

Boletim Técnico da Escola Politécnica da USP
Departamento de Engenharia Eletrônica

ISSN 1413-2206

BT/PEE/9936

**Reconhecimento Automático do
Locutor com Redes Neurais
Artificiais do Tipo Radial Basis
Function (RBF) e Minimal Temporal
Information (MTI)**

Antonio P. Timoszczuk
Euvaldo F. Cabral Jr.

São Paulo – 1999

O presente trabalho é um resumo da dissertação de mestrado apresentada por Antonio P. Timoszczuk sob orientação do Prof. Dr. Euvaldo F. Cabral Jr.: "Reconhecimento Automático do Locutor com Redes Neurais Artificiais do Tipo Radial Basis Function (RBF) e Minimal Temporal Information (MTI)", defendida em novembro de 1998, na Escola Politécnica.

A íntegra da dissertação encontra-se à disposição com o autor e na Biblioteca de Engenharia de Eletricidade da Escola Politécnica/USP.

FICHA CATALOGRÁFICA

Timoszczuk, Antonio Pedro

Reconhecimento automático do locutor com redes neurais artificiais do tipo radial basis function (RBF) e minimal temporal information (MTI) / A.P. Timoszczuk, E.F. Cabral Jr. -- São Paulo : EPUSP, 1999.

25 p. -- (Boletim Técnico da Escola Politécnica da USP, Departamento de Engenharia Eletrônica, BT/PEE/9936)

1. Redes neurais (Inteligência artificial) 2. Reconhecimento de vozes (Computação) I. Cabral Jr., Euvaldo Ferreira II. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Eletrônica III. Título IV. Série

ISSN 1413-2206

CDD 006.3
006.454

AUTOMATIC SPEAKER RECOGNITION USING RADIAL BASIS FUNCTION NEURAL NETWORK AND MINIMAL TEMPORAL INFORMATION (MTI)

Antonio P. Timoszczuk and Euvaldo F. Cabral Jr.
Man Machine Communication - MMC group
Laboratório de Comunicações e Sinais (LCS), DEE, EPUSP
Caixa Postal 61458 - CEP 05424-970 - São Paulo - S.P. - Brasil
e-mail: pedro@lcs.poli.usp.br / euvaldo@lcs.poli.usp.br

Abstract

This work aims to verify the capabilities of the Radial Basis Function neural network when applied as a classifier to the speaker recognition task.

A speaker recognition system is defined using the Mel-Frequency Cepstral Coefficients to represent the speech segments. The Minimal Temporal Information - MTI - is defined as a novel way to organize the speech segments in order to be presented to the ANN classifier.

Text dependent and text independent tests are performed using two different data bases with sixteen and ten speakers. A multilayer perceptron is used as a classical method in order to compare the results obtained with the RBF.

The use of RBF neural networks as classifiers in ASR has been found promising. The main advantages of the RBF are the short training time required and the reduced network complexity. The use of MTI strategy to organize the speech segments improved the recognition rate and discrimination capability of the RBF, indicating that the MTI representation is useful and deserves further investigations.

RECONHECIMENTO AUTOMÁTICO DO LOCUTOR COM REDES NEURAIS ARTIFICIAIS DO TIPO *RADIAL BASIS FUNCTION (RBF)* E *MINIMAL TEMPORAL INFORMATION (MTI)*

Antonio P. Timoszczuk e Euvaldo F. Cabral Jr.
grupo CHM - Comunicação Homem Máquina
Laboratório de Comunicações e Sinais (LCS), DEE, EPUSP
Caixa Postal 61458 - CEP 05424-970 - São Paulo - S.P. - Brasil
e-mail: pedro@lcs.poli.usp.br / euvaldo@lcs.poli.usp.br

Resumo

Neste trabalho é avaliado o potencial de aplicação da rede neural do tipo *Radial Basis Function - RBF* - como classificador na tarefa de Reconhecimento Automático do Locutor - RAL -.

O reconhecedor utiliza como características extraídas das locuções, os coeficientes *Mel-Cepstrais*. Uma nova forma de organização dos segmentos temporais do sinal de voz denominada de *Minimal Temporal Information - MTI* - é definida e utilizada para a montagem dos padrões apresentados para o classificador.

Para a realização dos testes são utilizadas duas bases de dados distintas, contendo frases foneticamente balanceadas e formadas por dezesseis e dez locutores respectivamente. Uma rede neural do tipo perceptron multicamada (*MLP*) é utilizada como classificador clássico para a comparação dos resultados obtidos com a *RBF* nas tarefas de reconhecimento do locutor de forma dependente e independente do texto.

A viabilidade da *RBF* como classificador foi constatada. Os resultados das *MTIs* em conjunto com a *RBF* proporcionaram uma melhoria nas taxas de acerto do reconhecedor e na capacidade de discriminação, tornando as *MTIs* promissoras como representações temporais dos locutores.

1 Introdução

A tarefa do RAL consiste em reconhecer automaticamente a identidade de um indivíduo, a partir de comparações de características extraídas do sinal de voz.

A evolução dos sistemas computacionais permitiu que as técnicas de reconhecimento automático do locutor até então desenvolvidas apenas do ponto de vista teórico, pudessem ser implementadas de forma eficiente, fazendo com que os sistemas de RAL atingissem níveis adequados de desempenho e proporcionando sua utilização em aplicações reais. A partir de então pôde ser observada uma ampliação nas aplicações do RAL¹ e que a principal área de aplicação passou a ser a dos sistemas de segurança, com o propósito de restrição de acesso, seja a informações ou instalações físicas.

Os sistemas de RAL podem ser divididos segundo duas tarefas a serem executadas: **a identificação** do locutor e **a verificação** do locutor. Na verificação a locução de entrada é comparada apenas contra o padrão da identidade informada pelo indivíduo que fornece a locução de entrada, fornecendo uma decisão binária do tipo SIM / NÃO. No caso da identificação, é assumido um conjunto de elementos cada qual com o seu padrão, e a locução de entrada é comparada contra todos os elementos do conjunto fornecendo uma indicação de qual padrão melhor representa a locução de entrada; aqui a decisão não é do tipo binário podendo inclusive resultar em nenhum dos elementos do conjunto. De uma forma resumida, a verificação e a identificação são problemas de reconhecimento de padrões que compartilham as mesmas etapas de processamento, diferindo essencialmente no estágio final de classificação².

Uma outra classificação dos sistemas de RAL, é feita com relação ao texto pronunciado pelo locutor. Os sistemas que trabalham com textos fixos, ou seja, para o reconhecimento é necessário que o locutor pronuncie um texto pré-definido, são denominados **dependentes do texto**. Por outro lado os sistemas nos quais não existe a necessidade da utilização de textos pré-definidos são denominados **independentes do texto**. Neste trabalho é desenvolvido e testado um sistema para **identificação automática do locutor** de forma **dependente e independente do texto**.

Revisões abrangentes sobre a tarefa do RAL podem ser encontradas em Atal³, Corsi⁴ e Gish⁵.

2 A representação dos sinais de voz

A forma de representação dos sinais é fundamental para o sucesso de qualquer sistema de processamento de voz, entendendo-se aqui não só o reconhecimento da fala e do locutor, mas também qualquer sistema de comunicação que utilize a voz como principal meio de transmissão das informações. Atualmente, a grande maioria das aplicações como transmissão de voz, sistemas de resposta automática, sistemas de síntese de voz e sistemas de reconhecimento do locutor, vêm utilizando a representação digital do sinal de voz.

Uma representação digital para a geração dos sinais de voz, pode ser feita considerando um filtro digital variável no tempo controlado por coeficientes que representam os parâmetros do trato vocal, conforme mostrado na Fig. 2.1. A excitação deste filtro é feita através de um trem de impulsos quase periódico representando os sons vozeados* ou através de uma fonte de ruído branco responsável pelos sons não vozeados.

* O termo vocálico também é utilizado.

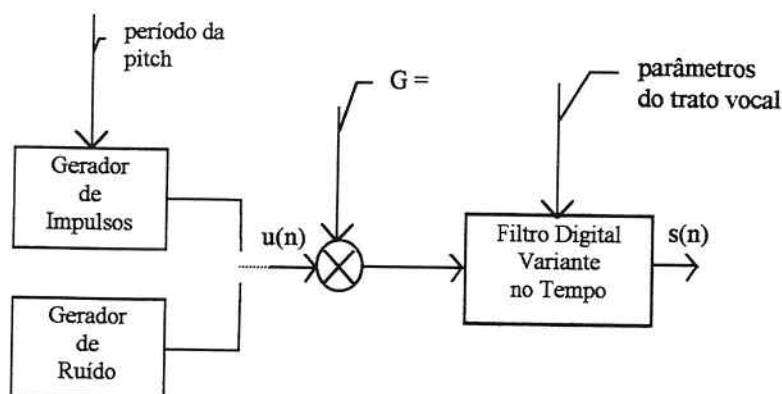


Fig. 2.1 - Diagrama simplificado de um modelo de produção da voz.

Processamento homomórfico do sinal de voz

O processamento homomórfico ⁷ é um método de análise do sinal de voz, que faz parte das técnicas não-lineares de processamento de sinal, sendo baseada na generalização do princípio de superposição que define os sistemas lineares. Sistemas que atendam a generalização do princípio de superposição, são denominados **sistemas homomórficos**, uma vez que podem ser representados através de um mapeamento linear (homomórfico, ou que preserva a forma) da entrada para a saída.

A partir do modelo utilizado para a geração do sinal de voz descrito anteriormente, o problema de análise do sinal de voz, em segmentos curtos o suficiente no tempo, pode ser encarado como um problema de separação das componentes de uma convolução da função excitação (pulsos quase periódicos ou ruído branco) com a função de resposta impulsiva do trato vocal, ou seja, um problema de deconvolução proveniente do modelo da Fig.2.1.

A partir destas considerações, a deconvolução homomórfica pode ser aplicada para separar estes sinais. O processo consiste em tomar o resultado da transformada de Fourier do sinal de voz e calcular a transformada inversa de Fourier do logaritmo deste sinal. O logaritmo é aplicado com o objetivo de transformar o produto entre a função de excitação e o trato, em uma soma.

$$\log(S) = \log(EV) = \log(E) + \log(V) \quad (2.1)$$

onde: S = transformada de Fourier do sinal de voz

E = parcela correspondente a função de excitação

V = parcela correspondente a função de transferência do trato vocal

$$c(n) = \mathfrak{F}^{-1} \log(S) = \mathfrak{F}^{-1} \log(E) + \mathfrak{F}^{-1} \log(V) \quad (2.2)$$

O resultado desta transformação, em função das partes real e imaginária da transformada de Fourier, é o chamado **cepstrum complexo** (Fig. 2.2).

A forma de cepstrum comumente utilizada e aplicada neste trabalho é o chamado **cepstrum real** ou simplesmente **cepstrum**. A diferença básica entre estas duas formas de cepstrum é descrita em ⁷, e consiste em que o cepstrum descarta a informação de fase. Embora o cepstrum complexo seja mais preciso, apresentando informações que são perdidas no cepstrum real, sua aplicação prática é mais onerosa em termos computacionais.

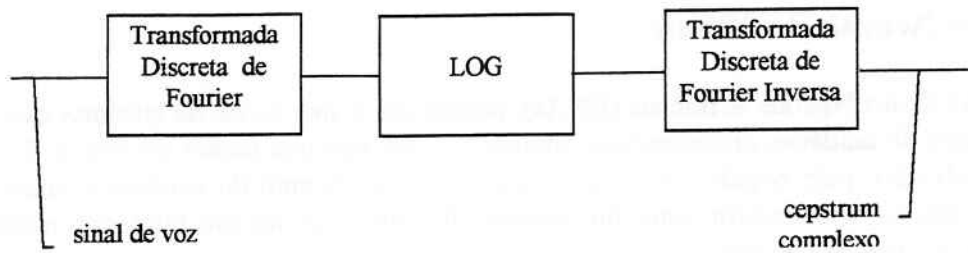


Fig. 2.2 - Operações básicas para obtenção do cepstrum complexo.

Uma avaliação importante dos parâmetros mais utilizados nos sistemas de processamento de voz atuais é apresentada em PICONÉ⁸, onde as representações utilizadas compartilham algumas importantes características. Em primeiro lugar são constituídas por informações de origem temporal e em segundo lugar tendem a ser perceptualmente compatíveis com o sistema auditivo humano, por exemplo o uso do log da energia (potência em decibéis) ou o uso de bancos de filtros distribuídos segundo escalas perceptuais (escala Bark ou Mel) na análise espectral, procuram emular a resposta do ouvido humano. As expressões a seguir definem o mapeamento das frequências acústicas f para as escalas perceptuais Bark, definida como *critical band rate*, e Mel.

$$Bark = 13 \operatorname{atan}\left(\frac{0.76f}{1000}\right) + 3.5 \operatorname{atan}\left(\frac{f^2}{7500^2}\right) \quad (2.3)$$

$$Mel = 2595 \log_{10}\left(1 + \frac{f}{700.0}\right) \quad (2.4)$$

Atualmente os parâmetros tendem a ser normalizados segundo alguma forma estatística. Os coeficientes Mel-cepstrais são os mais utilizados, colocando os coeficientes de predição linear LPC em segundo plano. O vetor de parâmetros mais utilizado consiste de coeficientes cepstrais, coeficientes delta-cepstrais, a energia e a delta-energia. A técnica de normalização mais utilizada é a da normalização pela variância⁹.

Outro estudo recente que ratifica esta conclusão é apresentado em RUNSTEIN¹⁰, onde diversos parâmetros são analisados para aplicação em um sistema classificador baseado em redes neurais artificiais. Os parâmetros que apresentam melhor desempenho são os Mel-cepstrais, destacando que os coeficientes gerados utilizando a DCT apresentam resultados melhores. Isto provavelmente ocorra devido a uma propriedade da DCT, que é ordenar os coeficientes obtidos por ordem de variância, colocando aqueles que apresentam maior variância em primeiro lugar; assim os coeficientes obtidos via DCT provavelmente carreguem mais informação do que os seus correspondentes obtidos através da FFT inversa. Um diagrama bastante objetivo de como gerar os coeficientes Mel-cepstrais utilizando a DCT é apresentado em REYNOLDS¹¹ e aqui reproduzido na Fig. 2.3.

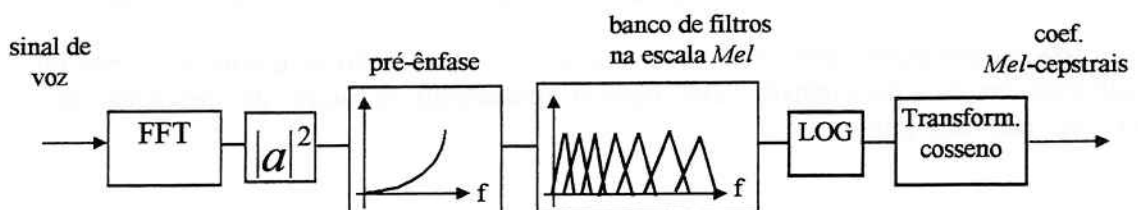


Fig. 2.3 - Sequência para geração dos coeficientes Mel-cepstrais.

3 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs), pertencem a uma classe de sistemas que são constituídos de unidades processadoras simples. Os estudos que deram origem às RNAs, foram motivados pelo objetivo de compreender o funcionamento do cérebro humano e a idéia de simular o comportamento do cérebro, fez com que fossem buscados modelos matemáticos para os neurônios.

As RNAs procuram explorar os princípios adotados pelo cérebro, especialmente o paralelismo, na busca de implementar sistemas que possam solucionar problemas da mesma forma que este último.

Assim uma RNA é constituída por neurônios ou elementos processadores, conexões e pesos associados a estas, sendo estes últimos ponderadores das conexões. De uma forma geral, na literatura, é citado que os pesos são os responsáveis pelo armazenamento do conhecimento da rede. Esta afirmação não é errônea, porém existem arquiteturas nas quais os neurônios, por meio de sua função de ativação também armazenam parte do conhecimento da rede, como por exemplo a rede do tipo *RBF*, partilhando esta responsabilidade com os pesos das conexões.

Topologia da rede

Um exemplo de modelo de RNA, é apresentado na Fig. 3.1. Este modelo é baseado em camadas, embora existam redes que não apresentem este tipo de estrutura. Do ponto de vista de topologia, a forma mais comum de rede, é aquela organizada em camadas, as quais incluem:

- a camada de entrada;
- a camada oculta (uma ou mais de uma);
- a camada de saída;

Com relação a conectividade as redes podem ser:

- propagação para adiante (*feedforward*): todas as conexões apontam em uma única direção, da entrada para a saída;
- recorrentes: quando existem conexões de realimentação.

e quanto ao tipo de conexão, podem ser do tipo simétrico ou assimétrico.

A ordem das conexões é determinada pelo número de saídas de neurônios, que são combinadas em uma única conexão; tipicamente todas as RNAs são de primeira ordem.

Com relação aos pesos, estes podem ser números reais ou inteiros, e podem ou não ser confinados em uma determinada faixa. Após o treinamento os valores dos pesos não são, em geral, mais alterados.

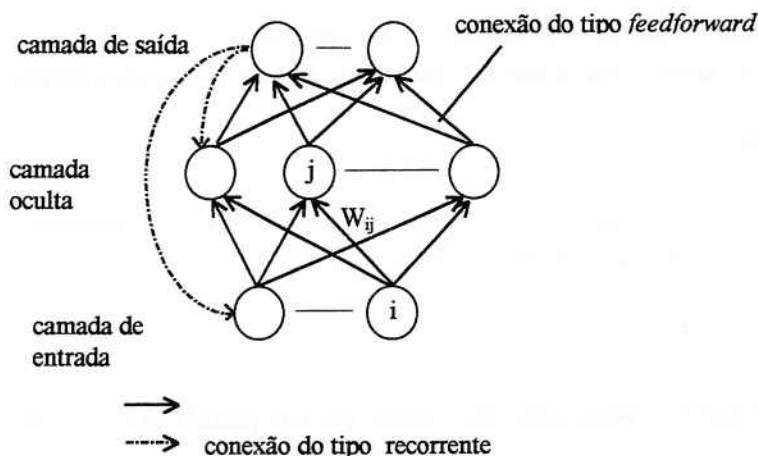


Fig. 3.1 - Exemplo de modelo de uma RNA.

A construção de uma RNA envolve os seguintes passos¹²:

1. determinação das propriedades da rede: a topologia (ou conectividade), o tipo de conexões, a ordem das conexões e a faixa dinâmica dos pesos;
2. determinação das propriedades dos neurônios que compõem a rede: a função de neurônio (ou função de transferência) e a faixa dinâmica de ativação dos neurônios;
3. determinação da dinâmica do sistema: envolve o esquema de inicialização dos pesos, o cálculo da função de cada neurônio e a regra de aprendizado.

O Perceptron

Conforme citado em HUSH¹³, um dos modelos de neurônio mais utilizado atualmente como bloco básico para a construção de redes mais complexas, surgiu através de Rosenblatt em 1958. A sua arquitetura é essencialmente a mesma do neurônio de McCulloch e Pitts e a grande diferença com relação ao modelo anterior, é o estabelecimento de um algoritmo de aprendizado para o ajuste dos pesos⁴⁷, que é apresentado a seguir:

passo 1: inicializar pesos (w_i) ($0 \leq i \leq N-1$) e limiar (θ) com valores aleatórios e pequenos

passo 2: apresentar o primeiro par entrada ($i_1, i_2, \dots, i_n$)/saída desejada (d)

passo 3: calcular a saída atual

$$o(t) = F\left(\sum_0^{N-1} w_i i_i(t) - \theta\right) \quad (3.1)$$

passo 4: atualizar os pesos, segundo a regra

$$w_i(t+1) = w_i(t) + \alpha e(t) x_i(t), \quad 0 \leq i \leq N-1 \quad (3.2)$$

onde:

$e(t)$, corresponde ao erro (no caso de Rosenblatt $e(t) = d(t) - o(t)$)

α = ganho positivo < 1 , corresponde ao coeficiente de aprendizado

passo 5: se o erro, desejado foi atingido, parar;
caso contrário, ir para o **passo 2**

Um comentário importante a ser feito aqui, é o fato de que o *perceptron* funciona como um classificador linear e o exemplo a seguir ilustra esta afirmação.

Exemplo:

Supondo um *perceptron* com duas entradas (i_1 e i_2), deseja-se separar duas classes distintas A e B, descritas em função de i_1 e i_2 . A equação que descreve a saída fica:

$$o(t) = w_1 i_1(t) + w_2 i_2(t) + \theta \quad (3.3)$$

a região limite de separação das classes, ocorre quando $o(t) = 0$, ou

$$w_1 i_1(t) + w_2 i_2(t) + \theta = 0 \quad (3.4)$$

$$\text{ou } i_2(t) = -\frac{w_1}{w_2} i_1(t) - \frac{\theta}{w_2} \quad (3.5)$$

que é a equação de uma reta

a Fig. 3.3, ilustra uma possível separação destes dois padrões.

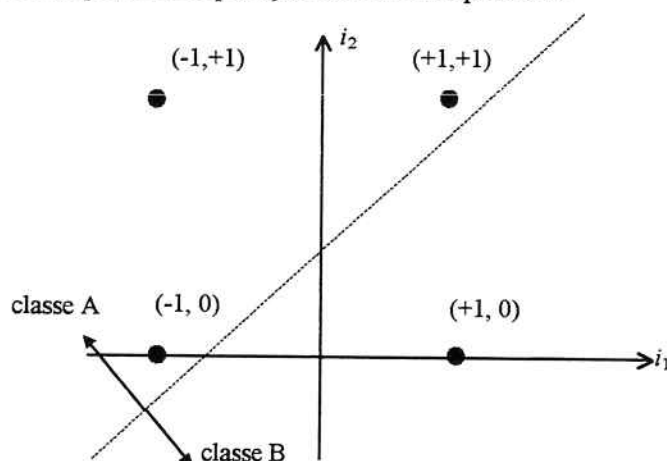


Fig. 3.2 - Separação linear de dois padrões no espaço.

Deve ser observado que até aqui, a função do neurônio corresponde ao degrau unitário (*hard limiting*) conforme descrito originalmente por Rosenblatt. Contudo na prática, a função *sigmoidal* do tipo:

$$o(t) = \frac{1}{1 + e^{-\beta x(t)}} \quad (3.6)$$

é mais utilizada, como não linearidade do *perceptron*, porque a ativação se “suaviza”. Devido a isso CABRAL¹⁴ usa o termo *Neuron Function* para essa função. A função da equação 3.7 é contínua e monotônica, com $0 \leq o(t) \leq 1$, para $-\infty \leq x(t) \leq +\infty$. O fator β , corresponde a inclinação da *sigmóide*, e para valores de β muito grandes, a função tende para um degrau unitário. Em Hush¹³, são apresentadas algumas vantagens que justificam a utilização da função *sigmóide*, como função de ativação do neurônio, dentre elas:

1. é uma função diferenciável, permitindo a implementação de algoritmos de aprendizado do tipo gradiente descendente (*gradient descent*), para redes de múltiplas camadas, conforme será visto mais adiante neste trabalho;
2. é adequada para aplicações que necessitem de uma saída com valores contínuos;
3. produz uma saída com valores entre 0 e 1, que pode ser interpretada como uma probabilidade estimada, dependendo do tipo de rede e algoritmo de treinamento utilizado.

O neurônio RBF

Conforme descrito no trabalho de COVER¹⁵, a capacidade de separar classes aumenta com a complexidade da função de separação. Em BROOMHEAD¹⁶ é examinada uma classe de funções que apresenta simetria radial, isto é, um caso particular de funções quadráticas que apresentem um eixo de simetria no ponto de inflexão. Utilizando como função do neurônio esta classe particular de funções, pode ser contruído um neurônio cuja saída (o) seja proporcional à menor distância entre a entrada (i) e este eixo de simetria da função do neurônio no espaço, definindo assim uma região de ativação.

As regiões com mesmo valor da saída (o), formam assim superfícies de decisão paralelas no espaço sob a forma de elipsóides concêntricas, este é o conceito do classificador por região também chamado de *kernel classifier*¹⁷. Uma função de neurônio muito utilizada para este tipo de classificador, é a função Gaussiana, conforme abaixo:

$$O = e^{-\frac{\|I - P\|^2}{2\sigma^2}} \quad (3.7)$$

onde $\|I - P\|$ é a distância Euclideana entre o vetor de entrada I e o centro da Gaussiana P , e σ corresponde a abertura da Gaussiana (ou raio da função).

A RNA do tipo Perceptron Multicamadas ou Multi Layer Perceptron (MLP)

O *perceptron* simples, só é útil para a separação de classes linearmente separáveis, o que limita em muito a sua aplicação prática. A crítica mais contundente sobre esta deficiência foi apresentada através de Minsky¹⁸, embora soluções já fossem apontadas para contornar este problema utilizando múltiplos *perceptrons* distribuídos em camadas totalmente conectadas. Na Fig. 3.3 pode ser observado um exemplo de *MLP* com três camadas totalmente conectadas, utilizando a nomenclatura criada por CABRAL¹⁴.

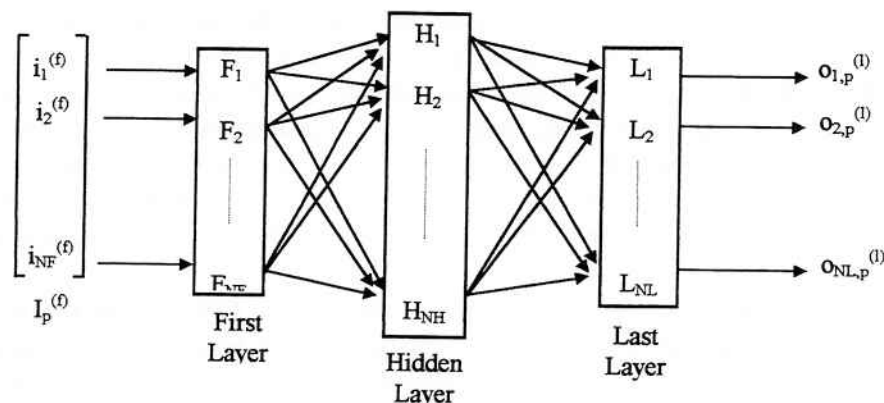


Fig. 3.3 - Diagrama de um *MLP* de três camadas clássico.

Somente em 1986 com a divulgação do trabalho de RUMELHART¹⁹, é que o algoritmo de retropropagação ou *backpropagation* tornou-se efetivamente conhecido permitindo o uso efetivo do *MLP*.

O algoritmo de retropropagação (*backpropagation*)

O algoritmo de *backpropagation*, consiste na propagação do erro da saída da rede para a sua entrada. Este algoritmo foi desenvolvido a partir da generalização da regra delta, citada anteriormente neste trabalho, utilizando a técnica do gradiente descendente (*gradient descent*) para efetuar a correção dos pesos em função da derivada do erro com relação aos pesos anteriores^{20, 13, 19}.

As expressões genéricas para a atualização dos pesos, são as seguintes:

$$w(t+1) = w(t) + \Delta w(t) \quad (3.8)$$

para neurônios pertencentes à qualquer camada:

$$\Delta w(t) = \alpha \delta(t) i(t) \quad (3.9)$$

para um neurônio de saída:

$$\delta(t) = (d(t) - o(t)) \partial f(nf) \quad (3.10)$$

para um neurônio que não pertença a saída:

$$\delta(t) = \partial f(nf) \sum_k \delta_k(t) w_k(t) \quad (3.11)$$

onde:

Δw é o valor da correção aplicada aos pesos;

$\partial f(nf)$, é a derivada da função de ativação do neurônio; e

$\delta_k(t)$ e $w_k(t)$, são respectivamente a derivada da função de erro com relação aos pesos de um neurônio da camada subsequente e o peso da ligação entre o neurônio atual e o seu subsequente.

O termo $\delta(t)$, corresponde a derivada da função de erro com relação aos pesos, sendo responsável pela propagação do erro da saída da rede para as suas camadas intermediárias, através de um cálculo recursivo.

A RNA do tipo Função de Base Radial ou Radial Basis Function (RBF)

Esta é uma classe de redes multicamada, construídas a partir dos modelos de neurônios do tipo *RBF* descritos anteriormente. Os neurônios *RBF* são utilizados na camada oculta, sendo o resultado destes combinado linearmente na camada de saída.

O interesse pelas redes neurais do tipo *RBF* tem crescido em várias áreas dentre as quais podem ser citadas o reconhecimento de padrões, o processamento digital de sinais, o modelamento de sistemas e controle, dentre outras. Os motivos para tal interesse residem na sua estrutura simples, na existência de uma base teórica bem estabelecida e principalmente pela velocidade de aprendizado desta rede, fator decisivo para aplicações em tempo real.

A rede do tipo *RBF* enquadra-se na classe das redes *feed forward*, sendo formada geralmente por uma camada de entrada, uma camada oculta e uma camada de saída. A diferença com relação as outras redes do tipo *feed forward*, como o *MLP* por exemplo, está no fato dos neurônios da camada oculta apresentarem uma resposta localizada que é radialmente simétrica em relação a um centro. Desta forma são formadas regiões denominadas campos receptivos e a camada de saída efetua uma combinação linear da

resposta destas regiões. A escolha do tipo de não linearidade utilizado como função de neurón da camada oculta, recai sobre alguns tipos de funções que serão vistos mais adiante, porém tipicamente a função utilizada é do tipo Gaussiana.

Esta rede pode ser utilizada tanto para tarefas de classificação como aproximação de funções. No caso da classificação, os campos receptivos representam padrões e a rede determina a proximidade de um padrão apresentado em sua entrada com relação aos padrões representados pelos campos receptivos. Sob este ponto de vista as redes do tipo *RBF* apresentam um desempenho muito bom, especialmente se o problema apresenta-se em espaços de grande dimensionalidade. Para a aproximação de funções, a rede efetua o mapeamento da entrada para a saída através de uma função $f: R^n \rightarrow R^m$ que é uma combinação linear das saídas das unidades da camada oculta.

O mecanismo das RBFs

A utilização das *RBFs* na camada oculta desta rede, provoca um mapeamento das amostras de entrada para um espaço não linear, buscando tornar estas linearmente separáveis pela camada de saída. Esta propriedade é descrita pelo *Teorema de Cover para a separação de padrões*¹⁵, onde é determinado que: “um problema de classificação de padrões complexo, tem maior chance de ser linearmente separável, se for mapeado não linearmente para um outro espaço diferente daquele utilizado originalmente pelo problema”, ou seja, levar um problema de classificação para um espaço, eventualmente, de maior dimensão pode fazer com que sejam observadas propriedades que não são observáveis no espaço original do problema.

A arquitetura usual da RBF

As redes do tipo *RBF*, na sua forma clássica, utilizam uma estrutura composta por três camadas de neurónios que são dispostas conforme mostrado na Fig. 3.4.

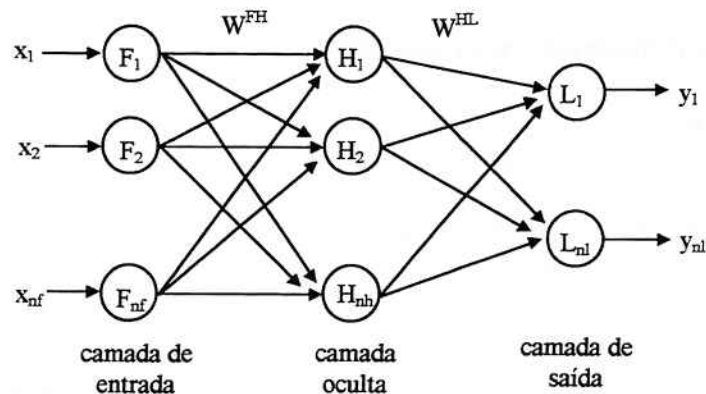
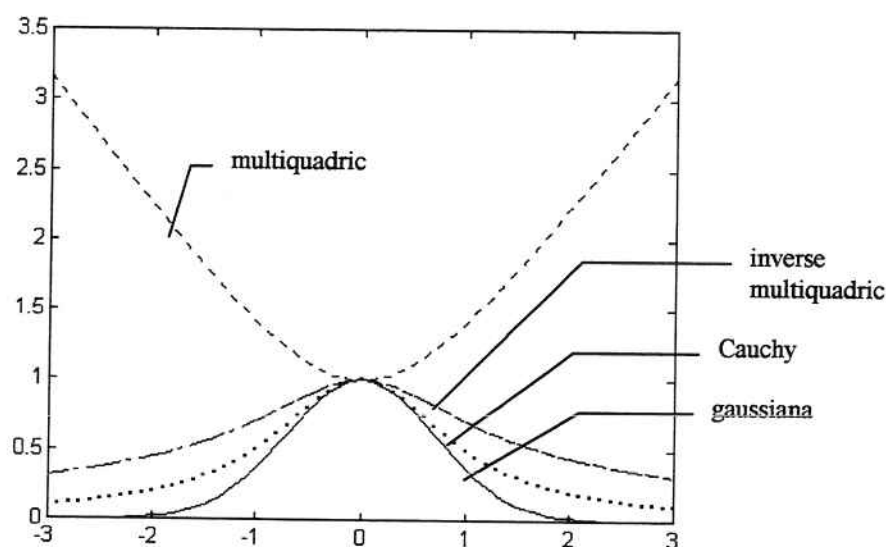


Fig. 3.4 - Estrutura clássica de uma rede do tipo *RBF*; os índices *nf*, *nh* e *nl* representam o número de neurónios da camada de entrada, da camada oculta e da camada de saída respectivamente.

A camada de entrada (neurónios *F*) é totalmente conectada através da matriz de pesos W^{FH} com a camada oculta (neurónios *H*) e esta por sua vez, totalmente conectada através da matriz de pesos W^{HL} com a camada de saída (neurónios *L*). A entrada da rede é dada por um vetor $X=[x_1, x_2, \dots, x_n]$ e a saída da rede é dada pelo vetor $Y=[y_1, \dots, y_n]$. Cada uma destas camadas será examinada em detalhes nos itens a seguir

Tab. 3.1 - Funções de base radial (*RBF*) mais comuns.

Função	Expressão
gaussiana	$\phi(z) = e^{-z}$
multiquadric	$\phi(z) = (1+z)^{\frac{1}{2}}$
inverse multiquadric	$\phi(z) = (1+z)^{-\frac{1}{2}}$
Cauchy	$\phi(z) = (1+z)^{-1}$

**Fig. 3.5** - Representação gráfica das funções de base radial mais comuns.

A Construção e o treinamento das redes *RBF*

Construção da rede

A definição da rede consiste em estipular qual será o número de neurônios em cada camada da rede e as funções de neurôn que serão utilizadas para cada uma das camadas. O número de neurônios da camada de entrada *nf* deve ser adequado ao vetor de entrada que pretende-se utilizar, sendo este dado dependente do problema e da forma de representação dos dados de entrada utilizados. Da mesma forma o número de neurônios na camada de saída *nl* deve ser adequado à resposta desejada para o problema; por exemplo caso esteja-se lidando com um problema de classificação o número de neurônios da camada de saída poderia ser igual ao número de classes existentes no problema.

O treinamento da rede

Típicamente, o aprendizado da rede do tipo *RBF*, é feito em duas fases distintas:

- a primeira consiste em posicionar as *RBFs*; e
- na segunda fase é efetuado o treinamento dos pesos da rede.

posicionamento dos centros das RBFs

O posicionamento dos centros, consiste na verdade em sintonizar a função de neurôn, o que no caso de funções Gaussianas, consiste em determinar a posição dos centros das

funções e o raio destas funções. A forma mais simples de posicionar os centros das RBFs, é escolher os mesmos de forma aleatória, a partir do conjunto de dados de treinamento. Porém esta não é uma boa prática, pois pode conduzir a uma “polarização” da rede sobre uma parte do espaço de entrada. As soluções apontadas para evitar este tipo de problema, consistem em utilizar algum método de quantização, de forma a construir um conjunto de vetores representativos das classes do problema, a partir dos dados de entrada. Na prática têm sido aplicados:

- o algoritmo K-Means;
- o LBG;
- redes de quantização vetorial, do tipo LVQ (*Learning Vector Quantization*).

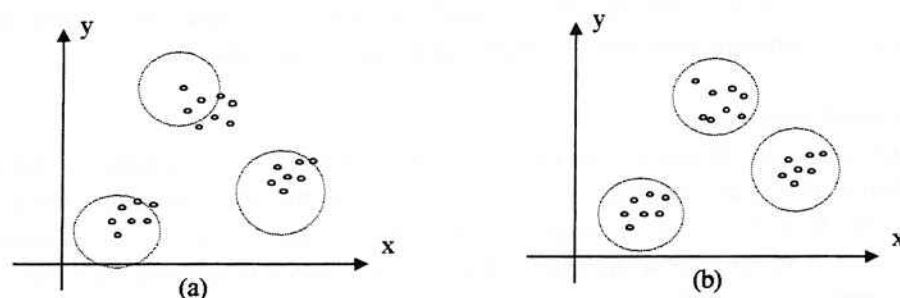


Fig 3.5 - Exemplo do posicionamento das funções de base. (a) posicionamento aleatório pode não cobrir adequadamente todas as classes; (b) posicionamento adequado cobrindo todas as classes completamente.

atualização dos pesos

Na forma mais comum de rede, são atualizados apenas os pesos entre a camada oculta e a camada de saída, sendo que os pesos entre a camada de entrada e a camada oculta são fixos e iguais a unidade. O algoritmo mais utilizado é o de retropropagação, porém dada a topologia de rede descrita inicialmente, é possível efetuar-se o por meio de regras mais simples, como por exemplo regra do *perceptron* (utilizada inicialmente neste trabalho).

4 Descrição do sistema de reconhecimento implementado

Dois sistemas de reconhecimento foram implementados neste trabalho, um deles utilizando as redes do tipo *RBF* na etapa de classificação e o outro utilizando um *MLP*. O sistema que utiliza o *MLP* foi utilizado como referência durante os testes para estabelecer o desempenho das *RBFs* na tarefa do reconhecimento automático do locutor.

Pré-processamento

A etapa de pré-processamento tem por objetivo preparar as amostras do sinal de voz para a extração dos parâmetros que serão utilizados no reconhecimento, eliminando trechos indesejáveis (como silêncio), compensando algumas propriedades do sinal de voz e distorções introduzidas durante a digitalização do sinal. As etapas aplicadas no pré-processamento são apresentadas na Fig. 4.1 sendo detalhadas nos itens seguintes.



Fig. 4.1 - Pré-processamento dos sinais de voz utilizado no reconhecedor.

Amostragem dos sinais de voz

Para a aquisição dos sinais de voz, foi utilizado um microcomputador do tipo PC equipado com uma placa de som modelo *Sound Blaster 16* e os sinais foram gravados em arquivos no formato *wave* (arquivos .WAV). Os sinais foram adquiridos em ambiente de laboratório com baixo ruído ambiente, utilizando um microfone em modo mono, como forma de reduzir o tamanho dos arquivos gerados. A frequência de amostragem adotada foi de 11 KHz (11.025 Hz) e as amostras foram codificadas em PCM - 16 bits, tomando-se o cuidado de normalizar a amplitude máxima de forma que não ocorresse saturação.

Deteção dos pontos terminais de uma locução

A detecção dos pontos terminais (*end-points*) de uma locução, tem por objetivo determinar os pontos de início e fim de uma locução permitindo eliminar os trechos de silêncio que situam-se imediatamente antes e depois da locução a ser analisada.

Segmentação do sinal amostrado

Para a aplicação do processo de geração dos parâmetros de análise, o sinal amostrado resultante de cada locução foi dividido em segmentos menores. Foram utilizados segmentos de 256, 512 e 1024 pontos com uma superposição de 50% entre segmentos sucessivos. A superposição de segmentos consecutivos é necessária para evitar a perda de informação quando da aplicação da etapa de janelamento que será vista adiante. A Tab. 4.1 apresenta a duração de cada segmento e a superposição no tempo em cada um dos casos.

Tab. 4.1 - Duração resultante e superposição dos segmentos de voz amostrados.

tamanho de cada segmento		superposição de segmentos consecutivos	
número de pontos	tempo (mseg)	número de pontos	tempo (mseg)
256	23,22	128	11,61
512	46,44	256	23,22
1024	92,88	512	46,44

Pré-ênfase

O sinal de voz concentra a maior parte de sua energia nas frequências mais baixas (atenuação de cerca de 20 dB/década na medida em que a frequência aumenta) e esta energia tende a predominar durante a fase de extração de parâmetros. A pré-ênfase é aplicada como uma forma de compensar este efeito por meio de um filtro de compensação de primeira ordem, $H(z)$ dado por: $H(z) = 1 - \alpha z^{-1}$, com $\alpha = 0,94$.

Janelamento

Ao efetuar a divisão do sinal amostrado de uma locução em segmentos sucessivos, equivale a aplicar uma janela retangular. A transição abrupta nos extremos de cada segmento, faz com que surjam componentes de frequências elevadas durante a análise e que não fazem parte do sinal original (*spectral leakage*).

Extração das características do sinal de voz

Nesta fase são gerados os parâmetros utilizados para o reconhecimento do locutor sendo feita a adequação dos mesmos para a fase de classificação. Os coeficientes *Mel-cepstrais* (MFCCs - *Mel Frequency Cepstral Coefficients*) foram escolhidos como parâmetros a serem utilizados em função da sua ampla aplicação nos sistemas de reconhecimento de voz desenvolvidos atualmente e superioridade do ponto de vista de

robustez em relação a outros parâmetros como os coeficientes LPC, cepstrais, Energia, conforme citado anteriormente. A Fig. 4.2 apresenta as etapas da extração de características, destacando as variações existentes para cada tipo de estratégia de apresentação dos vetores de parâmetros para o classificador, conforme descrito no início deste capítulo.

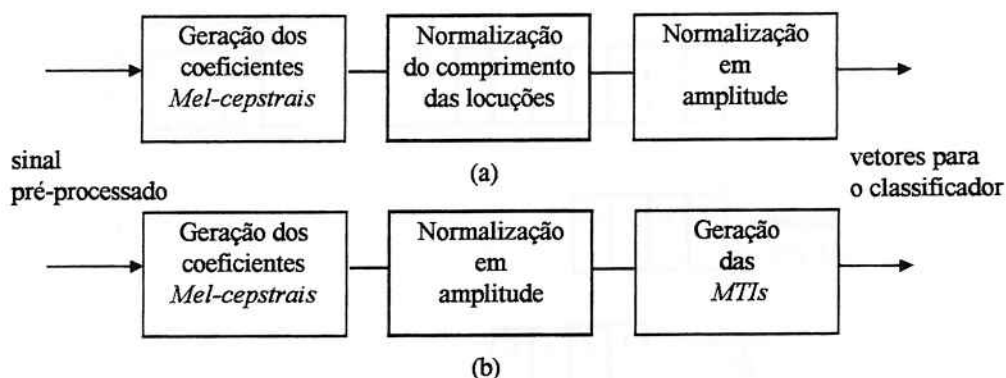


Fig. 4.2 - Etapas para extração de parâmetros e preparação dos vetores para o classificador onde (a) gera vetor contendo toda a locução e (b) gera vetor segundo o conceito de *MTI*.

Geração dos MFCCs

Os coeficientes *Mel-cepstrais* foram gerados utilizando a transformada discreta cosseno (*Discrete Cosine Transform - DCT*).

Normalização em amplitude

A estratégia adotada neste caso, foi a de normalizar os vetores pela maior magnitude global do conjunto, de maneira que os *MFCCs* tenham seus valores situados entre -1 e +1.

A MTI - Minimal Temporal Information

A *MTI* proposta e utilizada neste trabalho, tem por objetivo pesquisar a possibilidade de caracterização do locutor por meio de estruturas temporais obtidas a partir do vetor de *MFCCs* utilizados como parâmetros para o reconhecimento do locutor.

A forma de obtenção destas estruturas é simples e direta, e aliada a ferramentas de agrupamento (*clustering*) apresenta um grande potencial para ser aplicada em conjunto com as RNAs para a modelagem de locutores. A seguir é apresentada a forma de se obter as *MTIs* juntamente com a denominação utilizada para identificá-las.

As *MTIs* são conjuntos de segmentos consecutivos extraídos a partir da sequência de segmentos que representam uma locução. Ao longo de uma locução as *MTIs* são consecutivas e podem apresentar repetição ou não de segmentos utilizados na *MTI* imediatamente anterior. A Fig. 4.3 apresenta de forma gráfica o método de construção das *MTIs*.

Na Fig. 4.3 é apresentado o processo para obtenção da *MTI*₄₂. Os dois algarismos após a sigla *MTI* determinam o modo pelo qual a mesma foi montada, neste caso uma *MTI* é formada por 4 (quatro) segmentos consecutivos com 2 (dois) segmentos de repetição. A *MTI* seguinte, por sua vez, é obtida com a sequência que utiliza os dois

últimos segmentos da *MTI* anterior justapostos aos dois segmentos seguintes. Desta forma, a denominação geral para as *MTIs* passa a ser:

MTIsr, onde *s* indica o número de segmentos utilizados por uma *MTI*; e
r indica o número de segmentos repetidos da *MTI* imediatamente anterior.

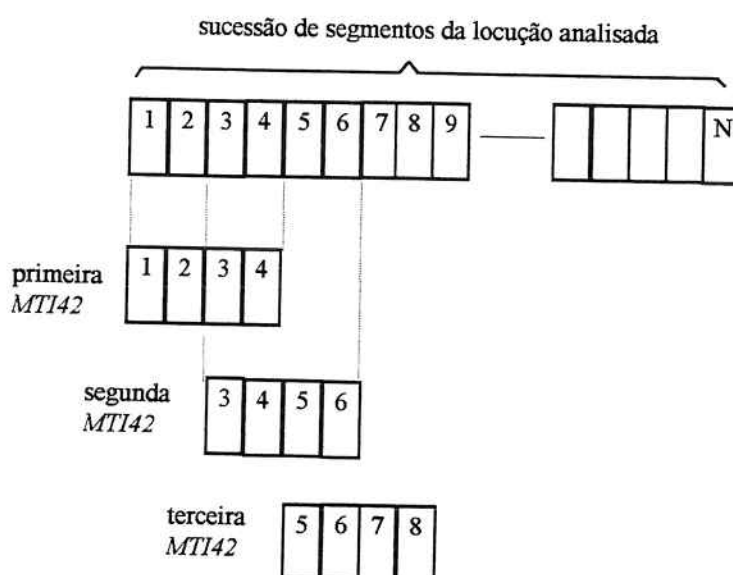


Fig. 4.3 - Processo de obtenção da *MTI42*.

Acredita-se que deva existir um número máximo de segmentos que componha uma *MTI*, a partir do qual passa-se a trabalhar com informação dependente do texto (fonemas por exemplo) e a partir deste ponto é lógico esperar-se que sua utilização deixe de trazer benefícios.

Tab. 4.2 - Extensão temporal de cada *MTI* utilizada.

MTI	Segmento 256 ptos (mseg)	Segmento 512 ptos (mseg)	Segmento 1024 ptos (mseg)
MTI1X	23,22	46,44	92,88
MTI2X	34,83	69,66	139,32
MTI3X	46,44	92,88	185,76
MTI4X	58,05	116,10	232,20
MTI5X	69,66	139,32	278,64

MTInX: indica todas as *MTIs* de *n* segmentos.

5 Experimentos

Bases de dados utilizadas

Para a realização dos experimentos deste trabalho foram utilizadas duas bases de dados distintas. A primeira base de dados, denominada base de dados 1, foi utilizada para o reconhecimento do locutor na forma dependente do texto e a segunda base de dados, denominada de base de dados 2, foi utilizada para o reconhecimento do locutor na forma independente do texto.

Base de dados 1

A base de dados 1 foi construída a partir de frases foneticamente balanceadas, as quais foram originalmente avaliadas para o projeto *NESPER (Neural Speaker Recognition)*⁷⁶. A relação de frases sugeridas por este estudo e utilizadas nesta base de dados são apresentadas na Tab. 5.1.

Tab. 5.1 - Frases utilizadas na base de dados 1.

frase 1	O prazo tá terminando.
frase 2	Amanhã ligo de novo.
frase 3	Aguarde próximo contato.

Dentre as frases sugeridas, a frase 2 foi escolhida em função do conteúdo nasal da mesma, fato que tem sido comprovado de grande utilidade para o RAL, conforme examinado no capítulo 2.

Dezesseis locutores, identificados pelas siglas LOC01 a LOC16, participaram das sessões de gravação, sendo nove do sexo masculino e sete do sexo feminino e com idades variando entre 20 e 35 anos. Cada um dos locutores gravou cinco repetições da frase escolhida em uma única sessão de gravação, sendo as repetições identificadas pelas siglas R1 a R5. Desta forma foram gravadas no total 80 locuções.

Em seguida foi efetuada a determinação dos conjuntos de treino e teste através de uma análise dos agrupamentos (*clustering analysis*) utilizando o algoritmo *MKM (Modified K-Means Algorithm)* descrito em WILPON²¹.

Foram utilizados também 6 locutores escolhidos aleatoriamente a partir da base de dados *POLIDATA*, que foi desenvolvida pelo grupo CHM - Comunicação Homem Máquina* - no âmbito do Laboratório de Comunicações e Sinais, para utilização em projetos na área de voz. Foi utilizada uma única repetição da frase 2 para cada um dos locutores, que foram identificados como IMP01 a IMP06. Estas locuções foram utilizadas para efetuar testes em conjunto aberto (*open set*), onde são apresentadas locuções pertencentes a locutores que não fazem parte da base de dados original.

Base de dados 2

A base de dados 2 foi construída a partir de frases foneticamente balanceadas que são encontradas em ALCAIM²². No trabalho de Alcaim foi feita uma análise estatística da ocorrência dos fones no português falado no Rio de Janeiro tendo sido montadas vinte listas contendo dez frases cada uma, de forma que em qualquer uma delas são encontrados os 37 fones constatados no levantamento estatístico efetuado.

A partir destas 20 listas disponíveis, foram montadas 4 listas de dez frases segundo o seguinte os critérios descritos a seguir.

lista 1

Para a lista 1 foram escolhidas as frases de número 1 a 6 da lista 3 do trabalho de Alcaim. Estas frases foram escolhidas pois segundo Alcaim concentram alguns sons específicos, a saber:

- frase 1 - vogais e semivogais
- frase 2 - vogais e consoantes nasais
- frase 3 - fricativas sonoras e surdas

* Esta base de dados está disponível via ftp.

frase 4 - oclusivas e africadas

frase 5 - laterais

frase 6 - vibrantes

A Tab. 5.2 apresenta as frases consideradas para a lista 1.

Tab. 5.2 - Frases da base de dados 2 - lista 1.

frase 1	Eu vi logo a Iôlô e o Léo.
frase 2	Um homem não caminha sem um fim.
frase 3	Vi Zé fazer essas viagens seis vezes.
frase 4	O atabaque do Tito é coberto com pele de gato.
frase 5	Ele lê no leito de palha.
frase 6	Paira um ar de arara rara no Rio Real.

lista 2

Foi escolhida de forma aleatória a lista 2 do trabalho de Alcaim contendo dez frases. As frases que fazem parte desta lista são apresentadas na Tab. 5.3.

Tab. 5.3 - Frases da base de dados 2 - lista 2.

frase 1	Nosso telefone quebrou.
frase 2	Desculpe se magoei o velho.
frase 3	Queremos discutir o orçamento.
frase 4	Ela tem muita fome.
frase 5	Uma índia andava na mata.
frase 6	Zé, vá mais rápido!
frase 7	Hoje dormirei bem.
frase 8	João deu pouco dinheiro.
frase 9	Ainda são seis horas.
frase 10	Ela saía discretamente.

lista 3

Para esta lista foram escolhidas de forma aleatória, dez frases dentre aquelas disponíveis nas vinte listas do trabalho de Alcaim. As frases que fazem parte desta lista são apresentadas na Tab. 5.4, onde pode ser verificado que aparecem repetidas 2 frases da lista 1 e 2 frases da lista 2. Estas coincidências surgiram pelo fato das listas 3 e 4 terem sido aproveitadas de um trabalho anterior, pois as locuções foram proferidas pelos mesmos locutores das listas 1 e 2.

Tab. 5.4 - Frases da base de dados 2 - lista 3.

frase 1	Leila tem um lindo jardim.
frase 2	Sei que atingiremos o objetivo.
frase 3	Ela tem muita fome.
frase 4	João deu pouco dinheiro.
frase 5	Vi Zé fazer essas viagens seis vezes.
frase 6	Esses são nossos times.
frase 7	Espero te achar bem quando voltar.
frase 8	A paixão dele é a natureza.
frase 9	É hora do homem se humanizar mais.
frase 10	A escuridão da garagem assustou a criança.

lista 4

Para esta lista foi escolhida de forma aleatória a lista de número 11 do trabalho de Alcaim. As frases que fazem parte desta lista são apresentadas na Tab. 5.5.

Tab. 5.5 - Frases da base de dados 2 - lista 4.

frase 1	Um casal de gatos come no telhado.
frase 2	A cantora foi apresentar seu grande sucesso.
frase 3	Lá é um lugar ótimo para tomar uns chopinhos.
frase 4	O musical consumiu sete meses de ensaio.
frase 5	Nosso baile inicia após as nove.
frase 6	Apesar desses resultados, tomarei uma decisão.
frase 7	A verdade não poupa nem as celebridades.
frase 8	As queimadas devem diminuir este ano.
frase 9	O vão entre o trem e a plataforma é muito grande.
frase 10	Infelizmente não compareci ao encontro.

Dez locutores, identificados pelas siglas E01 a E10, participaram das sessões de gravação, todos do sexo masculino e com idade variando entre 20 e 30 anos. Cada um dos locutores gravou uma única repetição de cada uma das frases que são identificadas pelas siglas Fnn, onde nn corresponde ao número da frase.

As listas 3 e 4 foram gravadas em uma única sessão e as listas 1 e 2 foram gravadas em outra sessão 15 dias após a primeira sessão. Foram totalizadas 360 locuções distribuídas conforme descrito a seguir:

- locuções da lista 1: utilizada para o treinamento dos reconhecedores no modo independente do texto;
- locuções das listas 2, 3 e 4: utilizadas para o teste dos reconhecedores no modo independente do texto.

Experimento 1 - reconhecimento dependente do texto

O experimento 1 consistiu em utilizar a base de dados 1 para efetuar a identificação do locutor na forma dependente do texto, utilizando como vetor de entrada dos classificadores a locução inteira de uma única vez.

Para que isto fosse possível foi necessário efetuar a normalização do comprimento das locuções, conforme descrito no capítulo 4 na etapa de extração dos parâmetros. Após a normalização todas as locuções passaram a apresentar 78 segmentos de comprimento e cada segmento apresentando dimensão 8 (correspondente aos 8 MFCCs utilizados para representar cada segmento). Desta forma a dimensão da entrada dos classificadores ficou definida como 624 neurônios.

Reconhecimento com o classificador MLP

A configuração do número de neurônios adotado para cada uma das camadas deste classificador foi o seguinte:

- 624 neurônios na camada de entrada;
- 16 a 512 neurônios foram testados na camada oculta; e
- 16 neurônios na camada de saída, correspondendo a um neurônio para cada classe de locutor.

O resultado para o conjunto de teste da base de dados 1 e, de acordo com o critério adotado verifica-se que a rede identificou 100% dos locutores à ela apresentados. É importante observar que o segundo teste efetuado com o locutor 2 - LOC02B - apresentou um baixo fator de discriminação.

Foi efetuado em seguida, um teste apresentando o conjunto de impostores que faz parte da base de dados 1. O critério adotado para a classificação ou rejeição foi de que o fator de discriminação para aceitar a classificação como válida, deveria ser maior do que o menor fator obtido no teste anterior. Após a apresentação do conjunto de impostores foi verificado que o classificador aceitou 50% dos impostores como pertencentes ao conjunto válido, indicando que o *MLP* não conseguiu produzir um modelo suficientemente robusto para alguns locutores.

Reconhecimento com o classificador *RBF*

Na configuração adotada para as diversas camadas da rede, a camada de entrada e de saída foram configuradas com o mesmo número de neurônios utilizado para o *MLP*. Quanto a camada oculta, foi feita uma nova análise de agrupamentos, utilizando o algoritmo *MKM*, sobre todo o conjunto de treinamento. Foi obtida a separação completa em 16 agrupamentos indicando que existiria uma grande possibilidade de conseguir-se a classificação correta utilizando 16 neurônios na camada oculta. Assim sendo, o número de neurônios adotados para cada camada da rede foi o seguinte:

- 624 neurônios na camada de entrada;
- 16 neurônios na camada oculta (16 funções de base); e
- 16 neurônios na camada de saída, sendo um neurônio para cada classe de locutor.

Neste teste o classificador *RBF* identificou corretamente todos os locutores e apresentou um fator de discriminação superior àquele demonstrado pelo classificador *MLP*.

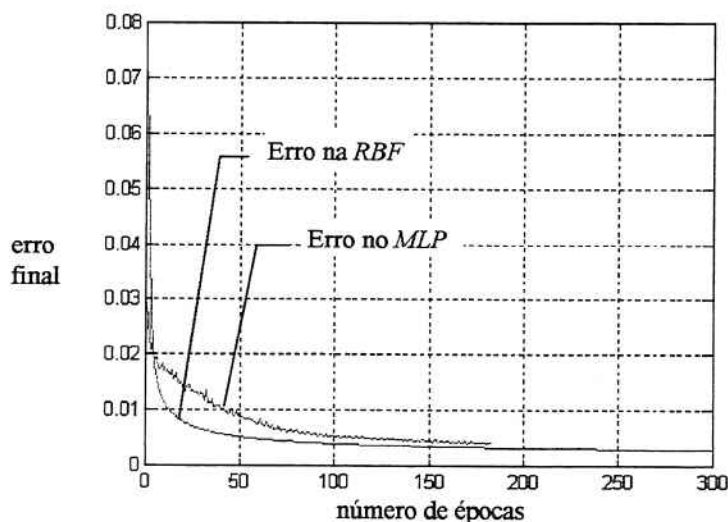


Fig. 5.1 - Comparação do erro durante o treinamento do *MLP* e *RBF*.

Experimento 2 - reconhecimento dependente do texto utilizando MTIs

Neste experimento foram realizados testes utilizando o classificador *RBF* em conjunto com a base de dados 1, tendo sido alterada a estratégia de montagem dos vetores apresentados para o classificador passando a utilizar as *MTIs*. Com isso foi feita uma

avaliação inicial do comportamento das *MTIs* em conjunto com o classificador *RBF* para o reconhecimento do locutor na forma dependente do texto.

O primeiro teste foi realizado considerando a largura da região de ativação das funções de base com mesmo valor para todos os centros, tendo sido utilizadas a *MTI10* e *MTI21* como objeto de comparação. Foram treinadas 21 redes com diversas configurações de neurônios e diferentes valores de largura da região de ativação foram avaliados. Os resultados obtidos podem ser observados na Tab. 5.6.

Tab. 5.6 - Taxa de acerto da *RBF* para *MTI10* e *MTI21* com largura da região de ativação igual para todos os centros.

MTI	NREP	Sigma=0,10	Sigma=0,15	Sigma=0,20	Sigma=0,25
MTI10	008	90,63%	84,37%	28,13%	-
	016	93,75%	87,50%	43,75%	-
	032	100,0%	93,75%	56,25%	-
MTI21	008	93,75%	96,87%	87,50%	53,12%
	016	100,0%	100,0%	96,87%	68,75
	032	100,0%	100,0%	100,0%	90,62%

MTI = tipo de MTI utilizado

NREP = indica o número de representantes por locutor

Sigma = indica o raio ou abertura das *Gaussianas*

De imediato pode ser observado que a *MTI21* proporciona uma melhoria na taxa de reconhecimento para uma rede com mesmo número de representantes por locutor. Nota-se também que a variação no valor da abertura das funções *Gaussianas* (sigma), proporciona uma variação na taxa de reconhecimento. No caso da *MTI10*, o valor sigma igual a 0,10 proporciona as melhores taxas de acerto, e para a *MTI21* o valor de sigma igual a 0,15 apresenta-se com o melhor resultado.

Foi buscada então uma forma de individualizar a largura da região de ativação para cada uma das classes existentes. A alternativa examinada foi utilizar como parâmetro algum valor relacionado com a medida de distância entre o conjunto de referência de uma classe e todos os demais conjuntos. Foram calculados os valores de distância mínima, distância média e variância das distâncias de cada classe para as demais, individualizando por classe a largura da região de ativação das funções de base. Foram realizados alguns testes preliminares que levaram a abandonar a distância média, tendo sido mantidos a distância mínima e a variância das distâncias de uma classe para todas as demais. Os testes finais foram realizados para um conjunto mais amplo de *MTIs* e 29 redes foram treinadas. Os resultados obtidos nos testes encontram-se na Tab. 5.7

Observando os testes que utilizaram a distância mínima, pode ser notado que as *MTI3x* melhoram a taxa de reconhecimento. Nos testes utilizando a variância das distâncias, na medida em que a taxa de 100% de acerto foi obtida em praticamente todos os testes é necessário utilizar outro parâmetro para avaliação do reconhecedor.

Foi calculado então, o fator de discriminação para os resultados obtidos com a variância das distâncias, sendo obtida a Tab. 5.8.

Tab. 5.7 - Taxa de acerto da *RBF*, largura da região de ativação individual por classe.

MTI	NREP	Sigma=Dmin	Sigma=Var
MTI10	008	75,00%	40,62%
	016		81,25%
	032		100,0%
MTI20	008	93,75%	100,0%
	016	84,44%	100,0%
	032	100,0%	100,0%
MTI21	008		100,0%
	016		100,0%
	032		100,0%
MTI30	008	93,75%	100,0%
	016	93,75%	100,0%
	032		
MTI31	008	96,87%	100,0%
	016	100,0%	100,0%
	032		
MTI32	008	100,0%	100,0%
	016	90,63%	100,0%
	032	100,0%	100,0%
	064	71,87%	100,0%

MTI = tipo de MTI utilizado

NREP = indica o número de representantes por locutor

Sigma = indica o raio ou abertura das *Gaussianas*

Tab. 5.8 - Fator de discriminação para a *RBF* com largura da região de ativação individual para cada classe.

MTI	NREP	Discrimin. média	Discrim. média (%)
MTI10	032	28,33	58,58
MTI20	008	10,98	54,16
	016	14,02	62,44
	032	13,87	63,94
MTI21	008	22,20	51,60
	016	27,98	61,38
	032	30,94	65,23
MTI30	008	6,83	50,69
	016	9,01	60,40
	032		
MTI31	008	11,50	53,45
	016	14,41	62,23
	032		
MTI32	008	35,36	51,74
	016	41,40	59,06
	032	33,34	65,73
	064	25,70	59,84

MTI = tipo de MTI utilizado

NREP = indica o número de representantes por locutor

Discrim. = indica o fator de discriminação

Experimento 3 - reconhecimento independente do texto com MTIs

Neste experimento foi utilizada a base de dados 2, para verificar o desempenho do classificador *RBF* na tarefa de reconhecimento do locutor na forma independente do texto. Foi feita uma pesquisa sobre a influência das *MTIs* na taxa de reconhecimento, de forma e

após determinar o melhor desempenho sobre algumas *MTIs*, foram implementadas melhorias nos classificadores e uma comparação direta foi efetuada com o *MLP*.

teste de verificação do comportamento das *MTIs*

Este primeiro teste foi efetuado para verificar se as *MTIs* influenciam de alguma maneira o índice de reconhecimento do classificador *RBF*. Para o treinamento foi utilizada a lista 1 da base de dados 2 e para o teste foi utilizada a lista 2. No total foram treinadas e testadas 122 redes nas diversas configurações de *MTI*, variando o tamanho dos segmentos que compõem a *MTI* e o número de centros (neurônios na camada oculta). A largura da região de ativação dos centros foi adotada sempre em função das variâncias das distâncias entre os centros, sem efetuar-se nenhuma otimização e o treinamento foi limitado a 30 épocas, pois após este número de épocas o erro médio quadrático não apresentava alterações significativas.

Considerando os melhores resultados obtidos na Tab. 5.8 para NREP008, foram montados os gráficos da Fig. 5.2, onde a influência das *MTIs* na taxa de acerto pode ser observada de forma mais clara.

Por meio da Fig. 5.2, pode ser observado que as *MTIs* apresentam melhor desempenho para as janelas de 256 e 512 pontos, e que a *MTI41* (janela com 256 pontos) e a *MTI32* (janela com 512 pontos) proporcionaram as melhores taxas de reconhecimento.

teste comparativo *RBF* x *MLP*

A partir do primeiro teste, foram escolhidas as *MTIs* que apresentaram as maiores taxas de acerto e foi feita uma comparação direta *RBF/MLP* introduzindo algumas melhorias nas redes visando aumentar a taxa de reconhecimento. As *MTIs* escolhidas para este teste foram:

- janela W0256: *MTI10* e *MTI41*; e
- janela W0512: *MTI10* e *MTI32*.

A utilização das *MTI10* é justificada pelo fato da mesma equivaler a não usar *MTI* alguma, servindo portanto como referência para verificar se a utilização das *MTIs* provoca alguma melhoria na taxa de reconhecimento. A *MTI31* foi incluída nos testes como forma de ampliar a avaliação.

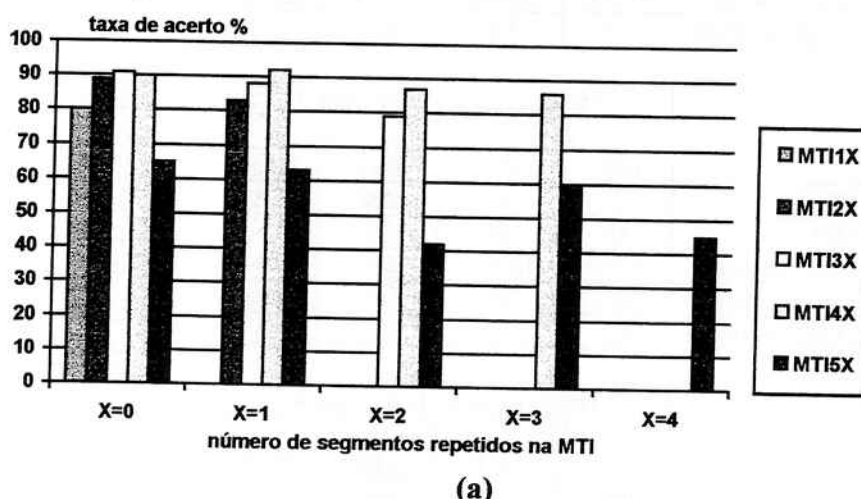


Fig 5.2 - Taxa de acerto para janela de 256 pontos.

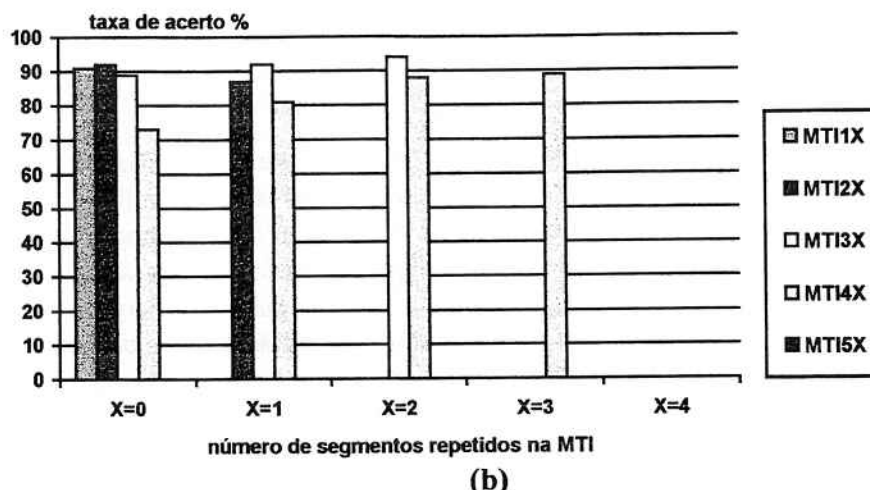


Fig. 5.2 - Taxa de acerto para janela de 512 pontos.

Tab. 5.9 - Melhores taxas de acerto de cada teste (experimento 3).

	Treino Lista 1 / Teste Lista 2			
	256		512	
	RBF	MLP	RBF	MLP
NREP	008	064	008	008
MTI	10	10	10	10
Pesos treinados	800	16.640	800	2.080
Taxa de acerto	96%	100%	95%	96%
Tempo de treino	0.4	4.5	0.3	0.6

	Treino Lista 2 / Teste Lista 1			
	256		512	
	RBF	MLP	RBF	MLP
NREP	008	008	008	008
MTI	41	31	32	10
Pesos treinados	800	4.640	800	2.080
Taxa de acerto	93,3%	100%	95%	100%
Tempo de treino	0,4	1,0	0,3	0,6

	Treino Lista 2 / Teste Lista 3			
	256		512	
	RBF	MLP	RBF	MLP
NREP	016	064	008	008
MTI	10	10	32	10
Pesos treinados	1.600	16.640	800	2.080
Taxa de acerto	64%	74%	64%	66%
Tempo de treino	1,2	4,5	0,3	0,6

	Treino Lista 2 / Teste Lista 4			
	256		512	
	RBF	MLP	RBF	MLP
NREP	008	064	008	008
MTI	41	10	32	10
Pesos treinados	800	16.640	800	2.080
Taxa de acerto	71%	75%	70%	66%
Tempo de treino	0,4	4,5	0,3	0,6

Tab. 5.10 - Comparação dos melhores resultados da *RBF* com *MLP* de complexidade equivalente (experimento 3).

	Treino Lista 1 / Teste Lista 2			
	256		512	
	RBF	MLP	RBF	MLP
NREP	008	008	008	008
MTI	10	10	10	10
Pesos treinados	800	2.080	800	2.080
Taxa de acerto	96%	96%	95%	96%
Tempo de treino	0,4	0,6	0,3	0,6

	Treino Lista 2 / Teste Lista 1			
	256		512	
	RBF	MLP	RBF	MLP
NREP	008	008	008	008
MTI	41	41	32	32
Pesos treinados	800	2.080	800	2.080
Taxa de acerto	93,3%	98,3%	95%	98,3%
Tempo de treino	0,4	1,5	0,3	1,2

	Treino Lista 2 / Teste Lista 3			
	256		512	
	RBF	MLP	RBF	MLP
NREP	016	016	008	008
MTI	10	10	32	32
Pesos treinados	1.600	4.160	800	2.080
Taxa de acerto	64%	67%	64%	65%
Tempo de treino	1,2	4,5	0,3	1,2

	Treino Lista 2 / Teste Lista 4			
	256		512	
	RBF	MLP	RBF	MLP
NREP	008	008	008	008
MTI	41	41	32	32
Pesos treinados	800	2.080	800	2.080
Taxa de acerto	71%	69%	70%	66%
Tempo de treino	0,4	1,5	0,3	1,2

6 Conclusão

Com relação ao reconhecimento dependente do texto, ao montar o banco de dados foram incluídos propositalmente locutores que do ponto de vista perceptual, apresentassem vozes bastante parecidas, tentando com isso dificultar a tarefa de reconhecimento. O sucesso obtido no reconhecimento dependente do texto ratificou a importância dos sons nasalizados aliada a robustez dos MFCCs utilizados como parâmetros para montagem das amostras submetidas à rede. As duas redes treinadas executaram o reconhecimento de forma adequada obtendo-se 100% para a taxa de acerto. Foi observado porém que ao efetuar-se o teste com impostores apenas a *RBF* comportou-se de forma adequada, rejeitando-os na sua totalidade. Este fato pode ser atribuído à característica de localidade do mapeamento efetuado pela *RBF* que “aprende” apenas o espaço à ela mostrado apresentando resposta nula para amostras que estejam localizados fora da região de ativação. Por outro lado o *MLP* além de “aprender” o espaço de

amostras desejado, acaba produzindo respostas inadequadas para amostras que estejam fora do espaço desejado.

Nos resultados obtidos com os experimentos na forma dependente do texto pode ser verificado que na primeira avaliação efetuada com a *RBF* (vide Tab. 5.13) a utilização das *MTIs*, no caso a *MTI21*, causou um aumento na taxa de reconhecimento para classificadores de mesma dimensão. Porém como no teste com a *MTI10*, que equivale ao segmento original sem *MTI* foi obtida uma taxa de acerto de 100%, a utilização das *MTIs* não é justificável. Na Tab. 5.15 tem-se outro indício de que as *MTIs* devem melhorar o reconhecimento, desta feita foi observado um aumento no fator de discriminação para a *MTI21* e *MTI32* indicando que apesar de todas as redes terem atingido 100% de taxa de acerto (vide Tab 5.14) as redes que utilizam a *MTI21* e *MTI32* separam os locutores de forma mais destacada.

Após os estudos e experimentos realizados sobre a aplicação das *RBFs* ao reconhecimento do locutor e os experimentos realizados com este novo conceito que são as *MTIs*, podem ser oferecidas as conclusões citadas a seguir.

- As *RBFs* constituem um paradigma neural que pode ser aplicado com sucesso para o RAL;
- Os fatores cruciais para o desempenho das *RBFs* são o posicionamento adequado dos centros das funções de base e definição adequada da extensão da região de ativação destes centros;
- Do ponto de vista de desempenho, as *RBFs* são comparáveis ao *MLP* demonstrando uma pequena desvantagem na taxa de reconhecimento, apresentando porém a vantagem de um menor tempo de treinamento e complexidade da rede final em termos de número de pesos que devem ser treinados;
- As *MTIs* apresentam-se promissoras como nova forma de representação para capturar características dos locutores e auxiliar na sua identificação;
- Foi verificado que a utilização de *RBFs* em conjunto com as *MTIs* produz desempenho comparável ao *MLP*.

São colocadas a seguir algumas sugestões para o prosseguimento dos trabalhos.

- Testar outras alternativas para o posicionamento dos centros, tais como o algoritmo MKM.
- Com relação a *RBF* devem ser investigados métodos de treinamento que englobem o posicionamento dos centros e dimensionamento das regiões de ativação de forma adaptativa durante o treinamento.
- A investigação da robustez do reconhecedor com *RBF* ao ruído, na medida em que sistemas reais estão sujeitos a estes efeitos.
- Quanto as *MTIs* este conceito deve ser aplicado a outros problemas onde o conjunto de informações contidas no treinamento seja pobre, de forma a se avaliar melhor os efeitos das *MTIs*.

Testar as *MTIs* em casos forenses reais, os quais sempre apresentam grandes diferenças entre o conjunto de treino e o possível teste.

Bibliografia

-
- 1 - RABINER, L.R. Applications of voice processing to telecommunications.
Proceedings of the IEEE, v.82, n.2, p.197-228, Feb. 1994.

- 2 - ROSENBERG, A.E. Automatic speaker verification: a review. **Proceedings of the IEEE**, v.64, n.4, p.475-86, Apr. 1976.
- 3 - ATAL, B.S. Automatic recognition of speakers from their voices. **Proceedings of the IEEE**, v.64, n.4, p.460-75, Apr. 1976.
- 4 - CORSI, P. Speaker recognition: a survey. In: PROCEEDINGS OF THE NATO ADVANCED STUDY INSTITUTE, Bonas, 1981. **Automatic Speech**
- 5 - GISH, H.; SCHMIDT, M. Text-independent speaker identification. **IEEE Signal Processing Magazine**, v.11, n.4, p.18-32, Oct. 1994.
- 6 - RABINER, L.R.; SCHAFER, R.W. **Digital processing of speech signals**. Englewood Cliffs, Prentice Hall, 1978.
- 7 - DELLER JR., J.R. et al. **Discrete-time processing of speech signals**. New York, MacMillan, 1993.
- 8 - PICONE, J.W. Signal modeling techniques in speech recognition. **Proceedings of the IEEE**, v.81, n.9, p.1215-47, Sept. 1993.
- 10 - RUNSTEIN, F.; VIOLARO, F.; NUNES, H.F. Uso de diferentes parâmetros de entrada em u sistema de reconhecimento de fala baseado em redes neurais. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES, 13°, Aguas de Lindóia, 1995. **Anais**. Campinas, Unicamp, 1995, p.155-60
- 11 - REYNOLDS, D.A.; ROSE, R.A. Robust text-independent speaker identification using gaussian mixture speaker models. **IEEE Transactions on Speech and Audio Processing**, v.3, n.1, p.72-83, Jan.1995.
- 12 - FU, L. **Neural networks in computer intelligence**. New York, MacGraw Hill, 1994.
- 13 - HUSH, D.R.; HORNE, B.G. Progress in supervised neural networks: what's new since Lippmann? **IEEE Signal Processing Magazine**, v.10, n.1, p.8-39, Jan. 1993.
- 14 - CABRAL JR., E.F. **Reconhecimento automático do locutor com redes neurais artificiais: um curso teórico/prático para engenheiros e cientistas**. A ser publicado, 1998.
- 15 - COVER, T.M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. **IEEE Transactions on Electronic Computers**, v.14, n.3, p.326-334, June 1965.
- 16 - BROOMHEAD, D.S.; LOWE, D. Multi-variable function interpolation and adaptive networks. **Complex Systems**, v.2, p.269-303, 1988.
- 17 - LIPPMANN, R.P. An introduction to computing with neural nets. **IEEE ASSP Magazine**, v.4, n.2, p.4-22, Apr. 1987.
- 18 - MINSKY, M.L.; PAPERT, S.A. **Perceptrons**. Cambridge, MIT Press, 1969.
- 19 - RUMELHART, D.E.; MCCLELLAND, J.L. **Parallel distributed processing**. San Diego, MIT Press, 1986.
- 20 - HAYKIN, S. **Neural networks: a comprehensive foundation**. New York, MacMillan, 1994.
- 21 - WILPON, J.G.; RABINER, L.R. A modified K-means clustering algorithm for use in isolated word recognition. **IEEE Transactions on Acoustics, Speech and Signal Processing**, v.33, n.3, p.587-94, June 1985.
- 22 - ALCAIM, A.; SOLEWICZ, J.A.; MORAES, J.A. Frequência de ocorrência dos fones e listas de frases foneticamente balanceadas no português fala no Rio de Janeiro. **Revista da Sociedade Brasileira de Telecomunicações**, v.7, n.1, p.23-41 dez. 1992.

BOLETINS TÉCNICOS - TEXTOS PUBLICADOS

- BT/PEE/9301 - Oscilador a HEMT - 10 GHz - FÁTIMA S. CORRERA, EDMAR CAMARGO
- T/PEE/9302 - Representação Senoidal da Voz através dos Polos do Filtro Preditor - MARCELO B. JOAQUIM, NORMONDS ALENS
- BT/PEE/9303 - Blindagens por Grades Conductoras: Cálculo do Campo Próximo - LUIZ CEZAR TRINTINALIA, ANTONIO ROBERTO PANICALI
- BT/PEE/9304 - Sistema de Otimização e Controle de Produção em Minas de Pequeno e Médio Porte - TSEN CHUNG KANG, VITOR MARQUES PINTO LEITE
- BT/PEE/9401 - Determinação das Frases de Aplicação Forense para o projeto NESPER e Tese de Mestrado IME/94, com Base em Estudos Fonéticos - MARCONI DOS REIS BEZERRA, EUVALDO F. CABRAL JUNIOR
- BT/PEE/9402 - Implementação e Teste de uma Rede Neural Artificial do Tipo KSON (Kohonen Self-Organizing Network) com Entradas Bidimensionais - MARCELO YASSUNORI MATUDA, EUVALDO F. CABRAL JR.
- BT/PEE/9403 - Transformada de Walsh e Haar Aplicadas no Processamento de Voz - ALEXANDRE AUGUSTO OTTATI NOGUEIRA, THIAGO ANTONIO GRANDI DE TOLOSA, EUVALDO F. CABRAL JÚNIOR
- BT/PEE/9404 - Aplicação de Redes Neurais ao Problema de Reconhecimento de Padrões por um Sonar Ativo - ALEXANDRE RIBEIRO MORRONE, CRISTINA COELHO DE ABREU, EDUARDO KOITI KIUKAWA, EUVALDO F. CABRAL JR.
- BT/PEE/9405 - Tudo que se Precisa Saber sobre a Prática da FFT - Transformada Rápida de Fourier (Inclui Software) - ROGÉRIO CASAGRANDE, EUVALDO F. CABRAL JR.
- BT/PEE/9406 - A Survey on Speech Enhancement Techniques of Interest to Speaker Recognition - CELSO S. KURASHIMA, EUVALDO F. CABRAL JR.
- BT/PEE/9407 - Identificação de Pulsos Decádicos em Linhas Telefônicas - ANTONIO P. TIMOSZCZUK, MÁRCIO A. MATHIAS, EUVALDO F. CABRAL JR.
- BT/PEE/9408 - Implementação e Teste de Filtros do Tipo Adaptativo e ©Notch^a para a Remoção de Interferência de 60 Hz em Sinais de Eletrocardiograma - FLÁVIO ANTÔNIO MENEGOLA, JOSÉ AUGUSTO DE MATTOS, JOSÉ GOMES G. FILHO, SIDNEY SILVA VIANA, EUVALDO F. CABRAL JR.
- BT/PEE/9409 - Compressão de Sinais de Voz utilizando Transformadas de Karhunen-Loève, Fourier e Hadamard - IVAN LUIS VIEIRA, LUIZ FERNANDO STEIN WETZEL, EUVALDO F. CABRAL JR.
- BT/PEE/9410 - ©Ray Tracing^a Paralelo - EDUARDO TOLEDO SANTOS, JOÃO ANTONIO ZUFFO
- BT/PEE/9411 - Implementação de uma Ferramenta Posicionador para ©Gate-Arrays^a Tipo Mar de Portas - JORGE W. PERLAZA PRADO, WILHELMUS A. M. VAN NOIJE
- BT/PEE/9412 - Tudo que se Precisa Saber Sobre a Teoria da FFT - Transformada Rápida de Fourier - FÁBIO LUÍS ROMÃO, REINALDO SILVEIRA, ROGÉRIO CASAGRANDE, EUVALDO CABRAL JR.
- BT/PEE/9413 - Análise do Ruído Sonoro em uma Sala de Aquisição de Amostras de Som com Microcomputador - FÁBIO LUÍS ROMÃO, REINALDO SILVEIRA, EUVALDO CABRAL JR.
- BT/PEE/9414 - Cor: Aspectos Relevantes para Visualização de Dados - SÍLVIA DELGADO OLABARRIAGA
- BT/PEE/9415 - Projeto de Filtros Digitais IIR com Fase Aproximadamente Linear Utilizando Redução de Ordem - IVAN F. J. RODRIGUES, MAX GERKEN
- BT/PEE/9416 - GERAFILTRO: Sistema para Projeto Automático de Filtros Digitais ©IIR^a (da especificação em alto nível ao leiaute do ©ASIC^a) - RICARDO PIRES, JOSÉ VIEIRA DO VALE NETO
- BT/PEE/9417 - Redes Neurais Artificiais Aplicadas à Identificação de Pulsos Decádicos em Linhas Telefônicas - ANTONIO P. TIMOSZCZUK, EUVALDO F. CABRAL JR.
- BT/PEE/9501 - Estudo Comparativo de Métodos de Cálculo da Frequência Fundamental - MARCOS COSTA HUNOLD, EUVALDO F. CABRAL JR.
- BT/PEE/9502 - Combinando Técnicas de Redes Neurais Artificiais e Informações de Excitação no Reconhecimento Automático do Locutor - ANDRÉ BORDIN MAGNI, EUVALDO F. CABRAL JR.
- BT/PEE/9503 - Utilização de Redes Neurais Artificiais para Detecção e Identificação de Falhas em Circuitos - MÁRCIO YUKIO TERUYA, ROBERTO AMILTON BERNARDES SÓRIA, EUVALDO CABRAL JR.
- BT/PEE/9504 - Uso de Redes Neurais Artificiais no Reconhecimento de Locutores no Domínio Temporal - BENEDITO JOSÉ BARRETO FONSECA JÚNIOR, EUVALDO CABRAL JÚNIOR
- BT/PEE/9505 - Projeto de Filtros Passivos e Ativos em Técnicas de Circuitos Integrados de Microondas - DAVID VIVEIROS JÚNIOR, DENISE CONSONNI
- BT/PEE/9506 - Uma Análise de Clustering para as Frases de Projeto NESPER - RONALDO OLIVEIRA MESSINA, EUVALDO F. CABRAL JR.
- BT/PEE/9507 - Controle com Estrutura Variável e Modos Deslizantes - Um Estudo para Aplicação em Controle Carga-frequência da Geração - JOSE PAULO F. GARCIA, JOCELYN FREITAS BENNATON

- BT/PEE/9508 - Recuperação das Margens de Ganho e de Fase para Sistemas de Fase Não Mínima por Realimentação da Saída - MARCO H. TERRA, VITOR M. P. LEITE
- BT/PEE/9509 - Sistema de Inspeção Óptica de Dispositivos Bi-Dimensionais - CASIMIRO DE ALMEIDA BARRETO, PEDRO LUÍS PRÓSPERO SANCHEZ
- T/PEE/9510 - Sistema de Partículas Uma Poderosa Técnica de Animação em Computação Gráfica - RENATO CURTO RODRIGUES, JOÃO ANTÔNIO ZUFFO
- BT/PEE/9511- Efeito de Ruídos em Sinais de Voz Visualizados em Trajetórias Neurais de Kohonen - CELSO S. KURASHIMA, EUVALDO F. CABRAL JR.
- BT/PEE/9601 - "Um Reconhecedor de Sinais Sonoros Utilizando LVQ" - ALEXANDRE TORNICE, EUVALDO CABRAL JR.
- BT/PEE/9602 - "Coleção Artificial Neural Networks: Uma Visão Geral dos Sistemas Neurais Artificiais de Stephen Grossberg" - CHIU HSIUNG HUANG
- BT/PEE/9603 - "Reactively-Sputtered TiN Formation Using a RF Magnetron System"- SÉRGIO PAULO AMARAL OSÓRIO, LUIZ SÉRGIO ZASNICOFF
- BT/PEE/9604 - Aspectos em Tradução de Linguagens Naturais Através de Redes Neurais Artificiais - CARLOS EDUARDO DANTAS DE MENEZES, EUVALDO F. CABRAL JR.
- BT/PEE/9605 - Implementação de Blocos Passa-Tudo Utilizando Realimentação de Erro - SÉRGIO JOSÉ CARNEIRO LEÃO, MAX GERKEN
- BT/PEE/9606 - Coleção SANN group Redes Neurais Artificiais: A Rede Neural de Sakoe - ANDRÉ BORDIN MAGNI, EUVALDO F. CABRAL JR.
- BT/PEE/9607 - Coleção SANN group Redes Neurais Artificiais: A Rede Neural de Steinbuch - ROBERTO AMILTON BERNARDES SÓRIA, EUVALDO F. CABRAL JR.
- BT/PEE/9608 - Desenvolvimento de uma Estrutura de Duplo Nível de Metal para a Confeção de Interconexões em Circuitos Integrados - JOSÉ AUGUSTO DE ALENCAR PEREIRA, LUIZ CARLOS MOLINA TORRES
- BT/PEE/9609 - Determinação de Parâmetros de Processo para Fotomáscara "Balzers" Utilizando Gerador de Padrões - JORGE SÉKI, MEGUMI SAITO
- BT/PEE/9610 - Um Ambiente para Desenvolvimento de Sistemas Distribuídos - PEDRO F. ROSA, JOÃO A. ZUFFO
- BT/PEE/9611 - Interpretações Teóricas do Funcionamento Cerebelar: Uma Revisão - MARCUS FRAGA VIEIRA, ANDRÉ FÁBIO KOHN
- BT/PEE/9612 - Marcapasso Cardíaco Temporário Microcontrolado de Demanda e Baixo Consumo - FLAVIO ANTONIO MENEGOLA, JOSÉ CARLOS TEIXEIRA DE BARROS MORAES
- BT/PEE/9613 - Um Sistema de Planejamento de Ação Baseado em Casos para uma Célula Flexível de Manufatura - RICARDO LUÍS DE FREITAS, MÁRCIO RILLO
- BT/PEE/9614 - Aplicações do Boundary-Scan para o Teste de Módulos Multichip - ROBERTO C. COSSI JR., JOSÉ ROBERTO DE A. AMAZONAS
- BT/PEE/9615 - A 2.488 Gb/s GaAs 1:4/1:16 Demultiplexer IC with Skip Circuit for Sonet STS-12/48 Systems - TAUFIK ABRÃO, FATIMA S. CORRERA
- BT/PEE/9616 - Uma Contribuição para a Construção de Algoritmos em Projetos de Redes - ALLAN DE SOUZA, JOSÉ ROBERTO CASTILHO PIQUEIRA
- BT/PEE/9617 - Análise Crítica dos Métodos de Medição do Intervalo QT do Eletrocardiograma - SÍDNEY DA SILVA VIANA, JOSÉ CARLOS TEIXEIRA DE BARROS MORAES
- BT/PEE/9618 - Deposição e Caracterização de Filmes de SiO₂ Crescidos pela Técnica de PECVD a Baixa Temperatura - MARCO ALAYO CHÁVEZ, INÉS PEREYRA
- BT/PEE/9619 - PARSTOOL: Uma Ferramenta de Auxílio à Simulação de Sistemas Paralelos - LI KUAN CHING, LIRIA MATSUMOTO SATO
- BT/PEE/9620 - Análise de um Método de Otimização por Malha no Treinamento de Robôs - OLÍMPIO MURILO CAPELI, JOSÉ CARLOS T. B. MORAES, SADA O ISOTANI
- BT/PEE/9701 - Identification of Unstable Mechanical Systems - ROBERTO MOURA SALES, ANSELMO BITTAR, MICHAEL PORSCH, LAÉRCIO LUCCHESI
- BT/PEE/9702 - Analysis of the Subthreshold Slope Transition Region in SOI nMOSFET - VICTOR SONNENBERG, JOÃO ANTONIO MARTINO
- BT/PEE/9703 - Introduction of the SOI MOSFET Dimensions in the High-Temperature Leakage Drain Current Model - MARCELO BELLODI, JOÃO ANTONIO MARTINO, DENIS FLANDRE
- BT/PEE/9704 - Controle de Largura de Banda Dinâmica para Transmissões Multicast para Redes de Alta Velocidade - SANG SOON LEE, SERGIO TAKEO KOFUJI
- BT/PEE/9705 - Uma Modificação Proposta para o Controle Preditivo Generalizado com Filtro de Kalman - JAIME QUINTERO R., OSWALDO L. V. COSTA
- BT/PEE/9706 - Aplicações de Redes Neurais em Previsões Financeiras - OLÍMPIO MURILO CAPELI, EUVALDO F. CABRAL JR.

- BT/PEE/9707 - Sistema Microcontrolado, Multicanal e Portátil para Estimulação Neuromuscular Funcional - ROGÉRIO QUIARIM ZARZA, JOSÉ CARLOS TEIXEIRA DE BARROS MORAES
- BT/PEE/9708 - Requisitos para o Mapeamento Tecnológico em Projetos de Microeletrônica - LUCIANO DE OLIVEIRA CORRÊA DE BRITO, JOSÉ ROBERTO DE ALMEIDA AMAZONAS
- BT/PEE/9709 - Sistemas PRMA com Dados Acoplados - JOSÉ AUGUSTO DE LIMA, PAUL JEAN ETIENNE JESZENSKY
- BT/PEE/9710 - Algoritmos Genéticos (AG's) para a Otimização de Controladores Nebulosos - JULIO CESAR CEBALLOS AYA, OSWALDO L. V. COSTA
- BT/PEE/9711 - Um Estudo Sobre a Redução de Ruídos em Sinais Caóticos - ERNANE JOSÉ XAVIER COSTA, EUVALDO F. CABRAL JR.
- BT/PEE/9712 - Geradores não Lineares de Sequência para uso em Sistemas Spread Spectrum - ANGEL ANTONIO GONZALEZ MARTINEZ, PAUL JEAN ETIENNE JESZENSZY
- BT/PEE/9713 - Modelamento Físico do Sistema Heteroestrutura - Metal - CECÍLIA WETTERLE RODRIGUES, MEGUMI SAITO
- BT/PEE/9714 - Tensões Induzidas em Linhas Aéreas por Descargas Atmosféricas Indiretas - Modelagem e Aplicação ao Cálculo de Interrupções - ALEXANDRE PIANTINI, JORGE M. JANISZEWSKI
- BT/PEE/9715 - RECMAP - Uma Ferramenta para Otimização em Síntese de Alto Nível Baseada em Reconhecimento Funcional e Mapeamento de Componentes - ANDRÉ GERHARD, JOSÉ VIEIRA DO VALE NETO
- BT/PEE/9716 - Estudo da Sinterização de Contatos Al/Ti por Recozimento Térmico Rápido Visando a Aplicação em Circuitos Integrados - ANGELO EDUARDO BATTISTINI MARQUES, ROGÉRIO FURLAN
- BT/PEE/9717 - Mixed H2/H- Control of Discrete-Time Markovian Jump Linear Systems - OSWALDO L. V. COSTA, RICARDO P. MARQUES
- BT/PEE/9718 - Aluminium Etching with CCl4-N2 Plasmas - ANGELA MAKIE MAKAZAWA, PATRICK VERDONCK
- BT/PEE/9719 - O Uso de Resistes Amplificados Quimicamente e de Sililação em Litrografia por Feixe de Elétrons - ANTONIO C. SEABRA, PATRICK B. VERDONCK
- BT/PEE/9720 - Implementação de um Simulador de um Circuito Neuro-Medular que Atua no Controle da Força Motora - LUIZ JURANDIR SIMÕES DE ARAÚJO, ANDRÉ FÁBIO KOHN
- BT/PEE/9721 - Avaliação das Características Físico-Químicas e Elétricas de Filmes de SiO2 Depositados por PECVD a Partir da Reação entre O2 e TEOS - ALVARO ROMANELLI CARDOSO E CLAUS MARTIN HASENACK
- BT/PEE/9722 - Controle e Simulação Dinâmica de Colunas de Destilação: Aplicação Prática em uma Coluna com Refluxo por Gravidade
- BT/PEE/9723 - Circuitos de Portas Lógicas Primitivas Implementados a Partir de uma Classe de Lógicas Paraconsistentes Anotadas - JOÃO INÁCIO DA SILVA FILHO, JAIR MINORO ABE, PEDRO LUÍS PRÓSPERO SANCHEZ
- BT/PEE/9724 - Lattice Heating and Energy Balance Consideration on the I-V Characteristics of Submicrometer Thin-Film Fully Depleted SOI NMOS Devices - CLAUDIA BRUNETTI, NELSON L. A. BRAGA, LUIZ S. ZASNICOFF
- BT/PEE/9725 - Identificação de um Processo de Neutralização de pH via Redes Neurais - SILVIO FLABOREA, CLAUDIO GARCIA
- BT/PEE/9726 - Uma Estratégia de Migração de Sistemas de Telefonia Móvel com Tecnologia AMPS para a Tecnologia CDMA - RONALD LUÍS CLARKSON EISNER, PAUL JEAN ETIENNE JESZENSKY
- BT/PEE/9727 - Controle de pH Usando Conceitos de Invariantes de Reações e Geometria Diferencial - OSCAR A. ZANABARIA S., CLAUDIO GARCIA
- BT/PEE/9728 - Estudo da Influência dos Parâmetros de Recozimento Térmico Rápido na Morfologia dos Filmes de TiSi2 Formados e sua Correlação com a Tensão Mecânica - SILVANA GASPAROTTO DE SOUZA, ARMANDO ANTONIO MARIA LAGANÁ, SEBASTIÃO GOMES DOS SANTOS FILHO
- BT/PEE/9729 - Analysis of Silicon Surface Microirregularities by LASER Light Scattering - JOSÉ CÂNDIDO DE SOUSA FILHO, SEBASTIÃO GOMES DOS SANTOS FILHO
- BT/PEE/9730 - Wavelets in Music Analysis and Synthesis: Timbres Analysis and Perspectives - REGIS ROSSI ALVES FARIA, RUGGERO ANDREA RUSCHIONI, JOÃO ANTONIO ZUFFO
- BT/PEE/9731 - Estudo de Efeitos Mútuos da Distribuição de Corrente em Condutores - AUGUSTO CARLOS PAVÃO, JORGE MIECZYSLAW JANISZEWSKI
- BT/PEE/9801 - Equivalência entre a Semântica da Lógica de Transações e a Semântica de sua Implementação Prolog - PAULO E. SANTOS, FLÁVIO S. C. DA SILVA
- BT/PEE/9802 - Nash Game in Mixed H2/Hoo Control Theory a Convex Optimization Approach- HELENICE OLIVEIRA FLORENTINO, ROBERTO MOURA SALES
- BT/PEE/9803 - Text-Independent Speaker Recognition Using Vector Quantization and Gaussian Mixture Models - THOMAS E. FILGUEIRAS F., RONALDO O. MESSINA E EUVALDO F. CABRAL JR.
- BT/PEE/9804 - Elementos Piezoresistivos para Sensores de Pressão com Tecnologia CMOS - LUIZ ANTONIO RASIA, E. C. RODRIGUEZ

- BT/PEE/9805 - Automação do Processo de Casamento de Impedância em Sistemas de Aquecimento por Microondas - J. C. DE SOUZA, J. T. SENISE, V. C. PARRO, F. M. PAIT
- BT/PEE/9806 - Considerações para o Projeto de Células de Memória SI com Transistores HEMT - JAIME H. LASSO, EDGAR CHARRY R.
- BT/PEE/9807 - Covariance Controller with Structure Constraint and Closed Loop H Bound - ANTÔNIO CARLOS DE LIMA, ROBERTO MOURA SALES
- BT/PEE/9808 - Controle Unidimensional de Objetos: Uma Aplicação Prática em Fresadora Automática - MÁRCIO A. F. MURATORE, OSWALDO L. V. COSTA
- BT/PEE/9809 - Redes Neurais com Retardos Temporais Aplicadas ao Reconhecimento Automático do Locutor - ROGÉRIO CASAGRANDE, EUVALDO F. CABRAL JR.
- BT/PEE/9810 - Topological Computation and Voluntary Control - HENRIQUE SCHÜTZER DEL NERO, JOSÉ ROBERTO CASTILHO PIQUEIRA, ALFREDO PORTINARI MARANCA
- BT/PEE/9811 - Casamento de Impedância em Guia de Onda Retangular Utilizando como Variável Medida a Potência Refletida - J. C. DE SOUZA JR., J. T. SENISE
- BT/PEE/9812 - Applied Surface Science - RONALDO D. MANSANO, PATRICK VERDONCK, HOMERO S. MACIEL
- BT/PEE/9813 - Contribuição ao Estudo da Morfologia da Superfície e da Interface do Siliceto de Titânio Formado sobre Si (100) Empregando a Técnica de Microscopia de Força Atômica (AFM) - N. M. HASAN, A. A. M. LAGANÁ, S. G. SANTOS FILHO
- BT/PEE/9814 - Estudo Experimental da Tensão Mecânica em Filmes Finos de Cobre Obtidos por Evaporação ou Deposição Eletroquímica Espontânea - A. I. HASHIMOTO, S. G. FILHO
- BT/PEE/9815 - Controle Híbrido de Manipuladores Robóticos - WEBER ALLEGRI, JOSÉ JAIME DA CRUZ
- BT/PEE/9816 - Entropia Informacional e Cronobiologia - ANA AMÉLIA BENEDITO SILVA, JOSÉ ROBERTO CASTILHO PIQUEIRA
- BT/PEE/9817 - Estabilidade de Lyapunov e Controle de Atitude - FERNANDO SOUSA, FREITAS JÚNIOR, PAULO SÉRGIO PEREIRA DA SILVA
- BT/PEE/9818 - Projeto de um Conversor de Frequências Resistivo em Tecnologia MMIC - CLÁUDIA C. A. APARÍCIO, DENISE CONSONNI
- BT/PEE/9819 - Estudo de Sensibilidade de um Sistema de Modulação Digital via Simulação - RONALDO DI MAURO, LUIZ ANTONIO BACCALÁ
- BT/PEE/9820 - Engenharia do Conhecimento Aplicada ao Conhecimento Aplicada ao Domínio de Gerenciamento de Falhas em Redes de Comunicação: Uma Abordagem Baseada em Modelo - MARILZA ANTUNES DE LEMOS, MARCIO RILLO
- BT/PEE/9821 - Uma Biblioteca de Métodos de Resolução de Problemas de Planejamento para Fornecer Apoio à Aquisição de Conhecimento - LELIANE NUNES DE BARROS, MARCIO RILLO
- BT/PEE/9822 - Um Estudo de Técnicas de Aprendizado por Reforço Livre de Modelo - Aplicação ao Pêndulo Invertido - SÉRGIO RIBEIRO AUGUSTO, ADEMAR FERREIRA
- BT/PEE/9823 - Identificação de Sistemas Dinâmicos com Redes Neurais - FRANCISCO CARLOS PONS, CLÁUDIO GARCIA
- BT/PEE/9824 - Comparison between Single and Double Langmuir Probe Techniques for Analysis of Inductively Coupled Plasmas - RAUL M. DE CASTRO, GIUSEPPE A. CIRINO, PATRICK VERDONCK, HOMERO S. MACIEL, MARCOS MASSI, MARCELO B. PISANI, RONALDO D. MANSANO
- BT/PEE/9825 - DECMEF: Um Sistema de Decomposição Aplicada à Síntese de Máquinas de Estados Finitos - CARLOS HUMBERTO LLANOS QUINTERO, MARIUS STRUM
- BT/PEE/9826 - Controladores Preditivos Através de Desigualdade Matriciais Lineares - RENATO C. BARÃO, RICARDO P. MARQUES, OSWALDO L. V. COSTA
- BT/PEE/9827 - Parametrization of all H2 Optimal Output Feedback Controllers - JOÃO YOSHIYUKI ISHIHARA, ROBERTO MOURA SALES
- BT/PEE/9901 - Digital Equalization Using Time Delay Neural Network (TDNN) - MARIO ANDRÉS VERGARA ESCOBAR, EUVALDO F. CABRAL JR.
- BT/PEE/9902 - Simulação e Análise Estrutural do Modelo de Deposição da Publicação ICRP 66 - JOAQUIM CARLOS SANCHES CARDOSO, JOSÉ CARLOS T. DE BARROS MOARAES
- BT/PEE/9903 - Modelamento e Ajuste por Regra Sub-Ótica de Maximização da Informação da Mútua de um Sistema Reconhecedor de Palavras Isoladas, Independente do Falante, para Ambientes Adversos, com Redundância de Informação - MÁRIO MINAMI, IVANDRO SANCHES
- BT/PEE/9904 - Modelagem Matemática e Controle Multivariável do Tanque de Contato Empregado no Processo de Lixiviação dos Minerais Niquelíferos - DANIEL GUZMÁN DEL RÍO, CLAUDIO GARCIA
- BT/PEE/9905 - Avaliação de Transdutores para Análise Metabólica Humana - HENRIQUE TAKACHI MORIYA, JOSÉ CARLOS T. DE BARROS MORAES
- BT/PEE/9906 - Estudo da Aplicação do Método TLM-2D à Análise de Campos Eletromagnéticos - MARCO ANTONIO MATHIAS, JORGE M. JANISZEWSKI

- BT/PEE/9907 – Comportamento Transitório de Campos Eletromagnéticos em Meios Condutores: Simulação Computacional Eficiente por Diferenças Finitas no Domínio do Tempo – ALEXANDRE AUGUSTO OTTATI NOGUEIRA, JORGE MIECZYSLAW JANISZEWSKI
- BT/PEE/9908 – H₂ and H_∞ Control for Maglev Vehicles – ANSELMO BITTAR, ROBERTO MOURA SALES
- BT/PEE/9909 – Amplificador Distribuído em GaAs – 1 a 17 GHz – CRISTIANE FERREIRA DE ARAÚJO, FATIMA SALETE CORRERA
- BT/PEE/9910 – Modelo Não-Linear de MESFET para Simulação de Amplificadores de Alta Eficiência – ANTONIO SANDRO VERRI, FATIMA SALETE CORRERA
- BT/PEE/9911 – Projeto de um Sistema de Propulsão e Levitação Magnética com Dois Graus de Liberdade – ALEXANDRE BRINCALEPE, FELIPE MIGUEL PAIT
- BT/PEE/9912 – A New Technique to Obtain the MOS Gate Oxide Thickness and Electric Breakdown Field Distributions From Fowler-Nordheim Tunneling Current – WILLIAN AURÉLIO NOGUEIRA, SEBASTIÃO GOMES DOS SANTOS FILHO
- BT/PEE/9913 – Polarization Effects on the Raman and Photoluminescence Spectra of Porous Silicon Layers – WALTER JAIMES SALCEDO, FRANCISCO JAVIER RAMIREZ FERNANDEZ
- BT/PEE/9914 – Aproximação Gaussiana Melhorada Aplicada na Análise de um Método de Aquisição em Dois Estágios para Sistemas DS/CDMA – IVAN ROBERTO SANTANA CASELLA, PAUL JEAN ETIENNE JESZENSKY
- BT/PEE/9915 – Simulação e Análise de Soluções de Comunicação entre Sub-Redes IP sobre ATM – MARCELO ZANONI SANTOS, JOÃO ANTONIO ZUFFO
- BT/PEE/9916 – Uma Nova Abordagem para a Análise Computacional de Movimento – RAMONA M. STRAUBE, JOÃO ANTONIO ZUFFO
- BT/PEE/9917 – Caches Remotos e Prefetching em Sistemas Multiprocessadores de Alto Desempenho – Considerações Arquiteturais – EDWARD DAVID MORENO, SERGIO TAKEO KOFUJI
- BT/PEE/9918 – Um Modelo de Referência para o Controle do Processo de Lodo Ativado – OSCAR A. ZANABRIA SOTOMAYOR, SONG WON PARK, CLAUDIO GARCIA
- BT/PEE/9919 – Identificação Não-Linear de um Processo de Neutralização de pH Multivariável Utilizando Modelos Narmax Polinomiais com Tempo Morto – ROSIMEIRE APARECIDA JERÔNIMO, CLAUDIO GARCIA
- BT/PEE/9920 – Avaliação do U-Net em Clusters com Rede Myrinet – PAULO A. GEROMEL, SERGIO T. KOFUJI
- BT/PEE/9921 – Implementação de Finos Diafragmas em Lâminas de Silício Monocristalino, Altamente Dopadas com Boro, Visando Fabricação de Microsensores de Pressão CMOS Utilizando Pós-Processamento – HUMBER FURLAN, EDGAR R. CHARRY
- BT/PEE/9922 – A5GHz Continuous Time Sigma-Delta Modulador Implemented in 0.4μm InGaP/InGaAs – A. OLMOS, E. CHERRY, M. NIHEI, Y. WATANABE
- BT/PEE/9923 – Dimensionamento de Unidades Remotas de Sistemas VSAT – CLAUDIO TRÁPAGA FAGUNDES DO NASCIMENTO FILHO, FATIMA SALETE CARRERA
- BT/PEE/9924 – Implementação de um Pós-Filtro Adaptativo para a Melhora de Qualidade Perceptual de Sinais de Voz com Ruído – CELSO SETSUO KURASHIMA, IVANDRO SANCHES
- BT/PEE/9925 – Redes Neurais Artificiais Aplicadas a Estimação de Processos Biotecnológicos – PEDRO S. PEREIRALIMA, ZSOLT L. KOVÁCS
- BT/PEE/9926 – The Utilization of Low Temperature Co-Fired Ceramics (LTCC-ML) Technology for Meso-Scale SEM, a Simple Thermistor Based Flow Sensor – M. GONGORA-RUBIO, L. M. SOLÁ-LAGUNA, P. J. MOFFETT, J. J. SANTIAGO-AVILÉS
- BT/PEE/9927 – Análise do Desempenho de uma Rede ATM sob Tráfego Heterogêneo – HANS MARCELO GAMARRA DELGADO, JOSÉ ROBERTO DE ALMEIDA AMAZONAS
- BT/PEE/9928 – Ambiente de Síntese de Circuitos CMOS de Alto Desempenho – FABIO LUÍS ROMÃO, WILHELMUS A. M. VAN NOIJE
- BT/PEE/9929 – Polarization Effect on Raman Scattering and Photoluminescence Spectra in Porous Silicon Layer – WALTER JAIMES SALCEDO, FRANCISCO JAVIER RAMIREZ FERNANDEZ
- BT/PEE/9930 – Photonic Band Structure of Periodic-Like Porous Silicon – WALTER JAIMES SALCEDO, FRANCISCO JAVIER RAMIREZ FERNANDEZ
- BT/PEE/9931 – Investigação da Aplicação de Redes Neurais Artificiais à Solução de Problemas Inversos em Eletromagnetismo – THIAGO ANTONIO GRANDI DE TOLOSA, JORGE MIECZYSLAW JANISZEWSKI
- BT/PEE/9932 – Células Solares Fotovoltaicas de Silício Mono e Multicristalino com Camadas Antirefletoras Simples (SiN:H) e Duplas (SiN:H/SiO₂ fino) Depositadas por PECVD – KÁTIA APARECIDA ABADE, FERNANDO JOSEPETTI FONSECA
- BT/PEE/9933 – Controle de Posição de Manipuladores Pré-Compensação com Filtro de Kalman – RENATO VENTURA BAYAN HENRIQUES, JOSÉ JAIME DA CRUZ
- BT/PEE/9934 – Controle Preditivo Adaptativo Não-Linear Aplicado a um Processo de Neutralização de pH – JOSÉ MARÍA VARGAS LARA, CLAUDIO GARCIA

BT/PEE/9935 – Aperfeiçoamento de um Sistema de Medição de Pressão Arterial por Método Oscilométrico – MARCELO
CERULLI, JOSÉ CARLOS TEIXEIRA DE BARROS MORAES

