medida em que reflete a taxa com que os ganhos são alterados em função dos erros, e d sendo o valor de saída esperado, o ajuste sináptico dos pesos no i -ésimo componente do vetor de entrada, Δw_i , é dado por:

$$\Delta w_i = c \cdot (d - \text{sign}(\sum x_i w_i)) \cdot x_i \quad (5.2)$$

A expressão $\text{sign}(\sum x_i w_i)$ é o valor de saída do perceptron, que pode assumir os valores $+1$ ou -1 . A diferença entre o valor desejado d e a saída atual, poderá ser 0 , 2 ou -2 . Logo, para cada componente do vetor de entrada, conclui-se que [LUG 98]:

- se, a saída desejada e a atual possuírem valores iguais, não haverá alteração alguma no peso;
- se o valor atual da saída for -1 e a saída desejada for 1 , o peso deverá ser ajustado na i -ésima linha da rede por $2 \cdot c \cdot x_i$;
- se o valor atual de saída for 1 e a saída desejada for -1 , o peso deverá ser ajustado por $-2 \cdot c \cdot x_i$.

Logo,

$$\begin{cases} \Delta w_i = 0, & \text{para } y_l^d = y_l \\ \Delta w_i = 2 \cdot c \cdot x_i, & \text{para } y_l^d \neq y_l \end{cases} \quad (5.3)$$

$$\Downarrow$$

$$\Delta w_i = c \cdot (1 - y_l^d \cdot y_l) \cdot y_l^d \cdot x_{i,l}$$

O procedimento mostrado acima tem como principal função, ajustar o conjunto de pesos da rede, a fim de minimizar o erro médio do conjunto de treinamento.

Como os perceptrons são utilizados em problemas de classificação, eles possuem a capacidade de aprender e classificar os dados de entrada em grupos ou classes.

Supondo-se uma rede perceptron, teremos a seguinte unidade sigma (Σ):

$$\sum = x_0 \cdot w_0 + x_1 \cdot w_1 + x_2 \cdot w_2 + \dots + x_n \cdot w_n \quad (5.4)$$

onde, o vetor $[x_0, x_1, x_2, \dots, x_n]$ são os sinais de entrada, e o vetor $[w_0, w_1, w_2, \dots, w_n]$ são os pesos respectivos ao vetor de entrada.

Teremos então:

$$\sum = \sum_{i=0}^n x_i \cdot w_i \quad (5.5)$$

Se a camada de entrada possuir dois elementos, x_1 e x_2 , a unidade sigma da rede será representada por:

$$\sum = x_0 \cdot w_0 + x_1 \cdot w_1 + x_2 \cdot w_2 \quad (5.6)$$

Como o viés, é representado por $x_0 = 1$, teremos:

$$\sum = w_0 + x_1 \cdot w_1 + x_2 \cdot w_2 \quad (5.7)$$

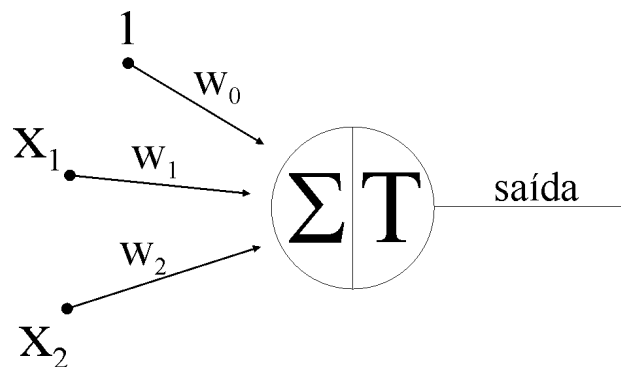


Figura 5.11 - Perceptron de duas entradas

Na equação (5.1), usando o limiar $t = 0$, tem-se que a saída é 1 se, $\sum_{i=0}^n x_i w_i \leq 0$, caso contrário, teremos -1.

A separação entre duas classes é chamada de superfície de decisão. Como só existem duas entradas, a superfície de decisão é uma reta. Se existirem mais do que duas entradas, haverá então um hiperplano [TAF 96].

As redes perceptrons foram inicialmente elogiadas, entretanto, Nils Nilson em 1965, Minsky e outros pesquisadores, analisaram as limitações das redes perceptrons. Eles demonstraram que os perceptrons não podiam resolver uma certa classe de problemas, normalmente chamados de problemas linearmente não separáveis [LUG 98].

Estes problemas foram solucionados a partir da década de 80, onde houve o surgimento de outras técnicas de aprendizado, como por exemplo, o *backpropagation*.

5.9 Limitações: O problema do OU-EXCLUSIVO

Um dos problemas que o perceptron não seria capaz de resolver era o do ou-exclusivo. Foi baseado neste exemplo que Minsky e Papert mostraram à comunidade científica que o modelo de Rosenblatt não era tão eficiente e promissor.

Tabela 5.1 - Tabela verdade do ou-exclusivo

x_1	x_2	saída
1	1	0
1	0	1
0	1	1
0	0	0

Considerando uma rede perceptron com duas entradas $[x_1, x_2]$, dois pesos $[w_1, w_2]$, e um limiar t , a rede, para aprender com estes fatos, deveria encontrar os pesos designados para satisfazer a tabela verdade e as seguintes premissas [LUG 98]:

- para a linha 1 da tabela verdade: $w_1 \cdot 1 + w_2 \cdot 1 < t$
- para a linha 2 da tabela verdade: $w_1 \cdot 1 + 0 > t$
- para a linha 3 da tabela verdade: $0 + w_2 \cdot 1 > t$
- para a linha 4 da tabela verdade: $0 + 0 < t$

As premissas apresentadas, baseadas nos pesos $[w_1, w_2]$ e no limiar t , não possuem solução. Logo, o perceptron de uma única camada é incapaz de resolver este tipo de problema.

O motivo pelo qual torna o problema do ou-exclusivo impossível para as redes do tipo perceptron é que as duas classes que precisam ser distinguidas não são linearmente separáveis.

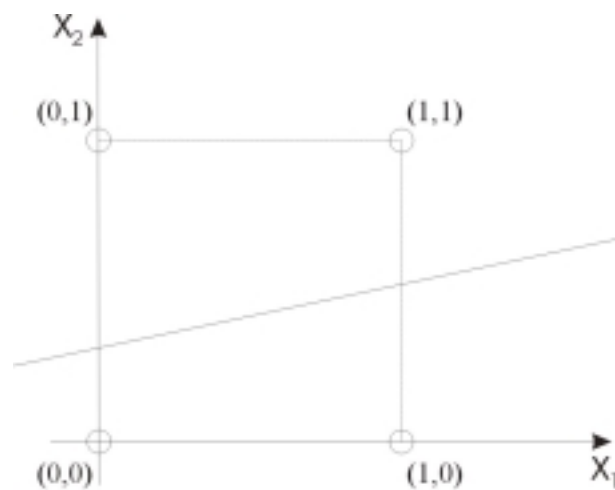


Figura 5.12 - Plano que representa as combinações possíveis do XOR

Percebe-se que é impossível plotar uma linha reta que separe em duas dimensões os pontos $\{(0,0), (1,1)\}$ de $\{(0,1), (1,0)\}$.

Cada parâmetro dos dados de entrada corresponde a uma dimensão, com cada valor de entrada definindo um ponto no espaço [LUG 98].

5.10 Redes Multilayer Perceptron

Os problemas não linearmente separáveis podem ser resolvidos através das redes com uma ou mais camadas intermediárias. A alteração da arquitetura da rede, como a inserção de camadas ocultas e/ou o número de neurônios, a princípio, não parece ser

problema, pois um dos principais agravantes passa a ser o algoritmo de treinamento para as redes multicamadas. Fator este que, devido à inexistência ou desconhecimento, causou uma atenuação nas pesquisas em redes neurais artificiais em meados da década de 70. Uma das alternativas adotadas é dividir a rede em um conjunto de subredes, sendo uma subrede para cada camada, com um treinamento independente. Este método de subdivisão, muitas vezes, ou não é possível ou é muito complicado. Outra possibilidade seria realizar um treinamento completo, isto é, de uma só vez. O problema encontrado para este segundo método está em como realizar o treinamento dos nodos que pertencem à camada intermediária, visto que é extremamente complicado determinar que tipo de resposta desejada estes teriam, ou seja, como determinar o erro. A aplicabilidade deste método está restrita a definição do erro nos nodos pertencentes às camadas intermediárias da rede. Se for utilizada uma função do tipo limiar, a avaliação do erro será complexa, visto que, os nodos das camadas intermediárias e de saída não terão como saber a margem de erro ou a diferença entre as respostas de seus nodos com relação às respostas desejadas. Uma das soluções para o problema apresentado seria a utilização de uma função de ativação não linear, a qual resolve o mesmo em parte, visto que a utilização deste tipo de função em redes multicamada resultaria na equivalência de uma rede de uma única camada [CAR 98].

Adotou-se então treinar as redes com mais de uma camada através de métodos baseados no gradiente descendente. Métodos baseados no gradiente descendente precisam ter a função de ativação contínua, diferenciável e não decrescente. A função adotada precisa informar os erros que a rede cometeu para as camadas anteriores, com uma boa precisão. Logo a função que mais se adapta a estas características é a função do tipo sigmóide [CAR 98].

O processamento atribuído a cada neurônio pertencente à rede é resultante da combinação do processamento realizado pelos neurônios da camada anterior, que por sua vez estão atribuídos a este nodo da próxima camada. A medida com que cada camada intermediária da rede se aproxima da camada de saída há uma delimitação do espaço de decisão dos dados que está recebendo. Para uma rede com duas camadas intermediárias, teríamos a primeira camada oculta, delimitando o espaço de padrões de treinamento através das “retas traçadas” pelos neurônios. A segunda camada forma regiões convexas, onde o número de lados que compõe tal região é determinado pela quantidade de unidades conectadas a este neurônio, que por sua vez combina as retas

que surgiram da camada anterior. Cada neurônio da camada de saída forma regiões, provenientes das combinações das regiões convexas [CAR 98]. Conclui-se que cada neurônio que compõe uma rede Multilayer Perceptron contribui para a detecção de características dos dados apresentados.

A determinação do número de camadas a ser utilizada influi de forma crucial no aprendizado da rede. O uso de um grande número de camadas intermediárias não é recomendado, visto que o erro ocorrido em uma camada é propagado a camadas anteriores da rede. A determinação do número de neurônios que pertence a camadas intermediárias é definida de forma empírica, e normalmente depende da distribuição dos padrões de treinamento e validação da rede. Um uso excessivo de neurônios levará a rede a decorar o conjunto de treinamento, ao invés de extrair as características gerais (generalizar). Ao processo de memorização do conjunto de treinamento, dá-se o nome de *overfitting*. Um número razoavelmente pequeno de neurônios levará a rede a aumentar o tempo de treinamento, dificultando a determinação da representação ótima do problema proposto. Neste caso, alguns neurônios poderão ficar sobrecarregados, pois estes precisam lidar com um número elevado de restrições a serem analisadas.

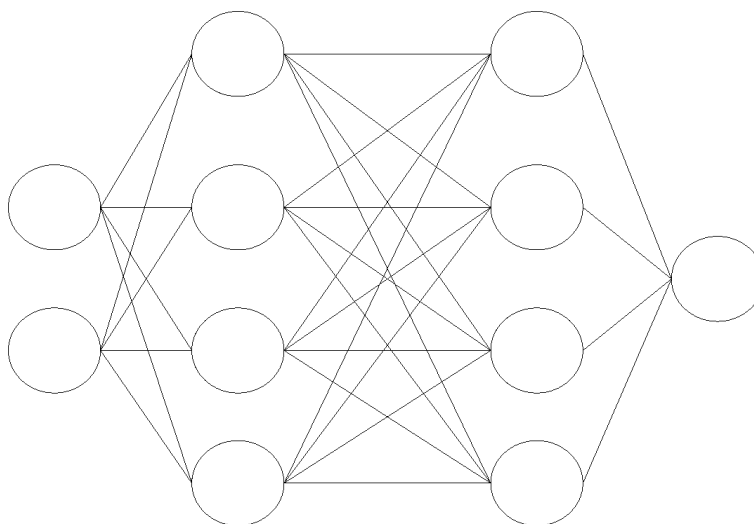


Figura 5.13 - Uma rede MLP

5.11 Algoritmo de treinamento das redes MLP

O algoritmo de aprendizado mais conhecido para a realização do treinamento das redes neurais multicamadas é o *backpropagation*. Cabe ressaltar que existem vários

algoritmos de aprendizado para as redes MLP, e estes normalmente possuem um aprendizado supervisionado. Pode-se ainda classificar os mesmos em dois grupos:

- estáticos;
- dinâmicos.

Os algoritmos de aprendizado estáticos não alteram a estrutura da rede, mudando somente o valor dos pesos sinápticos. Já os dinâmicos, podem mudar a arquitetura da rede, isto é, podem alterar o número de camadas, o número de neurônios da camada intermediária e o número de conexões da rede.

O método de aprendizado *backpropagation* foi descoberto através de inúmeras linhas pesquisas. Werbos, em 1974, foi um dos primeiros a propor o uso deste método de aprendizado na Universidade de Harvard em sua teste de doutorado “*Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*”. Parker em 1985 redescobriu as técnicas utilizadas por Werbos no relatório do MIT, “*Learning Logic*”. Até então, as pesquisas e principalmente os créditos eram dados a Rumelhart e aos outros membros do Grupo de Processamento Paralelo e Distribuído, por popularizar e desenvolver procedimentos que realmente pudessem ser utilizados. Este grupo publicou dois volumes que relatavam vários tipos de arquiteturas de redes neurais artificiais, incluindo um tratamento considerável sobre o procedimento de treinamento da regra delta generalizada, o *backpropagation* propriamente dito e alguns tópicos relacionados as RNA's.

O método de aprendizado *backpropagation* pode ser aplicado a qualquer rede que usufrui de uma função de ativação diferencial e aprendizado supervisionado. Assim como a regra delta, sua otimização é baseada no gradiente descendente, que ajusta os pesos para reduzir o erro da rede. O nome *backpropagation* surgiu do método na qual as correções da rede são realizadas nos pesos das conexões.

Durante a fase de treinamento, os sinais/padrões de entrada são apresentados a rede artificial em uma determinada ordem. Cada padrão de treinamento é propagado adiante, camada após camada, até a produção do sinal/padrão de saída. A saída computada pela rede é então comparada com uma saída desejada. Esta comparação irá gerar um valor que determinará o erro. Este erro será utilizado como uma realimentação para as conexões, que resultará no ajuste dos pesos sinápticos de cada camada num sentido oposto à propagação dos sinais de treinamento. Os acoplamentos retrógrados

somente existirão na fase de treinamento, considerando que as conexões adiante (sentido entrada → saída) serão usadas durante a fase de treinamento e uso da rede.

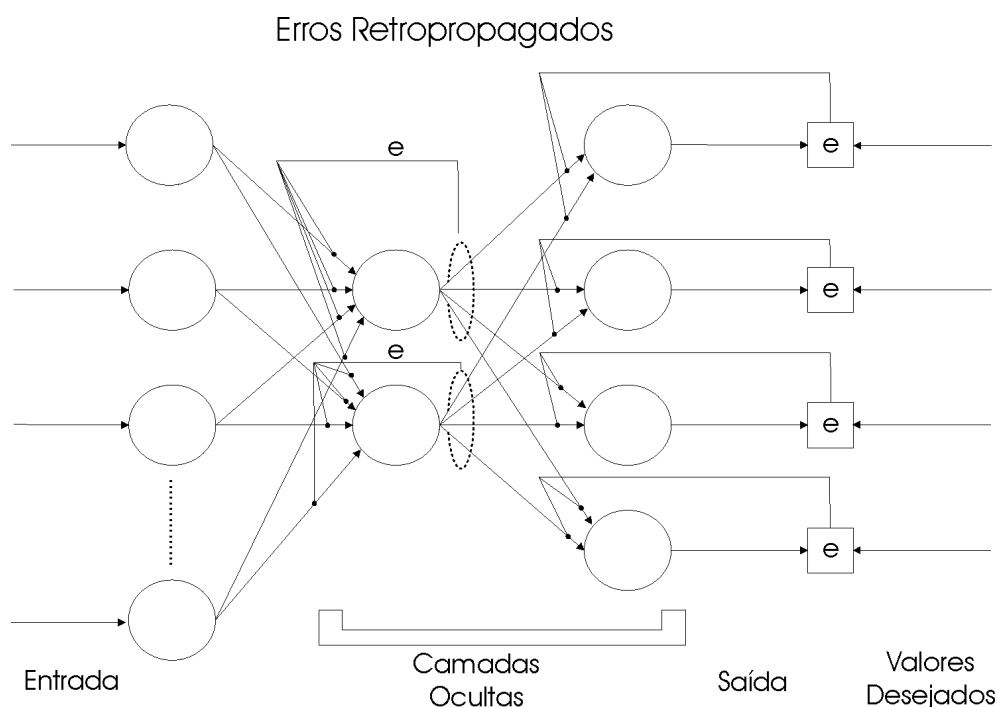


Figura 5.14 - Rede MLP com os acoplamentos retrógrados para os ajustes sinápticos [PAT 95]

Fazendo o uso do *backpropagation*, as camadas ocultas terão os seus pesos ajustados de acordo com as camadas subseqüentes, i.e, com as camadas seguintes. Deste modo, os erros computados na camada de saída serão usados para ajustar os pesos entre a última camada escondida ou oculta com a camada de saída. Assim, o erro calculado de uma camada escondida será usado para ajustar os pesos da camada oculta anterior. Este processo será repetido até que a primeira camada oculta seja ajustada. Desta forma, os erros serão retropropagados camada a camada com as devidas correções. Rotina esta que será realizada de uma maneira repetitiva, ajustando os pesos das respectivas camadas.

O processo é repetido por um número determinado de vezes para cada padrão de dados durante o treinamento até que o erro total da saída tenha convergido a um valor mínimo, ou até que algum limite predeterminado de iterações tenha sido completado.

Pode-se então criar duas fases para o algoritmo de treinamento do *backpropagation*. Cada fase percorre um sentido da rede. A primeira fase, chamada de *forward*, define a saída da rede para um determinado padrão de dados de entrada. A

segunda e última está incumbida de utilizar a saída desejada/esperada e a saída fornecida pela última camada da rede para ajustar os pesos sinápticos da rede neural.

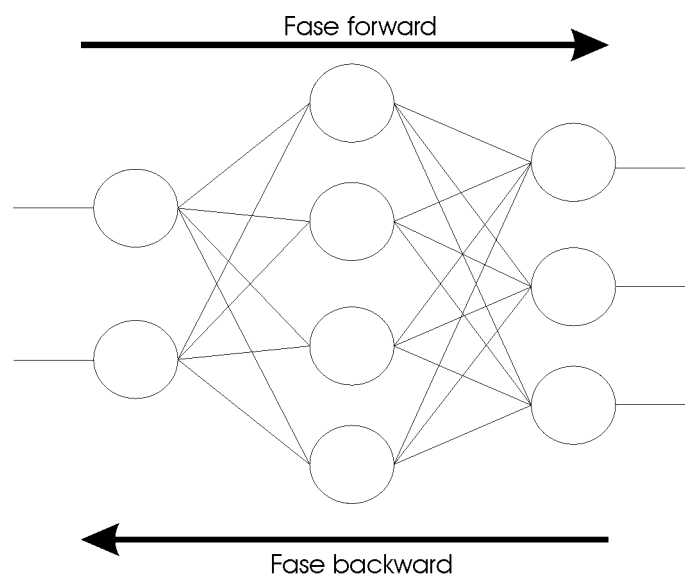


Figura 5.15 - Fluxo de treinamento de uma MLP com *backpropagation*

Segundo Carvalho [CAR 98], podemos definir os passos da seguinte maneira:

Fase *forward*

1. Os padrões de saída são apresentados a primeira camada c_1 que compõe a rede;
2. Para cada camada c_i a partir da camada de entrada:
 - 2.1. Os sinais de saída do neurônio da camada c_i irão alimentar a entrada da camada c_{i+1} , i.e, serão os sinais de entrada para a próxima camada;
3. Os sinais provenientes da última camada serão comparados com os sinais desejados;

Fase *backward*

1. Da última camada da rede até a primeira [$n \rightarrow 1$];
 - 1.1. Os neurônios artificiais da camada c_a (camada atual) devem ajustar seus pesos sinápticos de forma a reduzir seus erros;
 - 1.2. O erro de um neurônio das camadas intermediárias, $c_{[2, n-1]}$, por exemplo, c_i , será calculado utilizando os erros retropropagados dos neurônios que pertencem às camadas seguintes conectados a ele, no caso, c_{i+1} , os quais serão ponderados pelos pesos das conexões entre eles;

O *backpropagation*

1. Inicializar a rede, ou seja, pesos e parâmetros;
2. Repita
 - 2.1. Para cada padrão/dados de treinamento $P[x]$, para todo $x \in [1, n]$, sendo n o número total de amostras do conjunto de treinamento.
 - 2.1.1. Calcular a saída (S) da rede utilizando a fase *forward*;
 - 2.1.2. Comparar a saída (S), calculada no item 2.1.1, com as saídas desejadas;
 - 2.1.3. Realizar atualização dos pesos sinápticos fazendo o uso da fase *backward*;
3. Até o erro ser mínimo ou até x ciclos;

O algoritmo *backpropagation* também é chamado de regra delta generalizada, pois é baseado na regra delta apresentada por Widrow e Hoff.

6 O SISTEMA MAPA FAN

6.1 Introdução

O tratamento com padrões que possuem uma quantidade muito grande de características é objeto de estudo e preocupação para muitos pesquisadores há várias décadas [BEL 61, DUD 73]. Porém, nas duas últimas décadas a busca, por métodos mais sofisticados e eficientes, tornou-se mais intensa, pois se tornou necessário explorar as informações contidas em dados de alta dimensão, como as imagens digitais e os sinais digitais [SCH 97, SOU 99].

Os estudos são focados na busca por representações em espaços de baixa dimensão, ou seja, extração de características que minimizam a quantidade de características, mas que mantenham as informações relevantes. O grande desafio é o equilíbrio entre a quantidade de características e a quantidade de informações. Um dos métodos mais usados é a Análise de Componentes Principais [JOH 98].

Na maioria dos casos, um estudo detalhado dos padrões é requerido para buscar formas de minimizar esse problema. Devido ao problema da multidimensionalidade dos padrões [CAR 97], muitos pesquisadores, buscam métodos para extrair as características relevantes antes de verificar a possibilidade de desenvolver métodos de classificação que trabalhem com os padrões em seu espaço original. Neste trabalho, o método escolhido utiliza os dados no espaço de alta dimensão, preocupando-se também com a forma de representar os dados (seção 6.3), e tendo em vista o problema e o método utilizado para a classificação.

Um exemplo de representação em alta dimensão são os sinais de EEG. Para o reconhecimento de voz, a alta dimensionalidade se verifica, por exemplo, quando uma palavra é gravada, por exemplo, a 11025 Hz com uma duração de 2 segundos, produzindo então um padrão com 22050 características. Para a classificação, esta palavra é dividida em amostras, chamadas de padrões. Nesse caso, tem-se um padrão com dimensão de 500 a 1500 características, dependendo de como é feita a divisão inicial. Isso caracteriza um problema de alta dimensão. Analisando os padrões (amostras), vê-se que o que caracteriza o padrão é o contexto, neste caso, frequência,

amplitude e periodicidade. Cada uma das palavras deve ser classificada, ou seja, o conjunto de padrões/amostras que definem aquela palavra.

O objetivo deste capítulo é verificar a utilização do método de pré-processamento (ordenação) e do método de classificação (Mapa FAN) desenvolvido pelo trabalho de doutorado de Dandolini [DAN 00].

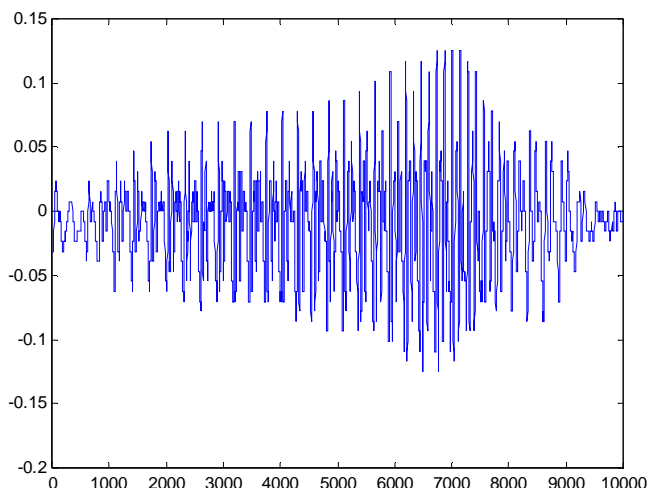


Figura 6.1 - Sinal de voz da palavra cálculo

O desenvolvimento do pré-processamento – ordenação – foi baseado na teoria da Informação. Já a elaboração do modelo de classificação está relacionada a aspectos da inteligência artificial, devido sua inspiração estar vinculada a aspectos de redes neurais artificiais, bem como aos sistemas difusos, mais especificamente na rede híbrida FAN [RAI 98].

A seção seguinte apresenta a arquitetura do sistema Mapa FAN que é composto de dois módulos: o pré-processamento (que inclui a ordenação) e a classificação (Mapa FAN). Essa arquitetura poderá ser utilizada para a classificação de sinais digitais por amostras.

6.2 Arquitetura

O sistema Mapa FAN foi desenvolvido com o objetivo de classificar sinais digitais de alta dimensão, como os sinais para o reconhecimento de voz. Sinais que são

classificados utilizando trechos do sinal (amostras), como mostrado na Figura 6.2. A dimensão das amostras depende da frequência que o sinal é capturado.

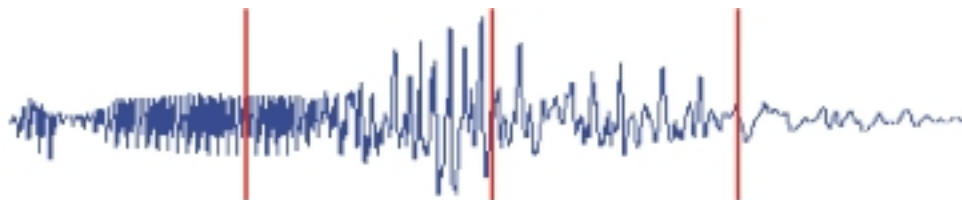


Figura 6.2 - Trechos do sinal (amostras)

Como o objeto de estudo é o sinal digital, geralmente, necessita-se realizar um pré-processamento. Muitos pesquisadores utilizam os filtros digitais e as transformadas [PAN 95]. Neste trabalho optou-se por utilizar uma forma alternativa de pré-processamento [RAI 99] baseada na ordenação dos sinais (seção 6.3) como um passo anterior à extração de características.

O sistema é composto então, de dois módulos distintos: o pré-processamento e o classificador. A Figura 6.3 mostra a arquitetura do sistema Mapa FAN.

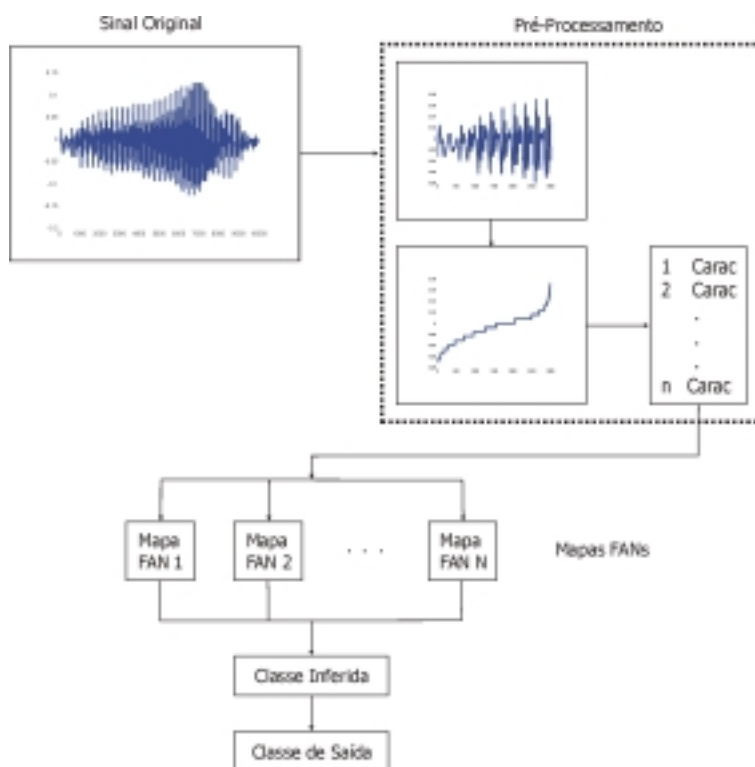


Figura 6.3 - Arquitetura do Mapa FAN

6.3 Pré-Processamento: Ordenação

O pré-processamento dos dados representa um papel importante em todo sistema de classificação, sendo esse em *hardware* ou *software* (analógico ou digital).

Esta seção descreve, então, uma forma alternativa de pré-processamento de sinais digitais através da ordenação das amplitudes dos sinais. O método foi baseado na Teoria da Informação (TI). A ordenação do sinal fornece a forma, que segundo a TI, auxilia no processo de reconhecimento.

6.3.1 Teoria da Informação

A Teoria da Informação (TI) foi criada para auxiliar a solução de certos problemas de otimização do custo da transmissão de sinais [SHA 75]. Hoje, por ser uma teoria atomística [MOL 69] (isto é, ela expõe seus conceitos básicos, demonstra seus teoremas fundamentais, pretendendo com isto dar conta da realidade ou pelo menos parte dela), está sendo usada em outras áreas como psicolingüística, estética, economia, desenho industrial, psicologia, biologia, pedagogia, entre outras.

Segundo a TI, quanto maior a desordem (a incerteza) maior a quantidade de informação. A própria percepção, segundo a teoria gestáltica, só é possível na medida da emergência de uma forma que, vista à luz dos conceitos da TI, é forma porque contém menos variedade e, portanto, menos informação do que o estado caótico (desordenado).

A forma desempenha um papel importante na percepção, no reconhecimento dos objetos, e conseqüentemente, na memorização e no aprendizado [EPS 88]. Por outro lado, a forma equivale a uma diminuição da informação contida no estado desordenado.

Dentro da TI, a forma está diretamente relacionada com a redundância nos dados. Por exemplo, uma tela de TV, quando está sintonizada num canal fora do ar, aparece uma chuva de pontos (totalmente desordenados). Neste caso, tem-se a quantidade máxima de informação que o canal pode transmitir¹. A quantidade de informação é a medida da complexidade (então quanto mais complexo um sinal mais informação ele possui). Porém, para se ter uma imagem nítida na tela precisa-se da

¹ Por exemplo, se a emissora de TV tem a capacidade de transmitir 30 imagens por segundos, se cada imagem contém 525 linhas, e se cada linha permite uma resolução correspondente de 630 bits de informação, a capacidade de informação é de $30 \times 525 \times 630 = 9922500$ bits/seg.

forma. Então, nota-se que a informação difere essencialmente da significação. Para ter significado necessita-se da redundância, da forma.

No caso dos sinais de fala, os dados são muito complexos, a quantidade de informação que eles transmitem é muito grande. Por isto, para entendê-los melhor, necessita-se da forma.

Uma maneira encontrada para obter-se esta forma foi através da ordenação das amplitudes das ondas (Figura 6.2). Uma outra forma de trabalhar com a forma seria utilizando a envoltória, forma essa não aplicada neste trabalho.

6.3.2 Ordenação

Uma forma encontrada para auxiliar o entendimento de sinais complexos foi transformar o sinal original no sinal ordenado (ordenação das amplitudes) [RAI 99].

O sinal é dividido em amostras (o tamanho da amostra depende dos dados e do problema) e, então cada amostra é ordenada. A ordenação do sinal de cada uma dessas amostras não obedece a nenhum outro critério a não ser a ordenação em ordem crescente segundo as amplitudes do sinal.

Seja $\mathbf{A} = (a_1, a_2, \dots, a_n)$ o conjunto de amplitudes de uma amostra do sinal. A ordenação dispõe os elementos de \mathbf{A} em ordem crescente, ou seja, é uma permutação dos elementos de \mathbf{A} ($Perm(\mathbf{A})$), tal que, para quaisquer a_i e a_{i+1} em \mathbf{A} , $a_i \leq a_{i+1}$. Formalmente:

$$TOrdem(\mathbf{A}) = \{ Perm(\mathbf{A}) \mid a_i \leq a_{i+1}, \forall i \} \quad (6.1)$$

Esta transformação é não linear e pode ser inversível desde que se mantenha a posição dos elementos onde se encontravam antes da permutação.

A ordenação do sinal vem em busca da forma para facilitar o reconhecimento e a percepção. Porém, não se pode esquecer que a forma equivale a uma diminuição da informação contida no estado desordenado.

Por exemplo, nos sinais de fala, os dados são muito complexos, a quantidade de informação que eles transmitem é muito grande. Mas, ao ordenar os sinais encontra-se a *forma*, requerida pela TI. Na Figura 6.4, tem-se uma amostra do sinal de voz. Estes sinais são bastante complexos, aperiódicos. No momento em que se ordenam esses sinais (Figura 6.5) obtém-se uma curva (forma).

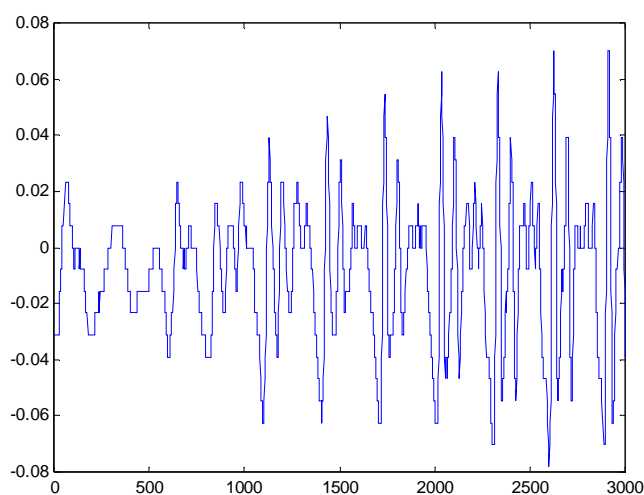


Figura 6.4 - Amostra do sinal de voz

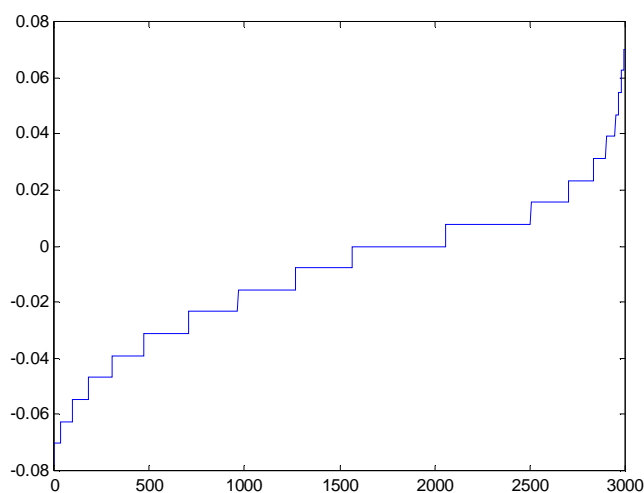


Figura 6.5 - Ordenação do sinal de voz

Ao aplicar a *TOrdem* nos sinais referentes a cada palavra observou-se que os gráficos dos sinais ordenados diferiam de uma palavra para outra. A ordenação fornece uma padronização aparentemente distinta dos sinais para cada palavra falada.

Observando os gráficos acima, conclui-se que a *TOrdem*, poderá ajudar no reconhecimento de palavras faladas. Ela servirá como um pré-processamento para o uso das redes neurais para a classificação.

Observando as figuras do sinal e a sua ordenação, fica claro que o sinal ordenado é mais fácil de reconhecer que o sinal original, principalmente para os seres humanos.

Isto vêm de encontro com a TI, quanto maior a desordem maior quantidade de informação. Mas por outro lado para se ter o significado necessita-se da forma, da redundância. Como no caso dos sinais a quantidade de informação que eles contém é muito maior do que o que se necessita, a idéia de buscar a forma através da ordenação realmente poderá ser muito útil para a classificação dos sinais (reconhecimento).

O pré-processamento de sinais é um trabalho difícil, pois, como se pôde observar, existem muitas informações a disposição e deve-se extrair o suficiente para habilitar a tarefa desejada. Esta seção ressalta a importância da forma na informação para a sua apresentação às redes neurais.

A *TOrdem* simplifica a forma do sinal e fornece uma padronização diferente para cada classe. Ela é uma transformação não linear e inversível que pode prover um bom pré-processamento para sinais digitais.

Esta abordagem será utilizada como parte do pré-processamento do sinal de voz.

6.4 O classificador Mapa FAN

O Mapa FAN foi desenvolvido com intuito de sanar a principal limitação do FAN [RAI 98]: a explosão combinatorial decorrente de problemas que apresentam alta dimensão. As filosofias subjacentes do FAN e do Mapa FAN são muito semelhantes. Por isso, antes de apresentar o mapa FAN, descreve-se basicamente, nesta seção, o sistema híbrido FAN, o qual serviu de base para o surgimento do classificador Mapa FAN.

6.4.1 FAN

Free Associative Neurons (FAN) é um exemplo de sistema híbrido. A base do FAN é as redes conexionistas, modelagem difusa e representação de padrões. FAN é baseado na noção do aprendizado neural; utiliza técnicas de conjuntos difusos para promover a granularidade de informações; e permite incluir diferentes métodos de associação de padrões para melhorar a sua capacidade de aprendizado [RAI 98].

A motivação para o desenvolvimento de FAN foi elaborar um método capaz de simular abstrações. Devido a esta capacidade cognitiva, os seres humanos compreendem um cenário completo baseados apenas nas partes que o compõem. De fato, na linguagem natural as pessoas resumem a informação, classificando objetos em

estruturas de categorias de similares (palavras), evitando a sobrecarga de dados [ZAD 96]. Em FAN, a noção de classificação de padrões é feita por três processos: (a) expansão do espaço de entrada; (b) projeção da vizinhança do padrão resultante; (c) quantificação e aprendizado do grau de similaridade entre as classes FAN e as projeções dos padrões.

Em FAN, cada padrão de entrada é expandido em uma vizinhança difusa. Cada conjunto suporte dessa vizinhança é a combinação dos valores das características próximos dos originais. O grau de similaridade entre a vizinhança difusa e o padrão original de entrada é feito através das técnicas utilizadas na teoria dos conjuntos difusos. O aprendizado acontece através da projeção de toda a vizinhança difusa no espaço FAN. Existe uma unidade FAN (representado por uma matriz) para cada classe do domínio do problema. Cada unidade é uma grade composta por todas as combinações de características observadas em sua classe correspondente. Durante o treinamento, cada combinação é representada por uma célula difusa que contém um peso correspondente à sua frequência de ocorrência e grau de pertinência. O treinamento é baseado no reforço na célula (se a classificação foi correta) ou em esquecimento (se houve uma classificação incorreta) [RAI 97].

FAN foi aplicado em diversos problemas práticos. Os resultados obtidos em três problemas de classificação [RAI 97]: (a) Os dados da IRIS de Fisher; (b) Classificação de cromossomos, segundo o grupo de Denver, com os dados de Copenhagen; e (c) Diagnóstico financeiro de empresas baseados em testes financeiros [DAN 97], foram comparáveis, aos melhores resultados encontrados com outros sistemas.

A principal característica observada nestas aplicações é a capacidade de descobrir gradualmente as classes. FAN define as classes durante a primeira época de treinamento. Ou seja, ele reconhece uma classe antes mesmo de processar todo o conjunto de treinamento. Esse comportamento parece indicar que o treinamento em FAN ocorre com crescente capacidade de generalização. As unidades FAN acumulam conhecimento durante o processo de treinamento de forma que as modificações nos pesos se tornam progressivamente menores. Em grandes conjuntos de treinamento, os últimos padrões são vistos mais como dados de teste, causando modificações muito pequenas no sistema.

Os resultados obtidos usando FAN são comparáveis aos métodos tradicionais de redes neurais, em termos de classificação e superior em termos de estabilidade no aprendizado e capacidade de generalização [DAN 97].

A grande limitação do FAN é a explosão combinatorial dos dados que decorre da granularização das características. Por exemplo, para cada padrão de entrada com três características podem ser gerados 3^3 sub-padrões² com três características. Devido a esse problema é inviável usar FAN em problemas de alta dimensão, como é o caso das imagens.

6.4.2 O Mapa FAN

O FAN foi desenvolvido com a finalidade de resolver problemas de reconhecimento de padrões onde a quantidade de características, que representa cada padrão, é pequena. O Mapa FAN é uma variação do FAN em relação à forma de tratar o padrão de entrada. O objetivo é evitar a explosão combinatorial que ocorre no FAN, tornando possível a sua utilização em problemas multidimensionais (como, em reconhecimento de voz, em classificação de imagens de satélite e na classificação dos estágios do sono) [DAN 00].

Os Mapas FAN são matrizes bidimensionais ($d \times n$), onde uma dimensão se refere às amplitudes dos sinais e a outra a ordem do sinal amostrado. Para cada classe é criado um Mapa FAN. Cada padrão de entrada é apresentado ao Mapa FAN que representa a classe do padrão. Desta forma tem-se um processo de aprendizagem supervisionado. O objetivo é modelar cada mapa com aspectos da classe que ele representa.

Um padrão a ser testado é apresentado a cada mapa treinado e este retorna o grau de pertinência daquele padrão em relação à classe que aquele mapa representa. O padrão de entrada será classificado de acordo com o maior grau de similaridade (pertinência).

6.4.2.1 Treinamento dos Mapas FAN

Defini-se em primeiro lugar as dimensões dos mapas, d , o número de linhas e n , o número de colunas. Esses parâmetros dependem dos dados com o qual se está trabalhando. O número de linhas (d) refere-se à amplitude do sinal, ou mais

² Sub-padrão é um padrão com pertinência variada em relação ao padrão que o gerou.

especificamente da normalização realizada no pré-processamento. O número de colunas (n) é o número de entradas de cada mapa (características que representam o padrão de entrada).

As entradas para o Mapa FAN são P e $Clas$, onde P é a matriz ($n \times 2$) que representa o padrão de entrada e $Clas$ a classe para o qual o padrão P pertence. Na matriz P , a segunda coluna se refere às amplitudes do sinal, que deverão ser normalizadas e truncadas entre os valores inteiros 1 e d . A primeira coluna, indica a ordem das amplitudes (de 1 a n).

Definida as dimensões, inicia-se o processo de aprendizagem dos mapas. O padrão de entrada P define as posições no mapa (posições da matriz), e a classe correspondente $Clas$ indica em qual mapa estas posições serão acessadas. Por exemplo, seja o padrão de entrada a matriz $P = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 6 & 8 \end{bmatrix}^t$ e $Clas = C$ a classe que ele pertence, então as posições setadas no Mapa_C são aquelas mostradas na Figura 6.6. A aprendizagem ocorre acrescentado-se um valor $\delta > 0$ fixo nas posições definidas pelo padrão de entrada (P). Este acréscimo refere-se ao estímulo ou reforço no processo de aprendizagem.

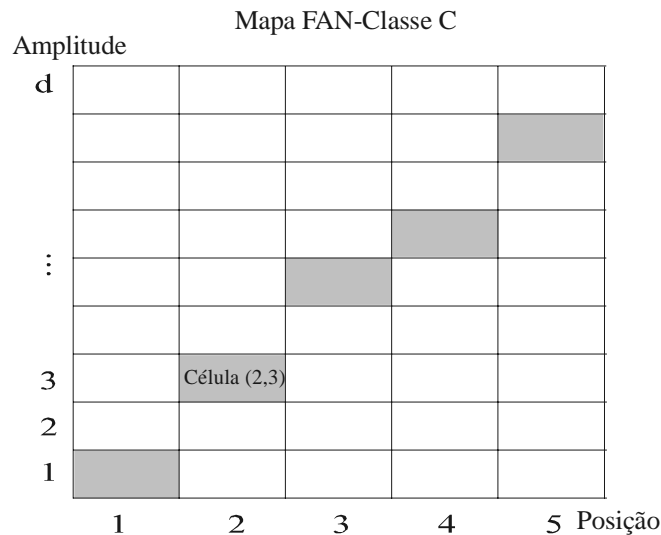


Figura 6.6 - Aprendizagem no Mapa_C indicada pela matriz do padrão de entrada

No final do processo de aprendizagem uma normalização é realizada em cada mapa segundo a quantidade de acessos realizados. Cada padrão de entrada é apresentado ao Mapa FAN somente uma vez.

6.4.2.2 Teste

A fase de testes, segue o mesmo raciocínio utilizado no FAN [RAI 98]. Para cada padrão a ser testado é gerada uma matriz de entrada P a qual definirá as posições dos Mapas que serão setadas. Definidas as posições, para cada Mapa treinado é feito uma verificação de pertinência com esta matriz de entrada de acordo com o nível de ativação que cada célula acessada possui, ou nível de reforço adquirido no processo de aprendizagem. A similaridade é calculada segundo a intersecção da ativação que cada mapa contém nas posições setadas pelo padrão de entrada (matriz de entrada). A quantidade de reforço adquirida pelo mapa no processo de aprendizagem determina a resposta do mapa para com a entrada. Assim, tem-se um vetor de respostas, chamado vetor de pertinência da entrada P com os Mapas treinados.

O cálculo da pertinência entre uma matriz de entrada P (gerada a partir de um padrão de entrada) e o Mapa_C, podem variar entre uma aplicação e outra, ou depender da matriz de entrada P.

A pertinência entre P e o Mapa_C é dada por [DAN 00]:

$$Pert(P, Mapa_C) = 1 - \prod_{j=1}^n \left(1 - \frac{P^{(i,j)} * Mapa_C^{(i,j)}}{\sum_{i,j} Mapa_C^{(i,j)}} \right) \quad (6.2)$$

onde $P^{(i,j)}$ é uma célula da matriz de entrada (posição da matriz), $Mapa_C^{(i,j)}$ são as respectivas posições no Mapa_C e o $\sum_{i,j} Mapa_C^{(i,j)}$ é todo o acúmulo de ativação adquirido pelo Mapa_C durante o processo de treinamento supervisionado.

A Equação (6.2) mede o grau de pertinência ou similaridade da matriz de entrada P com o Mapa FAN representante da classe C. Ela representa o uso da definição de complemento de conjuntos para escolha entre classes. A idéia está em concordância com o fato que: é mais fácil dizer o que não é, do que afirmar o que realmente é, ou seja, é mais fácil eliminar o errado que escolher o certo. O produto $P^{(i,j)} * Mapa_C^{(i,j)}$ realizado na fórmula (6.2) representa uma forma de realizar a intersecção de acordo com a teoria dos conjuntos difusos. Assim para cada entrada P, obtém-se o vetor de pertinências $VetPert(P)$, com relação a cada Mapa treinado, dado por:

$$VetPert_k = Pert(P, Mapa_k) \quad (6.3)$$

onde $k = 1, \dots, M$ e M é o número de mapas.

Assim, a classe do padrão P indicada pelo modelo é o índice do vetor $VetPert$ com maior pertinência. Ou seja, a classe do padrão P indicada pelo modelo é dada por:

$$ClasInd_p = \text{Índice} \left[\max_k (VetPert_k) \right] \quad (6.4)$$

onde k varia segundo a quantidade de classes.

A Figura 6.7 mostra a idéia do procedimento de verificação do nível de similaridade entre uma entrada P e um Mapa treinado. Quanto mais escuras forem as células, maior é o grau de reforço (ativação) que esta célula recebeu durante o treinamento. Os quadrados com borda são as posições indicadas pelo padrão P para a verificação da similaridade.

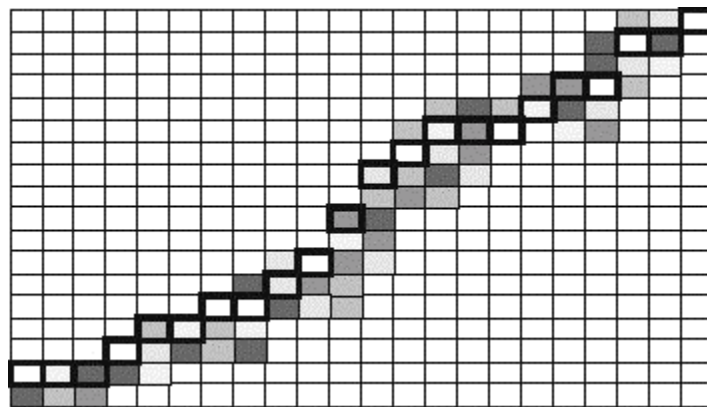


Figura 6.7 - Intersecção entre um padrão de entrada (quadrados pequenos com bordas) e um mapa treinado (onde o nível de cinza caracteriza o grau de reforço de uma célula)

Observando a Figura 6.7, pode-se concluir que cada célula no mapa pode ser considerada como um neurônio e o tom de cinza seu nível de ativação. A Figura 6.9 mostra a situação real para dois mapas treinados (duas palavras).

Nota-se também que com o Mapa pode-se tratar com padrões de dimensão tão alta quanto o limite de máquina (tempo de processamento). Para isso, basta definir n (número de colunas da matriz) como sendo a dimensão do problema em questão.

A Figura 6.8 mostra uma visão tridimensional do Mapa FAN treinado. Observe a diferença no nível de ativação referente a cada palavra (classe).

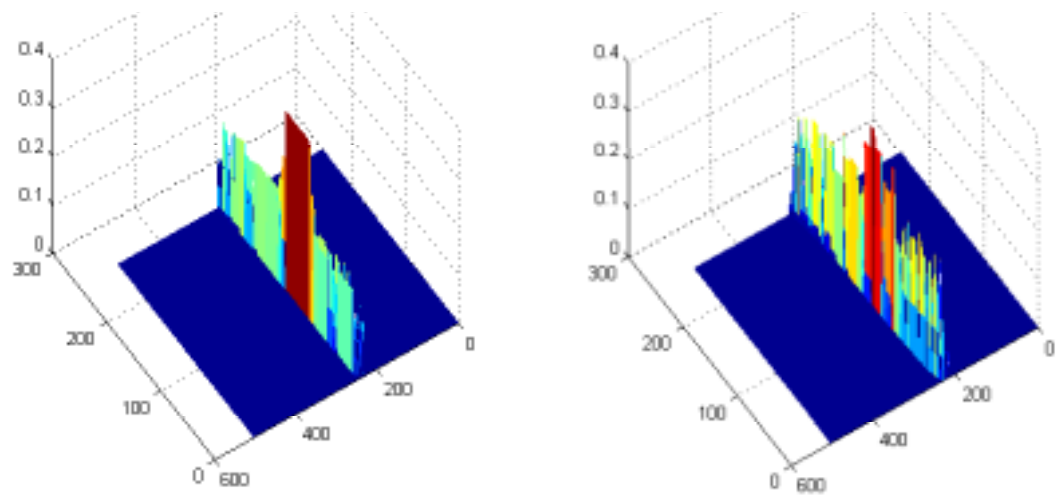


Figura 6.8 – Visualização tridimensional de duas palavras treinadas (dois mapas)

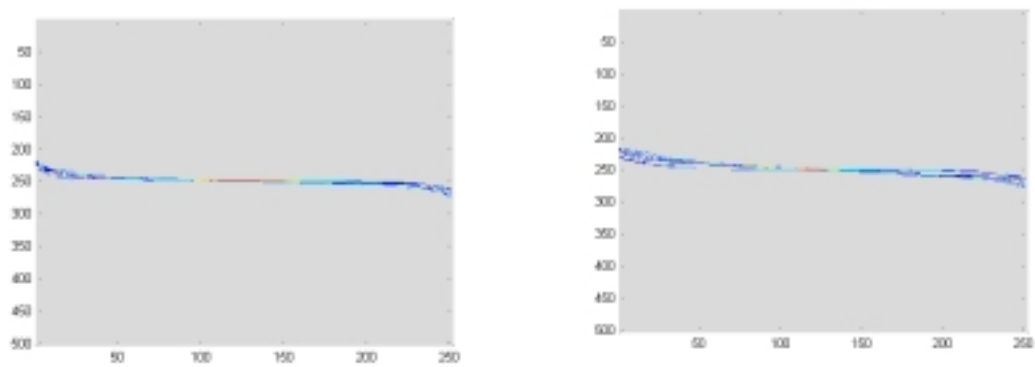


Figura 6.9 - Visão bidimensional de duas palavras treinadas

7 PROTÓTIPO DESENVOLVIDO

Foi utilizado como plataforma de Sistema Operacional para o desenvolvimento do protótipo, o software Microsoft Windows 9x, devido a sua ampla utilização no mercado e a grande variedade de programas disponíveis para este ambiente, o que possibilita ao usuário uma boa integração com todo o sistema.

Como um dos objetivos deste trabalho é o desenvolvimento de um protótipo para a realização do reconhecimento de palavras isoladas, não será abordada a especificação formal do processo de desenvolvimento do projeto. Fazendo-se uma análise comparativa entre os modelos de Redes Neurais citados, chegou-se a conclusão que o Mapa FAN, em virtude de tratar com padrões de alta dimensão, como os sinais de voz, o modelo citado possui um maior desempenho, tanto no reconhecimento como no tempo de convergência.

O protótipo para a captura do som foi implementado usando-se a Linguagem de Programação Inprise Borland Delphi, devido a grande versatilidade que o mesmo proporciona ao desenvolvedor, obtendo-se resultados satisfatórios em pouco tempo de desenvolvimento. Também foi utilizada a ferramenta matemática MatLab para a construção do Mapa FAN, pois a mesma proporciona uma maior flexibilidade quanto ao uso e manuseio de bibliotecas matemáticas disponíveis, minimizando o tempo de implementação e teste do protótipo.

7.1 A captura do sinal de voz

Para a captura do sinal das palavras do locutor, foi utilizada uma placa de som *Audio Excel PnP 310*. Esta é compatível com as placas de som da *Sound Blaster 16/PRO/2.0*, *Microsoft Windows Sound System*, *AdLib*. A taxa de amostragem máxima desta placa é de 44.100 Hz com uma resolução de 16 bits, em estéreo.

Além da placa de som usou-se um aplicativo desenvolvido no Delphi (Figura 7.2) para realizar a captura do sinal, possibilitando ao locutor gravar o conjunto de palavras que a rede neural artificial (Mapa FAN) posteriormente utilizará para treinamento e teste. Também é possível através deste realizar a delimitação do início e fim das palavras. Após a realização de vários testes com diversas palavras, as quais

possuem características sonoras distintas, verificou-se que era necessário eliminar a região de silêncio formada no início e fim de cada palavra pronunciada (Figura 7.1).

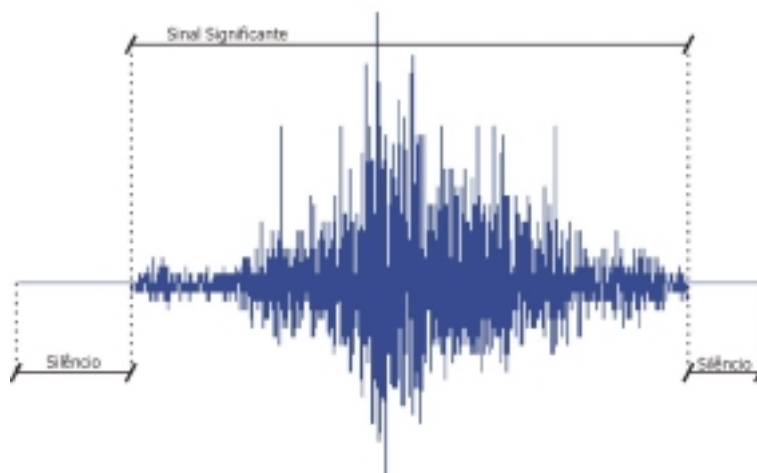


Figura 7.1 - Regiões de silêncio e voz da palavra capturada

A precisão de conversão da placa de som utilizada foi de 8 bits, visto que 256 sinais de diferenciação são suficientes para representar a intensidade do sinal sonoro. A região de silêncio é representada pelo valor 127, sendo os valores acima (128 - 255) representam o ciclo positivo, e os valores abaixo (0 - 126), o ciclo negativo.

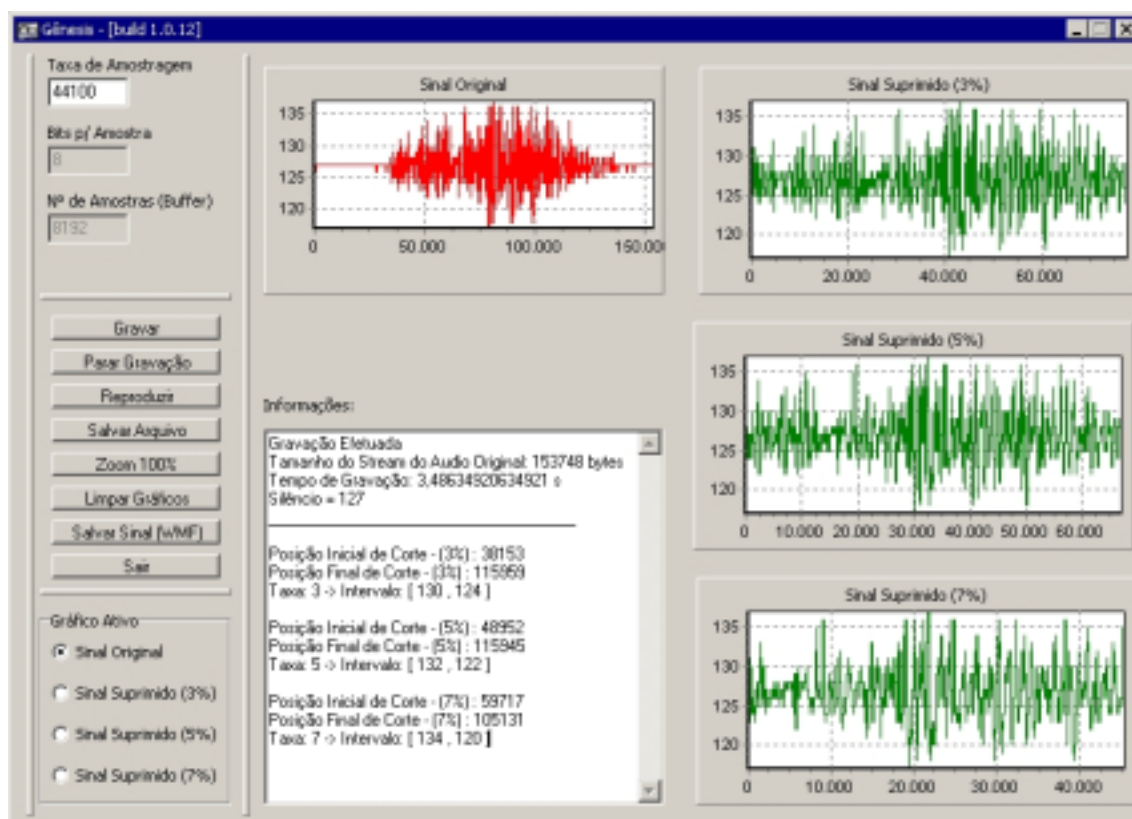


Figura 7.2 - Aplicativo para a captura do sinal de som

O aplicativo representado na Figura 7.2 também possibilita ao locutor escolher, de forma manual, o melhor sinal a ser gravado. Há três opções de delimitação da região de silêncio: 3, 5 e 7% aplicados sobre o valor 127 (silêncio).

7.1.1 Componentes e características de uma placa de som [TOR 99]

- **Bits:** Define a resolução do áudio capturado e reproduzido pela placa de som. As placas de som atualmente são de 16 bits, a mesma qualidade do CD.
- **Taxa de amostragem:** É a quantidade de pontos por segundo que é capturada ou reproduzida pela placa de som. A maioria das placas de som trabalha com uma taxa máxima de 44.100 Hz, a mesma qualidade do CD de áudio. Diversas placas de som conseguem trabalhar com uma taxa de amostragem maior, em geral 48 KHz.
- **Resposta de frequência:** É a faixa de frequência que uma placa de som consegue capturar ou reproduzir. A resposta de frequência padrão adotada mundialmente é a faixa de 20 Hz a 20 KHz, que é a faixa de frequência que o ouvido humano é capaz de escutar.
- **Relação sinal/ruído:** Essa característica mede o nível de ruído gerado pela placa de som. A maioria das placas de som possui uma péssima relação sinal/ruído, não sendo indicadas para o uso de áudio profissional.
- **Sintetizador:** O sintetizador é responsável pela reprodução de arquivos MIDI. Existem dois tipos de sintetizador: FM e Wave Table.
- **Memória RAM:** Placas de som com sintetizador de *Wave Table* em geral vêm com uma pequena quantidade de memória RAM *on board* para a criação de novos instrumentos.
- **Amplificador:** A placa de som geralmente possui duas saídas, uma de linha, chamada *line out*, que não é amplificada, e uma saída amplificada, chamada *speaker out*. A maioria das placas de som possui um amplificador *on board*, para a conexão de caixas acústicas não-amplificadas.
- **Mixer:** O mixer da placa de som é controlado por software e controla os níveis das entradas e das saídas da placa de som.

- **Entradas e saídas:** As placas de som normalmente possuem duas entradas e uma saída. As entradas são chamadas *line in* e *mic in* e servem, respectivamente, para a conexão de uma entrada de linha (*tape deck*, toca-discos, aparelho de CD, MD, DVD, televisão, etc) e uma entrada para microfone. Já a saída é chamada de *speaker out* e serve, como uma saída amplificada, utilizada para a conexão de caixinhas acústicas.
- **Entrada para CD:** Todas as placas de som possuem um conector para a entrada do áudio da unidade de CD-ROM do micro.
- **Joystick:** Todas as placas de som possuem uma porta para a conexão de um joystick.
- **MIDI:** Todas as placas de som possuem uma interface MIDI, que permite a conexão do micro com instrumentos musicais que possuem essa interface, como teclados e baterias eletrônicas.

7.2 Testes Realizados

Para verificar a eficiência do reconhecedor, foram estabelecidos diversos conjuntos de treinamentos e testes, todos distintos dois a dois.

O primeiro experimento realizado utilizou o conjunto de palavras {um, dois, três}, mostrado na Figura 7.3. A Tabela 7.1 mostra os resultados obtidos utilizando 4 conjuntos distintos para treinamento e dois para teste.

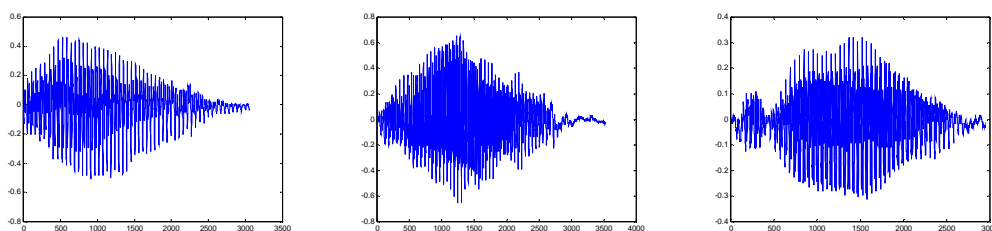


Figura 7.3 – Conjunto de palavras de teste com um percentual de 100% de acerto

Tabela 7.1 - Resultados para um conjunto de 3 palavras

	CONJUNTOS					
	Treinamento 1	Treinamento 2	Treinamento 3	Treinamento 4	Teste 1	Teste 2
TESTE	100%	100%	100%	100%	100%	66,6667%

O resultado de alguns testes realizados não foi tão bom, pois o resultado obtido foi da ordem de 66,6667%. O elemento causador de tal resposta é decorrente da semelhança entre a forma de onda da palavra um e dois, conforme a Figura 7.3. Comparando os resultados obtidos em Malcom [TAF 96], estes não são tão divergentes.

Para o segundo experimento realizado, utilizou-se um conjunto de palavras {esquerda, direita, frente, atrás}, mostrado respectivamente na Figura 7.4. A Tabela 7.2 mostra os resultados obtidos utilizando-se 5 conjuntos distintos para treinamento e um para teste.

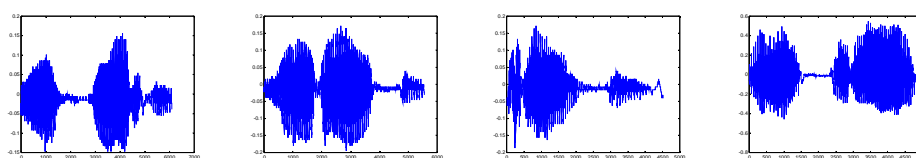


Figura 7.4 - Conjunto de palavras de teste com um percentual de 75% de acerto

Tabela 7.2 - Resultado para um conjunto de 4 palavras

	CONJUNTOS					
	Treinamento 1	Treinamento 2	Treinamento 3	Treinamento 4	Treinamento 5	Teste 1
TESTE	75%	100%	100%	100%	75%	75%

O resultado obtido para um total de 4 palavras também foi bom, visto que a taxa de aprendizado foi verdadeira perante o resultado obtido.

8 CONCLUSÕES E RECOMENDAÇÕES FINAIS

A principal meta deste trabalho é desenvolver um protótipo capaz de reconhecer palavras faladas da língua portuguesa, através do uso das redes neurais artificiais. O modelo proposto (Mapa FAN) proporcionou que este objetivo fosse alcançado, conforme demonstrado nos testes realizados, com resultados que variaram entre 66 a 100%, dependendo dos conjuntos de palavras utilizados para treinamento e teste. O desempenho do sistema utilizado pode ser melhorado. Para tanto, podem ser aplicadas técnicas, além da ordenação e da supressão do silêncio, capazes de realizar um refinamento mais detalhado das características ocultas da palavra a ser processada. Citamos por exemplo, a FFT (*Fast Fourier Transform*), *Wavelet*, conjuntos difusos, no entanto, a aplicação destas técnicas fica como sugestão para novos trabalhos.

Quanto aos conjuntos de treinamento utilizados para a realização dos testes, notou-se que, quanto maior for o número destes, maior será a percentual de acertos da rede. Também se verificou a dificuldade encontrada pela rede em reconhecer palavras homônimas, visto que as formas de onda dos sinais que estas apresentam, serem muito parecidas. Uma solução para este problema seria realizar um estudo mais aprofundado sobre as estruturas morfológicas das palavras, i.e, decompor unidades em seus elementos componentes (Análise Sintática em Reconhecimento de Padrões).

Para a continuação deste trabalho, sugere-se a criação de um ambiente, que trabalhe em *background*, capaz de realizar uma interconexão com outros aplicativos desenvolvidos para a plataforma Windows 9x, a fim de diminuir a frequência de uso de outros periféricos. Também pode ser realizado um estudo de minimização do conjunto de dados tratados pelo Mapa FAN para uma possível utilização deste em dispositivos móveis ou de pequeno porte, como por exemplo, os robôs. Esta minimização refere-se a um tratamento nos mapas treinados, visto que a matriz utilizada é muito esparsa, podendo haver um estudo de otimização utilizando matriz de banda.

9 REFERÊNCIAS BIBLIOGRÁFICAS

- [ALB 86] Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D. 1986. 23 ed. Barcelona, Espanha, Edições Omega, p. 1112
- [BAR 98] BARROS, Carlos, PAULINO, Wilson. **O Corpo Humano**. Editora Ática, 1998.
- [BAR99] BARONE, Dante Augusto Couto, “**Projeto Revox**”, versão eletrônica em <http://www.ucs.tc.br/revox>, 1999.
- [BEL 61] BELLMAN, R. **Adaptive Control Processes: A Guided Tour**. Princeton University Press, 1961.
- [BIS 95] BISHOP, C. M. **Neural Networks for Pattern Recognition**. Oxford University Press, 1995.
- [CAR 97] CARREIRA-PERPIÑÁN, M. Á. **A Review of Dimension Reduction Techniques**, Technical Report CS-96-09, 1997. <http://www.dcs.shef.ac.uk/~miguel/papers/ps/cs-96-09.ps.gz> (09 de Dezembro de 2000, 01:00).
- [CAR 98] CARVALHO, André, LUDEMIR, Antônio. **Fundamentos de Redes Neurais Artificiais: 11ª Escola de Computação**. Imprinta Gráfica e Editora Ltda, 1998.
- [CHU 92] Churchland, P.S. , Sjenowski, T.J. 1992. **The computacional Brain**. Cambridge, Mass,: MIT Press.
- [COO 84] Cooter, R. 1984. **The Cultural Meaning of Popular Science: Phrenology and yhe Organization of Consent in Nineteenth.Century Britain**. Cambridge, England: Cambridge University Press.
- [DAC 92] Data & Analysis Center for Software. **Artificial Neural Networks Technology**. Disponível por WWW em <http://www.dacs.dtic.mil> (20/09/2000).
- [DAN 96] DANDOLINI, Gertrudes A. **Um procedimento para avaliação da saúde financeira de pequenas empresas: estudo de um caso usando redes neuronais artificiais**. Dissertação de Mestrado, Universidade Federal de Santa Catarina, Departamento de Engenharia de Produção, 1996.
- [DAN 00] DANDOLINI, Gertrudes A. **Mapa FAN no Estagiamento Automático do Sono**. Tese de Doutorado. Programa de Pós-graduação em Engenharia de Produção. Universidade Federal de Santa Catarina. Florianópolis, 2000.

- [DEL 98] DELGADO, José M.^a et al. **Manual de Neurociencia**. Editorial Sintesis, 1998.
- [DUD 73] DUDA, O., HART, P. E. **Pattern classification and scene analysis**. John Wiley & Sons, Inc., 1973.
- [ESP 99] ESPAIN, Carlos. **Reconhecimento Computacional da Fala**. Cadernos do CEFAT, Porto, 1999.
- [EPS 98] EPSTEIN, I. **Teoria da Informação**. Segunda Edição. Editora Ática. São Paulo, 1988.
- [FAW 87] FAWCETT, D.W. 1987 **Bloom Fawcett – Tratado de Histología**. 11 ed. Madrid: Interamericana, p. 350.
- [GES 74] Geschwind, N. 1974. **Selected Papers on Language and the Brain**. Dordrecht, Holland: Eldeer.
- [GON 72] Gonçalves, Dalton. **Física**. Rio de Janeiro: “Ao Livro Técnico S.A.”, 1972.
- [HAL 96] HALLIDAY, David et al. **Física 2**. 4^a ed. Livros Técnicos e Científicos S.A., 1996.
- [HAY 94] HAYKIN, Simon. **Neural Networks: A comprehensive Foundation**. New York: Macmillan College Publish Company, 1994.
- [HAR 35] Harrison, R.G. 1935. **On the origin and development of the nervous system studied by methods of experimental embryology**. *Proc.R. Soc. Lon. [Biol.]*, v. 118, p. 155 – 196.
- [HAR 87] Harrington, A. 1987. **Medicine, Mind, and the Doble Brain: A Study in nineteenth-Century Thought** . Princeton, N.J.: Princeton University Press.
- [IFE 95] IFEACHOR, E. C. e JERVIS, B. W. **Digital Signal Processing – A Practical Approach**. Addison-Wesley, 1995.
- [JOH 98] JOHNSON, R. A., WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Jersey: Prentice Hall inc., 4 ed., 1998.
- [KAN 98] Kandell, E.R., Jessel, T.M., Schwartz, J.H. 1998. **Neurociencia y conducta**. Madrid: Prentice Hall, 812 p. Il.
- [KAS 96] KASABOV, Nikola K. **Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering**. The MIT Press, 1996.
- [KOR 98] KORTENKAMP, David et al. **Artificial Intelligence and Mobile Robots**. MIT Press, 1998.
- [KOV 96] KOVÁCS, Zsolt L. **Redes Neurais Artificiais: Fundamentos e Aplicações**. Segunda Edição, Collegium Cognitio, 1996.

- [LUF 95] LUFT, Celso Pedro. **Novo Manual de Português**. Editora Globo S.A, 1995.
- [LUG 98] LUGER, G. F., Stubblefield W. A., **Artificial Intelligence**. Addison Wesley, 1998.
- [LUN 00] LUNA, Paulo. **Utilização da Inteligência Artificial em Reconhecimento de Voz**. Disponível por WWW em <http://www.inf.furb.rct-sc.br/~alx/ia.html> (04/12/2000).
- [MAD 00] MADISETTI, Vijay K. **Signal Processing for the NII - Workshop/Panel Report**. <http://users.ece.gatech.edu/~vkm/nii/> (04 de abril de 2000, 02:00).
- [MAR 88a] Marshall, J.C. 1988. **Cognitive neurophysiology: The life-blood of language**. *Nature*, v. 331, p. 560 – 561
- [MAR 88b] Marshall, J.C. 1988. **Cognitive neuropsychology: Sensation and semantics**. *Nature*, v. 334 p. 378.
- [MCC 87] MCCOMB, G. **Robot Builder's Bonanza: 99 Inexpensive Robotics Projects**. MacGraw-Hill, 1987.
- [MINa 00] MINELLO, Luiz Fernando. **A voz Humana**. Disponível por E-mail em dbclfm@unileon.es (16 Dez. 2000).
- [MINb 00] MINELLO, Luiz Fernando. **O Sistema Nervoso**. Disponível por E-mail em dbclfm@unileon.es (17 Dez. 2000).
- [MOL 69] MOLES, A. **Teoria da Informação e percepção estética**. Rio de Janeiro. Tempo Brasileira, 1969. (Livro considerado clássico no tema. Traduzido para várias línguas.)
- [NAS 94] NASCIMENTO, Cairo L. **Artificial Neural Networks in Control and Optimization**. Doctor Thesis. University of Manchester. Manchester, 1994.
- [OKU 82] OKUNO, Emico, CALDAS, Iberê, CHOW, Cecil. **Física para Ciências Biológicas e Biomédicas**. Harper & Row do Brasil, 1982.
- [PAT 95] PATTERSON, Dan W. **Artificial Neural Networks: Theory and Applications**. Prentice Hall, 1995.
- [POS 94] Posner, M.I., Raichle, M.E. 1994. **Images of Mind**. New York: Scientific American Library.
- [RAI 97] RAITZ, R. T.; SOUZA, J.A.; DANDOLINI, G. A.; PACHECO, R. C. S.; MARTINS, A., GAUTHIER, F & BARCIA, R.. **Learning by Means of Free Associative Neurons**. NAFIPS 97: Annual Meeting Of The North American Fuzzy Information Processing Society, September 21-24, 1997, Syracuse, New York.

- [RAI 98] RAITTZ, R. T., SOUZA, J. A. DANDOLINI, G. A., et al., **FAN: Learning by Means of Free Associative Neurons**, WCCI98 - IEEE World Congress on Computational Intelligence, Anchorage Alaska, pp. 425-430, May 1998.
- [RAI 99] RAITTZ, R. T., SOUZA, J. A. DANDOLINI, G. A., et al. **Pré-Processamento de Sinais Digitais**. XIX ENEGEP - Encontro Nacional de Engenharia de Produção, 1999.
- [RIO 99] RIOS, Ana; RODRIGUES, Fabrício. **Reconhecimento da Fala para Sistemas de Caixas Eletrônicos**. Relatório de Estágio. Curso de Sistemas de Informação. Universidade Potiguar. Natal, 1999.
- [SAN 72] SANTOS, Udmyr. **Física (Acústica – Ótica – Eletricidade)**. Companhia Editora Nacional, 1972.
- [SCH 97] SCHOWENGERDT, Robert A. **Remote Sensing – Models and Methods for Image Processing**. Segunda Edição, Academic Press, pp. 522, 1997.
- [SHA 75] SHANNON, C.; WEAVER, W. A. **Teoria Matemática da Comunicação**. Rio de Janeiro, Difel, 1975.
- [SOU 99] SOUZA, João Artur. **Reconhecimento de Padrões Usando Indexação Recursiva**. Florianópolis, 1999. Tese (Doutorado em Engenharia de Produção). Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina.
- [TAFa 96] TAFNER, Malcon A. **Reconhecimento de palavras isoladas usando redes neurais artificiais**. Dissertação de Mestrado. Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina. Florianópolis, 1996.
- [TAFb 96] TAFNER, Malcon, XEREZ, Marcos, Rodrigues, Ilson. **Redes Neurais Artificiais: Introdução e Princípios de Neurocomputação**. EKO, 1996.
- [TOR 99] TORRES, Gabriel. **Placas de Som**. Disponível por WWW em <http://www.clubedohardware.com.br> (09/08/2000).
- [WIN 93] WINSTON, P. H. **Artificial Intelligence**. Third Edition. Addison-Wesley, 1993.
- [ZAD 65] ZADEH, L.A. **Fuzzy sets**. Information and Control, Vol. 8, 1965, pp. 338-353.