Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação
Departamento de Comunicações

Reconhecimento Automático de Fala Contínua Empregando Modelos Híbridos ANN + HMM

por

Edmilson S. Morais
Eng. Eletricista (Universidade de Brasília) 1995

Orientador : Prof. Dr. Fábio Violaro
DECOM - FEEC - UNICAMP

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da UNICAMP como requisito parcial para a obtenção do título de Mestre em Engenharia Elétrica.

Campinas, dezembro de 1997

Aos meus pais Espedito e Dirce.

Reconhecimento Automático de Fala Contínua

Empregando

Modelos Híbridos ANN + HMM

por

Edmilson S. Morais

Dissertação apresentada à Faculdade de
Engenharia Elétrica e de Computação da UNICAMP
em Dezembro de 1997, como requisito parcial para a obtenção do título de
Mestre em Engenharia Elétrica

Resumo

Atualmente, os sistemas que representam o estado-da-arte em reconhecimento de fala contínua baseiam-se em modelos ocultos de Markov - HMM ("Hidden Markov Models"), uma estrutura duplamente estocástica capaz de modelar tanto as *variabilidades acústicas* como *temporais* do sinal de fala. Porém, para viabilizar o modelamento matemático de um HMM, são realizadas inúmeras suposições simplificadoras que limitam o seu potencial efetivo. Redes neurais artificiais - ANN ("Artificial Neural Networks") não necessitam fazer uso de muitas destas suposições, podem aprender e generalizar superfícies complexas de decisão, tolerar ruídos e suportar paralelismo. Todas estas vantagens tornam as ANNs extremamente poderosas para modelar as variabilidades acústicas da fala. Entretanto, ao contrário dos HMMs, as ANNs não têm se mostrado eficientes para o modelamento das variabilidades temporais. Com o objetivo de unir em uma única estrutura o que há de melhor nas tecnologias de redes neurais artificiais e de modelos ocultos de Markov, têm sido estudados e avaliados nos últimos sete anos [36, 14, 3, 33, 21, 8, 1], modelos híbridos ANN-HMM nos quais o modelamento das variabilidades acústicas é confiado à ANN enquanto o HMM responsabiliza-se pela absorção das variabilidades temporais.

Os objetivos desta Tese foram o estudo e a implementação dos principais módulos de um sistema para reconhecimento de fala contínua baseado em modelos híbridos ANN-HMM. As análises desenvolvidas concentram-se em três tópicos fundamentais : (1) Capacidade das ANNs, treinadas a partir dos critérios MSE ("Minimum Square Error") ou entropia relativa, de estimarem verossimilhanças de emissão de símbolos de HMMs contínuos, (2) Algoritmo REMAP ("Recursive Estimation and Maximization of A Posteriori Probabilities") para a reestimação dos parâmetros dos modelos híbridos, (3) Algoritmos de busca para a determinação da seqüência mais provável de palavras que compõem uma determinada sentença a ser reconhecida.

Para análise de desempenho são apresentados resultados da avaliação do sistema no reconhecimento de 100 sentenças (constituídas de um total de 319 palavras distintas) dependentes de locutor e gravadas sob condições de estúdio. A taxa de acerto de palavras do sistema sem o uso de restrições gramaticais foi de 83,26% e, com o uso de restrições gramaticais do tipo *pares de palavras,* foi de 99,47%.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

1.1 Considerações Iniciais

Muitos problemas de reconhecimento de padrões, de fundamental importância hoje em dia, são de natureza inerentemente seqüencial. Alguns exemplos incluem o reconhecimento de movimentos humanos através de uma seqüência de imagens de vídeo, ou o reconhecimento de fala contínua a partir de uma seqüência de amostras de um sinal de voz. Nos últimos anos, vários pesquisadores têm apresentado soluções baseadas em modelos ocultos de Markov - HMM ("Hidden Markov Models") para o reconhecimento de fala contínua [16, 27]. Os modelos ocultos de Markov são estruturas duplamente estocásticas que têm se mostrado eficientes para modelar, tanto as *variabilidades acústicas* como *temporais do sinal de fala.* O critério de treinamento mais usual de um HMM baseia-se no método da máxima verossimilhança - ML ("Maximum Likelihood") [27], porém este método apresenta um baixo poder discriminativo, uma vez que procura maximizar a verossimilhança de um determinado modelo gerar uma dada seqüência observada, mas não se preocupa em minimizar a verossimilhança dos outros modelos gerarem esta mesma seqüência. Um critério mais adequado para o treinamento de um HMM seria o de maximização da informação mútua - MMI ("Maximum Mutual Information") o qual ressalta a capacidade de discriminação dos modelos que competem entre si [14]. Infelizmente o critério MMI não pode ser solucionado por análise direta ou por reestimação [14]. Por outro lado, pode-se mostrar que o critério MMI é equivalente à maximização a posteriori - MAP ("Maximum A Posteriori Probabilities"), que é o mesmo método utilizado por uma rede neural artificial - ANN ("Artificial Neural Network") otimizada a partir do critério de minimização do erro quadrático médio - MSE ("Minimum Square Error") [33, 30, 4].

Além de ser não-discriminativo, o critério ML necessita fazer uso de muitas suposições simplifi-

cadoras para viabilizar o modelamento matemático de um HMM. ANNs treinadas segundo o critério MSE não precisam fazer uso de várias destas suposições, e além disso são capazes de aprender e generalizar funções complexas, tolerar ruídos e suportar paralelismo. Estas vantagens tornam as ANNs extremamente poderosas para modelar as variabilidades acústicas da fala. Entretanto, ao contrário dos HMMs, as ANNs não têm se mostrado eficientes para o modelamento das variabilidades temporais. Com o objetivo de unir em uma única estrutura o que há de melhor na tecnologia de redes neurais artificiais e dos modelos ocultos de Markov, têm sido estudados e avaliados nos últimos sete anos [36, 14, 3, 33, 21, 8, 1], sistemas para reconhecimento de fala contínua baseados em modelos híbridos ANN-HMM, que utilizam ANNs para estimar os parâmetros, responsáveis pelo modelamento das variabilidades acústicas, de HMMs contínuos.

Os primeiros sistemas para reconhecimento de fala contínua baseados em modelos híbridos ANN-HMM treinavam a ANN uma única vez, utilizando uma base de dados (conjunto de sentenças) previamente segmentada em termos de unidades sub-lexicais (sílabas, polifones, fones, ou unidades subfônicas) para estimar $P(q_k|\mathbf{x}_n)$, a probabilidade a posteriori de um determinado estado q_k de um HMM contínuo, dado um vetor acústico $\mathbf{x}_n$ (segmento do sinal de fala parametrizado em termos de alguma transformação específica, por exemplo : coeficientes Mel Cepstrais [5]). Porém em 1996, Konig [14] estabeleceu as bases matemáticas de um algoritmo recursivo para a reestimação dos parâmetros desta ANN. Este algoritmo foi denominado "Recursive Estimation and Maximization of A Posteriori Probabilities" - REMAP e possui a propriedade de melhorar progressivamente a estimativa de $P(q_k|\mathbf{x}_n)$. O algoritmo REMAP é motivado por estudos perceptuais que mostram a elevada importância das informações transicionais entre unidades sub-lexicais para o reconhecimento da fala.

1.2 Objetivos

Os objetivos desta Tese foram o estudo e implementação dos principais módulos de um sistema para reconhecimento de fala contínua baseado em modelos híbridos ANN-HMM. As análises realizadas concentraram-se em três tópicos fundamentais :

- Estimação de verossimilhanças de emissão de símbolos de HMMs contínuos empregando ANN.
- Reestimação dos parâmetros dos modelos híbridos utilizando o algoritmo REMAP.
- Utilização do algoritmo "Level Building" [29] na determinação da seqüência mais provável de palavras que compõem uma determinada sentença a ser reconhecida.

A avaliação do sistema híbrido ANN-HMM implementado foi realizada com o uso de uma base de dados composta por 100 sentenças (construídas a partir de 319 palavras distintas) dependentes de locutor e gravadas sob condições de estúdio (sem ruído ambiente).

1.3 Estrutura da Tese

Capítulo 2 Apresenta a estrutura básica dos sistemas que representam o estado da arte em reconhecimento estatístico de fala contínua. Estabelece as definições e notações que serão utilizadas ao longo de toda a Tese. Faz uma breve revisão sobre modelos ocultos de Markov. Discute alguns problemas básicos do critério de ML como por exemplo : treinamento independente e suposição de independência estatística dos vetores acústicos. Anuncia os sistemas baseados em modelos híbridos ANN-HMM como uma possível alternativa aos sistemas que empregam apenas os modelos ocultos de Markov.

Capítulo 3 Apresenta uma breve revisão sobre as redes neurais artificiais do tipo multilayer perceptron - MLP ("Multilayer Perceptron") [10], redes recorrentes [31] e "Time Delay Neural Networks" - TDNN [40]. Introduz a ANN como um estimador estatístico e discute a estimação de verossimilhanças de emissão de símbolos de HMMs contínuos com o uso de ANNs. Lista algumas das vantagens das redes neurais artificiais na estimativa dos parâmetros que modelam a variabilidade acústica da fala. Apresenta os dois tipos básicos de modelos híbridos ANN-HMM segundo Konig, 1996 [14] : (1) Modelo híbrido ANN-HMM *padrão*, treinado a partir do critério MSE, que é equivalente ao critério MAP, e avaliado segundo o critério ML, (2) Modelo híbrido ANN-HMM *discriminativo*, treinado e avaliado segundo o critério MAP.

Capítulo 4 Relata a importância das informações transicionais entre unidades sub-lexicais no reconhecimento da fala. Descreve os procedimentos para a estimação (etapa Pré-REMAP) dos parâmetros de um modelo híbrido ANN-HMM a partir de uma base de dados previamente segmentada em termos de unidades sub-lexicais. Apresenta o algoritmo de reestimação de parâmetros - REMAP discutindo duas alternativas para a estimação dos alvos a serem utilizados no retreinamento da ANN: (1) *Alvos abruptos* - define transições instantâneas entre as unidades sub-lexicais, (2) *Alvos suaves* - define transições suaves entre as unidades sub-lexicais fornecendo informações sobre o grau de confusão acústica que existe nas fronteiras entre estas unidades.

Capítulo 5 Apresenta um breve relato sobre a evolução histórica dos algoritmos de busca baseados em programação dinâmica. Introduz formalmente o algoritmo "Level Building". Apresenta um modelo de duração para as palavras a ser empregado durante a realização da busca pela seqüência "ótima"

de palavras. Discorre alguns comentários sobre o algoritmo "One Stage".

Capitulo 6 Relata o sistema híbrido ANN-HMM implementado. Descreve detalhes sobre o conjunto de exemplos de treinamento utilizado e sobre o pré-processamento das amostras do sinal de fala para o cálculo de atributos. Apresenta a topologia e o treinamento da ANN utilizada assim como a topologia dos HMMs empregados. Descreve o processo de Estimação (Pré-REMAP) e reestimação (REMAP) dos parâmetros do sistema. Para finalizar, discute aspectos de implementação do algoritmo de busca incluindo a gramática do tipo *pares de palavras,* utilizada na avaliação final do sistema.

Capítulo 7 Apresenta uma análise dos resultados obtidos, tanto a nível de unidades sub-lexicais (fones) como a nível de sentenças. Uma ênfase especial é dada à evolução das taxas de acerto do sistema após cada etapa de reestimação.

Capítulo 8 Apresenta as considerações finais, concluindo a Tese e apresentando sugestões para trabalhos futuros.

Capítulo 2

Reconhecimento de Fala Contínua

2.1 Introdução

Atualmente os sistemas para reconhecimento de fala contínua baseiam-se fundamentalmente em princípios de reconhecimento estatístico de padrões. A tarefa destes sistemas é construir uma função probabilística capaz de mapear amostras de um sinal de fala - representadas através de uma seqüência de vetores acústicos - no espaço das possíveis sentenças (palavras ou frases). Em geral esta função probabilística é parametrizada em função das principais fontes de variabilidade da fala : a *variabilidade acústica* e a *variabilidade temporal.*

Sendo $s(n)$ a seqüência de amostras de uma sinal de fala desconhecido, este sinal pode ser submetido a um **pré-processamento** e representado por um conjunto de vetores acústicos $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. Seja M_i $(i = 1, 2, \ldots, I_M)$ o conjunto de todas as possíveis sentenças e Θ o conjunto de parâmetros responsáveis pelo modelamento das variabilidades *acústicas* e *temporais.* O objetivo de um sistema para reconhecimento de fala contínua consiste em estimar durante o treinamento, e utilizar durante o reconhecimento, a seguinte função probabilística, $P(M|\mathbf{X},\Theta)$, isto é, a probabilidade da sentença M dado o conjunto de vetores acústicos $\mathbf{X}$ e os parâmetros Θ.

Uma vez que uma sentença M pode ser construída a partir da concatenação de palavras $M = \{W_1, W_2, \ldots, W_{N_M}\}$, a tarefa de um sistema para reconhecimento de fala contínua também pode ser interpretada como a determinação da seqüência de palavras mais prováveis $\widehat{M}$, dada a seqüência de vetores acústicos $\mathbf{X}$ e o conjunto de parâmetros Θ. Se a regra de Bayes for utilizada para decompor $P(M|\mathbf{X},\Theta)$, então $\widehat{M}$ pode ser determinada a partir da seguinte expressão,

$$\widehat{M} = \arg\max_M P(M|\mathbf{X},\Theta) = \arg\max_M \frac{p(\mathbf{X}|M,\Theta) \cdot P(M|\Theta)}{p(\mathbf{X}|\Theta)} \tag{2.1}$$

Esta equação mostra que encontrar $\widehat{M}$, é equivalente a encontrar a seqüência de palavras que maximiza o produto entre $p(\mathbf{X}|M,\Theta)$ e $P(M|\Theta)$, uma vez que durante o reconhecimento o termo $p(\mathbf{X}|\Theta)$ será o mesmo para todas as sentenças avaliadas. O primeiro termo $p(\mathbf{X}|M,\Theta)$, representa a verossimilhança da seqüência de vetores acústicos $\mathbf{X}$ dada uma seqüência de palavras específica M e o conjunto de parâmetros Θ, e esta verossimilhança pode ser determinada a partir de um **modelo acústico** (modelo estatístico) da sentença M. O segundo termo $P(M|\Theta)$, representa a probabilidade da sentença M dado o conjunto de parâmetros Θ, e esta probabilidade pode ser determinada por um **modelo da língua**.

O processo de determinação de $\widehat{M}$ é denominado **decodificação** e projetos de decodificadores (algoritmos de busca) eficientes são cruciais para a realização prática de sistemas para reconhecimento de fala contínua. Portanto, um sistema estatístico para reconhecimento de fala contínua pode ser divido em quatro módulos principais : **(1) Pré-processamento**, **(2) Modelamento Acústico**, **(3) Modelo da língua**, **(4) Decodificação (Busca)**. A Figura 2-1 mostra um diagrama de blocos com estes quatro módulos principais e as ligações entre eles.

2.1.1 Pré-Processamento

O pré-processamento pode ser visto como um mapeamento da seqüência de amostras $s(n)$ (sinal de fala digitalizado) em uma seqüência de vetores acústicos $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N\}$. Este mapeamento possui dois objetivos básicos : (1) Ressaltar características, (2) Reduzir dimensionalidade sem perdas de informações cruciais ao reconhecimento. A forma mais usual para realização deste mapeamento consiste em janelar o sinal $s(n)$ com um certo grau de superposição, conforme Figura 2-3, e em seguida calcular um conjunto de atributos (através de uma transformação específica) para cada uma das janelas. Os atributos mais comumente utilizados, são : coeficientes Mel cepstrais [5] , log da energia normalizada [26], e as primeira e segunda derivadas destes coeficientes [6].

2.1.2 Modelamento Acústico

O objetivo deste módulo é a construção de um **modelo acústico** (modelo estatístico), para cada uma das possíveis sentenças M_i $(i = 1, 2, \ldots, I_M)$, que permita a avaliação da verossimilhança $p(\mathbf{X}|M_i,\Theta)$ durante a etapa de reconhecimento. Em geral estes modelos são construídos através de um procedimento hierárquico, que utiliza como base a decomposição de sentenças em palavras e estas em unidades sub-lexicais como sílabas, polifones, fones ou unidades sub-fônicas. A construção destes modelos de sentenças consiste de três etapas básicas :

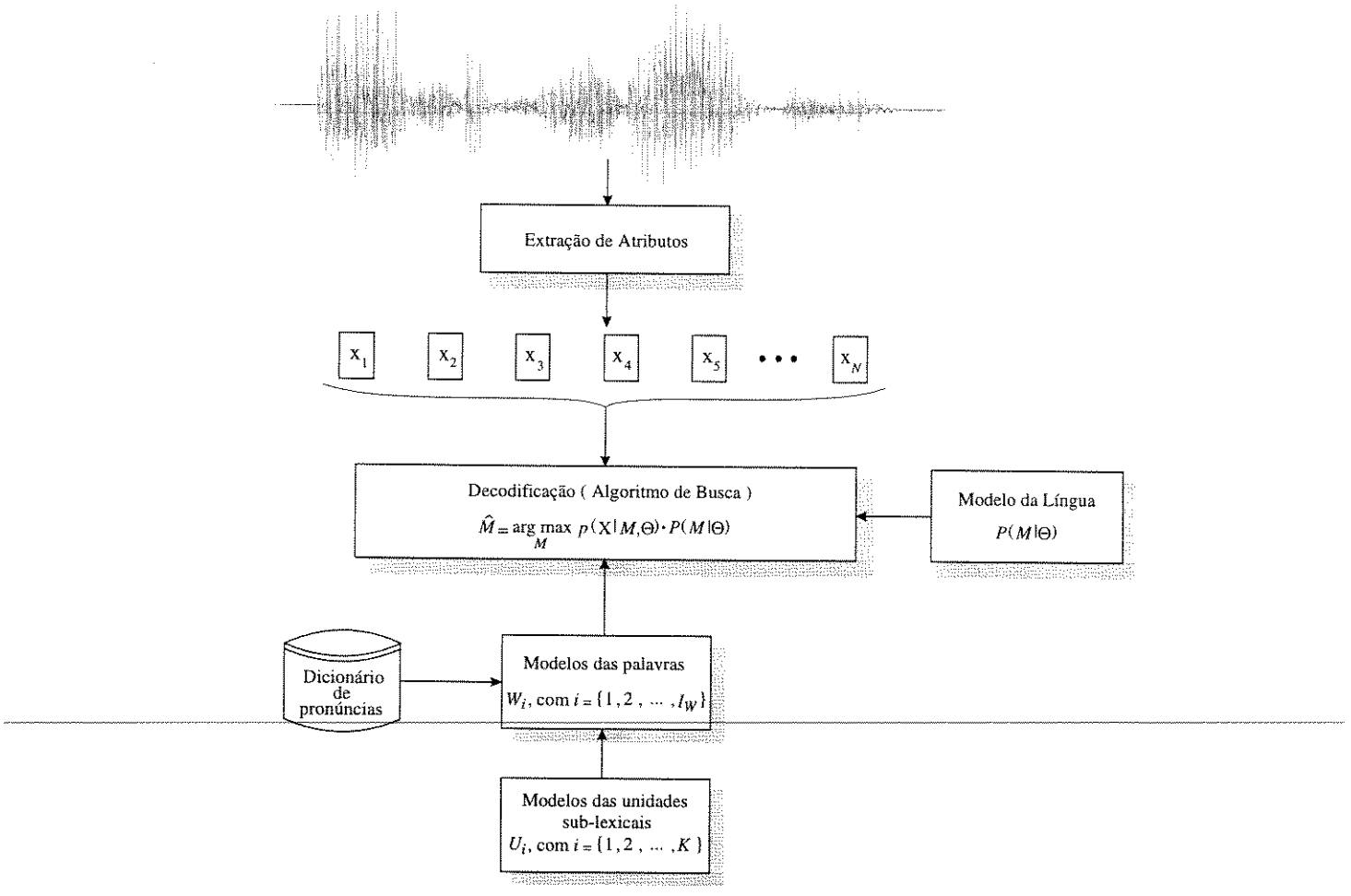

Figura 2-1: Diagrama de blocos de um sistema para reconhecimento estatístico de fala contínua.

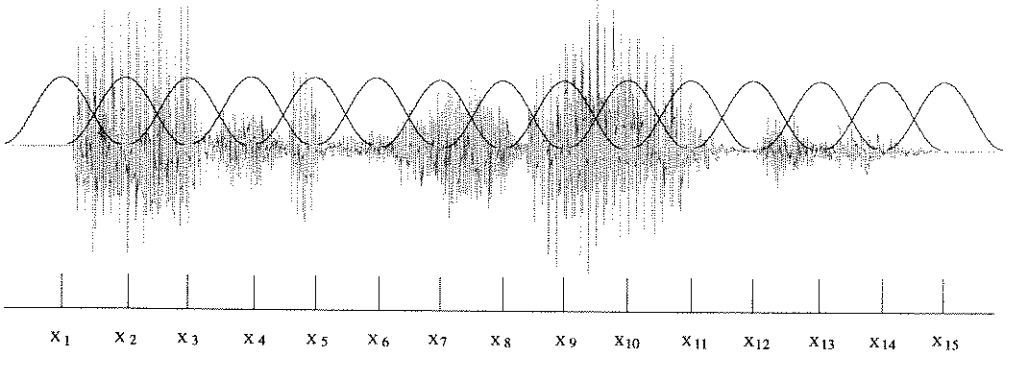

Figura 2-2: Estimativa da seqüência de atributos (vetores acústicos) associados a uma seqüência de amostras de um sinal de fala

1. Construir um modelo estatístico para cada uma das unidades sub-lexicais.

2. De posse de um **dicionário de pronúncias**, montar os modelos estatísticos para cada uma das palavras do léxico através da concatenação dos correspondentes modelos das unidades sub-lexicais. Por exemplo, o modelo da palavra "*casa*" pode ser obtido pela concatenação dos modelos dos fones $/k/ + /a/ + /z/ + /a/$.

3. Por último, através de um procedimento de concatenação, sujeito ou não a restrições gramaticais, montar os modelos das sentenças a partir dos modelos de palavras.

O modelamento acústico mais comumente utilizado baseia-se em modelos ocultos de Markov - HMMs, e as explicações para este fato são :

- Por serem máquinas de estado finito[1], os HMMs permitem facilmente a concatenação de modelos, possibilitando a construção hierárquica de modelos acústicos para as sentenças e também a aplicação de restrições gramaticais tanto a nível de palavras como de unidades sub-lexicais.

- Por serem estruturas duplamente estocásticas os HMMs são capazes de modelar, tanto a *variabilidade acústica* como *temporal* da fala, permitindo a avaliação de $p(\mathbf{X}|M,\Theta)$.

2.1.3 Modelo da língua

A tarefa do modelo da língua é estimar $P(M|\Theta)$, isto é, a probabilidade da sentença dado o conjunto de parâmetros Θ. O Modelo da língua também pode ser utilizado para determinar a probabilidade de uma palavra W_i, em uma sentença, dadas todas as palavras que a precedem, $W_1, W_2, \ldots, W_{i-1}$. Isto pode ser obtido, verificando que $P(M|\Theta)$ pode ser reescrito como

$$P(M|\Theta) = P(W_1|\Theta) \cdot \prod_{i=2}^{N_M} P(W_i|W_{i-1}, W_{i-2}, \ldots, W_1, \Theta) \tag{2.2}$$

sendo N_M o número total de palavras presentes na sentença M.

Em geral a probabilidade de um modelo M é considerada independente dos parâmetros Θ, logo,

$$P(M|\Theta) \approx P(M) = P(W_1|\Theta) \cdot \prod_{i=2}^{N_M} P(W_i|W_{i-1}, W_{i-2}, \ldots, W_1) \tag{2.3}$$

[1]Máquina de estados com um número finito de estados.

Uma maneira simples mas efetiva para calcular (2.3) é assumir que a probabilidade das palavras em um sentença dependem apenas das $N-1$ palavras anteriores,

$$P(M) \approx P(W_1|\Theta) \cdot \prod_{i=2}^{N_M} P(W_i|W_{i-1}, W_{i-2}, \ldots, W_{i-N+1}) \tag{2.4}$$

Esta solução é denominada modelamento estatístico $N-gram$ (ou gramática $N-gram$). Os modelos mais utilizados são os Bi-gram, $P(W_i|W_{i-1})$ e Tri-gram, $P(W_i|W_{i-1}, W_{i-2})$. Estas probabilidades podem ser calculadas através da leitura automática[2] de vários textos[3] e do levantamento das freqüências relativas de pares e triplas de palavras.

2.1.4 Decodificação

Dado um modelo da língua e os modelos acústicos de todas as palavras do léxico (construídos através da concatenação dos modelos acústicos de unidades sub-lexicais), a tarefa do decodificador (algoritmo de Busca) é estimar de maneira otimizada a seqüência de palavras $\widehat{M}$ que maximiza a Equação (2.1) , isto é, deve realizar uma busca da sentença $\widehat{M}$ que maximiza o produto entre $p(\mathbf{X}|M,\Theta)$ e $P(M|\Theta)$. Os algoritmos de busca podem ser classificados, quanto à forma de implementação, em duas categorias: Algoritmos síncronos por palavras, destacando-se o "Level Building" [29] e os algoritmos síncronos por quadro de análise (síncrono por vetor acústico), destacando-se o "One Stage", [23, 17]. Estes algoritmos serão discutidos em detalhe no Capítulo 6.

2.2 Definições e Notações

A seguir são apresentadas algumas definições e notações que serão utilizadas nas seções seguintes e em todos os Capítulos subseqüentes.

- $Q = \{q_1, q_2, \ldots, q_K\}$ - Conjunto que contém todos os estados com os quais serão construídos os modelos de unidades sub-lexicais, palavras e sentenças. Por exemplo, considere que o conjunto de unidade sub-lexicais seja igual a 36 fones, então no caso de se modelar cada fone por um estado, o número de estados será $K = 36$. Se for desejado modelar a produção acústica de início, meio e final dos diferentes fones, cada fone deverá ser modelado por 3 estados e portanto $K = 3 \times 36 = 108$ estados.

[2]Detecção automática de palavras ao longo de um texto escrito.
[3]Escolhidos criteriosamente segundo critérios lingüísticos.

- $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ - Seqüência de vetores acústicos associados a uma sentença específica.
- $\mathbf{X}_{n-c}^{n+d} = \{\mathbf{x}_{n-c}, \mathbf{x}_{n-c+1}, \ldots, \mathbf{x}_n, \ldots, \mathbf{x}_{n+d}\}$ - Sub-seqüência de $\mathbf{X}$ com $c+d+1$ vetores acústicos.
- M_i - Modelo oculto de Markov da $i-$*ésima* sentença. Com $i \in \mathcal{I}_\mathcal{M} = \{1, 2, \ldots, I_M\}$, sendo I_M o número total de seqüências possíveis de serem construídas a partir das palavras presentes no léxico e permitidas pelo modelamento da língua. M_i também pode ser visto como um grafo direcionado com C_i estados, todos eles pertencentes a Q.
- W_i - Modelo oculto de Markov da $i-$*ésima* palavra. Com $i \in \mathcal{I}_\mathcal{W} = \{1, 2, \ldots, I_W\}$, sendo I_W o número total de palavras presentes no léxico.
- $\mathbf{X}_{M_j} = \left\{\mathbf{X}_{M_j}^{(1)}, \mathbf{X}_{M_j}^{(2)}, \ldots, \mathbf{X}_{M_j}^{(N_{e_j})}\right\}$ - Conjunto de seqüências de vetores acústicos associadas ao modelo M_j (seqüências de vetores acústicos utilizadas no treinamento de M_j).
- q^n - Denota um estado no instante n.
- q_k^n - Indica que o estado k ocorreu no instante n. Rigorosamento falando, q_k^n denota a associação do valor k à variável aleatória q^n.

- Γ - Seqüência de estados de comprimento N. Cada Γ é a realização de um processo estocástico, onde os valores associados a este processo em cada instante de tempo pertencem a Q. Ao longo desta dissertação também será usual escrever explicitamente $\Gamma_j = \left\{q_{\Gamma_j}^1, \ldots, q_{\Gamma_j}^n, \ldots, q_{\Gamma_j}^N\right\}$. Uma seqüência de estados permitida por um determinado modelo M_i também será denominada de *caminho* ao longo de M_i.
- τ_i - Conjunto de todas as seqüências de estado Γ permitidas pelo modelo M_i.
- $\Im$ - Conjunto de todas as seqüências de estado Γ permitidas por todos os modelos M_i, com $i \in \mathcal{I}_\mathcal{M}$.
- $\Theta = \{\theta_1, \theta_2, \ldots, \theta_I\}$ - Conjunto de parâmetros que descrevem todos os modelos M_i com $i \in \mathcal{I}_\mathcal{M}$. θ_i representa apenas os parâmetros presentes em M_i. Nos modelos híbridos discutidos nesta dissertação, $\Theta = \{\Psi \cup \mathbf{A} \cup \boldsymbol{\pi}\}$, sendo Ψ o conjunto de parâmetros de uma rede neural artificial (conjunto de pesos sinápticos), $\mathbf{A}$ o conjunto com as probabilidades de transição de cada um dos estados $q \in Q$, e $\boldsymbol{\pi}$ as probabilidades dos estados iniciais q^1 de cada um dos modelos M_i com $i \in \mathcal{I}_\mathcal{M}$.
- $P(\cdot)$ - Será utilizado para representar probabilidade ou distribuição de probabilidade.

- $p(\cdot)$ - Será utilizado para representar função densidade de probabilidade e verossimilhança.

2.3 Modelos Ocultos de Markov - HMM

2.3.1 Introdução

Atualmente os sistemas que representam o estado-da-arte em reconhecimento de fala contínua constroem seus *modelos acústicos* baseados em modelos ocultos de Markov. O sucesso destas estruturas deve-se, principalmente, à sua capacidade de modelar tanto as *variabilidades acústicas* como *temporais* do sinal fala e também por permitir a construção hierárquica dos modelos acústicos das sentenças.

A teoria de modelos ocultos de Markov apóia-se na suposição de que um processo contínuo pode ser aproximado por uma sucessão de curtos estados estacionários e modela uma seqüência de vetores acústicos como um processo estacionário por partes. Isto é, um HMM modela uma seqüência de vetores acústicos $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ como se estes tivessem sido emitidos (gerados) por uma sucessão de L *estados estacionários discretos* $q_k \in Q$, com *transições instantâneas* entre eles. Neste caso, um HMM é definido (e representado) como uma máquina de estados finito com uma topologia particular (usualmente *left-right*, uma vez que a fala é estritamente seqüencial). Esta abordagem define dois processos estocásticos concorrentes : A seqüência dos L estados estacionários discretos (responsável pelo modelamento da estrutura temporal da fala), e a seqüência de símbolos observados (seqüência de vetores acústicos) formada pela emissão de um símbolo em cada um dos L estados (responsável pelo modelamento acústico da fala). O HMM é chamado de "oculto" porque existe um processo estocástico escondido, a seqüência de estados, que não é observável mas que afeta a seqüência de símbolos emitidos. O modelo é chamado de "Markov" porque as estatísticas do estado corrente q^n são modeladas como sendo dependentes apenas do símbolo atual $\mathbf{x}_n$ e do estado prévio q^{n-1} (no caso de modelo de Markov de primeira ordem).

2.3.2 Definição

Um HMM pode ser definido formalmente por :

- $\mathbf{A} = \left\{a_{ij} | a_{ij} = P(q^{n+1} = j | q^n = i) = P(q_j^{n+1} | q_i^n)\right\}$ - Função distribuição de probabilidade de transição, sendo que a_{ij} denota a probabilidade de transição do estado i para o estado j. Geralmente assume-se que esta distribuição de probabilidade é a mesma para todo instante de tempo.

- $\mathbf{B} = \{b_j(\mathbf{x}_i) | b_j(\mathbf{x}_i) = P(\mathbf{x}_i|q = j) = P(\mathbf{x}_i|q_j)\}$ ou $\mathbf{B} = \{b_j(\mathbf{x}_i) | b_j(\mathbf{x}_i) = p(\mathbf{x}_i|q = j) = p(\mathbf{x}_i|q_j)\}$ - Para cada estado existe uma correspondente distribuição de probabilidade de saída no caso de um HMM discreto ou uma função densidade de probabilidade contínua para o caso de um HMM contínuo. $b_j(\mathbf{x}_i)$ se refere à probabilidade (no caso discreto) ou verossimilhança (no caso contínuo) do símbolo $\mathbf{x}_i$ ser gerado pelo estado q_j. Usualmente $b_j(\mathbf{x}_i)$ é chamado de probabilidade (ou verossimilhança) de emissão de símbolos.

- $\boldsymbol{\pi} = \{\pi_i | \pi_i = P(q^1 = i) = P(q_i^1)\}$ - Distribuição de probabilidade do estado inicial. Nos modelos $left - right$, normalmente é assumido $\pi_1 = 1$ e $\pi_i = 0$ para todo $i \neq 1$.

De posse destas definições pode-se representar um modelo oculto de Markov de forma compacta por $M = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$.

2.3.3 Algoritmos Básicos

Em geral [27], três problemas são levantados quando utilizam-se HMMs para modelar seqüências de observações :

1. **Avaliação** Qual a verossimilhança de um HMM gerar uma dada seqüência de observações ?

2. **Decodificação** Dada uma seqüência de observações e um HMM, qual é a seqüência de estados, ao longo deste HMM, com maior verossimilhança de gerar estas observações ?

3. **Estimação de Parâmetros** Dado um HMM e um conjunto de seqüências de observações (seqüências de vetores acústicos) a serem modeladas por este HMM, como adaptar seus parâmetros ($\mathbf{A}$, $\mathbf{B}$ e $\boldsymbol{\pi}$) para que este modelo maximize a verossimilhança de geração da seqüência de observações ?

Todos estes três problemas apresentam soluções bastante eficientes baseadas em casos particulares de algoritmos de programação dinâmica. Por exemplo, o problema de **avaliação** ocorre no reconhecimento de palavras isoladas em que se deseja avaliar diferentes modelos ocultos de Markov de palavras para verificar qual deles apresenta a maior verossimilhança de ter gerado a palavra a ser reconhecida. Este problema pode ser solucionado utilizando-se o algoritmo *Forward* [27]. O problema de **decodificação** é utilizado no reconhecimento de fala contínua em que se deseja descobrir qual o caminho (seqüência de estados) com maior verossimilhança, ao longo de um HMM bastante extenso (HMM de uma sentença), de ter gerado uma determinada seqüência de vetores acústicos. Este problema de

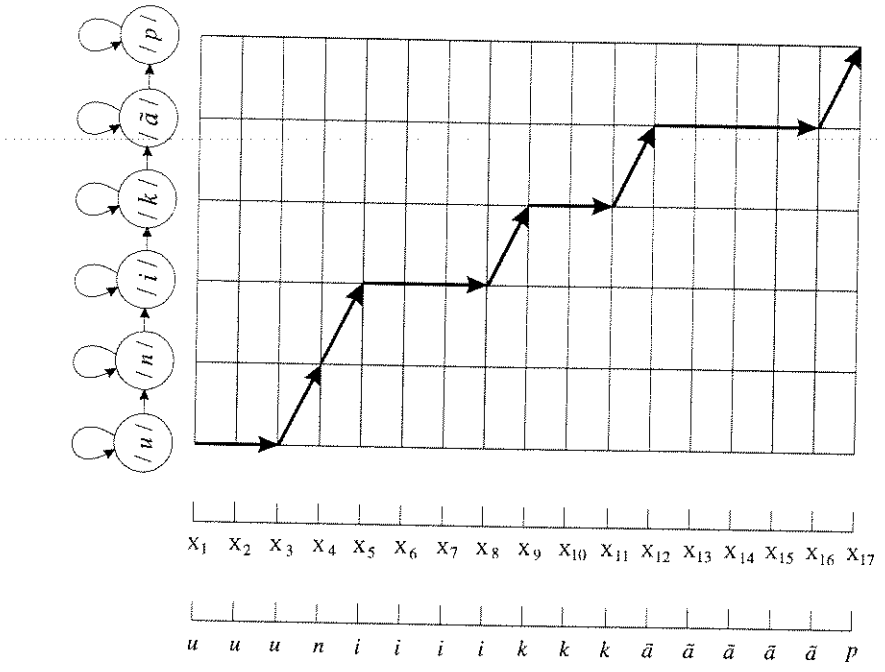

Figura 2-3: Trajetória de Viterbi : define a seqüência de estados com maior verossimilhança de ter gerado a seqüência de símbolos X.

decodificação pode ser solucionado utilizando-se o algoritmo de *Viterbi* [27]. A Figura 2-3 mostra um diagrama de treliça típico[4] com o caminho ótimo produzido pelo algoritmo de Viterbi.

O último problema, também chamado de **treinamento** (estimação de parâmetros), pode ser solucionado pelo algoritmo *Forward-Backward* [27] também conhecido como *Baum-Welch*, que é essencialmente uma versão do algoritmo EM ("Expectation-Maximization") [14]. No caso de HMMs estritamente *left-right* com transições de estado constantes no tempo, todos estes três algoritmos têm uma complexidade da ordem de apenas $(C_i \cdot N)$, sendo C_i o número de estados do HMM em questão e N o número de símbolos observados.

2.4 Discussão

2.4.1 Critérios de Treinamento - Problemas com o Critério ML

Como foi discutido no início deste Capítulo, o objetivo de um sistema para reconhecimento de fala contínua é, estimar durante o treinamento e utilizar durante o reconhecimento, a seguinte função

[4]Trajetória de Viterbi produzida pelo modelo da palavra "UNICAMP". Neste modelo foi utilizado o fone como unidade sub-lexical e estabelecida uma *correspondência um-a-um* entre fones e estados.

probabilística, $P(M|\mathbf{X},\Theta)$, que pode ser reescrita através da regra de Bayes como,

$$P(M|\mathbf{X},\Theta) = \frac{p(\mathbf{X}|M,\Theta)\cdot P(M|\Theta)}{p(\mathbf{X}|\Theta)} \tag{2.5}$$

Na fase de treinamento, a probabilidade $P(M|\mathbf{X},\Theta)$ do modelo M dado a seqüência de vetores acústicos $\mathbf{X}$, deve ser maximizada. O espaço de parâmetros onde esta otimização é realizada pode fazer a diferença entre *modelos treinados independentemente* (não-discriminativos) e *modelos discriminativos.*

Durante a fase de reconhecimento, $p(\mathbf{X}|\Theta)$ é uma constante pois os parâmetros do modelo são fixados. Entretanto, durante o treinamento, esta verossimilhança depende dos parâmetros de todos os possíveis modelos. Escrevendo $p(\mathbf{X}|\Theta)$ em termos de probabilidade marginal,

$$\begin{aligned} p(\mathbf{X}|\Theta) &= \sum_i p(\mathbf{X},M|\Theta) && (2.6)\\ &= \sum_i p(\mathbf{X}|M,\Theta)\cdot P(M|\Theta) && (2.7)\\ &\approx \sum_i p(\mathbf{X}|M,\Theta)\cdot P(M) && (2.8) \end{aligned}$$

A Equação (2.7) estende o somatório em i sobre todos os possíveis modelos de sentenças e assume a independência das probabilidades a priori dos modelos versus os parâmetros ($P(M|\Theta) \approx P(M)$). Substituindo (2.8) em (2.5) , e tratando do caso de um modelo particular M_j, tem-se :

$$P\left(M_j|\mathbf{X}_{M_j}^{(s)},\Theta\right) \approx \frac{p(\mathbf{X}_{M_j}^{(s)}|M_j,\Theta)\cdot P(M_j)}{p(\mathbf{X}_{M_j}^{(s)}|M_j,\Theta)\cdot P(M_j) + \sum_{i\neq j} p(\mathbf{X}_{M_j}^{(s)}|M_i,\Theta)\cdot P(M_i)} \tag{2.9}$$

sendo $\mathbf{X}_{M_j}^{(s)}$ com $s \in \left\{=1,2,\ldots,N_{e_j}\right\}$ as seqüências de treinamento associadas ao modelo M_j.

A maximização de $P\left(M_j|\mathbf{X}_{M_j}^{(s)},\Theta\right)$ como dado por (2.9) é usualmente simplificada restringindo a otimização ao subespaço de parâmetros de M_j. Esta restrição dá origem ao critério de máxima verossimilhança - ML ("Maximum Likelihood"). Neste critério mantém-se o somatório no denominador constante sobre todo o espaço de parâmetros de M_j. Como $P(M_j)$ é uma constante determinada pelo modelo da língua, então maximizar $P\left(M_j|\mathbf{X}_{M_j}^{(s)},\Theta\right)$ segundo o critério ML é semelhante a maximizar $p(\mathbf{X}_{M_j}^{(s)}|M_j,\Theta)$. Portanto, o treinamento de HMMs de acordo com o critério ML procura encontrar o

melhor conjunto de parâmetros $\widehat{\Theta}$, tal que[5],

$$\widehat{\Theta} = \arg\max_{\Theta} \prod_{j=1}^{I_M} \prod_{s=1}^{N_{e_j}} p\left(\mathbf{X}_{M_j}^{(s)} | M_j, \Theta\right) \tag{2.10}$$

sendo I_M o número total de sentenças.

Esta otimização "*modelo – por – modelo*" permite importantes simplificações no algoritmo de treinamento, porém pagando o preço de um baixo poder de discriminação entre os modelos.

Uma outra forma de maximizar $P\left(M_j | \mathbf{X}_{M_j}^{(s)}, \Theta\right)$ é considerar todo o espaço de parâmetros (isto é, os parâmetros de todos os modelos possíveis). Este critério é, de fato, discriminativo, uma vez que a contribuição de $p(\mathbf{X}_{M_j}^{(s)} | M_j, \Theta) \cdot P(M_j)$ deve ser maximizada enquanto que a de todos os modelos rivais, representada por,

$$\sum_{i \neq_j} p(\mathbf{X}_{M_j}^{(s)} | M_i, \Theta) \cdot P(M_i) \tag{2.11}$$

deve ser minimizada. O Apêndice A mostra que um treinamento baseado no critério de maximização da informação mútua - MMI ("Maximum Mutual Information") é capaz de maximizar a probabilidade a posteriori $P\left(M_j | \mathbf{X}_{M_j}^{(s)}, \Theta\right)$ com relação a todo o espaço de parâmetros Θ.

2.4.2 Mais Problemas com o Critério ML

Para tornar a estimação de $p\left(\mathbf{X}_{M_j}^{(s)} | M_j, \Theta\right)$ segundo o critério ML, tratável matematicamente, são realizadas algumas suposições simplificadoras que reduzem ainda mais a eficiência do critério de máxima verossimilhança - ML :

- **Suposição de independência das saídas** Os vetores acústicos não são correlacionados (independência das observações). O vetor acústico corrente $\mathbf{x}_n$ é assumido ser condicionalmente independente dos vetores acústicos prévios ($\mathbf{X}_1^{n-1} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-1}\}$). Esta suposição assume que observações pertencentes a um mesmo segmento da fala (geradas pelo mesmo estado) são independentes, uma suposição não realista dada a natureza não estacionária do sinal de fala. Para reduzir os problemas causados por esta suposição, o vetor acústico no instante n é geralmente complementado por suas primeira e segunda derivadas [6] calculadas sobre uma janela com alguns quadros, permitindo modelar um limitado contexto acústico. Por exemplo, a primeira derivada de $\mathbf{x}_n$, também conhecida como coeficiente Delta, $\Delta\mathbf{x}_n$, é comumente representada por

[5]Na Equação (2.10) também foi assumido que cada uma das seqüências de vetores acústicos é independente das demais.

$\Delta \mathbf{x}_n = a_c \cdot \mathbf{x}_{n-c} + \ldots + a_n \cdot \mathbf{x}_n + \ldots + a_d \cdot \mathbf{x}_{n+d}$, sendo que estes coeficientes a_i's, com $c \leq i \leq d$ são, em geral, fixados segundo critérios heurísticos. Da mesma forma a segunda derivada, coeficiente Delta-Delta, é geralmente definida por : $\Delta^2 \mathbf{x}_n = a_c \cdot \Delta \mathbf{x}_{n-c} + \ldots + a_n \cdot \Delta \mathbf{x}_n + \ldots + a_d \cdot \Delta \mathbf{x}_{n+d}$.

- **Suposição de Markov** Modelos de Markov são geralmente cadeias de Markov de primeira ordem. Explicitamente, a probabilidade da cadeia de Markov estar no estado q_k no instante n depende apenas do estado da cadeia de Markov no instante $n-1$, e é condicionalmente independente de todos os outros parâmetros passados (tanto do vetor acústico passado, $\mathbf{x}_{n-1}$, como dos estados anteriores ao estado q^{n-1}).

- **Suposição de superposição de *f.d.p's* gaussianas** A aplicação do critério ML a HMMs contínuos assume que as funções densidade de probabilidade de emissão de símbolos podem ser expressas pela superposição de um número finito de gaussianas, necessitando fixar a priori quantas gaussianas devem ser superpostas.

2.4.3 Alternativa - Modelos Híbridos ANN-HMM

Como alternativa aos sistemas baseados apenas em modelos ocultos de Markov, alguns autores, [36, 14, 3, 33, 21, 8, 1] têm apresentado nos últimos sete anos sistemas híbridos baseados tanto em redes neurais artificiais-ANN como em modelos ocultos de Markov. Nestes sistemas as ANNs são utilizadas para estimar, de forma mais eficiente que o tradicional algoritmo de reestimação de *Baum-Welch*, baseado no critério ML, parâmetros de HMMs contínuos. O Capítulo 3 apresenta a formulação matemática dos modelos híbridos assim como as suas principais vantagens em relação aos sistemas baseados apenas em HMM.

Capítulo 3

Modelos Híbridos ANN-HMM

3.1 Redes Neurais Artificiais - ANNs

3.1.1 Multilayer Perceptrons - MLPs

Nesta dissertação, a discussão sobre redes neurais será focalizada principalmente sobre o Perceptron multicamadas - MLP ("Multilayer Perceptron"), uma forma de ANN que é comumente utilizada em reconhecimento de fala. Entretanto, todas as conclusões básicas sobre a utilidade destas estruturas para estimar parâmetros (densidade de probabilidade de emissão de símbolos) para um HMM são perfeitamente estendidas para outros tipos de ANN, como por exemplo redes neurais recorrentes [31], ou "Time Delay Neural Network" (TDNN) [40].

Tipicamente, MLPs possuem uma arquitetura do tipo "$feed - forward$" com uma camada de entrada (consistindo do vetor de entrada), zero ou mais camadas escondidas, e uma camada de saída. Cada camada calcula um conjunto de funções discriminantes lineares [10] (via uma matriz de pesos sinápticos) e em seguida aplica à saída de cada uma destas funções uma não-linearidade, freqüentemente uma função do tipo sigmóide. As mais comuns são :

$$\varphi_l(v) = \frac{a}{1 + \exp(-b \cdot v)} \tag{3.1}$$

assim como,

$$\varphi_t(v) = \frac{2 \cdot a}{1 + \exp(-b \cdot v)} - a \tag{3.2}$$

Se $a = b = 1$, $\varphi_l(v)$ e $\varphi_t(v)$ se transformam, respectivamente, nas funções logística e tangente

hiperbólica.

Como discutido em [10], estas não-linearidades executam diferentes funções para as unidades escondidas e de saída. Nas unidades escondidas, suas funções são gerar momentos de ordem elevada do vetor de entrada; isto pode ser feito efetivamente com o uso de vários tipos de funções não lineares, não apenas por sigmóides. Nas unidades de saída, as não-linearidades podem ser vistas como uma aproximação diferencial de um limiar de decisão. Por este motivo, as funções não lineares das saídas devem ser sigmóides.

Pode ser provado que MLPs com um número elevado de unidades escondidas podem (em princípio) fornecer um mapeamento arbitrário $g(\mathbf{x})$ entre a entrada e a saída. O conjunto de parâmetros (elementos das matrizes de pesos) de uma MLP será denominado Ψ. Estes parâmetros são treinados para associar o vetor de saída "desejado" ao vetor de entrada. Isto é geralmente alcançado via algoritmo de treinamento EBP - "Error Back-Propagation" [10] que usa um procedimento baseado em gradientes para iterativamente minimizar a função de custo em seu espaço de parâmetros.

As funções de custo mais utilizadas, entre outras, são o critério de minimização do erro quadrático médio - MSE "Minimum Square Error",

$$\mathcal{E} = \sum_{n=1}^{N_e} \|g(\mathbf{x}_n, \Psi) - \mathbf{d}(\mathbf{x}_n)\|^2 \tag{3.3}$$

e o critério de entropia relativa. Uma versão discreta segundo Konig [14] da versão clássica de entropia relativa é,

$$\mathcal{E}_e = \sum_{n=1}^{N_e} \sum_{k=1}^{K} d_k(\mathbf{x}_n) \ln \frac{d_k(\mathbf{x}_n)}{g_k(\mathbf{x}_n, \Psi)} \tag{3.4}$$

onde $\mathbf{g}(\mathbf{x}_n, \Psi) = (g_1(\mathbf{x}_n, \Psi), \ldots, g_k(\mathbf{x}_n, \Psi), \ldots, g_K(\mathbf{x}_n, \Psi))^t$ representa o vetor de saída atual da MLP (em função do vetor de entrada $\mathbf{x}_n$ e dos parâmetros Ψ), $\mathbf{d}(\mathbf{x}_n) = (d_1(\mathbf{x}_n), \ldots, d_k(\mathbf{x}_n), \ldots, d_K(\mathbf{x}_n))^t$ representa o vetor de alvos desejados na saída (obtidos através de uma base de treinamento devidamente rotulada em termos de unidades sub-lexicais), K é o número total de classes e N_e o total de exemplos de treinamento.

MLPs, e outras estruturas conexionistas, têm sido utilizadas para uma grande variedade de tarefas relacionadas com o reconhecimento de fala [33]. O caso mais comum é o de reconhecimento de palavras isoladas, em que uma seqüência temporal de vetores acústicos é tratada como um padrão espacial. A função da rede neste caso é mapear um vetor de entrada (palavra a ser reconhecida) em uma das N_W unidades de saída, sendo N_W o número de palavras possíveis. A aplicação deste tipo de estrutura

no reconhecimento de fala contínua exigiria inicialmente uma segmentação da sentença em palavras, porém isto é extremamente complexo, uma vez que as palavras encontram-se, em geral, altamente coarticuladas.

3.1.2 Outras Arquiteturas Conexionistas

Como será visto na próxima seção, um modelo híbrido ANN-HMM é baseado em uma perspectiva estatística que é válida para qualquer arquitetura conexionista, dependendo apenas de algumas suposições sobre o procedimento de treinamento. Por este motivo vários trabalhos têm sido realizados utilizando outros tipos de estruturas conexionistas em sistemas híbridos ANN-HMM, destacando-se :

- **Redes Neurais Recorrentes - RNN** Vários trabalhos têm sido publicados pelo grupo de processamento digital de fala da Universidade de Cambrige descrevendo sistemas híbridos, cujas verossimilhanças de emissão de símbolos de HMMs contínuos são geradas por redes neurais recorrentes - RNN ("Recurrent Neural Network") [31] . Estas redes utilizam um conjunto de unidades de entrada para receber vetores acústicos e têm um conexionismo recorrente da saída para algumas unidades de entrada. A rede é treinada utilizando o algoritmo de retropropagação do erro ao longo do tempo - EBPTT ("Error Back Propagation Through Time") [10]. Resultados experimentais [31], mostram que este tipo de rede pode apresentar resultados similares aos de uma MLP, com o uso de um número significativamente menor de parâmetros (conexões sinápticas). Uma das principais limitações deste tipo de estrutura refere-se a alguns problemas inerentes à instabilidade.

- **Time-Delay-Neural Network - TDNN** Redes recorrentes podem ser aproximadas em um intervalo de tempo finito por uma rede do tipo feed-forward, onde os laços (realimentações) são substituídos pelo uso explícito de valores de ativações precedentes. Em 1989, Waibel et al [40] utilizaram, com sucesso, um rede como esta para o reconhecimento de fones, e em 1993 Waibel et al[39], aplicaram-na em um sistema híbrido denominado MS-TDNN ("Multi-State-TDNN") [39], que utilizava além de uma TDNN um procedimento de alinhamento temporal dinâmico - DTW ("Dinamic Time Warping"). A TDNN é uma estrutura bastante flexível e na prática, muitos laboratórios de pesquisa (inclusive o Laboratório de Processamento Digital de Fala da Universidade Estadual de Campinas - LPDF-UNICAMP) estão realizando pesquisas sobre este tipo de ANN para estimar verossimilhanças de emissão de símbolos para modelos ocultos de Markov em sistemas híbridos ANN-HMM.

Assim como em uma MLP, tanto a RNN como o TDNN podem utilizar as funções do tipo sigmóide em suas unidades escondidas e de saída e também podem fazer uso de treinamentos baseados nos critérios MSE e em entropia relativa.

3.2 ANNs como Estimadores Estatísticos

Redes neurais artificiais e modelos estatísticos para a análise de dados não constituem, em absoluto, metodologias disjuntas. Existe uma considerável superposição entre estas duas áreas do conhecimento. A metodologia estatística é diretamente aplicável a muitos modelos de redes neurais, resultando em uma maior eficiência tanto na estimação de parâmetros como dos algoritmos de otimização (aprendizado). Adicionalmente, métodos estatísticos providenciam ferramentas tais como intervalos de confiança e testes de hipótese que são ausentes no campo das redes neurais.

Recentemente, muitos trabalhos estatísticos têm sido publicados estabelecendo conexões entre modelos estatísticos e redes neurais, muitos deles mostrando não só a complementariedade mas também a equivalência entre as duas metodologias [9].

Foi descoberto recentemente [4], que se uma ANN for projetada como um classificador de padrões e otimizada segundo o critério da minimização do erro quadrático médio ou entropia relativa, então suas saídas podem aproximar a probabilidade a posteriori da classe dada a entrada, isto é, $P(classe|entrada)$. Os modelos denominados híbridos ANN-HMM têm utilizado com sucesso esta propriedade para estimar probabilidades a posteriori de um estado q_k de um HMM dado um vetor acústico $\mathbf{x}_n$, $P(q_k|\mathbf{x}_n)$. Na verdade, um parâmetro útil para um HMM seria a verossimilhança de um vetor acústico $\mathbf{x}_n$ ser emitido por um determinado estado q_k, isto é, $p(\mathbf{x}_n|q_k)$. Felizmente, $p(\mathbf{x}_n|q_k)$ pode ser facilmente obtido a partir de $P(q_k|\mathbf{x}_n)$ com o uso da regra de Bayes, como será visto na seção 3.3.

Como a ANN será utilizada para estimar probabilidades associadas a cada estado q_k de um HMM, então deve existir uma *equivalência um-a-um* entre as classes c_k's com $k = 1, \ldots, K$, e os estados q_k's $\in Q$ (conjunto de estados correspondentes a cada uma das unidades sub-lexicais).

O treinamento desta ANN deve ser realizado de forma supervisionada, e os exemplos de treinamento (vetor de entrada , vetor de saída (alvo desejado)) podem ser obtidos através da segmentação da base de dados (sentenças de treinamento) em termos de unidades sub-lexicais. Existem duas formas básicas para a realização desta segmentação : (1) segmentação manual, realizada por especialistas (pode ser bastante precisa, mas é muito demorada); (2) segmentação automática, em geral baseada em HMMs e alinhamentos de Viterbi (não é tão precisa quanto a manual, mas é muito mais rápida). Uma ANN

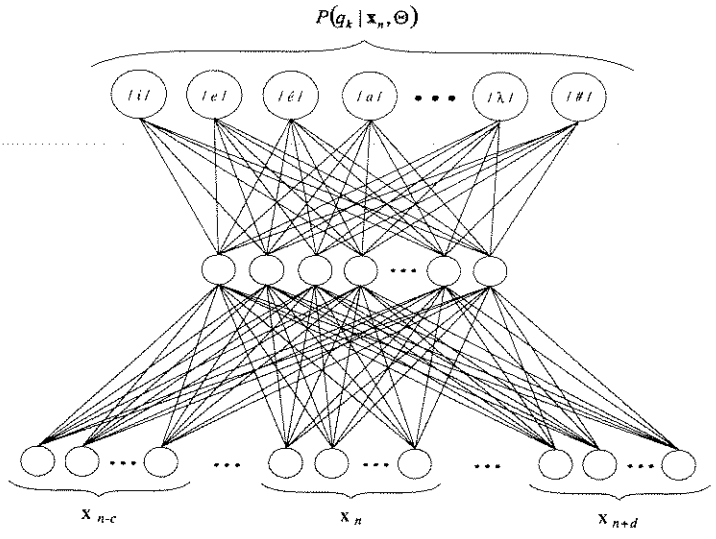

Figura 3-1: Rede MLP para estimativa das probabilidades das classes dado o vetor acústico de entrada e mais alguma informação contextual.

típica para estimar $P(q_k|\mathbf{x}_n)$ é mostrada na Figura 3-1.

A seguir é apresentada a demonstração formal, realizada por Bourlard, 1995 [22], de que uma ANN pode ser utilizada como um estimador da probabilidade da classe de saída dada a entrada, $P(classe|entrada)$.

Demonstração Seja $\mathbf{X}_e = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_e}\}$ o conjunto total de vetores acústicos para treinamento. Os pares entrada e saída para treinamento serão expressos por $(\mathbf{x}_i, q_c)$ com $\mathbf{x}_i \in \mathbf{X}_e$ e q_c = classe correta associada a $\mathbf{x}_i$, com $c \in \{1, 2, \ldots, K\}$. A $k-ésima$ saída da ANN dado o conjunto de parâmetros Θ e a entrada $\mathbf{x}_i$ será denotada por $g_k(\mathbf{x}_i, \Psi)$. O vetor de saída desejado (vetor alvo) para a entrada $\mathbf{x}_i$ será expresso por $\mathbf{d}(\mathbf{x}_i) = (d_1(\mathbf{x}_i), \ldots, d_k(\mathbf{x}_i), \ldots, d_K(\mathbf{x}_i))^t = \delta_{kc}(\mathbf{x}_i)$, sendo que $\delta_{kc}(\mathbf{x}_i) = 1$ para $k = c$ e $\delta_{kc}(\mathbf{x}_i) = 0$ para $k \neq c$, $k \in \{1, 2, \ldots, K\}$. A partir destas definições o valor médio do erro quadrático na saída da rede neural pode ser expresso por :

$$\mathcal{E} = \sum_{i=1}^{N_e} \sum_{c=1}^{K} \sum_{k=1}^{K} [\delta_{kc}(\mathbf{x}_i) - g_k(\mathbf{x}_i, \Psi)]^2 \cdot P(q_c, \mathbf{x}_i) \tag{3.5}$$

$$\mathcal{E} = \sum_{i=1}^{N_e} \sum_{k=1}^{K} \underbrace{\left\{ \sum_{c=1}^{K} [\delta_{kc}(\mathbf{x}_i) - g_k(\mathbf{x}_i, \Psi)]^2 \cdot P(q_c, \mathbf{x}_i) \right\}}_{\mathcal{E}_{kx}} \tag{3.6}$$

$$\mathcal{E} = \sum_{i=1}^{N_e}\sum_{k=1}^{K}\mathcal{E}_{kx} \tag{3.7}$$

De (3.5) e (3.9) pode ser verificado que minimizar $\mathcal{E}$ é equivalente a minimizar $\mathcal{E}_{kx}$.

Dividindo $\mathcal{E}_{kx}$ em duas partes, $k = c$ e $k \neq c$, resulta :

$$\begin{aligned}
\mathcal{E}_{kx} &= \sum_{c=1}^{K} P(q_c, \mathbf{x}_i)\cdot\delta_{kc}^2(\mathbf{x}_i) - \qquad (3.8)\\
&\quad 2\cdot P(q_c,\mathbf{x}_i)\cdot\delta_{kc}(\mathbf{x}_i)\cdot g_k(\mathbf{x}_i,\Psi) + P(q_c,\mathbf{x}_i)\cdot g_k^2(\mathbf{x}_i,\Psi)\\
&= P(q_k,\mathbf{x}_i) - 2\cdot P(q_k,\mathbf{x}_i)\cdot g_k(\mathbf{x}_i,\Psi) +\\
&\quad P(q_k,\mathbf{x}_i)\cdot g_k^2(\mathbf{x}_i,\Psi) + \sum_{\substack{c=1\\c\neq k}}^{K} P(q_c,\mathbf{x}_i)\cdot g_k^2(\mathbf{x}_i,\Psi) \qquad (3.9)\\
&= P(q_k,\mathbf{x}_i) - 2\cdot P(q_k,\mathbf{x}_i)\cdot g_k(\mathbf{x}_i,\Psi) +\\
&\quad P(q_k,\mathbf{x}_i)\cdot g_k^2(\mathbf{x}_i,\Psi) + (P(\mathbf{x}_i) - P(q_k,\mathbf{x}_i))\cdot g_k^2(\mathbf{x}_i,\Psi) \qquad (3.10)\\
&= P(q_k,\mathbf{x}_i) - 2\cdot P(q_k,\mathbf{x}_i)\cdot g_k(\mathbf{x}_i,\Psi) + P(\mathbf{x}_i)\cdot g_k^2(\mathbf{x}_i,\Psi) \qquad (3.11)
\end{aligned}$$

Utilizando o fato $P(q_k,\mathbf{x}_i) = P(q_k|\mathbf{x}_i)\cdot P(\mathbf{x}_i)$ e somando e subtraindo $P(q_k|\mathbf{x}_i)\cdot P(q_k,\mathbf{x}_i) = P(\mathbf{x}_i)\cdot P^2(q_k|\mathbf{x}_i)$ na expressão (3.11) tem-se,

$$\begin{aligned}
\mathcal{E}_{kx} &= P(\mathbf{x}_i)\cdot P^2(q_k|\mathbf{x}_i) - 2\cdot P(\mathbf{x}_i)\cdot P(q_k|\mathbf{x}_i)\cdot g_k(\mathbf{x}_i,\Psi) + \qquad (3.12)\\
&\quad P(\mathbf{x}_i)\cdot g_k^2(\mathbf{x}_i,\Psi) + P(q_k,\mathbf{x}_i) - P(q_k,\mathbf{x}_i)\cdot P(q_k|\mathbf{x}_i) \qquad (3.13)
\end{aligned}$$

que resulta em,

$$\mathcal{E}_{kx} = P(\mathbf{x}_i)\cdot(P(q_k|\mathbf{x}_i) - g_k(\mathbf{x}_i,\Psi))^2 + P(q_k,\mathbf{x}_i)\cdot(1 - P(q_k|\mathbf{x}_i)) \tag{3.14}$$

A partir de (3.14) pode ser verificado que $\mathcal{E}_{kx}$ será mínimo quando os parâmetros Ψ forem ajustados para que $g_k(\mathbf{x}_i,\Psi) = P(q_k|\mathbf{x}_i)$. Portanto, se uma ANN for otimizada segundo o critério MSE (ou entropia relativa, veja demonstração em [10]) suas **saídas ótimas** serão estimativas de probabilidades de classes (estados de uma HMM) condicionadas à entrada (vetor acústico), $\widehat{P}(q_k|\mathbf{x}_n)$:

$$g_k\left(\mathbf{x}_n, \Psi^{ótimo}\right) = \widehat{P}(q_k|\mathbf{x}_n) \tag{3.15}$$

Na Equação (3.15), $\Psi^{ótimo}$ representa o conjunto de parâmetros que minimiza tanto (3.3) quanto (3.4).

Porém, para que a ANN alcance estes valores ótimos, algumas condições de treinamento devem ser satisfeitas :

- A ANN deve possuir parâmetros "suficientes" para ser capaz de estimar com razoável aproximação a função de mapeamento *entrada* $\rightarrow$ *saída*.
- A rede não deve ser sobre-treinada. Isto pode ser evitado parando o treinamento antes do declínio da curva que expressa a capacidade de generalização da rede, [27].

Tem sido observado experimentalmente que, para sistemas treinados com grandes bases de dados, a saída de uma MLP treinada apropriadamente, de fato aproxima a probabilidade a posteriori, [22]. Esta conclusão pode facilmente ser estendida para outros casos. Por exemplo, se for fornecida à entrada de uma ANN o vetor acústico $\mathbf{x}_n$ no instante n, mais alguma informação contextual, expressa através do vetor, $\mathbf{X}_{n-c}^{n+d} = \{\mathbf{x}_{n-c}, \ldots, \mathbf{x}_n, \ldots, \mathbf{x}_{n+d}\}$, os valores de saída da MLP irão estimar

$$g_k\left(\mathbf{x}_n, \Psi^{ótimo}\right) = \widehat{P}\left(q_k|\mathbf{X}_{n-c}^{n+d}\right), \ \forall k = 1, \ldots, K \tag{3.16}$$

Este janelamento no tempo tem sido utilizado nos sistemas híbridos ANN-HMM para representar a correlação entre os vetores acústicos. Se a classe anterior é também fornecida à entrada da ANN, fornecendo uma rede recorrente, os valores na saída da MLP serão estimativas de

$$g_k\left(x_n, \Psi^{ótimo}\right) = \widehat{P}\left(q_k^n|\mathbf{X}_{n-c}^{n+d}, q_j^{n-1}\right), \ \forall k = 1, \ldots, K \tag{3.17}$$

sendo que q_j^{n-1} indica a classe vencedora no instante $n-1$ (saída da rede no instante anterior).

Será mostrado na Seção 3.5.2 que esta é uma forma de probabilidade local que pode ser utilizada na teoria de modelos híbridos ANN-HMM. Esta probabilidade será denotada como "*probabilidade de transição condicionada*".

Existe uma outra generalização importante desta propriedade que será essencial nesta tese. Se ANNs são treinadas para aproximar as probabilidades a posteriori dos estados de saída dada toda a seqüência de vetores acústicos $P\left(q_k|\mathbf{X}, \Theta\right)$ (em oposição aos alvos binários 0 e 1 impostos à K'*s* saídas da rede durante o período de treinamento), então (3.17) se mantém válido [14]. Em outras palavras, se algum sistema especialista independente fornecer alvos para uma ANN, a rede será capaz de aprender a produzir as probabilidades a posteriori destes novos alvos dado o vetor de entrada. Esta propriedade será utilizada no procedimento de reestimação de parâmetros dos modelos híbridos ANN-HMM, a ser discutido no Capítulo 4.

3.2.1 O Somatório das Saídas da ANN Deve Ser Igual a Um

Para que o vetor de saída da ANN $\mathbf{g}(\mathbf{x}_n, \Psi)$, para um determinado vetor acústico de entrada $\mathbf{x}_n$, possa ser intepretado como um função distribuição de probabilidade, então $\sum_k g_k(\mathbf{x}_n, \Psi) = 1$. Se a função logística for utilizada na camada de saída e se a rede convergir para um mínimo global então esta propriedade estará assegurada. Porém se a ANN convergir para um mínimo local então não haverá garantias que o somatório de suas saídas seja igual a 1. Para contornar este problema a literatura relata duas alternativas :

1. Após a otimização dos parâmetros Ψ realizar a seguinte normalização : $\overline{g}_k(\mathbf{x}_n, \Psi) = \frac{g_k(\mathbf{x}_n, \Psi)}{\sum_i g_i(\mathbf{x}_n, \Psi)}$, $\forall\, k \in \{1, 2, \ldots, K\}$.

2. Durante a otimização dos parâmetros, e também durante a avaliação das saídas da ANN, substituir a função logística da camada de saída pela função "$softmax$" [4] que possui a propriedade de garantir $\sum_k g_k(\mathbf{x}_n, \Psi) = 1$. Para a $k-ésima$ unidade (neurônio) da camada de saída a função "$softmax$" é definida como :

$$g_k(\mathbf{x}_n) = \frac{\exp(v_k(\mathbf{x}_n))}{\sum_{k=1}^{K} \exp(v_k(\mathbf{x}_n))} \tag{3.18}$$

e sua derivada pode ser calculada de forma extremamente simples, por :

$$g_k{}'(\mathbf{x}_n) = g_k(\mathbf{x}_n)\,(1 - g_k(\mathbf{x}_n)) \tag{3.19}$$

3.3 Estimação de Verossimilhanças para HMMs através de ANN

Uma vez que a saída da rede neural aproxima probabilidades Bayesianas, $g_k(\mathbf{x}_n, \Theta)$ é uma estimativa de

$$P(q_k|\mathbf{x}_n) = \frac{p(\mathbf{x}_n|q_k)}{p(\mathbf{x}_n)} \cdot P(q_k) \tag{3.20}$$

As verossimilhanças de emissão de símbolos $p(\mathbf{x}_n|q_k)$ a ser utilizadas pelos HMMs podem ser obtidas dividindo-se as saídas da rede, $g_k(\mathbf{x}_n)$, pela freqüência relativa da classe q_k no conjunto de treinamento $P(q_k)$, e multiplicando pela densidade de probabilidade dos símbolos, $p(\mathbf{x}_n)$.

$$p(\mathbf{x}_n|q_k) = \frac{P(q_k|\mathbf{x}_n)}{P(q_k)} \cdot p(\mathbf{x}_n) \tag{3.21}$$

Richard e Lippmann, [30], afirmam que como as probabilidades a priori das classes $P(q_k)$, que

aparecem em (3.20), representam apenas um termo multiplicativo, então estas probabilidades podem ser alteradas durante a fase de reconhecimento com o objetivo de compensar dados de treinamento com probabilidades de classes que não sejam representativas das condições de teste. No sistema implementado fez-se uso desta propriedade, porque a base de dados utilizada para treinamento do sistema não era balanceada, e o valor de $P(q_k)$ foi estimado segundo um outro procedimento conforme será discutido na Seção 6.3.2.

Durante o reconhecimento o valor de $p(\mathbf{x}_n)$ é constante para todas as classes (estados do HMM) e não exercerá qualquer influência sobre o processo de classificação. Portanto é usual utilizar em sistemas híbridos ANN-HMM a verossimilhança de emissão de símbolos normalizada

$$\frac{p(\mathbf{x}_n|q_k)}{p(\mathbf{x}_n)} = \frac{P(q_k|\mathbf{x}_n)}{P(q_k)} \tag{3.22}$$

A Equação (3.22) mostra que se forem conhecidas boas estimativas de $P(q_k)$ então o treinamento discriminativo da ANN (utilizando o critério MSE ou entropia relativa) irá garante, indiretamente, boas estimativas de $\frac{p(\mathbf{x}_n|q_k)}{p(\mathbf{x}_n)}$, isto é, boas estimativas das verossimilhanças de emissão de símbolos normalizadas $\frac{p(\mathbf{x}_n|q_k)}{p(\mathbf{x}_n)}$ estão condicionadas a bons estimadores tanto de probabiliades a posteriori $P(q_k|\mathbf{x}_n)$ quanto de probabilidades a priori $P(q_k)$.

3.4 Vantagens da ANN

ANNs possuem várias vantagens que as fazem particularmente atrativas para reconhecimento automático de fala, por exemplo :

- Podem providenciar aprendizado discriminativo entre unidades da fala ou estados de HMM que são representados por saídas de ANN treinadas como classificadoras de padrões. Isto é, quando treinadas para classificação (utilizando funções de custo como MSE ou entropia relativa), os parâmetros são ajustados para que as saídas desta ANN minimizem a taxa de erros e ao mesmo tempo maximizem a discriminação entre a classe de saída correta e as demais classes de saída. Em outras palavras, ANNs não se preocupam apenas em otimizar os parâmetros de cada classe a partir dos seus respectivos exemplos de treinamento, mas também em rejeitar os exemplos de treinamento não associados a ela. O critério de ML não apresenta este caráter discriminativo uma vez que não se preocupa em minimizar a taxa de erro.

- Uma vez que as ANNs podem incorporar múltiplas restrições e encontrar uma combinação ótima destas restrições para classificação, os vetores acústicos não necessitam ser assumidos independentes. Mais genericamente, não existe necessidade da forte suposição sobre as distribuições estatísticas dos dados de entrada, como usualmente é requerido por um HMM padrão.

- Ao contrário de HMMs treinados a partir do critério ML, as ANN não necessitam supor que as funções densidade de probabilidade de emissão de símbolos são formadas pela combinação de funções densidade de probabilidade gaussianas.

- Possuem arquiteturas bastante flexíveis que podem facilmente acomodar informações contextuais e realimentações, e também permitem o uso de entradas binárias ou contínuas.

- ANNs são estruturas altamente paralelas e regulares, o que fazem delas especialmente receptivas para implementações de arquitetura e hardware de alto-desempenho.

3.5 Tipos de Modelos Híbridos ANN-HMM

H. Bourlard e N. Morgan [22] classificam os sistemas híbridos em dois tipos básicos : (1) Modelos híbridos ANN-HMM padrão; (2) Modelos híbridos ANN-HMM discriminativo. Os modelos híbridos que têm sido discutidos até este momento e que foram simulados nesta Tese são do tipo padrão, mas com o objetivo de motivar trabalhos futuros com modelos híbridos discriminativos, também será apresentada uma breve discussão sobre este modelo. Para maiores informações sobre modelos híbridos ANN-HMM discriminativos, veja Konig [14].

3.5.1 Modelo Híbrido ANN-HMM Padrão

Consiste de um HMM contínuo com verossimilhanças de emissão de símbolos normalizadas $\frac{p(\mathbf{x}_n|q_k)}{p(\mathbf{x}_n)}$, estimadas por uma ANN. Neste caso o treinamento é realizado com um critério discriminativo, em geral a minimização do erro quadrático médio - MSE e estes HMMs permitem a avaliação de $\frac{p(\mathbf{X}|M_i,\Theta)}{p(\mathbf{X}|\Theta)}$,da verossimilhança normalizada de uma sentença $\mathbf{X}$ dado um determinado modelo M_i e o conjunto de parâmetros Θ. Este tipo de modelo é perfeitamente adequado ao modelamento acústico dos sistemas para reconhecimento de fala contínua discutidos no Capítulo 2.

O processo de avaliação dos modelos híbridos ANN-HMM padrão consiste em estimar o modelo oculto de Markov $\widehat{M}$, dados os parâmetros acústicos Θ, que apresente a maior verossimilhança de

gerar a seqüência de símbolos observados $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, isto é :

$$\widehat{M} = \arg\max_{M_i} p(\mathbf{X}|M_i, \Theta) \iff \arg\max_{M_i} \frac{p(\mathbf{X}|M_i, \Theta)}{p(\mathbf{X}|\Theta)} \tag{3.23}$$

Para expressar $\frac{p(\mathbf{X}|M_i,\Theta)}{p(\mathbf{X}|\Theta)}$ em função de parâmetros disponíveis, considere o desenvolvimento a seguir. Escrevendo $\frac{p(\mathbf{X}|M_i,\Theta)}{p(\mathbf{X}|\Theta)}$ em termos de probabilidade marginal, tem-se :

$$\frac{p(\mathbf{X}|M_i, \Theta)}{p(\mathbf{X}|\Theta)} = \frac{1}{p(\mathbf{X}|\Theta)} \cdot p(\mathbf{X}|M_i, \Theta) = \frac{1}{p(\mathbf{X}|\Theta)} \cdot \sum_{\Gamma \in \tau_i} p(\mathbf{X}, \Gamma|M_i, \Theta) \tag{3.24}$$

sendo τ_i o conjunto de todos as seqüências de estados permitidas pelo modelo M_i.

Manipulando (3.24) , tem-se :

$$\frac{p(\mathbf{X}|M_i, \Theta)}{p(\mathbf{X}|\Theta)} = \frac{1}{p(\mathbf{X}|\Theta)} \cdot \sum_{\Gamma \in \tau_i} p(\mathbf{X}|\Gamma, M_i, \Theta) \cdot P(\Gamma|M_i, \Theta) \tag{3.25}$$

$$= \sum_{\Gamma \in \tau_i} \frac{p(\mathbf{X}|\Gamma, M_i, \Theta)}{p(\mathbf{X}|\Theta)} \cdot P(\Gamma|M_i, \Theta) \tag{3.26}$$

$P(\Gamma|M_i, \Theta)$ pode ser determinado a partir das probabilidades de transição,

$$P(\Gamma|M_i, \Theta) = \pi_1 \cdot \prod_{n=1}^{N-1} P\left(q_\Gamma^{n+1}|q_\Gamma^n, M_i, \Theta\right) \tag{3.27}$$

$\frac{p(\mathbf{X}|\Gamma,M_i,\Theta)}{p(\mathbf{X}|\Theta)}$ pode ser reescrito como :

$$\frac{p(\mathbf{X}|\Gamma, M_i, \Theta)}{p(\mathbf{X}|\Theta)} = \frac{p(\mathbf{x}_1|\Gamma, M_i, \Theta) \cdot p(\mathbf{x}_2|\Gamma, M_i, \Theta, \mathbf{x}_1) \cdot \ldots \cdot p(\mathbf{x}_N|\Gamma, M_i, \Theta, \mathbf{x}_{N-1}, \ldots, \mathbf{x}_1)}{p(\mathbf{x}_1|\Theta) \cdot p(\mathbf{x}_2|\Theta, \mathbf{x}_1) \cdot \ldots \cdot p(\mathbf{x}_N|\Theta, \mathbf{x}_{N-1}, \ldots, \mathbf{x}_1)} \tag{3.28}$$

Considerando os símbolos não condicionados estatisticamente e lembrando que a verossimilhança de emissão de um símbolo normalizada no instante n depende apenas do estado neste mesmo instante n, então :

$$\frac{p(\mathbf{X}|\Gamma, M_i, \Theta)}{p(\mathbf{X}|\Theta)} = \prod_{n=1}^{N} \frac{p(\mathbf{x}_n|q_\Gamma^n, M_i, \Theta)}{p(\mathbf{x}_n|\Theta)} \tag{3.29}$$

Utilizando as Equações (3.27) e (3.29) $\frac{p(\mathbf{X}|M_i,\Theta)}{p(\mathbf{X}|\Theta)}$ pode ser expresso por :

$$\frac{p(\mathbf{X}|M_i, \Theta)}{p(\mathbf{X}|\Theta)} = \sum_{\Gamma \in \tau_i} \pi_1 \cdot \frac{p(\mathbf{x}_1|q_\Gamma^1, M_i, \Theta)}{p(\mathbf{x}_1|\Theta)} \cdot \prod_{n=1}^{N-1} P\left(q_\Gamma^{n+1}|q_\Gamma^n, M_i, \Theta\right) \cdot \frac{p\left(\mathbf{x}_{n+1}|q_\Gamma^{n+1}, M_i, \Theta\right)}{p(\mathbf{x}_{n+1}|\Theta)} \tag{3.30}$$

3.5.2 Modelo Híbrido ANN-HMM Discriminativo

Consiste de um "HMM" contínuo com *distribuições de probabilidades de transição condicionada* (veja Equação (3.17)) estimadas por uma ANN recursiva, conforme Figura 3.2. Neste caso o treinamento é realizado com um critério discriminativo, em geral a minimização do erro quadrático médio - MSE ou entropia relativa e estes "HMMs"[1] permitem a avaliação da probabilidade a posteriori do modelo M_i dada a sentença $\mathbf{X}$ e os parâmetros do modelo Θ, $P(M_i|\mathbf{X},\Theta)$. Este modelo é chamado discriminativo porque tanto o treinamento quanto a avaliação são realizados utilizando-se o método de maximização da probabilidade a posteriori MAP. Em sistemas de reconhecimento de fala contínua baseados em modelos híbridos ANN-HMM discriminativos utilizam-se explicitamente *informações da língua* L e neste caso o objetivo do sistema é determinar o modelo M_i com maior probabilidade a posteriori de ter gerado a seqüência de símbolos acústicos $\mathbf{X}$, dado os parâmetros Θ e L,

$$\widehat{M} = \arg\max_{M_i} P(M_i|\mathbf{X},\Theta,L) = \arg\max_{\Gamma\in\tau_i} \sum_{\Gamma} P(\Gamma|\mathbf{X},L,\Theta)\cdot P(M_i|\Gamma,\mathbf{X},L,\Theta) \tag{3.31}$$

Assumindo Γ independente da informação da língua L e M_i independente da seqüência acústica $\mathbf{X}$ se Γ for especificado, então,

$$P(M|\mathbf{X},\Theta,L) = \sum_{\Gamma\in\tau_i} P(\Gamma|\mathbf{X},\Theta)\cdot P(M_i|\Gamma,L,\Theta) \tag{3.32}$$

Neste caso $P(\Gamma|\mathbf{X},\Theta)$ é determinado a partir de um *modelamento acústico* de cada sentença e $P(M_i|\Gamma,L,\Theta)$ é determinado por um *modelo da língua.*

Desconsiderando as informações da língua L, o processo de avaliação dos modelos híbridos ANN-HMM discriminativos consiste em estimar o modelo oculto de Markov $\widehat{M}$ que maximiza a probabilidade a posteriori $P(M_i|\mathbf{X},\Theta)$, isto é, a probabilidade do modelo de Markov M dada a seqüência de símbolos $\mathbf{X} = \{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N\}$ e os parâmetros acústicos Θ,

$$\widehat{M} = \arg\max_{M_i} P(M_i|\mathbf{X},\Theta) \tag{3.33}$$

Considere o desenvolvimento proposto por H. Bourlard [4]. Escrevendo $P(M_i|\mathbf{X},\Theta)$ em termos de

[1] O termo HMMs foi colocado entre aspas por não se tratar exatamente de HMMs, pois fazem uso de probabilidades a posteriori e não de verossimilhanças de emissão de símbolos.

probabilidade marginal, resulta :

$$P(M_i|\mathbf{X},\Theta) = \sum_{\Gamma\in\Im} P(\Gamma, M_i|\mathbf{X},\Theta) \tag{3.34}$$

$$= \sum_{\Gamma\in\Im} P(M_i|\mathbf{X},\Gamma,\Theta)\cdot P(\Gamma|\mathbf{X},\Theta) \tag{3.35}$$

Sendo $\Im$ o conjunto de todas as seqüências de estado permitidas por todos os modelos $M_i = \{1,2,\ldots,I_M\}$.

Assumindo que M_i não precisa ser condicionado a $\mathbf{X}$ se Γ for especificado e que $\tau_i \cap \tau_j = \emptyset \ \forall\ i \neq j$, então :

$$P(M_i|\mathbf{X},\Theta) = \sum_{\Gamma\in\tau_i} P(\Gamma|\mathbf{X},\Theta) \tag{3.36}$$

porque neste caso $p(M_i|\Gamma,\Theta) = 0 \ \forall\ \Gamma \notin \tau_i$ e $P(M_i|\Gamma,\Theta) = 1 \ \forall\ \Gamma \in \tau_i$.

Reescrevendo $P(\Gamma|\mathbf{X},\Theta)$ como

$$P(\Gamma|\mathbf{X},\Theta) = P\left(q_\Gamma^2|\mathbf{X},q_\Gamma^1,\Theta\right)\cdot P\left(q_\Gamma^3|\mathbf{X},q_\Gamma^2,q_\Gamma^1,\Theta\right)\cdot\ldots \tag{3.37}$$

$$= \ldots\cdot P\left(q_\Gamma^N|\mathbf{X},q_\Gamma^{N-1},q_\Gamma^{N-2},\ldots,q_\Gamma^1,\Theta\right) \tag{3.38}$$

resulta,

$$P(\Gamma|\mathbf{X},\Theta) = \prod_{n=1}^{N} P\left(q_\Gamma^n|\mathbf{X},Q_1^{n-1},\Theta\right) \tag{3.39}$$

sendo que q_Γ^n representa o estado observado no tempo n e Q_1^N a seqüência de estados associada com $\mathbf{X}_1^N$. Portanto as probabilidades $P(\Gamma|\mathbf{X},\Theta) = P\left(q_\Gamma^1,q_\Gamma^2,\ldots,q_\Gamma^N|\mathbf{X},\Theta\right)$ podem ser calculadas em termos de probabilidades "locais" $P\left(q_\Gamma^n|\mathbf{X},Q_1^{n-1},\Theta\right)$. Estas probabilidades podem ser simplificadas se algumas restrições condicionais forem relaxadas : (1) Suposição de modelo de Markov de primeira ordem (estado atual dependente apenas do estado prévio), (2) Assumir que o estado q_Γ^n está condicionado apenas aos $c+d+1$ símbolos adjacentes a ele (o símbolo atual, os c anteriores e os d posteriores). Estas suposições permitem aproximar as probabilidades "locais" $P\left(q_\Gamma^n|\mathbf{X},Q_1^{n-1},\Theta\right)$ por:

$$P\left(q_\Gamma^n|\mathbf{X},Q_1^{n-1},\Theta\right) \approx P\left(q_\Gamma^n|\mathbf{X}_{n-c}^{n+d},q_\Gamma^{n-1},\Theta\right) \tag{3.40}$$

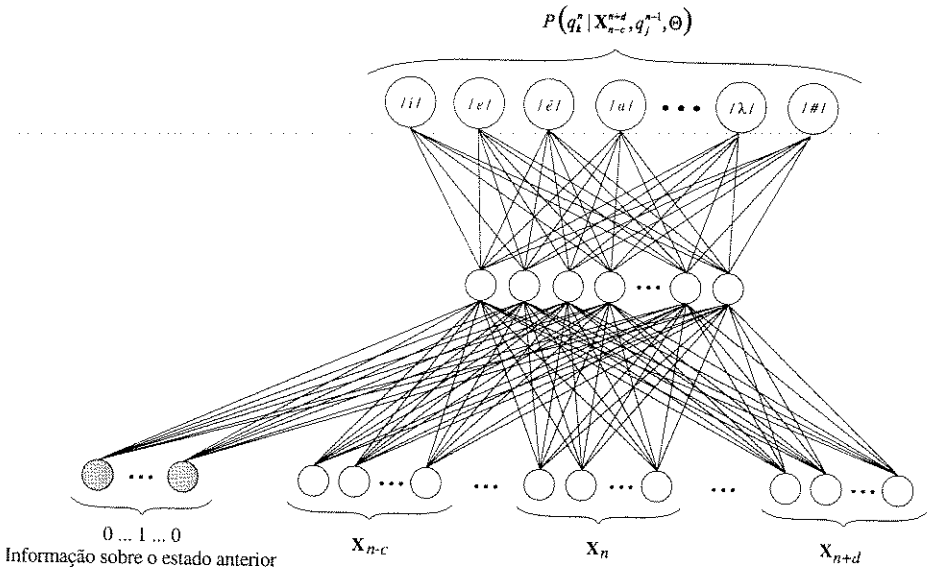

Figura 3-2: Rede neural recursiva para estimativa das "probabilidades de transição condicionadas" a serem utilizadas em um modelo híbrido ANN-HMM discriminativo.

Utilizando (3.39) e (3.40) em (3.36), $P(M_i|\mathbf{X},\Theta)$ pode ser aproximado por :

$$P(M_i|\mathbf{X},\Theta) \approx \sum_{\Gamma\in\tau_i}\prod_{n=1}^{N} P\left(q_\Gamma^n|\mathbf{X}_{n-c}^{n+d},q_\Gamma^{n-1},\Theta\right) \tag{3.41}$$

Estas probabilidades podem ser estimadas pelas saídas de uma ANN tendo com entrada $\mathbf{X}_{n-c}^{n+d}$ (o símbolo atual e mais informações contextuais) e mais uma realimentação (informações sobre o estado no instante anterior q_Γ^{n-1}). A Figura 3-2 mostra um esquema para esta rede neural.

Capítulo 4

REMAP

4.1 Introdução

A questão sobre o reconhecimento de unidades sub-lexicais ser sensível ou não a contexto tem sido estudada por Konig, [14] entre outros. Em seus experimentos Konig verificou o papel das transições dos formantes no reconhecimento de vogais. Especificamente foi testado se o reconhecimento de uma vogal é função, principalmente, da região estacionária dos formantes, ou do contexto acústico a curto-termo - direção e taxa de transição dos formantes adjacentes. Nos experimentos, ouvintes americanos foram submetidos à identificação de monossílabos sem sentido. Cada sílaba *consoante-vogal-consoante* (**CVC**) era constituída por uma seqüência de três elementos : *transição - vogal alvo - transição*. Os padrões de formantes das vogais alvo foram selecionados de segmentos estacionários das vogais [*i*] e [*u*]. A taxa e a direção das transições adjacentes foram variadas pela escolhas de dois contextos consonantais : [*w* - *w*] e [*j* - *j*]. Os resultados destes experimentos mostram que a identificação dos estímulos das vogais são determinados não apenas pelos padrões dos formantes na região estacionária, mas também pela direção das transições dos formantes adjacentes. Por exemplo o mesmo padrão de vogal : F1 = 350 Hz, F2 = 1578 Hz, F3 2604 Hz foi reconhecido por todos os ouvintes como [*i*] no contexto de [*j* **V** *j*] e foi reconhecido como [*u*] no contexto [*w* **V** *w*]. Em geral, foi mostrado que a categorização do contínuo é ajustada de acordo com os diferentes contextos para poder compensar o efeito de "Undershoot" no estímulo da vogal.

Motivado pelos resultados dos experimentos citados acima, Konig [14] propôs um método para reestimação dos parâmetros de um modelo híbrido ANN-HMM discriminativo baseado tanto em informações sobre as regiões estáveis das unidades sub-lexicais como nas regiões de transição entres estas unidades sub-lexicais. Em 1997 Cole [36] estendeu este método para a reestimação de parâmetros de

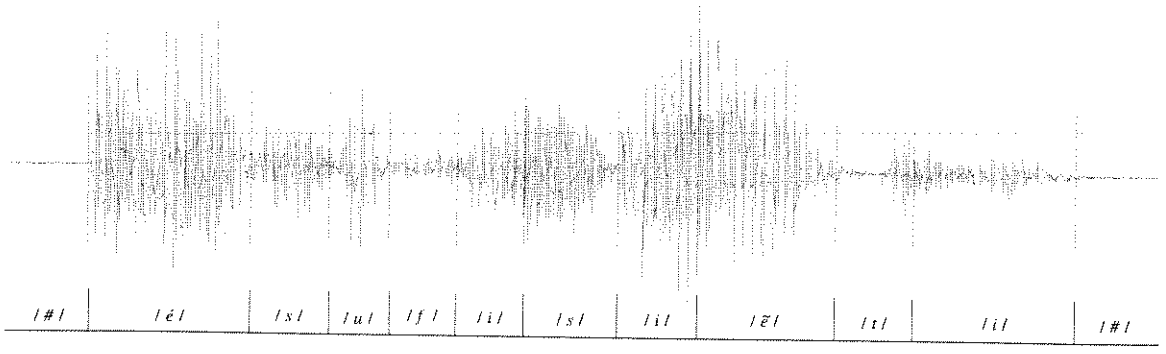

Figura 4-1: Sentença : " É suficiente " segmentada em termos de unidades sub-lexicais (fones).

um modelo híbrido ANN-HMM padrão. Nas próximas seções será apresentado e analisado o método proposto por Cole [36].

4.2 Estimação de Parâmetros - Pré-REMAP

Em um modelo híbrido ANN-HMM o treinamento da ANN é realizado de forma supervisionada, necessitando, portanto, de um conjunto de exemplos de treinamento $\mathbf{X}_e = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_e}\}$ e de seus correspondentes alvos desejados $\mathbf{d}_e(\mathbf{x}_n) = (d_{e_1}(\mathbf{x}_n), \ldots, d_{e_k}(\mathbf{x}_n), \ldots, d_{e_K}(\mathbf{x}_n))^t$, isto é, dos pares, $(\mathbf{x}_n, \mathbf{d}_e(\mathbf{x}_n))$. Em geral os exemplos de treinamento são acompanhados de informações contextuais e portanto a $n-$*éssima* entrada da ANN será $\{\mathbf{x}_{n-c}, \ldots, \mathbf{x}_n, \ldots, \mathbf{x}_{n+c}\} = \mathbf{X}_{n-c}^{n+c}$. Conforme já discutido do Capítulo 3 estes exemplos podem ser obtidos através da segmentação das sentenças de treinamento (base de dados) em termos de unidades sub-lexicais. Esta segmentação (manual ou automática) fornecerá marcas ao longo das sentenças indicando as posições inicial e final de cada uma das unidades sub-lexicais, conforme Figura 4-1.

Em geral as unidades sub-lexicais possuem durações distintas, porém o número de entradas da ANN é fixo, o que exige que os exemplos de treinamento $\mathbf{x}_i \in \mathbf{X}_e$ sejam extraídos com o uso de uma janela de tamanho fixo ($\mathbf{x}_i = \mathbf{X}_{n-c}^{n+c}$, com $c = cte$, $\forall i$). Um alternativa é tomar estas janelas com uma extensão igual à duração médias de todas as unidades sub-lexicais (previamente segmentadas) e centradas entre as marcas de início e fim de cada unidade (esta foi a abordagem utilizada neste trabalho). O conjunto de parâmetros da ANN treinada com o uso de $\mathbf{X}_e$ será denominado Ψ_e. A Figura 4-2 mostra um trecho da forma de onda da frase "É suficiente" e ilustra as posições onde devem ser retirados os exemplo de treinamento, para o caso de em que $c = 1$ (os exemplos de treinamentos correpondem a três janelas de análise).

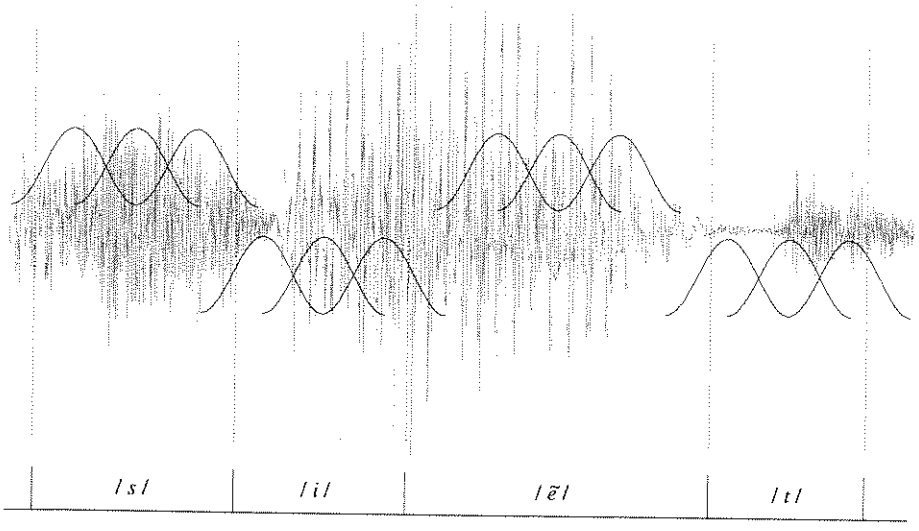

Figura 4-2: Ilustração do processo de retirada dos exemplos de treinamento da ANN, centrados entre as marcas de segmentação das unidades sub-lexicais.

Durante a fase de reconhecimento a ANN treinada com o conjunto de exemplos $\mathbf{X}_e$ descritos acima, terá que ser capaz de reconhecer vetores de entrada extraídos para cada janela de análise (em geral as janelas de análise são de 20 ms, mas tomadas a cada 10 ms). Muito provavelmente esta rede apresentará um bom desempenho para os vetores de entrada razoavelmente centrados em unidades sub-lexicais e um baixo desempenho para os vetores de entrada próximos a transições entre unidades sub-lexicais. Esta dificuldade em reconhecer regiões de transição é óbvia, pois tais informações não fazem parte do conjunto de treinamento. A Figura 4-3 mostra um trecho da sentença "É suficiente" e apresenta as posições dos vetores de entrada a serem utilizados durante a fase de reconhecimento (para o caso de $c = 1$). Para este exemplo a ANN deve apresentar um bom desempenho para as quadros de análise centrados nas unidades sub-lexicais, instantes : $n+1$, $n+4$, $n+7$, $n+8$, $n+9$ e $n+12$, e um desempenho ruim para os quadro de análise localizados nas transições entre estas unidades sub-lexicais, instantes $n+2$, $n+3$, $n+5$, $n+6$, $n+10$ e $n+11$. Uma solução para contornar este problema é retreinar a ANN com o uso de um novo conjunto de exemplos de treinamento $\mathbf{X}_e$', com exemplos extraídos a cada quadro de análise, obtendo assim exemplos de treinamento centrados nas marcas de segmentação e exemplos de treinamento localizados em regiões de transição entre as unidades sub-lexicais. A grande dificuldade para a definição deste novo conjunto de treinamento consiste na definição do conjunto dos sinais alvos correspondentes $\mathbf{d}_e$'.

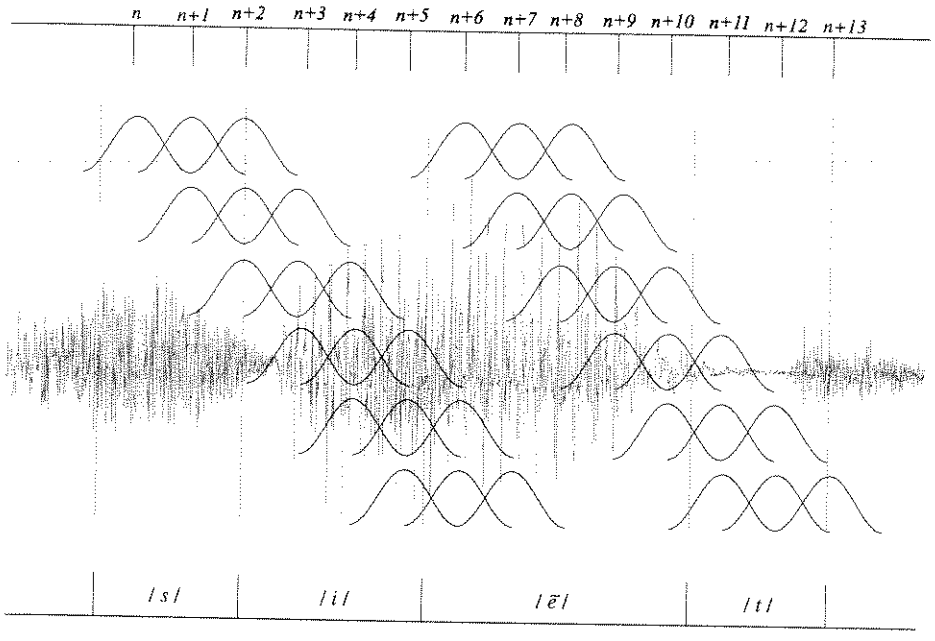

Figura 4-3: Exemplos utilizados durante a etapa de avaliação - exemplos tomados a cada quadro de análise.

4.3 Reestimação de Parâmetros - REMAP

4.3.1 Alvos Abruptos

A forma mais imediata para a determinação dos sinais alvos $\mathbf{d}_e$' a serem utilizados na reestimação dos parâmetros da ANN consiste em avaliar a ANN (com parâmetros Ψ_e) para cada uma das sentenças de treinamento e em seguida calcular a trajetória de Viterbi para cada uma destas sentenças e utilizar todos os pontos destas trajetórias, um para cada vetor acústico da sentença, como o conjunto de alvos $\mathbf{d}_e'(\mathbf{x}_n) = (d_{1_e}'(\mathbf{x}_n), \ldots, d_{k_e}'(\mathbf{x}_n), \ldots, d_{K_e}'(\mathbf{x}_n))^t$. O problema desta abordagem é que estes alvos representam decisões abruptas, isto é, um exemplo de treinamento $\mathbf{x}_i \in \mathbf{X}_e$' próximo a uma transição entre duas unidades sub-lexicais seria classificado pela ANN como tendo características pertencentes a apenas umas destas unidades sub-lexicais, desconsiderando quaisquer informações transicionais, que conforme foi discutido na introdução deste Capítulo, são cruciais para o processo de reconhecimento.

4.3.2 Alvos Suaves

Alvos mais adequados para a reestimação dos parâmetros da ANN devem ser capazes de apresentar uma mudança gradual de uma unidade sub-lexical para outra. Uma maneira de obter estes alvos é a proposta por Cole,1997 [36], e consiste basicamente na determinação das probabilidades a posteriori

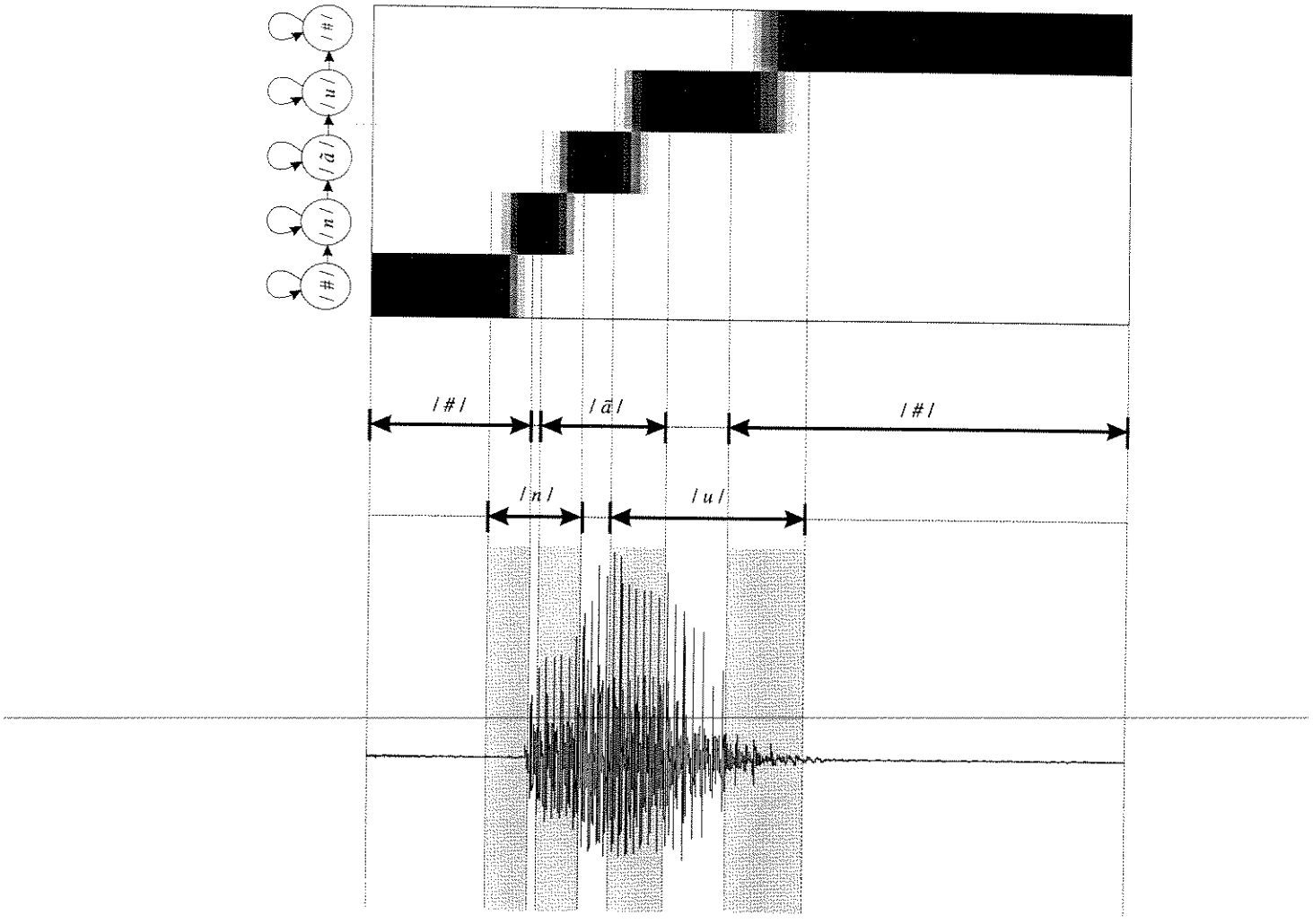

Figura 4-4: Alvos suaves obtidos segundo as probabilidades a posteriori dos estados dada toda a seqüência de vetores acústicos "probabilidades a posteriori globais".

globais de um estado q_k dada uma seqüência completa de vetores acústicos $\mathbf{X}$ correspondente a uma sentença M, isto é, o $k-ésimo$ alvo da ANN para o exemplo $\mathbf{x}_n$ será dado por $\gamma_n(k) = P(q_k^n|\mathbf{X}, M, \Theta)$. A Figura 4-4 mostra os alvos suaves[1] $\gamma_n(k)$ para o caso da sentença "Não"[2].

Para a obtenção destas probabilidades a posteriori globais basta avaliar a ANN (com parâmetros Ψ_e) para cada uma das sentenças e, de posse das probabilidades a posteriori locais, $P(q_k^n|\mathbf{x}_n, M, \Theta)$(saídas da ANN), utilizar o algoritmo $Foward-Backward$ para a determinação de $\gamma_n(k)$. A seguir será apresentado o algoritmo $Forward-Backward$ para a determinação destas probabilidades a posteriori

[1]Na Figura 4-4 os valores dos alvos suaves são dados em termos de tons de cinza : preto = 1; branco = 0.
[2]O símbolos /#/ corresponde ao silêncio.

globais.

Algoritmo Forward-Backward

O objetivo deste algoritmo é determinar as probabilidades a posteriori globais $\gamma_n(k) = P(q_k^n|\mathbf{X}, M, \Theta)$ em função das verossimilhanças de emissão de símbolos normalizada $\frac{p(\mathbf{x}_n|q_k^n, M, \Theta)}{p(\mathbf{x}_n|\Theta)}$. A seguir será apresentado um desenvolvimento detalhado para o cálculo de $\gamma_n(k)$ utilizando a verossimilhança de emissão de símbolos sem normalização e no final deste desenvolvimento será demonstrado que o uso do fator de normalização não acarreta qualquer alteração no cálculo de $\gamma_n(k)$.

$\gamma_n(k) = P(q_k^n|\mathbf{X}, M, \Theta)$ pode ser escrito como,

$$P(q_k^n|\mathbf{X}, M, \Theta) = \frac{P(q_k^n, \mathbf{X}|M, \Theta)}{p(\mathbf{X}|M, \Theta)} \tag{4.1}$$

$$= \frac{p\left(\mathbf{X}_1^n, \mathbf{X}_{n+1}^N, q_k^n|M, \Theta\right)}{p(\mathbf{X}|M, \Theta)} \tag{4.2}$$

$$= \frac{p(\mathbf{X}_1^n, q_k^n|M, \Theta) \cdot p\left(\mathbf{X}_{n+1}^N|\mathbf{X}_1^n, q_k^n, M, \Theta\right)}{p(\mathbf{X}|M, \Theta)} \tag{4.3}$$

Assumindo os vetores acústicos condicionalmente independentes, então :

$$P(q_k^n|\mathbf{X}, M, \Theta) = \frac{p(\mathbf{X}_1^n, q_k^n|M, \Theta) \cdot p\left(\mathbf{X}_{n+1}^N|q_k^n, M, \Theta\right)}{p(\mathbf{X}|M, \Theta)} \tag{4.4}$$

Definindo

$$\alpha_n(k) = p(\mathbf{X}_1^n, q_k^n|M, \Theta) \text{ e } \beta_n(k) = p\left(\mathbf{X}_{n+1}^N|, q_k^n, M, \Theta\right) \tag{4.5}$$

então :

$$\gamma_n(k) = P(q_k^n|\mathbf{X}, M, \Theta) = \frac{\alpha_n(k) \cdot \beta_n(k)}{p(\mathbf{X}|M, \Theta)} \tag{4.6}$$

Procedimento recursivo para estimar $\alpha_n(k)$:

$$\alpha_1(k) = \pi_k \cdot b_k(\mathbf{x}_1) \tag{4.7}$$

$$\alpha_n(k) = p(\mathbf{X}_1^n, q_k^n|M, \Theta) \tag{4.8}$$

$$\alpha_{n+1}(k) = p\left(\mathbf{X}_1^{n+1}, q_k^{n+1}|M, \Theta\right) \tag{4.9}$$

$$= p\left(\mathbf{X}_1^n, \mathbf{x}_{n+1}, q_k^{n+1}|M, \Theta\right) \tag{4.10}$$

$$= \sum_j p\left(\mathbf{X}_1^n, \mathbf{x}_{n+1}, q_j^n, q_k^{n+1}|M, \Theta\right) \tag{4.11}$$

$$= \sum_j p\left(\mathbf{x}_{n+1}, q_k^{n+1} | q_j^n, \mathbf{X}_1^n, M, \Theta\right) \cdot \overbrace{p\left(\mathbf{X}_1^n, q_j^n | M, \Theta\right)}^{\alpha_n(j)} \tag{4.12}$$

Como os vetores acústicos são condicionalmente independentes, tem-se :

$$\alpha_{n+1}(k) = \sum_j \alpha_n(j) \cdot p\left(\mathbf{x}_{n+1}, q_k^{n+1} | q_j^n, M, \Theta\right) \tag{4.13}$$

$$= \sum_j \alpha_n(j) \cdot p\left(\mathbf{x}_{n+1} | q_k^{n+1}, q_j^n, M, \Theta\right) \cdot P\left(q_k^{n+1} | q_j^n, M, \Theta\right) \tag{4.14}$$

$$= \sum_j \alpha_n(j) \cdot \underbrace{p\left(\mathbf{x}_{n+1} | q_k^{n+1}, M, \Theta\right)}_{b_k(\mathbf{x}_{n+1})} \cdot \underbrace{P\left(q_k^{n+1} | q_j^n, M, \Theta\right)}_{a_{jk}} \tag{4.15}$$

Portanto

$$\alpha_{n+1}(k) = \sum_j \alpha_n(j) \cdot a_{jk} \cdot b_k(\mathbf{x}_{n+1}) \tag{4.16}$$

Procedimento recursivo para estimar $\beta_n(k)$

$$\beta_n(k) = p\left(\mathbf{X}_{n+1}^N | q_k^n, M\right) \tag{4.17}$$

$$\beta_n(k) = p\left(\mathbf{X}_{n+2}^N, \mathbf{x}_{n+1} | q_k^n, M\right) \tag{4.18}$$

$$= \sum_j p\left(X_{n+2}^N, \mathbf{x}_{n+1}, q_j^{n+1} | q_k^n, M\right) \tag{4.19}$$

$$= \sum_j \frac{p\left(\mathbf{x}_{n+1}, q_k^n | q_j^{n+1}, \mathbf{X}_{n+2}^N, M\right) \cdot \overbrace{p\left(\mathbf{X}_{n+2}^N | q_j^{n+1}, M\right)}^{\beta_{n+1}(j)} \cdot P\left(q_j^{n+1}, M\right)}{P\left(q_k^n, M\right)} \tag{4.20}$$

Como os símbolos são não condicionados estatisticamente :

$$\beta_n(k) = \sum_j \beta_{n+1}(j) \cdot \frac{p\left(\mathbf{x}_{n+1}, q_k^n | q_j^{n+1}, M\right) \cdot p\left(q_j^{n+1}, M\right)}{P\left(q_k^n, M\right)} \tag{4.21}$$

$$= \sum_j \beta_{n+1}(j) \cdot \underbrace{p\left(\mathbf{x}_{n+1}, | q_j^{n+1}, M\right)}_{b_j(\mathbf{x}_{n+1})} \cdot \underbrace{p\left(q_j^{n+1} | q_k^n, M\right)}_{a_{kj}} \tag{4.22}$$

Trocando os índices j por k, apenas para melhorar a notação, tem-se :

$$\beta_n(j) = \sum_k \beta_{n+1}(k) \cdot a_{jk} \cdot b_k(\mathbf{x}_{n+1}) \tag{4.23}$$

De (4.16) é fácil observar que $p(\mathbf{X}|M,\Theta) = \sum_{\forall k} \alpha_N(k)$. Logo, resulta :

$$P(q_k^n|\mathbf{X},M,\Theta) = \frac{\alpha_n(k)\cdot\beta_n(k)}{\sum_k \alpha_N(k)} \tag{4.24}$$

A substituição de $b_k(\mathbf{x}_{n+1})$ (Equações (4.16) e (4.23) pela verossimilhança de emissão de símbolos normalizada $\frac{p(\mathbf{x}_n|q_k^n,M,\Theta)}{p(\mathbf{x}_n|\Theta)}$ transformam a probabilidade a posteriori global em,

$$\widehat{P}(q_k^n|\mathbf{X},M,\Theta) = \frac{\alpha_n(k)\cdot C_n\cdot\beta_n(k)\cdot D_n}{C_N\cdot\sum_k \alpha_N(k)} \tag{4.25}$$

onde : $C_n = \prod_{i=1}^{n} \frac{1}{p(\mathbf{x}_i)}$ e $D_n = \prod_{i=N}^{n+1} \frac{1}{p(\mathbf{x}_i)}$

Como $C_n \cdot D_n = C_N$, então pode ser verificado que o uso de verossimilhanças de emissão símbolos normalizadas não implica em qualquer alteração na probabilidade a posteriori global, isto é :

$$\widehat{P}(q_k^n|\mathbf{X},M,\Theta) = P(q_k^n|\mathbf{X},M,\Theta) \tag{4.26}$$

4.3.3 Algoritmo para Reestimação

O processo de reestimação de parâmetros de um modelo híbrido ANN-HMM pelo algoritmo REMAP pode ser sistematizado de acordo com os seguintes passos :

1. **Iteração $i = 0$ (Pré-REMAP)**. Iniciar com os parâmetros $\Psi^i = \Psi_e$, da ANN otimizada apenas com o uso dos exemplos de treinamento $\mathbf{X}_e$ centrados entre as marcas de segmentação das unidades sub-lexicais. Utilizar como probabilidades de transição entre os estados $P^{(i)}\left(q_k^n|q_j^{n-1}\right)$ e como probabilidades das classes $P^{(i)}(q_k)$, estimativas realizadas apenas em função de $\mathbf{X}_e$. Propostas para a estimativa destas probabilidades encontram-se no Capítulo 6, Equações (6.1) e (6.3), respectivamente.

2. Através da Equação (4.24) e dos parâmetros Ψ^i, $P^{(i)}\left(q_k^n|q_j^{n-1}\right)$ e $P^{(i)}(q_k)$ estimar os alvos suaves $P^{(i+1)}\left(q_k^n|\mathbf{X}_{M_j}^{(s)},M_j,\Theta^i\right)$ para cada uma das seqüências de treinamento $\mathbf{X}_{M_j}^{(s)}$ associadas a M_j. Realizar este procedimento para todos os modelos M_j.

3. Para cada uma das seqüências $\mathbf{X}_{M_j}^{(s)} = \left\{\mathbf{x}_1^{(j,s)},\mathbf{x}_2^{(j,s)},...,\mathbf{x}_N^{(j,s)}\right\}$ associadas a todas as sentenças M_j, com $j \in \mathcal{I}_{\mathcal{M}}$, otimizar os parâmetros $\Psi^{(i+1)}$ da ANN para minimizar o erro quadrático médio - MSE (ou entropia relativa), entre a saída da rede $g_k^{(i+1)}\left(\mathbf{x}_n^{(j,s)},\Psi^{(i+1)}\right)$ e o correspondente alvo

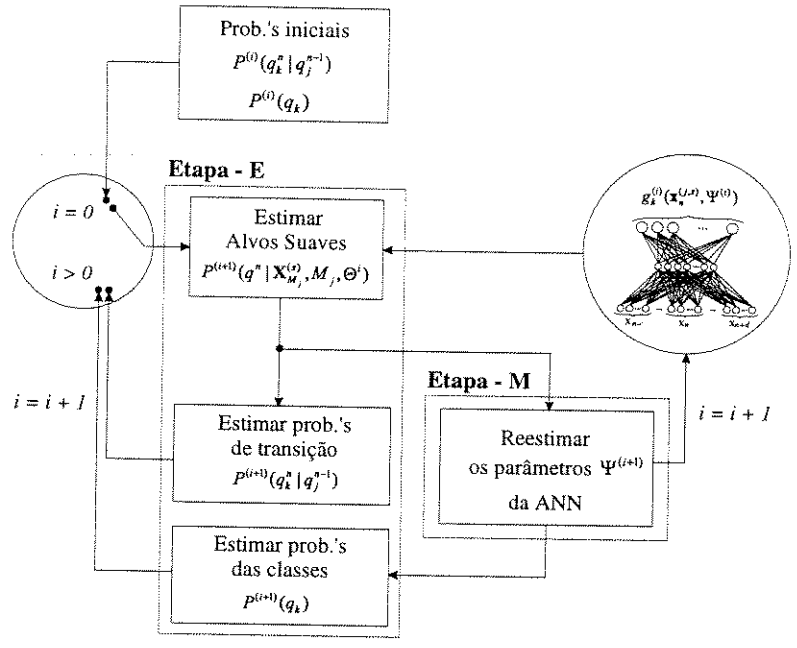

Figura 4-5: Diagrama de blocos do processo de reestimação de parâmetros REMAP.

suave $P^{(i+1)}\left(q_k^n|\mathbf{X}_{M_j}^{(s)}, M_j, \Theta^i\right)$. Este procedimento[3] fornecerá os novos parâmetros $\Psi^{(i+1)}$ da ANN.

4. A partir dos parâmetros $\Psi^{(i+1)}$ reestimados, calcular as novas probabilidades das classes $P^{(i+1)}(q_k)$, utilizando por exemplo, a Equação (6.4).

5. Ainda utilizando os parâmetros $\Psi^{(i)}$ e as probabilidades $P^{(i)}(q_k)$ e $P^{(i)}\left(q_k^n|q_j^{n-1}\right)$, estimar os valores das novas probabilidades de transição $P^{(i+1)}\left(q_k^n|q_j^{n-1}\right)$ conforme a Equação (4.34).

6. Se o sistema ainda *"não convergiu"*, fazer $i = i + 1$ e voltar ao passo 2.

Este procedimento pode ser dividido em duas etapas principais : (1) Estimação - **Etapa-E**, correspondente aos passos 2, 4 e 5; (2) Maximização - **Etapa-M**, correspondente ao passo 3. O algoritmo REMAP é ilustrado na Figura 4-5. O objetivo da **Etapa-M** é aumentar a cada iteração i, o número esperado de estados (saídas da ANN) individualmente "corretos"[4]. Para isto procura minimizar o erro quadrático médio entre a saída da rede, $g_k^{(i)}\left(\mathbf{x}_n, \Psi^{(i)}\right)$ e a função de probabilidade a posteriori global $P^{(i)}\left(q^n|\mathbf{X}_{M_j}^{(s)}, M_j, \Theta^{(i-1)}\right)$ que por sua vez maximiza este valor esperado [27].

[3]Este procedimento de treinamento *seqüência-por-seqüência* assume a independência de cada uma das seqüências de vetores acústicos em relação as demais.

[4]Corretos segundo o critério *"forward - backward"*.

4.3.4 Verossimilhanças de Emissão de Símbolos

A cada iteração do processo de reestimação, um novo conjunto de pesos sinápticos Ψ estará disponível, e as saídas da ANN, $g_k(\mathbf{x}_n, \Psi)$ serão novas estimativas das probabilidades a posteriori do estado q_k dado a entrada $\mathbf{x}_n$, isto é, $P(q_k|\mathbf{x}_n, \Theta)$. Estas estimativas poderão ser convertidas através da regra de Bayes e das novas probabilidades das classes $P(q_k)$ em verossimilhanças de emissão de símbolos normalizadas $\frac{p(\mathbf{x}_n|q_k,\Theta)}{p(\mathbf{x}_n|\Theta)}$. Porém neste caso as probabilidades das classes $P(q_k)$ não poderão ser obtidas através da freqüência relativa, uma vez que os alvos são suaves e não permitem a contagem direta da ocorrência ou não de q_k. Uma alternativa para a determinação de $P(q_k)$ é a proposta por Lippmann, 1991 [30], Seção 6.3.2, Equação (6.4).

4.3.5 Probabilidades de Transição

Considere que para um determinado modelo específico M_i,

$$\xi_n^{(i,s)}(j,k) = P\left(q_j^n, q_k^n | \mathbf{X}_{M_i}^{(s)}, M_i, \Theta\right) \tag{4.27}$$

sendo que $\mathbf{X}_{M_i}^{(s)} = \left\{\mathbf{x}_1^{(i,s)}, \mathbf{x}_2^{(i,s)}, \ldots, \mathbf{x}_{N_i}^{(i,s)}\right\}$ corresponde à *s-ésima* seqüência de vetores acústicos associada a M_i. Através de algumas manipulações simples $\xi_n^{(i,s)}(j,k)$ pode ser reescrito como[5] :

$$\xi_n^{(i,s)}(j,k) = \frac{P\left(q_j^n, q_k^n, \mathbf{X}_{M_i}^{(s)} | M_i, \Theta\right)}{p\left(\mathbf{X}_{M_i}^{(s)} | M_i, \Theta\right)} = \frac{\alpha_n^{(i,s)}(j) \cdot a_{jk} \cdot b_k\left(\mathbf{x}_{n+1}^{(i,s)}\right) \cdot \beta_{n+1}^{(i,s)}(k)}{\sum\limits_{\forall j} \alpha_N^{(i,s)}(j)} \tag{4.28}$$

sendo $\alpha_n^{(i,s)}(j)$ e $\beta_{n+1}^{(i,s)}(k)$ as variáveis *forward* e *backward* para o modelo M_i (conforme (4.5)) quando for apresentada a seqüência $\mathbf{X}_{M_i}^{(s)}$.

Considerando o somatório das probabilidades a posteriori globais, $\gamma_n^{(i,s)}(k) = P\left(q_k^n | \mathbf{X}_{M_i}^{(s)}, M_i, \Theta\right)$ ao longo do tempo, pode ser realizada a seguinte interpretação :

$$\sum_{n=1}^{N_i} \gamma_n^{(i,s)}(k) = \text{ número esperado de vezes em que o estado } q_k \text{ é visitado.} \tag{4.29}$$

[5] Assim como no caso de $\gamma_n^{(i,s)}$, o uso de verossimilhanças normalizadas no cálculo de $\xi_n^{(i)}$ implicará em fatores extras no numerador : C_n, $\frac{1}{p(\mathbf{x}_{n+1})}$ e D_{n+1} e também no denominador : C_{N_i} . Mas pode ser verificado que $C_n \cdot \frac{1}{p(\mathbf{x}_{n+1})} \cdot D_{n+1} = C_{N_i}$.

Se o somatório for até $N-1$, a Equação (4.29) pode ser interpretada da seguinte maneira,

$$\sum_{n=1}^{N_i-1} \gamma_n^{(i,s)}(k) \;=\; \text{número esperado de transições a partir de } q_k. \tag{4.30}$$

Esta afirmação baseia-se no fato que se o estado q_k é visitado, obrigatoriamente tem que haver uma transição a partir dele. Do mesmo modo, o somatório dos $\xi_n^{(i,s)}(j,k)$ pode ser interpretado da seguinte forma :

$$\sum_{n=1}^{N_i-1} \xi_n^{(i,s)}(j,k) = \text{número esperado de transições de } q_j \text{ para } q_k. \tag{4.31}$$

Utilizando-se as equações (4.30) , (4.31) e o conceito de freqüência relativa de ocorrência, a equação para reestimação das probabilidades de transição $P\left(q_k^n|q_j^n, M_i, \Theta\right)$ é dada por

$$P\left(q_k^n|q_j^n, M_i, \Theta\right) = \frac{\text{núm. esperado de transições do estado } q_j \text{ para } q_k}{\text{núm. esperado de transições a partir do estado } q_j} \tag{4.32}$$

Portanto, tem-se

$$P\left(q_k^n|q_j^n, M_i, \Theta\right) = \frac{\sum_{s=1}^{Ne_i}\left[\sum_{n=1}^{N_i-1} \xi_n^{(i,s)}(j,k)\right]}{\sum_{s=1}^{Ne_i}\left[\sum_{n=1}^{N_i-1} \gamma_n^{(i,s)}(j)\right]} \tag{4.33}$$

A Equação 4.33 refere-se apenas a uma única sentença M_i, porém durante a reestimação deseja-se realizar o cálculo das probabilidades de transição levando-se em consideração todas as sentenças de treinamento M_i com $i = \{1, 2, \ldots, I_M\}$. Para isto, pode ser utilizado o conceito de treinamento com múltiplas observações [27] que assume a independência entre as seqüências de vetores acústicos associadas aos modelos M_i. Como as fórmulas de reestimação das probabilidades de transição baseiam-se no conceito de freqüência relativa de ocorrência, a equação para reestimação a partir de múltiplas observações pode ser obtida realizando-se o quociente entre o número esperado de transições do estado q_j para o estado q_k e o número esperado de transições a partir do estado q_j , para todas as sentenças de treinamento,

$$a_{jk} = \frac{\sum_{i=1}^{I_M}\left\{\sum_{s=1}^{Ne_i}\left[\sum_{n=1}^{N_i-1} \xi_n^{(i,s)}(j,k)\right]\right\}}{\sum_{i=1}^{I_M}\left\{\sum_{s=1}^{Ne_i}\left[\sum_{n=1}^{N_i-1} \gamma_n^{(i,s)}(j)\right]\right\}} \tag{4.34}$$

Capítulo 5

Algoritmos de Busca

5.1 Introdução

O problema de reconhecimento de uma sentença falada de forma contínua baseado na concatenação de modelos de palavras (ou de modelos de unidades sub-lexicais) é extremamente importante para a tarefa de reconhecimento automático de fala. Uma grande variedade de algoritmos para a solução deste problema, em geral baseados em técnicas de programação dinâmica - DP ("Dynamic Program"), tem sido proposta e avaliada nos últimos anos [23, 29, 17]. O primeiro algoritmo para reconhecimento de sentenças a partir da concatenação de modelos de palavras foi proposto por Vintsyuk [35], que mostrou como a técnica de DP poderia ser utilizada para determinar a seqüência ótima de palavras associada a uma determinada seqüência de vetores acústicos observados. O procedimento proposto por Vintsyuk processava os sinais de fala de forma síncrona com os vetores acústicos, e seu trabalho pioneiro estabeleceu os fundamentos para as várias soluções baseadas em programação dinâmica para reconhecimento de fala. Vintsyuk também propôs um esquema rudimentar para incorporar restrições entre as palavras durante a pesquisa (regras gramaticais).

Atualmente existe uma variedade de algoritmos para encontrar a "melhor" seqüência de palavras associada a uma determinada seqüência de vetores acústicos observados. Entre eles destacam-se o "Stack Algorithm" desenvolvido por Jelinek [13]; o "Level Building" de Rabiner [29] e o "One Stage" (também conhecido como "Frame Synchronous Level Building") de Ney [23] e Lee & Rabiner [17]. Todos estes algoritmos são capazes de utilizar elaboradas regras de restrições gramaticais durante o procedimento de busca pela "melhor" seqüência de palavras.

As diferenças entre estes métodos são principalmente de implementação (por exemplo: síncrono por vetor acústico ou síncrono por palavras), e de características (por exemplo: diferentes formas de impor

restrições de duração, ou diferentes procedimentos para geração das sentenças a serem analisadas a partir dos modelos das palavras ou das unidades sub-lexicais).

5.2 Level Building

5.2.1 Motivação

Sendo $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ uma seqüência de vetores acústicos correspondente a uma sentença desconhecida e sendo M o modelo oculto de Markov associado a uma determinada sentença, então o objetivo do algoritmo "Level Building" - LB, é avaliar de forma eficiente a seguinte expressão

$$\widehat{M} = \arg\max_{M} p(\mathbf{X}|M,\Theta) \cdot P(M|\Theta) \tag{5.1}$$

Isto corresponde a estimar o modelo $\widehat{M}$ que maximiza o produto entre a verossimilhança da seqüência de símbolos $\mathbf{X}$ dado o modelo M, e o modelo da língua $P(M|\Theta)$.

Considerando o caso em que todas as palavras do léxico podem seguir quaisquer outras palavras com probabilidade igual a um (sem uso de restrições gramaticais),

$$P(M|\Theta) = P(W_1|\Theta) \cdot \prod_{i=2}^{N_M} P(W_i|W_{i-1}, W_{i-2}, \ldots, W_1, \Theta) = 1 \tag{5.2}$$

onde N_M é número de palavras presente em M, então o processo de decodificação resume-se a,

$$\widehat{M} = \arg\max_{M} p(\mathbf{X}|M,\Theta) \tag{5.3}$$

Neste caso será dito que a estimação de $\widehat{M}$ será realizada utilizando-se apenas a *decodificação acústica.*

Utilizando-se o conceito de probabilidade marginal, cada modelo M da Equação (5.3) pode ser avaliado pela seguinte expressão :

$$p(\mathbf{X}|M,\Theta) = \sum_{\Gamma \in \tau} p(\mathbf{X}, \mathbf{\Gamma}|M,\Theta) \tag{5.4}$$

sendo τ o conjunto de todas as seqüências de estado de comprimento N permitidas por M.

Porém o algoritmo LB avalia (5.4) fazendo uso da aproximação de Viterbi,

$$p(\mathbf{X}|M,\Theta) \approx p(\mathbf{X}, \mathbf{\Gamma}_v|M,\Theta) \tag{5.5}$$

sendo que,

$$\Gamma_v = \arg\max_{\Gamma \in \tau} p\left(\mathbf{X}, \mathbf{\Gamma}|M, \Theta\right) \tag{5.6}$$

onde Γ_v é denominada seqüência de estados (ou caminho) de Viterbi.

Além de utilizar a aproximação em (5.5) o algoritmo LB explora o fato dos modelos de sentenças poderem ser interpretados como a concatenação de modelos de palavras, $M = \{W_1, W_{2,\ldots,}W_{N_M}\}$. Para entender como isto pode ser feito, acompanhe o seguinte desenvolvimento.

Reescrevendo $p\left(\mathbf{X}, \mathbf{\Gamma}_v|M, \Theta\right)$ como,

$$p\left(\mathbf{X}, \mathbf{\Gamma}_v|M, \Theta\right) = p\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, q^1_{\Gamma_v}, q^2_{\Gamma_v}, \ldots, q^N_{\Gamma_v}|M, \Theta\right) \tag{5.7}$$

$$= \frac{p\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, q^1_{\Gamma_v}, q^2_{\Gamma_v}, \ldots, q^N_{\Gamma_v}, M|\Theta\right)}{P\left(M|\Theta\right)} \tag{5.8}$$

definindo : $1 < n_1 < n_2 < \ldots < n_{N_M} < N$, e realizando algumas manipulações simples com (5.8), tem-se :

$$\begin{aligned} p\left(\mathbf{X}, \mathbf{\Gamma}_v|M, \Theta\right) = \; & p\left(\mathbf{x}_1, \ldots, \mathbf{x}_{n_1}, q^1_{\Gamma_v}, \ldots, q^{n_1}_{\Gamma_v}|\mathbf{x}_{n_1+1}, \ldots, \mathbf{x}_N, q^{n_1+1}_{\Gamma_v}, \ldots, q^N_{\Gamma_v} M, \Theta\right) \cdot \\ & p\left(\mathbf{x}_{n_1+1}, \ldots, \mathbf{x}_{n_2}, q^{n_1+1}_{\Gamma_v}, \ldots, q^{n2}_{\Gamma_v}|\mathbf{x}_{n_2+1}, \ldots, \mathbf{x}_N, q^{n_2+1}_{\Gamma_v}, \ldots, q^N_{\Gamma_v} M, \Theta\right) \cdot \\ & \vdots \\ & p\left(\mathbf{x}_{n_{N_M}+1}, \ldots, \mathbf{x}_N, q^{n_{N_M}+1}_{\Gamma_v}, \ldots, q^N_{\Gamma_v}|M, \Theta\right) \end{aligned} \tag{5.9}$$

Assumindo a independência condicional dos vetores acústicos, e a hipótese de Markov de primeira ordem, então (5.10) pode ser reescrito como :

$$\begin{aligned} p\left(\mathbf{X}, \mathbf{\Gamma}_v|M, \Theta\right) = \; & p\left(\mathbf{x}_1, \ldots, \mathbf{x}_{n_1}, q^1_{\Gamma_v}, \ldots, q^{n_1}_{\Gamma_v}|M, \Theta\right) \cdot \\ & p\left(\mathbf{x}_{n_1+1}, \ldots, \mathbf{x}_{n_2}, q^{n_1+1}_{\Gamma_v}, \ldots, q^{n2}_{\Gamma_v}|M, \Theta\right) \cdot \\ & \vdots \\ & p\left(\mathbf{x}_{n_{N_M}+1}, \ldots, \mathbf{x}_N, q^{n_{N_M}+1}_{\Gamma_v}, \ldots, q^N_{\Gamma_v}|M, \Theta\right) \end{aligned} \tag{5.10}$$

Se n_1 for definido tal que a seqüência $q^1_{\Gamma_v}, \ldots, q^{n_i}_{\Gamma_v}$ corresponda aos estados percorridos ao longo do modelo W_1 e os demais n_i's sejam definidos de tal forma que as seqüências $q^{n_{i-1}+1}_{\Gamma_v}, \ldots, q^{n_i}_{\Gamma_v}$ correspondam aos estados percorridos ao longo dos modelos W_i 's, $2 \leq i \leq N_M$, então (5.10) pode ser reescrito

como :

$$p\left(\mathbf{X},\mathbf{\Gamma}_v|M,\Theta\right) = p\left(\mathbf{X}_1^{n_1},\mathbf{\Gamma}_{v_1}|W_1,\Theta\right)\cdot p\left(\mathbf{X}_{n_1+1}^{n_2},\mathbf{\Gamma}_{v_2}|W_2,\Theta\right)\cdot \ldots \cdot p\left(\mathbf{X}_{n_{N_M}+1}^{N},\mathbf{\Gamma}_{v_{N_M}}|W_{N_M},\Theta\right) \tag{5.11}$$

sendo $\mathbf{\Gamma}_{v_1}$ o caminho de Viterbi ao longo de W_1 começando em $n=1$ e finalizando em $n=n_1$; e $\mathbf{\Gamma}_{v_i}$ o caminho de Viterbi ao longo de W_i começando em $n=n_{i-1}+1$ e finalizando em $n=n_i$, com $2\leq i\leq N_M$.

O algoritmo LB explora justamente a propriedade expressa pela Equação (5.11), que afirma que $p\left(\mathbf{X},\mathbf{\Gamma}_v|M,\Theta\right)$ pode ser calculada como o produto das verossimilhanças, ao longo do caminho de Viterbi, de cada uma das N_M palavras presentes na sentença M.

5.2.2 Algoritmo Level Building

Nesta seção será apresentado o algoritmo LB para realização somente da *decodificação acústica*, (caso em que $P\left(M|\Theta\right)=1$). Durante o processo de decodificação o algoritmo LB desconhece a princípio o número de palavras que compõem a sentença a ser reconhecida e portanto torna-se necessário estabelecer uma quantidade mínima e máxima de palavras a serem procuradas. No desenvolvimento que será apresentado a seguir o número mínimo será de uma palavra e o número máximo será de L palavras. Esta faixa de procura de 1 à L será denominada de *níveis de procura.*

Nível 1 (Primeira palavra da sentença) : Calcula-se $p\left(\mathbf{X}|W_i,\Theta\right)$, para todas as palavras W_i do léxico, iniciando-se no instante $n=1$. Este cálculo é realizado utilizando-se o algoritmo de Viterbi, Figura 5-1, de acordo com os seguintes passos:

1. **Inicialização** - Definir $\delta_n\left(k\right)$ como a probabilidade conjunta das seqüências parciais de estado e de observação, $p\left(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n,q_{\Gamma_i}^1,q_{\Gamma_i}^2,\ldots,q_{\Gamma_i}^n|W_i,\Theta\right)$ e fazer

$$\begin{aligned}\delta_1\left(1\right) &= p\left(\mathbf{x}_1,q_1|W_i,\Theta\right)\\ \delta_1\left(k\right) &= 0,\ k=2,3,\ldots,N_{W_i}\\ \alpha_n\left(N_{W_i}\right) &= 1,\ n=1,2,\ldots,N\end{aligned} \tag{5.12}$$

sendo N_{W_i} o número de estados na palavra W_i, e $\alpha_n\left(N_{W_i}\right)=1$ a indicação de que todas as trajetórias de Viterbi no primeiro nível devem começar no instante $n=1$.

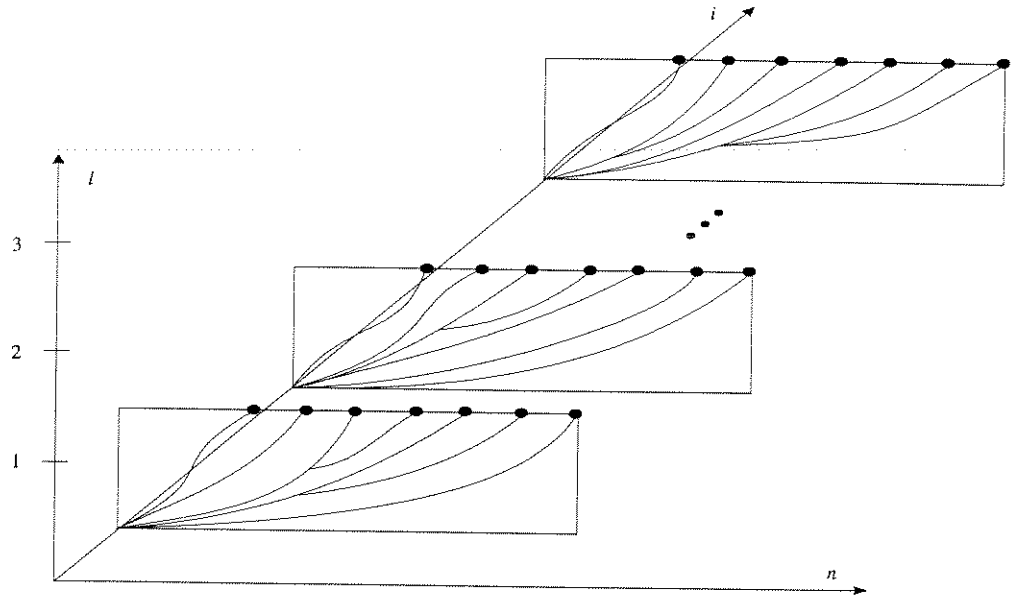

Figura 5-1: Avaliação de todas as palavras do léxico (eixo i) no primeiro nível de procura.

2. **Recursão** - para $2 \leq n \leq N$, e $1 \leq k \leq N_{W_i}$

$$\delta_n(k) = \max_{1 \leq j \leq N_{W_i}} \left[\delta_{n-1}(j) \cdot P\left(q_k^n | q_j^{n-1}, W_i, \Theta\right)\right] \cdot p\left(\mathbf{x}_n | q_k, W_i\right) \tag{5.13}$$

3. **Término** -

$$\begin{aligned} \mathbf{P}(l,n,i) &= \delta_n(N_{W_i}),\ 1 \leq n \leq N \\ \mathbf{C}(l,n,i) &= \alpha_n(N_{W_i}),\ 1 \leq n \leq N \\ \mathbf{T}(l,n,i) &= P\left(q_k^n | q_{N_{W_i}}^{n-1}, W_i, \Theta\right),\ 1 \leq n \leq N \end{aligned} \tag{5.14}$$

De acordo com o passo 3, as verossimilhanças acumuladas (para o último estado de cada palavra) para todos os instantes de tempo, devem ser guardadas em uma matriz tridimensional $\mathbf{P}$. As coordenadas desta matriz são : l - indicador de nível; n - instante de tempo (no espaço de símbolos) e i identificador da palavra analisada. A matriz $\mathbf{C}$ guardará os instantes iniciais (do nível anterior) dos caminhos de Viterbi finalizados em cada instante de tempo n (do nível atual) para todas as palavras avaliadas. Finalmente, para cada uma das palavras W_i avaliadas, a matriz $\mathbf{T}$ deve armazenar as probabilidades de transição do estado $q_{N_{W_i}}$ (probabilidade de transição do último estado da palavra W_i).

O nível $l = 1$ deve ser realizado para todas as palavras do léxico conforme descrito pelos 3 passos acima. Após todas as palavras serem analisadas deve ser realizado o que será chamado *redução de*

nível (eliminação do eixo i), formando os seguintes vetores,

$$\widehat{\mathbf{P}}(l,n) = \max_i \left[\mathbf{P}(l,n,i)\right] \tag{5.15}$$

$$\widehat{\mathbf{C}}(l,n) = \mathbf{C}\left[l,n,\arg\max_i \left(\mathbf{P}(l,n,i)\right)\right] \tag{5.16}$$

$$\widehat{\mathbf{W}}(l,n) = \arg\max_i \left[\mathbf{P}(l,n,i)\right] \tag{5.17}$$

$$\widehat{\mathbf{T}}(l,n) = \mathbf{T}\left(l,b,\widehat{\mathbf{W}}(l,n)\right) \tag{5.18}$$

sendo que $\widehat{\mathbf{P}}$ selecionará, para cada instante de tempo n, o valor da maior verossimilhança $\delta_n(N_{W_i})$ para todas as palavras W_i analisadas no nível l. $\widehat{\mathbf{W}}$ indicará, para cada instante de tempo n, qual a palavra W_i, que apresentou a maior verossimilhança $\delta_n(N_{W_i})$. $\widehat{\mathbf{C}}$ armazenará o instante n do nível anterior $(l-1)$ onde foi iniciado o caminho de Viterbi associado a palavra $\widehat{\mathbf{W}}(l,n)$ (no caso do primeiro nível este valor será igual a 1). $\widehat{\mathbf{T}}$ indicará, para cada instante de tempo n, a probabilidade de transição do último estado da palavra $\widehat{\mathbf{W}}(l,n)$.

Nível 2 (e demais níveis) : Os cálculos envolvidos no nível 2, e em todos os níveis restantes, diferem do nível 1 apenas no primeiro passo, correspondente ao procedimento de inicialização,

$$\delta_1(1) = 0, \tag{5.19}$$

$$\delta_n(1) = \max\left[\widehat{\mathbf{T}}(l-1,n-1)\cdot\widehat{\mathbf{P}}(l-1,n-1), P\left(q_1^n|q_1^{n-1},W_i,\Theta\right)\cdot\delta_{n-1}(1)\right]\cdot \cdot p(\mathbf{x}_n|q_k,W_i,\Theta) \tag{5.20}$$

com $2 \leq n \leq N$.

$$\alpha_n(1) = \begin{cases} n-1, \text{ se } \widehat{\mathbf{T}}(l-1,n-1)\cdot\widehat{\mathbf{P}}(l-1,n-1) > P\left(q_1^n|q_1^{n-1},W_i,\Theta\right)\cdot\delta_{n-1}(1) \\ \alpha_{n-1}(1), \text{ caso contrário} \end{cases} \tag{5.21}$$

A Equação (5.19) ajusta $\delta_1(1) = 0$ e a Equação (5.20) faz a ligação do primeiro estado do nível atual com o último estado do nível anterior. O vetor α_n representado pela Equação (5.21) guarda as informações necessárias para o resgate dos instantes do nível anterior onde foram iniciados os caminhos de Viterbi associados a cada instante n do nível atual. Durante o procedimento de recursão α_n deve ser ajustada por,

$$\alpha_n(k) = \alpha_{n-1}\cdot\left(\arg\max_{1\leq i\leq N}\left(\delta_{n-1}(i)\cdot P\left(q_k^n|q_j^{n-1},W_i,\Theta\right)\right)\right) \tag{5.22}$$

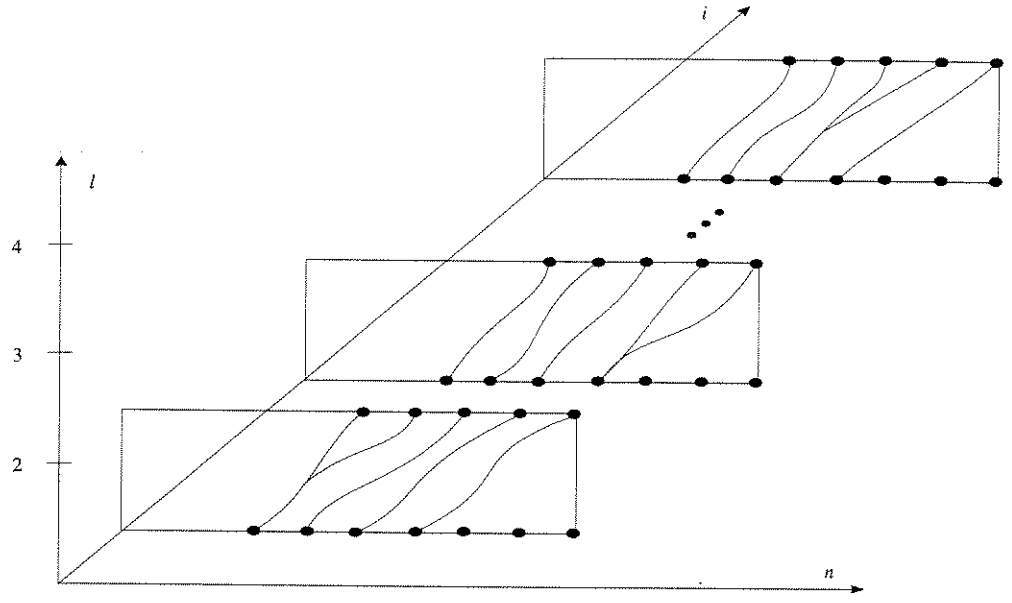

Figura 5-2: Avaliação de todas as palavras do léxico (eixo i) no segundo nível de procura.

Após a aplicação de todas as palavras neste nível, conforme ilustrado na Figura 5-2, os vetores $\widehat{\mathbf{P}}$, $\widehat{\mathbf{C}}$, $\widehat{\mathbf{W}}$ e $\widehat{\mathbf{T}}$ devem ser calculados (redução de nível), conforme as equações (5.15), (5.16), (5.17) e (5.18), e então seguir para análise do próximo nível.

Tecnicamente falando, ambos $\delta_n(k)$ e $\alpha_n(k)$ são funções do índice da palavra i. Entretanto, uma vez que as informações relevantes já foram salvas nos vetores $\mathbf{P}(l,n,i)$, $\mathbf{C}(l,n,i)$ e $\mathbf{T}(l,n,i)$, o uso explícito da dependência em relação à palavra i não é necessário nestas funções.

Todo este procedimento termina quando for realizado o último nível L, previamente especificado. Para identificar quantas palavras N_M existem na seqüência de palavras com maior verossimilhança, basta realizar a seguinte análise,

$$N_M = \arg\max_{1 \le l \le L} \widehat{\mathbf{P}}(l,N) \tag{5.23}$$

Em seguida para encontrar quais são as N_M palavras associadas a esta seqüência mais provável, basta realizar o caminho reverso ("*Backtracking*") com o uso do vetor $\widehat{\mathbf{B}}(l,n)$.

5.2.3 Incorporando Restrições de Duração ao Algoritmo Level Building

Durante a análise de um determinado *nível* o algoritmo LB avalia cada modelo de palavra sob inúmeras condições de compressão e de expansão. Porém, muitas destas compressões e expansões correspondem a durações que diferem demasiadamente das durações médias das palavras presentes nas sentenças de

treinamento do sistema. Desta maneira palavras como "suficiente" podem ser comprimidas excessivamente a ponto de serem confundidas com a palavra "em" e vice-versa. Uma possível solução para este problema é utilizar um modelo de duração para as palavras que leve em consideração a duração média e a variância destas palavras no conjunto de treinamento.

Rabiner, 1985 [29] propõe um modelo de duração bastante simples que associa à $i-$*ésima* palavra do léxico, uma função densidade de probabilidade gaussiana $f_i(D)$,

$$f_i(D) = \frac{1}{\sigma_i\sqrt{2\pi}} \cdot \exp\left(-\frac{\left(D-\overline{D_i}\right)^2}{2\sigma_i^2}\right)$$

onde , $\overline{D_i}$ e σ_i^2, representam, respectivamente, a média e a variância, da palavra W_i, obtidas a partir das sentenças de treinamento.

Este modelo de duração pode ser incorporado ao algoritmo LB no final de cada nível e antes do procedimento de redução de nível, de acordo com os seguintes passos :

1. Determinar a duração de cada palavra W_i através do procedimento de "Backtracking",

$$d_n = n - \mathbf{B}(l,n,i) \text{ para todo } n$$

2. Ponderar a verossimilhança acumulada $\mathbf{P}(l,n,i)$, utilizando a função densidade de probabilidade gaussiana $f_i(\cdot)$ ajustada em função da duração d_n determinda no passo 1,

$$\overline{\mathbf{P}}(l,n,i) = \mathbf{P}(l,n,i) \cdot f_i(d_n) \text{ para todo } n$$

Apesar deste método para incorporar informação duracional ser bastante heurístico, será comprovado, através dos resultados presentes na Tabela 7.2 e no Apêndice C, que na prática ele melhora significativamente a taxa de acertos de palavras do sistema.

5.2.4 Exemplo

Para facilitar a compreensão do algoritmo level building, considere o seguinte exemplo :

1. Número de níveis de busca igual a três (sentenças constituídas por no máximo três palavras).
2. $P(M|\Theta) = 1$ - Não será utilizado nenhum tipo de restrição gramatical.

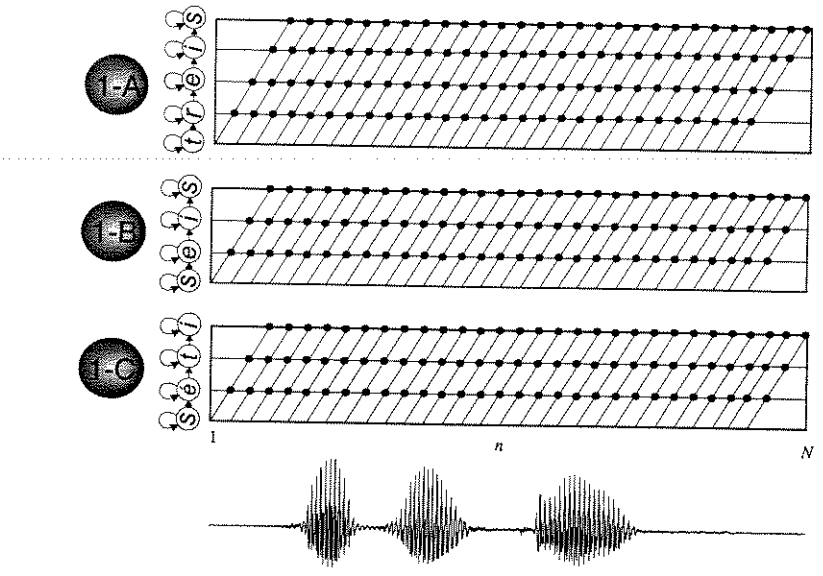

Figura 5-3: Avaliação de todas as sentenças constituídas por apenas uma única palavra.

3. Léxico constituído por apenas três palavras, correspondentes aos dígitos : Três, Seis e Sete. Estes dígitos também serão identificados, respectivamente, pelas letras **A**, **B** e **C**. Serão utilizadas a seguintes conversões ortográfico-fonética.

- Três : $t + r + e + i + s$ - **A**
- Seis : $s + e + i + s$ - **B**
- Sete : $s + e + t + i$ - **C**

A seguir são apresentados os procedimentos que devem ser realizados durante a análise de cada um dos três níveis de procura.

Primerio Nível : Avaliação de todas as sentenças constituídas por apenas uma única palavra As treliças mostradas na Figura 5-3 apresentam todos os possíveis caminhos pelos quais os HMMs das palavras **A**, **B** e **C** podem gerar a seqüência de vetores acústicos associada ao sinal de fala apresentado ao sistema. Em outras palavras estes caminhos descrevem todas as possíveis dilatações temporais (quantizadas em função do tamanho do quadro de análise) das palavras **A**, **B** e **C**.

Segundo Nível : Avaliação de todas as sentenças construídas através da concatenação de duas palavras A Figura 5-4 mostra três das nove sentenças possíveis de serem construídas a partir da combinação de duas palavras.

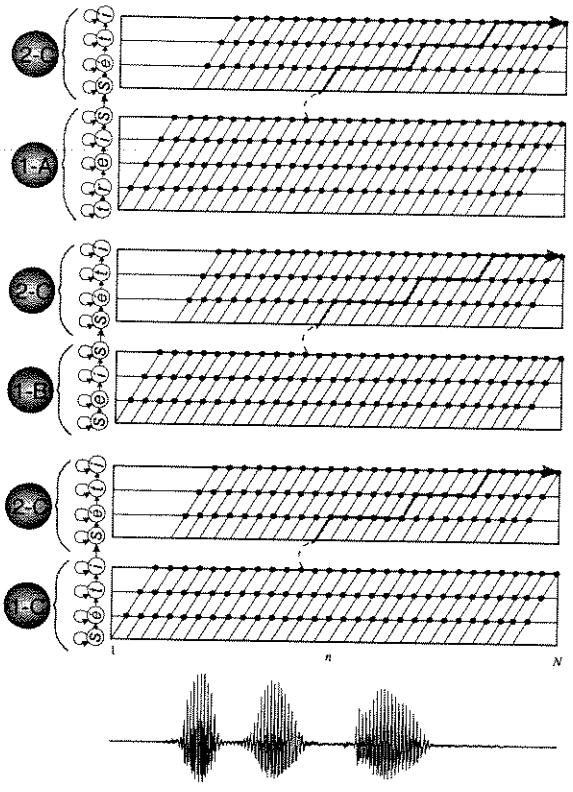

Figura 5-4: Três das nove sentenças possíveis de serem contruídas a partir da concatenação de duas palavras.

A verossimilhança acumulada, para um determinado caminho específico, ao longo da segunda palavra a ser concatenada, será determinada por três parâmetros básicos :

- Verossimilhanças de emissão de símbolos associadas a cada estado.
- Probabilidade de transição de cada estado.
- Probabilidade do estado inicial. Esta probabilidade corresponde a verossimilhança acumulada, no último estado da palavra do nível anterior, para um determinado instante n específico.

A Figura 5-4 mostra a mesma trajetória de Viterbi ao longo da palavra **C** nas sentenças **C-C**, **B-C** e **A-C**. Por se tratar da mesma trajetória de Viterbi, é óbvio que ela compartilha das mesmas verossimilhanças de emissão de símbolos e probabilidades de transição. Portanto, para esta trajetória específica, o parâmetro decisivo na determinação da verossimilhança acumulada com maior valor será a probabilidade do estado inicial. A partir desta análise pode-se concluir que após a realização do

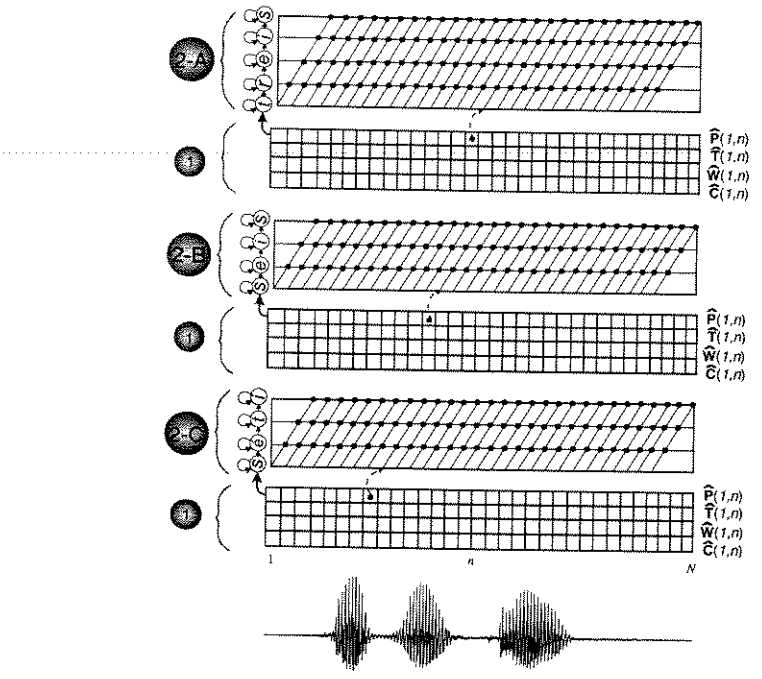

Figura 5-5: Avaliação das palavras candidatas ao segundo nível da sentença, após a redução do primeiro nível (utilizando apenas as informações realmente úteis no primeiro nível).

primeiro nível muitas das informações disponíveis, para cada instante de tempo n são desnecessárias e que a informações realmente úteis, e que portanto devem ser armazenadas, são :

- Entre todas as palavras avaliadas, a verossimilhança de emissão de símbolos acumulada com maior valor - $\widehat{P}(1,n)$.
- Indicação da palavra que apresentou esta maior verossimilhança - $\widehat{W}(1,n)$.
- Probabilidade de transição associada ao último estado desta palavra vencedora - $\widehat{T}(1,n)$.
- Instante inicial onde começou o caminho de Viterbi associado a esta maior verossimilhança acumulada - $\widehat{C}(1,n)$.

Este procedimento de análise e armazenamento das informações realmente úteis para cada nível é denominado *redução de nível.*

A Figura 5-5 mostra a avaliação das palavras candidatas ao segundo nível da sentença, porém para o caso em que já foi realizada a *redução do primeiro nível.* Neste caso apenas três combinação são necessárias para avaliar todas as concatenações de duas palavras que podem apresentar os maiores valores de verossimilhança acumulada.

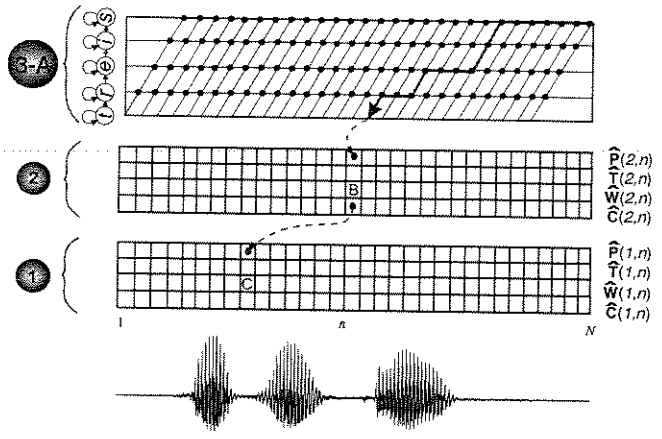

Figura 5-6: Ilustração do procedimento de "backtracing". Caso em que a sentença reconhecida é composta por três palavras e consiste na concatenação dos dígitos : Sete - Seis - Três.

Terceiro Nível : Avaliação de todas as sentenças construídas através da concatenação de três palavras Após a *redução do primeiro e segundo nível*, avaliar as palavras canditadas ao terceiro nível, verificar a vencedora e em seguinda, fazendo uso dos vetores ($\widehat{\mathbf{P}}$,$\widehat{\mathbf{T}}$,$\widehat{\mathbf{W}}$ e $\widehat{\mathbf{C}}$) com as informações dos níveis já reduzidos, realizar o "backtracing" para identificar a seqüência de palavras com maior verossimilhança de gerar a seqüência de vetores acústicos apresentados ao sistema. A Figura 5-6 ilustra a realização deste procedimento de "backtracing" para o caso em que a sentença com maior verossimilhança é composta por três palavras e consiste na concatenação dos dígitos : **C - B - A** $\Leftrightarrow$ Sete - Seis - Três.

5.3 One Stage

O algoritmo "One Stage" difere do algoritmo "Level Building" apenas em termos de implementação. Para cada *nível de procura* o algoritmo "Level Building" calcula os caminhos de Viterbi de todas as palavras para todos os instantes de tempo e somente depois de todos estes cálculos realiza a redução de nível e passa a análise do *nível* seguinte. Este tipo de procedimento é denominado síncrono por palavra e está limitado à execução serial dos *níveis de procura*. No caso do algoritmo "One Stage", para o *nível* de procura l calcula-se a trajetória de Viterbi de todas as palavras para um determinado instante de tempo n e, se os valores das verossimilhanças acumuladas nos últimos estados de todas as palavras W_i's estiverem disponíveis, então realiza-se a redução de *nível* (apenas para o instante de tempo n) e passa-se a executar simultaneamente o *nível* l e o *nível seguinte* $l+1$. O tipo de

procedimento utilizado pelo algoritmo "One Stage" é denominado síncrono por quadro (síncrono por janela de análise) e permite a execução paralela dos *níveis* de procura.

Outra vantagem do algoritmo "One Stage" é que, ao contrário do algoritmo "Level Building", ele não necessita esperar a chegada de toda a seqüência de vetores acústicos para iniciar o processo de decodificação.

Capítulo 6

Sistema Implementado

6.1 Base de Dados

6.1.1 Introdução

A base de dados utilizada para treinamento e avaliação do sistema implementado consiste de um total de 100 sentenças pronunciadas por um único locutor masculino, adquiridas sob condições de estúdio, amostradas a 16 kHz e quantizadas com 16 bits por amostra. Estas sentenças, que totalizam 5,21 minutos de fala, foram gravadas no CPQD TELEBRÁS e segmentadas por especialistas do Laboratório de Fonética Acústica e Psicolingüística Experimental - LAFAPE, do Instituto de Estudos da Linguagem - IEL, da UNICAMP.

Esta base foi gravada inicialmente para análise e levantamento de características prosódicas a serem utilizadas em um sistema para síntese concatenativa de fala a partir de texto. As grandes motivações para a utilização desta base para treinamento e avaliação de um sistema para reconhecimento de fala contínua foram :

- O fato dela se encontrar totalmente rotulada em termos de unidades sub-lexicais - 36 tipos de fones. Esta rotulação foi realizada manualmente por especialistas do LAFAPE.
- Apresentar uma diversidade relativamente grande de duração e de palavras, mostrando-se, portanto, bastante útil para a avaliação do sistema. A menor frase possui apenas uma única palavra e a maior 20 palavras. As 100 frases são constituídas de um total de 319 palavras distintas.

6.1.2 Unidades Sub-Lexicais (fones)

Para treinamento do sistema foram escolhidos aleatoriamente 80% (aproximadamente) dos exemplos de cada fone e os fones restantes foram utilizados para validação cruzada (monitoramento da capacidade de generalização da ANN). Não foi definido um conjunto de teste, porque a avaliação do sistema foi realizada a nível de sentenças e não de fones.

A Tabela 6.1 apresenta todos os tipos de fones que compõem a base de dados, seguidos de exemplos que ilustram a pronúncia de cada um deles. Também são mostradas as quantidades de exemplos de cada tipo de fone utilizado assim como suas durações médias. Na última linha da Tabela 6.1 são apresentados os totais de exemplos utilizados assim como a média das durações médias de todos estes exemplos.

6.1.3 Sentenças

A Tabela 6.2 apresenta uma lista com 20 das 100 sentenças que compõem a base de dados utilizada. Uma lista completa com as 100 encontra-se no Apêndice C.

6.1.4 Dicionário de Pronúncias

O Apêndice B apresenta a pronúncia[1] (conversão ortográfico-fonética) de todas as palavras que compõem o léxico utilizado. A pronúncia de cada uma destas palavras foi obtida a partir da conversão ortográfico-fonética de cada uma das 100 sentenças do Apêndice C. Com isto foi levado em consideração informações sobre o contexto em que cada uma destas palavras se encontrava.

6.2 Pré-Processamento

A parametrização do sinal de fala foi realizada utilizando-se apenas coeficientes Mel cepstrais - MCC ("Mel Cepstral Coeficientes") baseados em banco de filtros segundo Davis & Melmelstein [5]. Para cada vetor acústico $\mathbf{x}_n$ foram calculados 12 coeficientes Mel Cepstrais $\mathbf{x}_n = \{mcc_{n_1}, mcc_{n_2}, \ldots, mcc_{n_{12}}\}$. Estes vetores acústicos foram calculados para janelas de 20 ms tomadas a cada 10 ms (superposição de 50 %).

Para evitar problemas de saturação dos potenciais de ativação da ANN (efeito conhecido como congelamento de sinapses) é comum se realizar uma normalização dos coeficientes mcc_{n_j}, $j = \{1, 2, \ldots, 12\}$

[1] A definição da pronúncia de cada uma das palavras que compõem o léxico utilizado foi realizada por especialistas do Laboratório de Fonética Acústica e Psicolingüística Experimental - LAFAPE do Instituto de Estudos da Linguagem - IEL da UNICAMP.

Nº	Fones	Exemplos	Nº de fones	Duração Média (ms)
01	*i*	time,caí,mãe	411	57,05
02	*e*	besta	243	75,21
03	*é*	festa	56	129,08
04	*á*	pata	332	91,97
05	*ó*	porta	16	159,62
06	*o*	bolo	111	85,27
07	*u*	pão, pau, tudo	482	55,60
08	*a*	pata	137	63,06
09	*ã*	planta	89	86,17
10	*ẽ*	tenta	119	111,90
11	*ĩ*	pinta	74	109,82
12	*õ*	tonta	75	101,94
13	*ũ*	mundo	25	90,88
14	*p*	pata	101	82,96
15	*t*	pata	217	78,67
16	*k*	paca	178	85,97
17	*T*	tia	73	97,47
18	*b*	bata	51	65,27
19	*d*	dado	176	50,44
20	*g*	gata	29	51,96
21	*D*	dia	55	57,27
22	*f*	faca	50	98,02
23	*s*	sapo	320	112,59
24	*ʃ*	chato	18	104,61
25	*v*	vala	67	62,35
26	*z*	casa	145	61,82
27	*j*	jipe	11	73,09
28	*m*	maca	112	62,59
29	*n*	nada	113	45,04
30	*ɳ*	manhã	2	91,50
31	*ɾ̈*	caro	200	36,46
32	*r*	carta	32	51,93
33	*R*	carro	25	66,68
34	*l*	lata	56	44,48
35	*ʎ*	calha	6	57,83
36	#	(pausa)	37	314,67
		Total	**4244**	**76.54**

Tabela 6.1: Lista com os tipos de fones utilizados seguidos por exemplos, e suas respectivas quantidades e durações médias.

01	A cotação do dólar aumentou, e as bolsas fecharam em baixa.
02	A cotação do dólar aumentou, mas as bolsas fecharam em baixa.
03	A bolsa ficará estável, ou sofrerá uma pequena queda.
04	Não haverá ajustes nem modificações radicais no plano.
05	Foi detectado um problema em seu cartão. Ele deve ser substituído.
06	É necessário que o convênio permita o intercâmbio.
07	Posso afimar-lhes que o convênio permite o intercâmbio.
08	O convênio que foi assinado recentemente permite o intercâmbio.
09	O convênio permite o intercâmbio porque visa a integração entre alunos de culturas diferentes.
10	O convênio que foi assinado recentemente, permite o intercâmbio.
11	O convênio assinado na última reunião é mais interessante do que o anterior.
12	A medida que o tempo passa, mais nos convencemos da eficiência do convênio.
13	É suficiente.
14	Isto é suficiente.
15	O saldo é suficiente.
16	O saldo de sua conta é suficiente.
17	O saldo disponível é insuficiente.
18	O saldo disponível em sua conta é insuficiente.
19	Isto parece insuficiente.
20	O saldo parece insuficiente.

Tabela 6.2: Lista com 20 das 100 sentenças a serem utilizadas durante o treinamento e avaliação do sistema.

de todos os vetores acústicos $\mathbf{x}_n$ a serem utilizados tanto no treinamento como na avaliação do sistema. Uma normalização bastante comum é fazer $mcc_{n_j} = \frac{mcc_{n_j}}{f_N}$, sendo f_N o fator de normalização capaz de tornar o maior coeficiente, de todos os vetores acústicos utilizados durante o treinamento, igual a 1. Um problema associado a este tipo de normalização é que ele não leva em consideração a distribuição estatística dos coeficientes dos vetores acústicos de treinamento. Por exemplo, se o maior coeficiente de todos os vetores acústicos for muito maior que todos os coeficientes restantes, então este tipo de normalização reduzirá muito a magnitude da maioria dos dados de entrada da ANN, o que dificultará o processo de treinamento. Para contornar este problema montou-se uma seqüência $\mathbf{X}_{MCC}$, formada pela concatenação de todos os coeficientes mcc_{n_j}, de todos os vetores acústicos utilizados no treinamento e ajustou-se o valor de f_N para que 95% dos dados (coeficientes mcc_{n_j}) passassem a ter magnitudes entre -1.0 e 1.0. O valor encontrado para f_N foi equivalente a fazer o desvio padrão da seqüência $\mathbf{X}_{MCC}$ igual a $0,49$, $(\sigma = 0,49)$ mostrando que os dados de treinamento apresentavam uma função distribuição de probabilidade quase gaussiana. Este fato pôde ser comprovado através da análise do histograma da seqüência $\mathbf{X}_{MCC}$, ilustrado na Figura 6-1.

A Figura 6-2 mostra o histograma após a normalização dos coeficientes mcc_{n_j}

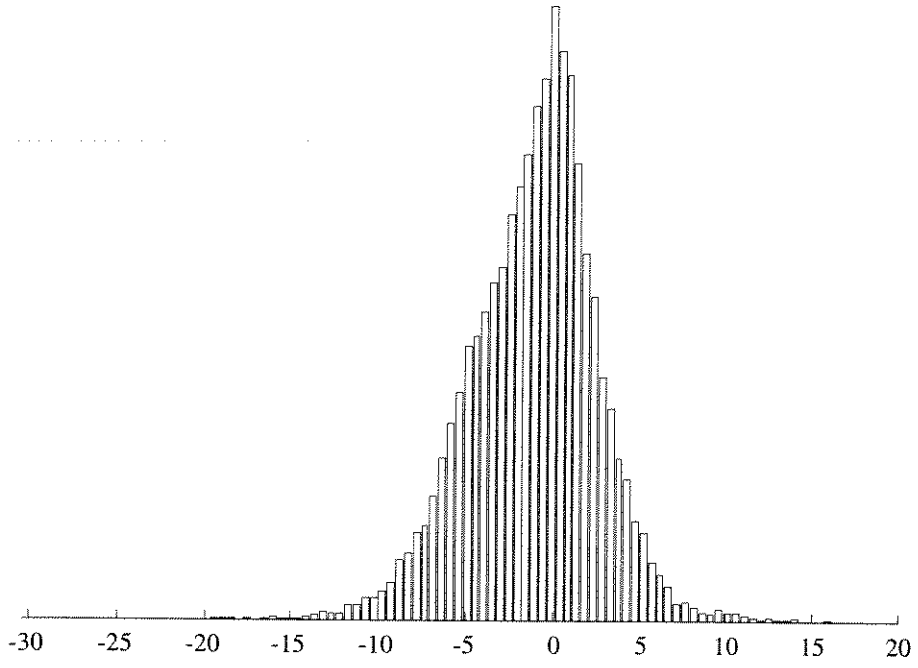

Figura 6-1: Histograma da amplitude dos componentes de todos os vetores acústicos que formam os exemplos de treinamento X_e

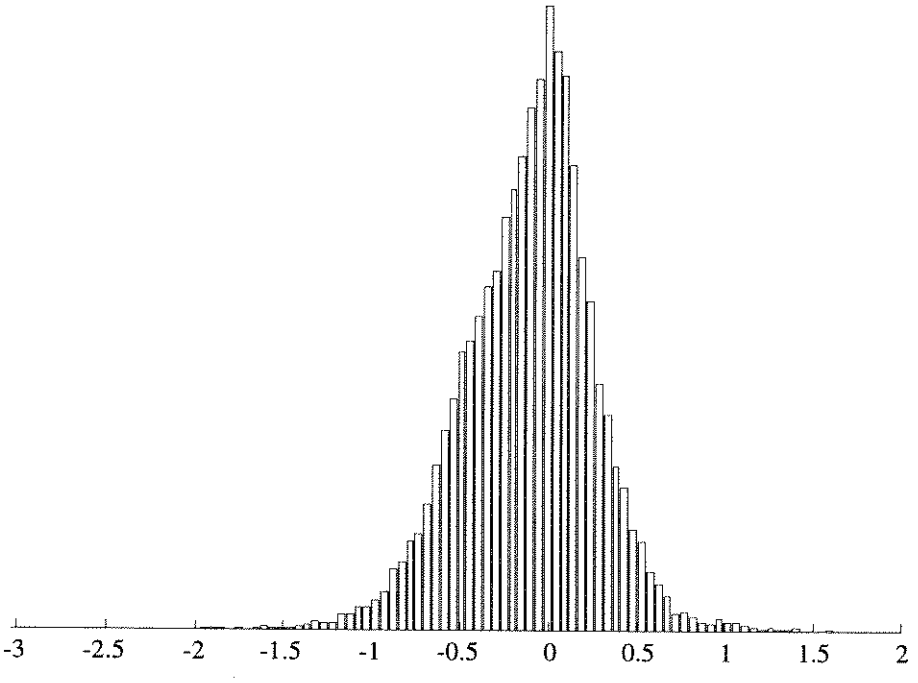

Figura 6-2: Histograma da amplitude dos componentes de todos os vetores acústicos utilizados no treinamento, após a normalização.

6.3 Modelos Híbridos ANN-HMM

6.3.1 Rede Neural Artificial

Foi utilizada uma ANN do tipo perceptron multicamadas - MLP ("Multilayer Perceptron") com apenas uma camada escondida e com a seguinte configuração :

- **Algoritmo de Treinamento** EBP - "Error Back-Propagation". Com atualização instantânea dos pesos sinápticos e sem o uso de momento [10].
- **Número de entradas** Igual a 84, correspondente a 7 vetores acústicos. Cada vetor acústico corresponde a 12 coeficientes Mel Cepstrais [5] extraídos a partir de janelas de 20 ms tomadas a cada 10 ms (superposição de 50%). A escolha de 7 vetores acústicos (correspondente a 70 ms de fala) foi motivada pelo fato de a média das durações médias de todos os fones ser igual a 76,57 ms. Também foram testadas MLPs com 108 entradas (9 vetores acústicos). Esta configuração, entretanto, além de ser mais difícil de ser treinada, apresentou piores resultados a nível de reconhecimento de fones.
- **Número de neurônios na camada escondida** Igual a 70, número este escolhido segundo um compromisso entre o desempenho da rede e o tempo de treinamento. Também foram treinadas MLPs com 80, 90 e 100 neurônios na camada escondida, mas para um mesmo número de épocas de treinamento, verificou-se que o aumento na quantidade de neurônios degradava o desempenho da ANN.
- **Número de saídas** Igual a 36, correspondente ao número total de fones necessários para realizar a transcrição ortográfico-fonética de qualquer palavra presente no léxico (cada fone foi modelado por um único estado). Este número de 36 saídas também pode ser interpretado como o conjunto estados $Q = \{q_1, q_2, \ldots, q_{36}\}$ necessários para montar o HMM de qualquer palavra presente no léxico. Isto porque existe uma *correspondência um-a-um* entre as classes de saída da MLP e os estados $q_k \in Q$.
- **Funções não lineares** Foi utilizada em todas as camadas, inclusive na camada de saída, a função logística, $\varphi(i) = \frac{1}{(1+\exp(-v_i))}$.
- **Normalização dos dados de entrada** Foram normalizados de modo a resultar um desvio padrão $\sigma = 0,49$, fazendo com que 95% dos dados estivessem entre -1.0 e 1.0.

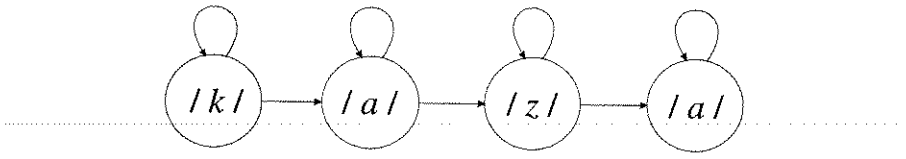

Figura 6-3: HMM "left-right" da palavra "casa" com um único estado por fone.

- **Normalização da saída** Pelo fato de a otimização não ter garantido um mínimo global e como não foi utilizada a função softmax na camada de saída da MLP, então para que a saída da rede $g_k(\mathbf{x}_n, \Psi)$ pudesse ser interpretada como uma medida de probabilidade, foi realizada uma normalização para que $\sum_{k=1}^{36} g_k(x_n, \Psi) = 1$.

6.3.2 Modelos Ocultos de Markov

A definição do HMM de cada uma das palavras do léxico consiste dos seguintes passos

1. **Seqüência de estados que compõem cada HMM** Definir, de acordo como o dicionário de pronúncias Apêndice B, a seqüência de N_{W_i} estados $q_k \in Q$, que devem ser concatenados para a construção do HMM de cada uma das palavras do léxico.

2. **Definir a topologia a ser utilizada** Assumiu-se o modelo *left-right*. Considere, por exemplo, o caso da palavra *"casa"* cujo correspondente HMM será formado pela concatenação dos estados associados aos fones $/k/ + /a/ + /z/ + /a/$, conforme ilustrado na Figura 6-3.

3. **Determinar as probabilidades de auto-transição de estados** a_{jj} Determinar a probabilidade de transição do estado q_j para ele próprio. Como o modelo é estritamente left-right, então a probabilidade de transição a_{jk}, do estado q_j para um estado q_k, será : $a_{jk} = 1 - a_{jj}$. Durante a estimação (Pré-REMAP) estas probabilidades foram definidas em função das durações médias dos fones, de acordo com a seguinte expressão :

$$a_{jj} = \left[\frac{\left(\frac{D_j}{10} - 1\right)}{\frac{D_j}{10}} \right] = \left[\frac{D_j - 10}{D_j} \right] \tag{6.1}$$

sendo D_j a duração média em milissegundos do fone associado ao estado q_j e o fator 10 corresponde ao intervalo de tempo em milissegundos em que são tomados as janelas de análise.

Durante a reestimação estas probabilidade de auto-transição (e transição) foram estimadas segundo a Equação (4.34) da Seção 4.3.5.

4. **Definir as verossimilhanças de emissão de símbolos normalizada $\frac{p(\mathbf{x}_n|q_k)}{p(\mathbf{x}_n)}$ associadas a cada estado** Pela regra de Bayes estas verossimilhanças podem ser obtidas a partir das probabilidades a posteriori estimadas pela ANN,

$$\frac{p(\mathbf{x}_n|q_k)}{p(\mathbf{x}_n)} = \frac{P(q_k|\mathbf{x}_n)}{P(q_k)} \tag{6.2}$$

Como já foi discutido na seção 3.3 o cálculo da probabilidade a priori da classe $P(q_k)$ pode ser realizado verificando-se quantas vezes esta classe ocorre no conjunto de treinamento e dividindo pelo número total de exemplos de treinamento (freqüência relativa). Porém, devido ao fato do conjunto de exemplos de treinamento, $\mathbf{X}_e = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_e}\}$, não ser balanceado (o fone /u/ possui 482 exemplares enquanto os fones /λ/ e /η/ possuem apenas 6 e 2 exemplares, respectivamente), a estimativa das verossimilhanças de emissão de símbolos normalizada $\frac{p(\mathbf{x}_n|q_k)}{p(\mathbf{x}_n)}$ não apresentou bons resultados quando $P(q_k)$ foi calculada segundo o método da freqüência relativa. Como foi citado na seção 3.3, Lippmann [30] afirma que durante a avaliação o valor de $P(q_k)$ pode ser ajustado para melhor representar o conjunto de teste, e com este objetivo decidiu-se adotar o seguinte procedimento para o cálculo da probabilidade a priori da classe $P(q_k)$,

$$P(q_k) \cong \sum_{\mathbf{x}_k \in \mathbf{X}_e} P(q_k, \mathbf{x}_k) = \sum_{\mathbf{x}_k \in \mathbf{X}_e} P(q_k|\mathbf{x}_k) \cdot p(\mathbf{x}_k) \tag{6.3}$$

Substituindo (6.3) em (6.2) e assumindo os símbolos $\mathbf{x}_k$ equiprováveis (suposição baseada no fato de que os vetores acústicos $\mathbf{x}_k$ não são quantizados e portanto podem ser considerados distintos entre si), obtém-se :

$$p(\mathbf{x}_n|q_k) = \frac{P(q_k|\mathbf{x}_n)}{\frac{1}{N_e} \cdot \sum_{\mathbf{x}_k \in \mathbf{X}_e} P(q_k|\mathbf{x}_k)} \cdot p(\mathbf{x}_n) \tag{6.4}$$

6.4 Treinamento

6.4.1 Estimação - Pré-REMAP

A rede neural foi treinada utilizando-se os exemplos de treinamento mostrados na Tabela 6.1. As probabilidades de transição $P\left(q_k^n|q_j^n\right)$ foram determinadas em função das durações médias dos fones, Equação (6.1) e as probabilidades das classes $P(q_k)$ foram estimadas a partir das saídas da ANN (sem

reestimação) segundo a Equação (6.3).

6.4.2 Reestimação - REMAP

Foram realizadas três reestimações utilizando alvos suaves. A reestimação dos parâmetros $P\left(q_k^n|q_j^{n-1}\right)$, $P\left(q_k^n\right)$ e Ψ foram realizadas de acordo com seguintes passos,

1. Estimou-se para cada uma das M_i sentenças de treinamento, com $i = \{1, 2, \ldots, 100\}$, os correspondentes alvos suaves, $P\left(q_k|\mathbf{X}_{M_i}, M_i, \Theta\right)$, sendo $\mathbf{X}_{M_i}$ a seqüência de vetores acústicos[2] correspondente a M_i.

2. Utilizando os 100 alvos suaves do item 1, reestimou-se dos parâmetros Ψ da ANN a partir das correspondentes seqüências de vetores acústicos $\mathbf{X}_{M_i}$. A apresentação dos pares de treinamento $(\mathbf{X}_{M_i}, P\left(q_k|\mathbf{X}_{M_i}, M_i, \Theta\right))$ foi realizada de forma aleatória, segundo uma *função distribuição de probabilidade uniforme.*

3. A partir da ANN com parâmetros Ψ reestimados e da seqüência $\{\mathbf{X}_{M_1}, \mathbf{X}_{M_2}, \ldots, \mathbf{X}_{M_{100}}\}$ (concatenação das seqüências de vetores acústicos correpondentes às 100 sentenças de treinamento) calculou-se as probabilidades a priori dos fones $P\left(q_k\right)$, com $k \in \{1, 2, \ldots, 36\}$, segundo a Equação (6.3).

4. Assumindo que cada uma das 100 sentenças de treinamento são independentes das 99 restantes, realizou-se a reestimação das probabilidades de transição $P\left(q_k^n|q_j^{n-1}\right)$ para todos os estados q_j, com $j \in \{1, 2, \ldots, 36\}$,conforme a Equação (4.34).

6.5 Decodificação

6.5.1 Algoritmo Level Building

Utilizou-se o algoritmo Level Building para realizar a decodificação das sentenças a partir dos modelos de palavras (determinação da seqüência da palavras M, com maior probabilidade dado a seqüência de vetores acústicos $\mathbf{X}$ apresentada ao sistema). Como entre as 100 sentenças a serem avaliadas, existiam sentenças com apenas uma única palavra e outras com até 20 palavras, Apêndice C, decidiu-se ajustar o Level Building para realizar buscas de sentenças compostas pela concatenação de 1 a 22 palavras, (número de níveis máximo, $L = 22$) . Este acréscimo de dois níveis ao número máximo de palavras

[2] A cada sentença M_i, do sistema implementado, foi associado uma única seqüência de vetores acústicos $\mathbf{X}_{M_i}$.

deve-se ao fato de estar sendo assumido que toda sentença pode ser precedida e seguida por um silêncio (pausa) que por sua vez será modelado como uma palavra adicional.

6.5.2 Modelo de Duração de Palavras

Foi incorporado ao algoritmo Level Building o modelo de duração de palavras discutido na Seção 5.2.3. Para a implementação deste modelo de duração foi realizado um levantamento das durações médias e variâncias de todas as palavras que compõem o léxico utilizado.

6.5.3 Restrições Gramaticais

No sistema implementado foi utilizado uma gramática do tipo Pares-de-palavras, *P-gram* [25]. Uma gramática do tipo *P-gram* faz uso da aproximação mostrada na Equação (6.5) e pode ser interpretada como uma versão determinística de uma gramática estocástica do tipo *Bi-gram.*

$$P(M) \approx G(W_1) \cdot \prod_{n=2}^{N_M} G(W_n|W_{n-1}) \tag{6.5}$$

onde $G(W_1)$ e $G(W_n|W_{n-1})$ são definidos como

$$\begin{cases} G(W_1) = 1, \text{ se for permitido que a palavra } W_1 \text{ inicie uma sentença.} \\ G(W_1) = 0, \text{ se não for permitido que a palavra } W_1 \text{ inicie uma sentença.} \end{cases} \tag{6.6}$$

$$\begin{cases} G(W_n|W_{n-1}) = 1, \text{ se a transição } W_{n-1} \Rightarrow W_n \text{ for permitida.} \\ G(W_n|W_{n-1}) = 0, \text{ se a transição } W_{n-1} \Rightarrow W_n \text{ não for permitida.} \end{cases} \tag{6.7}$$

Com as definições da Equação (6.6) e (6.7) pode ser verificado que, $P(M) = 1$ se for permitido que W_1 inicie a sentença e se todas as transições $W_{n-1} \Rightarrow W_n$ forem permitidas e $P(M) = 0$ se não for permitido que W_1 inicie a sentença ou se alguma das transições $W_{n-1} \Rightarrow W_n$ não forem permitidas.

A determinação da gramática *P-gram.* utilizada no sistema implementado foi realizada levando-se em consideração apenas as 100 sentenças do Apêndice C. Para o estabelecimento destas restrições gramaticais foram necessários os seguintes levantamentos :

- Determinação de todas as possíveis transições de palavras, $W_{n-1} \Rightarrow W_n$.
- Determinação de todas as possíveis palavras que podem iniciar as sentenças.

Além destes levantamentos foram estabelecidas as seguintes regras :

Regra 1 Todas as sentenças podem ser iniciadas por um "*silêncio*". Com isto foi definido que o "*silêncio*" (/ # /) consiste de uma das possíveis "palavras" que podem iniciar as sentenças.

Regra 2 Todas as palavas podem ser precedidas e seguidas por um "silêncio", isto é, as transições "*silêncio*" $\Rightarrow W_n$ e $W_n \Rightarrow$ "*silêncio*", são permitidas para todas as palavras W_n presentes no léxico.

A incorporação destas restrições gramaticais ao algoritmo Level Building foram realizadas da seguinte forma :

Nível 1 (Primeira palavra da sentença) Realizar este nível apenas para as palavras que puderem iniciar a sentença de acordo com a gramática *P-gram.*

Nível 2 (E demais níveis) Para cada uma das palavras W_j 's a serem analisadas, inicializar $\delta_n(1)$ e $\alpha_n(1)$ conforme as Equações (6.8) e (??).

$$\delta_n(1) = \max\left[\widehat{\mathbf{T}}(l-1,n-1)\cdot\widehat{\mathbf{P}}(l-1,n-1)\cdot G(W_j|W_i), P\left(q_1^n|q_1^{n-1},W_j,\Theta\right)\cdot\delta_{n-1}(1)\right] \cdot p(\mathbf{x}_n|q_k,W_j,\Theta) \quad (6.8)$$

$$\alpha_n(1) = \begin{cases} n-1, \text{ se } \widehat{\mathbf{T}}(l-1,n-1)\cdot\widehat{\mathbf{P}}(l-1,n-1)\cdot G(W_j|W_i) > \\ P\left(q_1^n|q_1^{n-1},W_i,\Theta\right)\cdot\delta_{n-1}(1) \\ \alpha_{n-1}(1), \text{ caso contrário} \end{cases} \quad (6.9)$$

Capítulo 7

Resultados Experimentais

7.1 Introdução

Neste capítulo são apresentados alguns resultados da avaliação do sistema implementado com o uso das unidades sub-lexicais (fones) definidas na Tabela 6.1 e das 100 sentenças apresentadas no Apêndice C. A partir destes resultados são realizadas quatro análises básicas :

1. **Reconhecimento a nível de unidades sub-lexicais** Desempenho da ANN como *classificadora de unidades sub-lexicais* (fones).

2. **Reconhecimento de sentenças com o uso de modelos de sentenças** Tem como objetivo avaliar o *modelamento acústico* realizado pela ANN e a absorção da *variabilidades temporal* realizada pelos HMMs.

3. **Algoritmo de reestimação de parâmetros - REMAP** Análise das reestimações dos parâmetros Ψ da ANN, das probabilidades de transição de estados $P\left(q_k^n | q_j^{n-1}\right)$ e das probabilidades a priori das classes (fones) $P\left(q_k\right)$.

4. **Reconhecimento de sentenças com o uso de modelos de palavras** Neste caso os objetivos principais são : avaliar o desempenho do *algoritmo de busca* - algoritmo "Level Building" - e comprovar a importância de um bom modelo da língua (restrições gramaticais) no desempenho final de um sistema para reconhecimento de fala contínua.

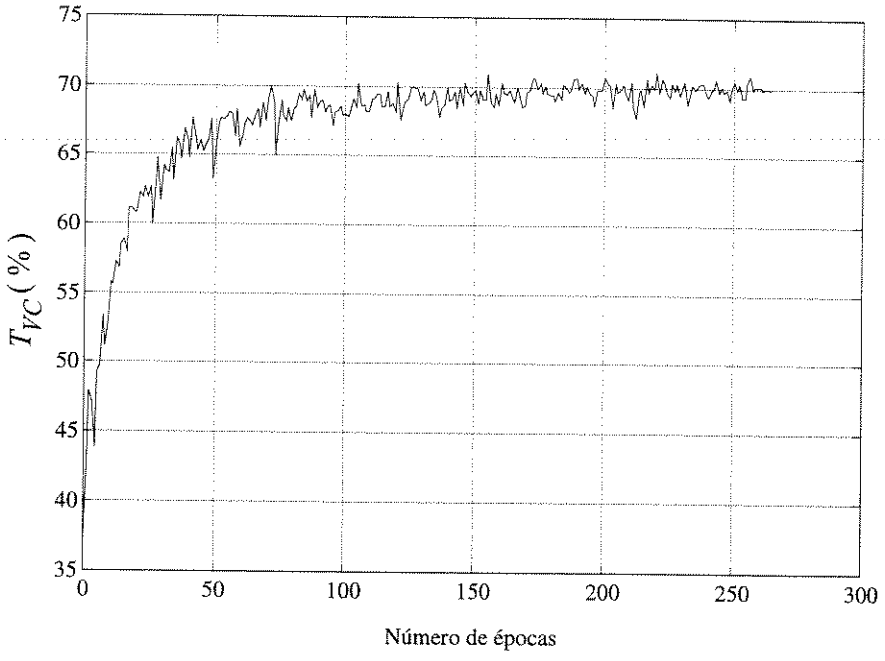

Figura 7-1: Taxas de acertos da ANN para os exemplos de validação cruzada em função do número de épocas de treinamento.

7.2 Reconhecimento a Nível de Unidades Sub-Lexicais (fones)

7.2.1 Taxas de Acertos da ANN ao Longo das Épocas de Treinamento

A Figura 7-1 apresenta as taxas totais de acertos de fones (considerando todos os 36 fones), para os exemplos de validação cruzada T_{VC}, ao longo do número de épocas de treinamento.

A Figura 7-2 apresenta as taxas totais de acertos de fones, tanto para os exemplos de validação cruzada T_{VC} , como para os exemplos de treinamento $T_{trein.}$, ao longo das épocas de 171 a 249.

A taxa de acertos global de fones da ANN T_G, foi definida como a média ponderada dos resultados para o conjunto de validação cruzada T_{VC} e para o conjunto de treinamento $T_{trein.}$, $T_G = 0,2 \cdot T_{VC} + 0,8 \cdot T_{trein.}$

A Figura 7-3 apresenta a taxa de acertos global da ANN T_G, para as épocas de 171 a 249 e mostra que o conjunto de pesos sinápticos Ψ que apresentou a melhor T_G, está associado à época 227.

7.2.2 Taxas de Acertos Finais da ANN

A Tabela 7.1 apresenta as taxas de acertos T_{VC} , $T_{trein.}$ e T_G da ANN, com pesos sinápticos Ψ associados à época número 227, para cada um dos 36 tipos de fones utilizados.

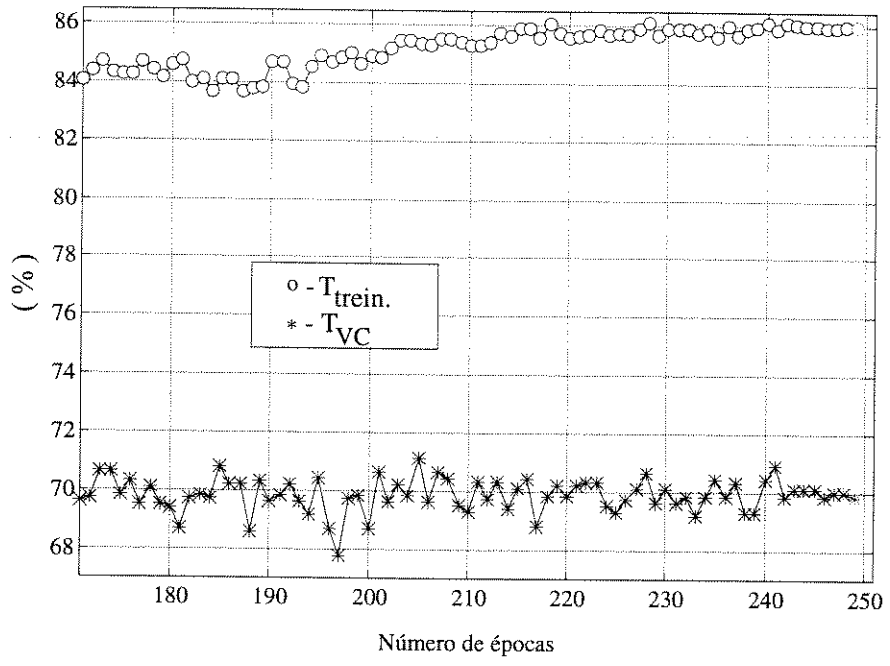

Figura 7-2: Taxa de acertos da ANN tanto para os exemplos de treinamento $T_{trein.}$, como para os exemplos de validação T_{VC}, em função do número de épocas de treinamento - da época 171 a 249.

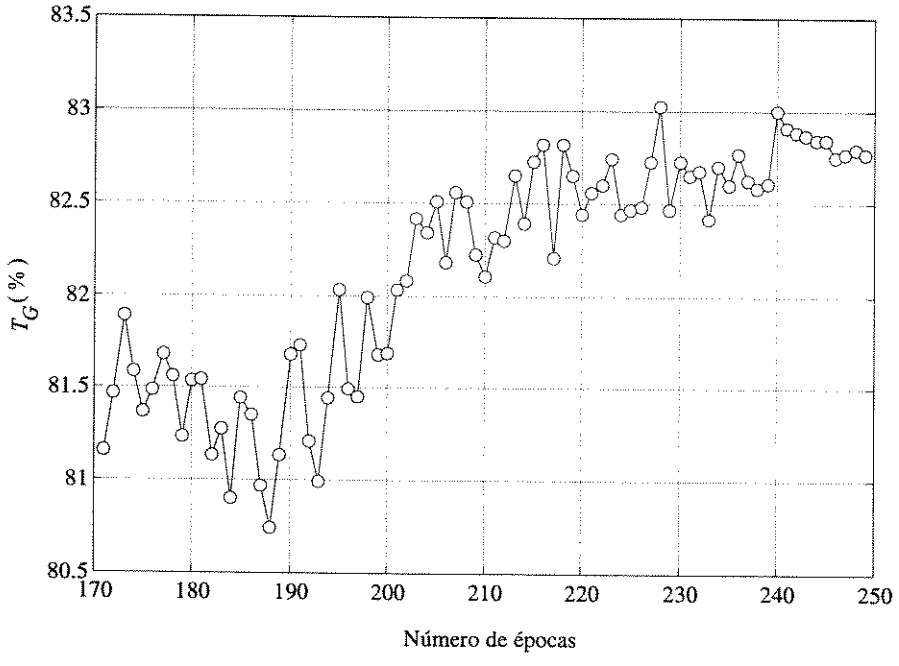

Figura 7-3: Taxa de acertos global, que é dada por uma média ponderada entre as taxas de acerto para os exemplos de validação cruzada e para os exemplos de treinamento, $T_G = 0,2 \cdot T_{VC} + 0,8 \cdot T_{trein.}$. - da época 171 a 249.

Nº	Fones	Exemplos	Nº de fones	T_{VC} (%)	$T_{trein.}$ (%)	T_G (%)
01	i	time,caí,mãe	411	74,70	97,14	92,65
02	e	besta	243	81,63	87,11	86,01
03	$é$	festa	56	72,73	95,56	90,99
04	$á$	pata	332	92,54	93,58	93,37
05	$ó$	porta	16	50,00	83,33	76,66
06	o	bolo	111	65,22	86,36	82,13
07	u	pão, pau, tudo	482	78,35	92,05	89,31
08	a	pata	137	46,43	77,98	71,67
09	$\tilde{a}$	planta	89	66,67	84,51	80,94
10	$\tilde{e}$	tenta	119	70,83	89,47	85,74
11	$\tilde{i}$	pinta	74	60,00	93,22	86,57
12	$\tilde{o}$	tonta	75	46,67	91,67	82,67
13	$\tilde{u}$	mundo	25	0,00	5,00	4,0
14	p	pata	101	75,00	93,82	90,05
15	t	pata	217	68,18	95,37	89,93
16	k	paca	178	66,67	94,37	88,83
17	T	tia	73	66,67	87,93	88,67
18	b	bata	51	60,00	85,37	80,29
19	d	dado	176	55,55	96,43	88,25
20	g	gata	29	50,00	47,83	48,26
21	D	dia	55	45,45	93,18	83,63
22	f	faca	50	100,00	92,50	94,00
23	s	sapo	320	93,75	98,44	97,50
24	$ʃ$	chato	18	0,00	42,86	34,29
25	v	vala	67	64,29	73,58	71,72
26	z	casa	145	82,76	92,24	90,34
27	j	jipe	11	0,00	0,00	0,00
28	m	maca	112	65,22	92,13	86,75
29	n	nada	113	78,26	84,44	83,20
30	$ɳ$	manhã	2	0,00	0,00	0,00
31	$ɾ̈$	caro	200	52,00	86,25	79,40
32	r	carta	32	16,67	57,69	49,48
33	R	carro	25	20,00	65,00	56,00
34	l	lata	56	41,46	90,91	81,02
35	$ʎ$	calha	6	0,00	0,00	0,00
36	#	(pausa)	37	87,50	93,10	91,98
		Total	**4244**	**70.67**	**86.12**	**83.02**

Tabela 7.1: Taxa de acertos da ANN após o Pré-REMAP, para cada um dos 36 tipos de fones.

7.3 Reconhecimento de Sentenças com uso de Modelos de Sentenças

7.3.1 Exemplo

Antes de realizar o reconhecimento de fala contínua propriamente dito, realizou-se um experimento para analisar o desempenho do sistema no reconhecimento de sentenças a partir de modelos de sentenças. Neste experimento montou-se um modelo híbrido ANN-HMM para cada uma das 100 sentenças a serem reconhecidas (uma única ANN e 100 HMMs). O processo de decodificação consistiu basicamente em verificar qual modelo de sentença M_i com $i = \{1, 2, \ldots, 100\}$ apresentava a maior verossimilhança de ter gerado a seqüência de vetores acústicos $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ apresentada ao sistema. Para ilustrar o procedimento considere o seguinte exemplo :

- Seqüência de vetores acústicos $\mathbf{X}$, apresentada ao sistema, correspondente à sentença : "O saldo é suficiente" .
- Modelos ocultos de Markov a serem analisados : (1) "É suficiente", (2) "Isto é suficiente", (3) "O saldo é suficiente" e (4) "O saldo de sua conta é suficiente".

A Figura 7-4 mostra a saída da ANN convertida em verossimilhanças de emissão de símbolos normalizada $\frac{p(\mathbf{x}_n|q_k)}{p(\mathbf{x}_n)}$, com $k = \{1, 2, \ldots, 36\}$, para a seqüência de vetores acústicos $\mathbf{X}$ (correspondente a sentença, "O saldo é suficiente"). Nesta figura os tons de cinza indicam o valor de $\frac{p(\mathbf{x}_n|q_k)}{p(\mathbf{x}_n)}$; quanto mais intenso (mais escuro) o tom de cinza, maior o valor da verossimilhança.

Nas Figuras 7.5a, 7.5b, 7.5c e 7.5d são mostrados os modelos e os alinhamentos realizados pelo algoritmo de Viterbi (seqüência de estados com maior verossimilhança acumulada $p(\mathbf{X}|M, \Theta)$, mais precisamente $\frac{p(\mathbf{X}|M,\Theta)}{p(\mathbf{X}|\Theta)}$), relativos, respectivamente, às sentenças "É suficiente" (12 fones), "Isto é suficiente" (16 fones), "O saldo é suficiente" (18 fones) e "O saldo de sua conta é suficiente" (27 fones), quando a seqüência de vetores acústicos apresentada ao sistema corresponde à sentença, "O saldo é suficiente".

As Figuras 7.5a, 7.5b, 7.5c e 7.5d, também mostram através das trajetórias de Viterbi, que a inclusão de estados correspondentes a pausas (silêncios) no início e no final dos HMMs tornou o sistema capaz de absorver os silêncios iniciais e finais das sentenças.

A Figura 7-6 apresenta os log's normalizados das verossimilhanças calculadas ao longo dos caminhos de Viterbi apresentados na Figura 7-5, e confirma que o modelo ccorrepondente à Figura 7.5c é o que tem a maior verossimilhança de gerar a seqüência de símbolos correpondentes à sentença : "O saldo é suficiente".

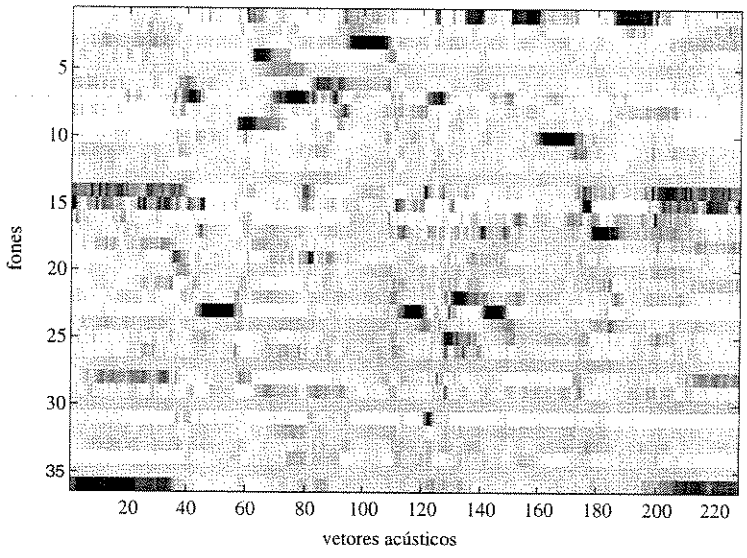

Figura 7-4: Saída da ANN convertida em verossimilhança de emissão de símbolos, para os vetores acústicos de entrada correspondentes à sentença, "O saldo é suficiente".

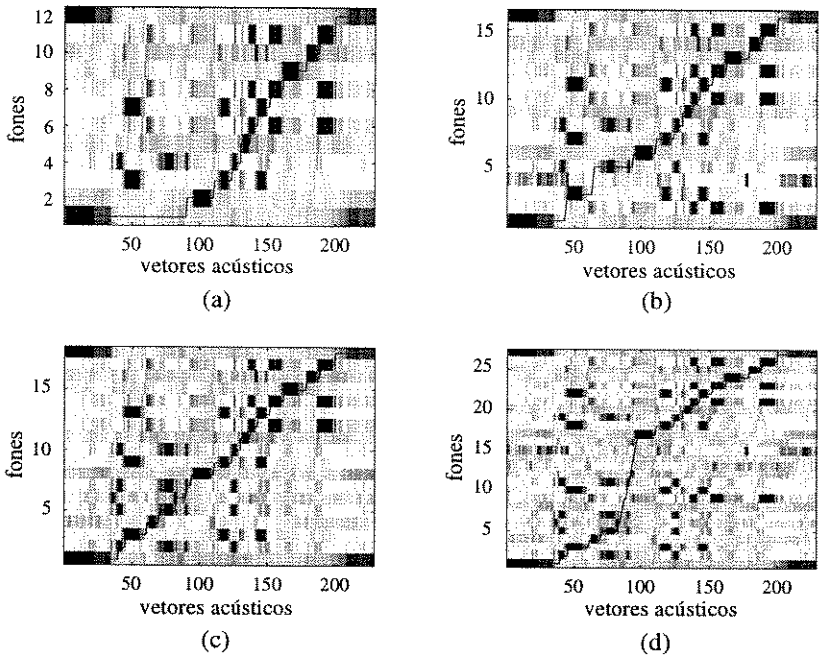

Figura 7-5: Modelos ocultos de Markov associados às sentenças : (a) "É suficiente", (b) "Isto é suficiente", (c) "O saldo é suficiente" e (d) "O saldo de sua conta é suficiente".

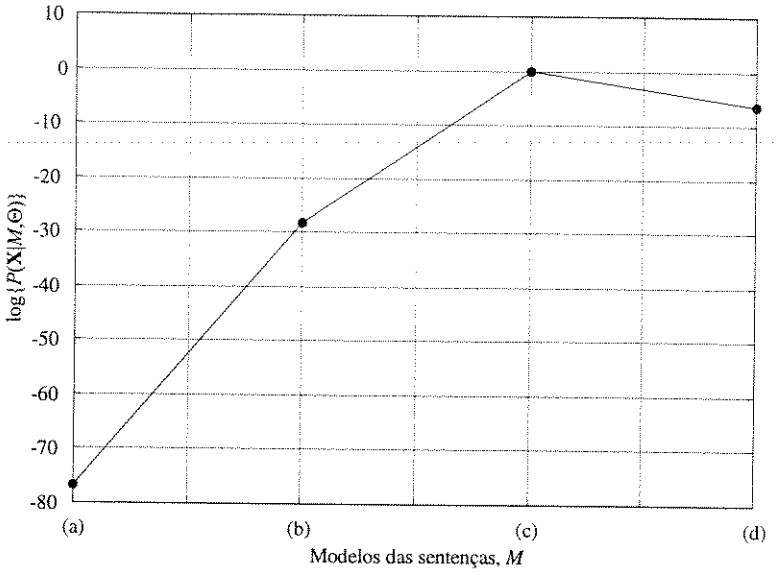

Figura 7-6: Log's das verossimilhanças normalizadas para os caminhos de Viterbi apresentados na Figura 7-5.

7.3.2 Discussão

Para avaliação de desempenho do sistema para reconhecimento de sentenças a partir de modelos de sentenças, utilizou-se as 100 frases presentes no Apêndice C e a taxa de acertos do sistema foi de 100%. Estes resultados comprovam a capacidade das ANNs de realizarem um bom modelamento acústico, uma vez que existem sentenças que diferem entre si em apenas uma única palavra, e também a boa capacidade de absorção das variabilidades temporais por parte dos HMMs, pois existem sentenças com apenas uma única palavra e outras com até vinte palavras.

Alguns autores [33] aplicam o algoritmo de Viterbi diretamente à saída da ANN, $P(q_k|\mathbf{x}_n, \Theta)$. Esta abordagem apresenta o inconveniente de não permitir a interpretação do segundo estágio do modelo híbrido como um HMM, pois este baseia-se em verossimilhanças de emissão de símbolos $p(\mathbf{x}_n|q_k, \Theta)$. Além deste inconveniente a Figura 7-7 mostra que o uso de $p(\mathbf{x}_n|q_k, \Theta)$ apresenta uma maior capacidade de discriminação do modelo que gerou a seqüência de vetores acústicos apresentada ao sistema. A Figura 7-7 mostra os resultados obtidos pelos HMMs correspondentes às 20 sentenças presentes na Tabela 6.2 quando foi apresentado ao sistema a seqüência de vetores acústicos $\mathbf{X}$ correspondente a sétima sentença desta mesma Tabela.

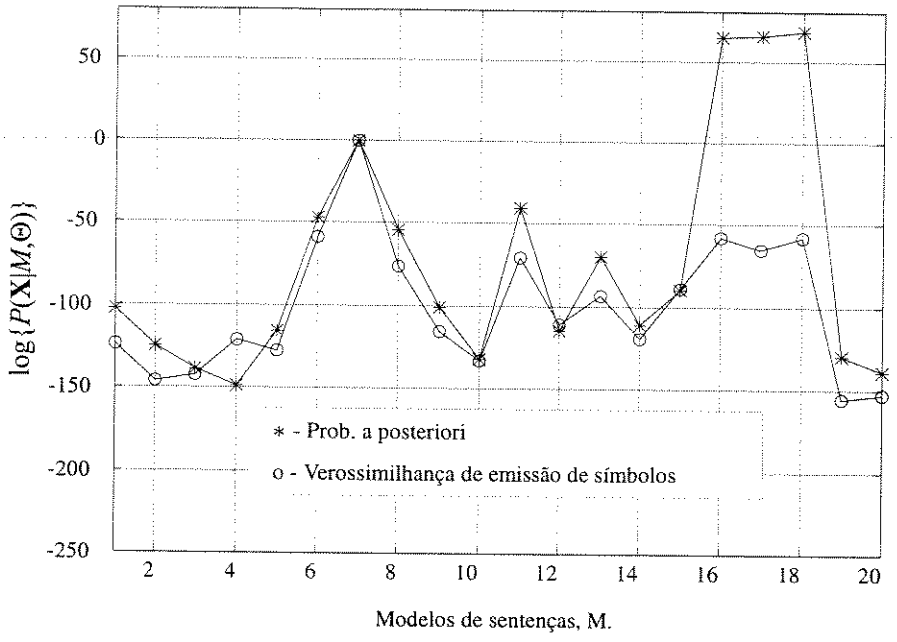

Figura 7-7: Resultados da apresentação da seqüência de vetores acústicos associados à sétima sentença da Tabela 6.2, aos modelos das vinte sentenças desta mesma tabela.

7.4 Desempenho do Algoritmo de Reestimação - REMAP

7.4.1 Reestimação dos parâmetros Ψ da ANN

A Figura 7-8 apresenta os alvos suaves $P(q_k|\mathbf{X}, M, \Theta)$ para a sentença "É suficiente". A Figura 7-9 apresenta os valores máximos de $P(q_k|\mathbf{X}, M, \Theta)$ para cada um dos vetores acústicos $\mathbf{x}_n$ presentes em $\mathbf{X}$. Estas duas figuras ilustram o grau de confusão acústica, nas fronteiras entre os fones, estabelecido pelos alvos suaves a serem utilizados na reestimação dos parâmetros Ψ da ANN.

A Figura 7-10 apresenta a saída da ANN após a estimação dos parâmetros Pré-REMAP e a Figura 7-11 apresenta a saída da ANN após o REMAP-2 (duas reestimações). Uma análise destas duas figuras permite as seguintes conclusões :

1. O processo de reestimação aumentou os valores das probabilidades a posteriori $P(q_k|\mathbf{x}_n)$ na saída da ANN.

2. O processo de reestimação reduziu as descontinuidades existentes entre as transições de um estado (fone) para outro.

O acréscimo dos valores das probabilidades a posteriori $P(q_k|\mathbf{x}_n)$ pode ser atribuído ao aumento

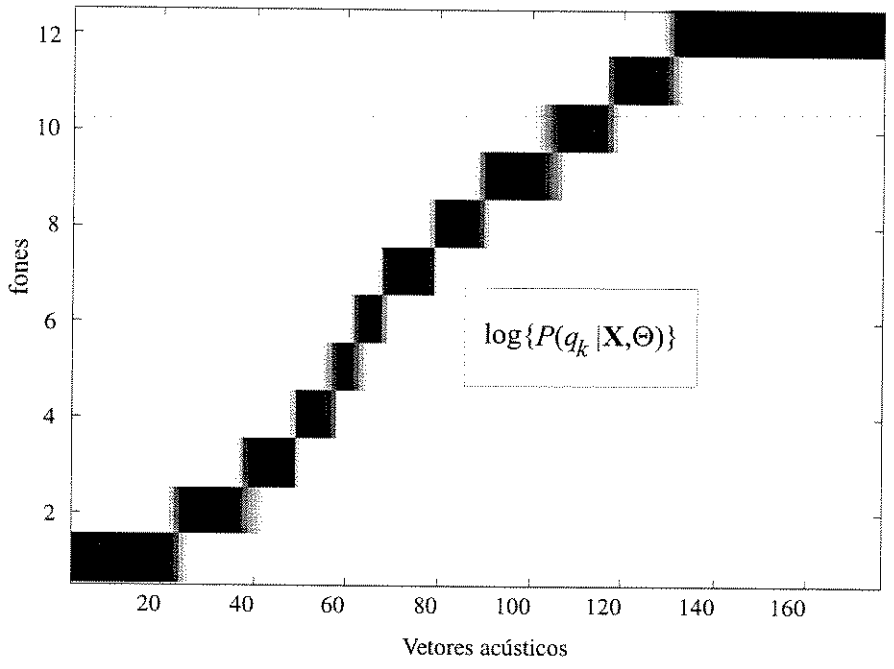

Figura 7-8: Alvos suaves $P(q_k|X, M, \Theta)$ para a sentença "É suficiente".

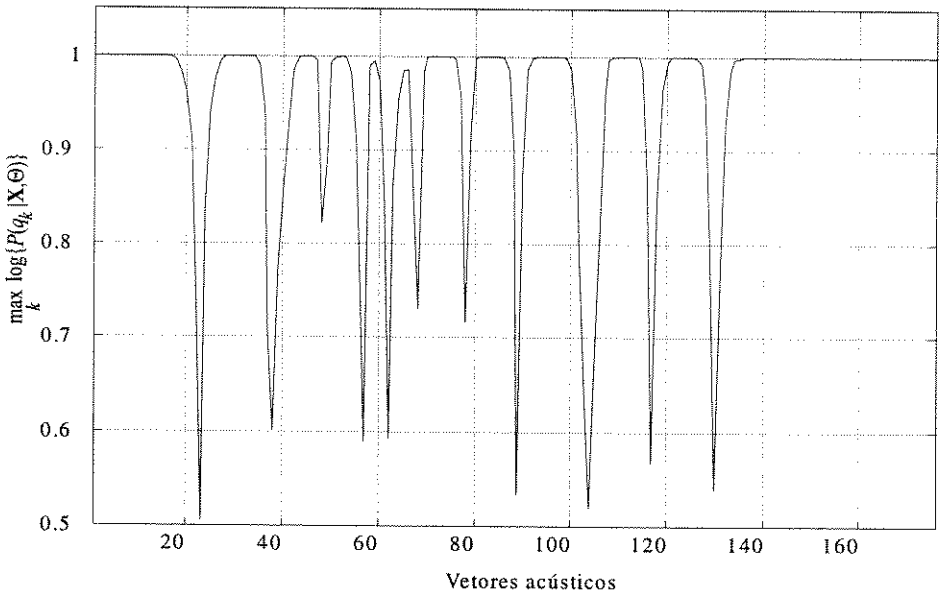

Figura 7-9: Valores máximos dos alvos $P(q_k|X, M, \Theta)$ para todos os estados k, para cada um dos vetores acústicos $x_n \in X$.

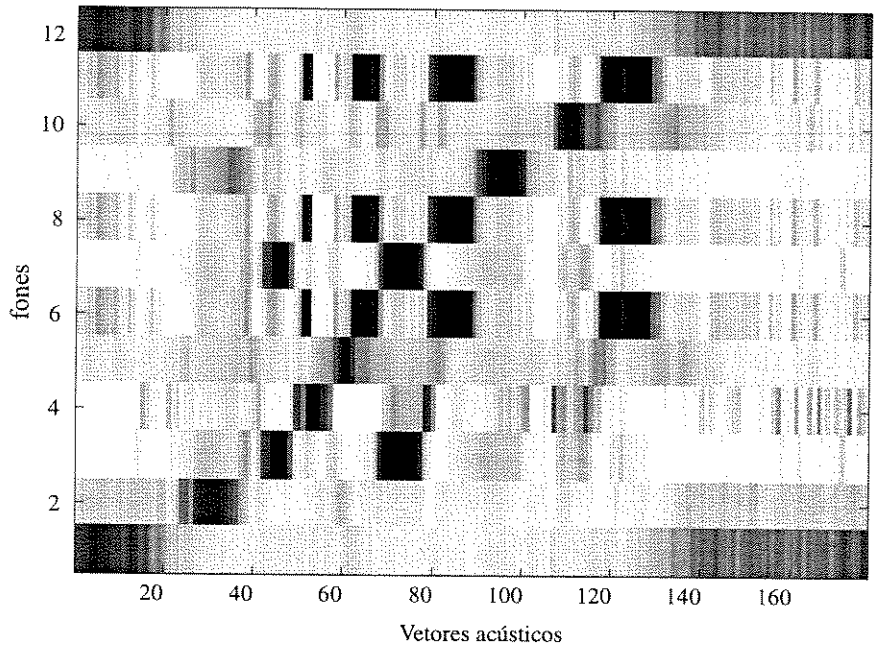

Figura 7-10: Saídas da ANN correspondentes aos fones presentes na sentença "É suficiente", antes da reestimação, Pré-REMAP.

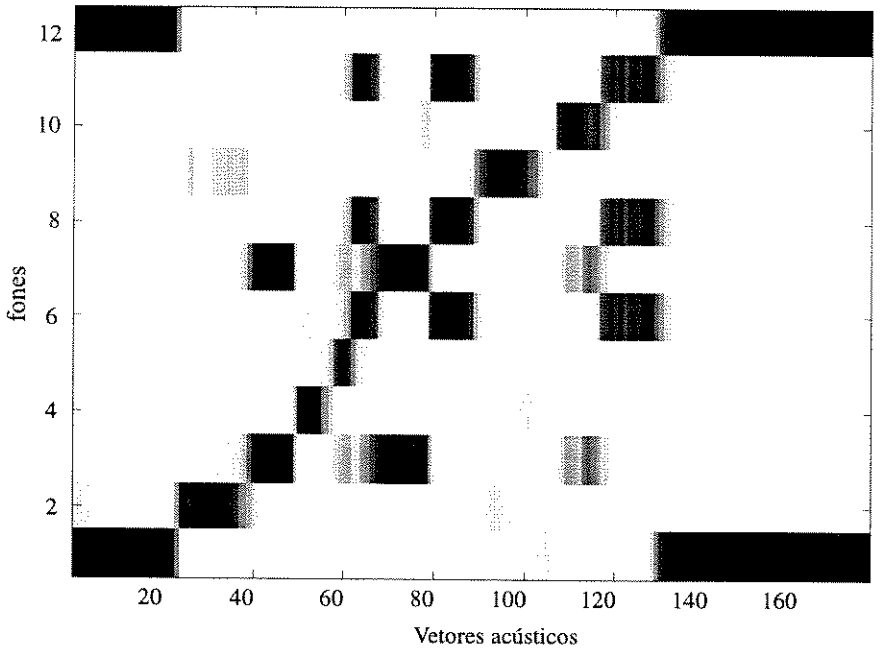

Figura 7-11: Saídas da ANN correspondentes aos fones presentes na sentença "É suficiente", após duas reestimações, REMAP-2.

na quantidade de exemplos de treinamento utilizados durante a fase de reestimação em relação aos exemplos de treinamento utilizados durante a fase de estimação[1]. A redução das descontinuidades deve-se ao fato que durante a fase de reestimação, ao contrário da fase de estimação, a ANN também é treinada com exemplos extraídos das regiões de fronteira entre as unidades sub-lexicais.

7.4.2 Reestimação das Probabilidades de Transição

A Figura 7-12 apresenta as probabilidades de autotransição de estados para três casos :

1. $P^{(0)}\left(q_k^n|q_k^{n-1}\right)$- Estimadas em função das durações médias das unidades sub-lexicais, conforme Equação (6.1).

2. $P^{(1)}\left(q_k^n|q_k^{n-1}\right)$- Estimadas a partir dos alvos suaves utilizados na 1ª reestimação - REMAP-1.

3. $P^{(2)}\left(q_k^n|q_k^{n-1}\right)$- Estimadas a partir dos alvos suaves utilizados na 2ª reestimação - REMAP-2.

A partir da Figura 7-12 pode ser verificado que após o REMAP-2 as probabilidades de autotransição entre estados, $P^{(2)}\left(q_k^n|q_k^{n-1}\right)$ tornaram-se muito próximas das probabilidades de autotransição estimadas em função das durações médias dos fones $P^{(0)}\left(q_k^n|q_k^{n-1}\right)$, com exceção dos estados, 30 e 35 (fones /η/ e /λ/). Como as estimativas destas probabilidades de autotransição são baseadas em levantamentos estatísticos, uma possível explicação para as diferenças nos casos dos fones /η/ e /λ/, pode ser associada ao fato destes fones ocorrerem em número extremamente reduzido.

A Figura 7-13 apresenta as probabilidades de transição de cada um dos 36 estados para os mesmos casos apresentados na Figura 7-12

7.4.3 Reestimação das Probabilidades a Priori das Classes

A Figura 7-14 apresenta as probabilidades a priori das classes (estados ou fones) para três casos :

1. $P(q_k)$ - Estimadas a partir das freqüências relativas (freqüência de ocorrência da classe no conjunto de treinamento $\mathbf{X}_e$).

2. $P^{(0)}(q_k)$ - Estimadas a partir das saídas da ANN sem reestimação - Pré-REMAP.

3. $P^{(2)}(q_k)$ - Estimadas a partir das saídas da ANN com reestimação - REMAP-2.

[1]Durante a etapa de reestimação os exemplos de treinamento são extraídos a cada quadro de análise e durante a etapa de estimação os exemplos são extraídos apenas centrados entre as marcas de segmentação dos fones.

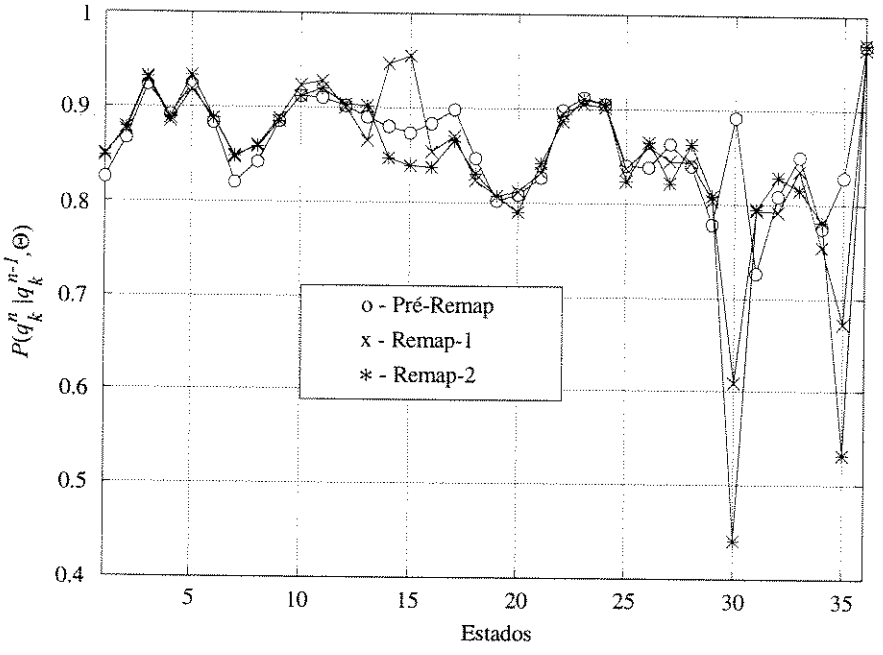

Figura 7-12: Probabilidades de permanência nos estados $P\left(q_k^n|q_k^{n-1},\Theta\right)$

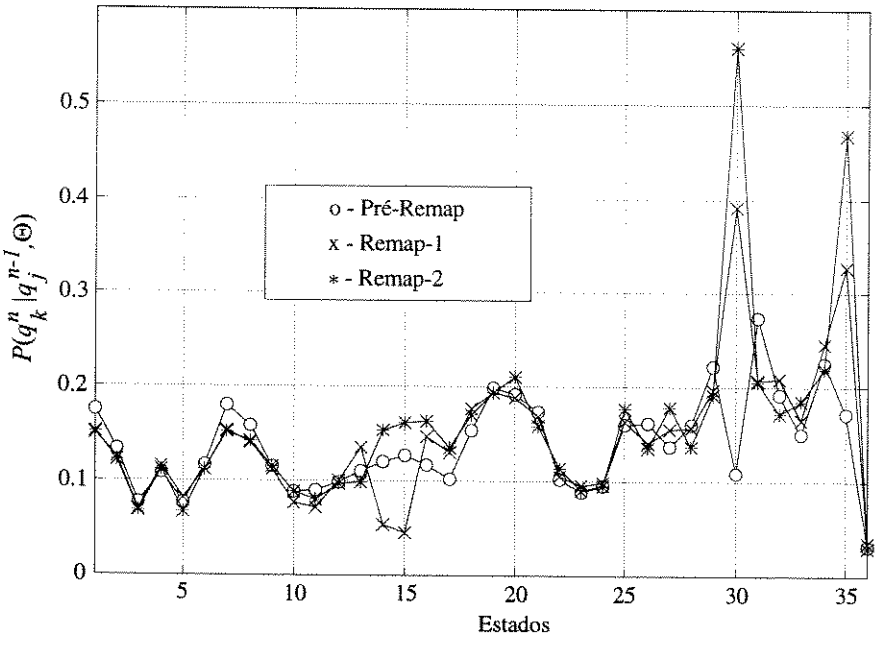

Figura 7-13: Probabilidade de transição de estados $P\left(q_k^n|q_j^{n-1},\Theta\right)$.

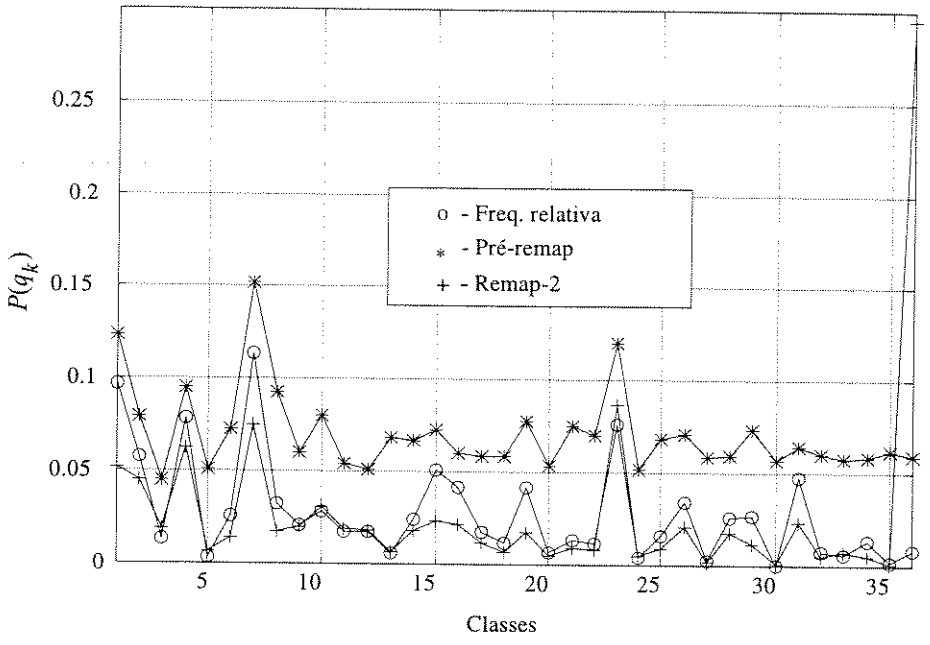

Figura 7-14: Probabilidades das 36 classes (fones) calculadas segundo a freqüência relativa e segundo a Equação 6.4 para os casos : Pré-REMAP e REMAP-2.

A partir da Figura 7-14 pode ser verificado que $P^{(2)}(q_k)$ aproxima $P(q_k)$ com uma precisão razoável, com exceção do estado 36 (fone /#/). Esta diferença está associada ao fato da estimativa através da freqüência relativa não levar em consideração as durações médias dos fones[2]. Através da Figura 7-14, também pode ser verificado que, a menos de um deslocamento, $P^{(2)}(q_k)$ aproxima com razoável precisão $P^{(1)}(q_k)$. Este deslocamento de $P^{(1)}(q_k)$ deve-se ao fato que durante o Pré-REMAP a ANN ainda apresenta uma elevada superposição entre as classes de saídas (fones).

7.4.4 Discussão

Para verificar a influência do processo de reestimação dos parâmetros dos modelos híbridos ANN-HMM no desempenho final do sistema, considere o exemplo da Figura 7-15 em que os vetores acústicos $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, correspondentes à sétima sentença da Tabela 6.2, são apresentados aos modelos das 20 sentenças desta mesma Tabela. Neste exemplo são apresentados os resultados (log da verossimilhança normalizada $\log(p(\mathbf{X}|M,\Theta))$, para três casos : (1) Pré-REMAP (sem reestimação), (2) REMAP-1 (primeira reestimação) e (3) REMAP-2 (segunda reestimação). A partir dos resultados

[2]Em geral as sentenças utilizadas são iniciadas e finalizadas por trechos de silêncio com durações médias em torno de 300 ms.

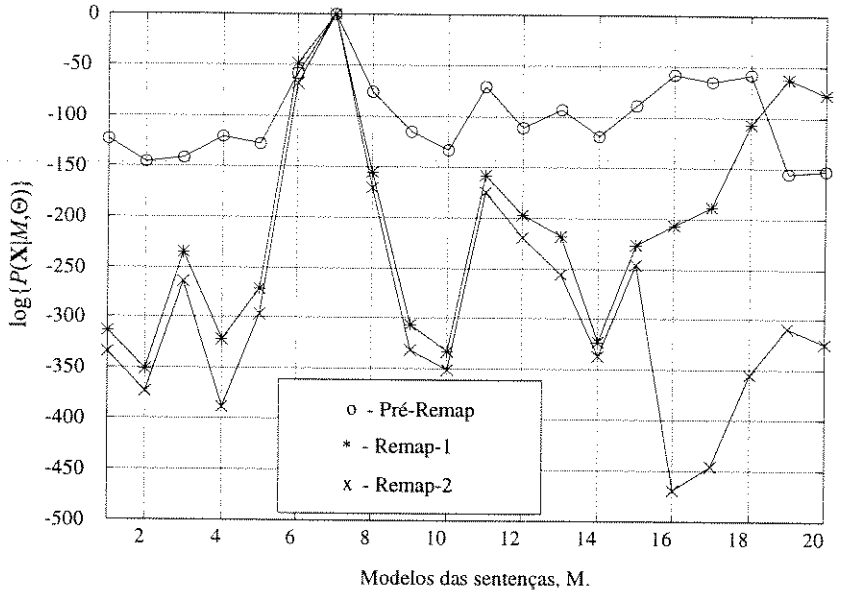

Figura 7-15: Log's das verossimilhanças normalizadas correspondentes a apresentação da seqüência de vetores acústicos associada à sétima sentença da Tabela 6.2, aos 20 modelos de sentenças desta mesma tabela.

da Figura 7-15 pode ser verificado que, em geral, a cada iteração os modelos das sentenças tornam-se mais discriminativos.

7.5 Reconhecimento de Sentenças com uso de Modelos de Palavras

Para avaliar o desempenho do sistema no reconhecimento de sentenças a partir de modelos de palavras foram realizados os seguintes experimentos :

1. Sem o uso de modelo da língua $P(M|\Theta) = 1$, e com o uso do modelo de duração apresentado na seção 5.2.3, foram avaliados os casos : Pré-REMAP, REMAP-1, REMAP-2 e REMAP-3.

2. Sem o uso de modelo da língua $P(M|\Theta) = 1$, e também sem o uso do modelo de duração avaliou-se o REMAP-3.

3. Como o uso de restrições gramaticais do tipo $P - Gram$, e com o uso do modelo de duração avaliou-se o REMAP-3.

Os resultados destes experimentos encontram-se na Tabela 7.2 e uma lista completa com as 100 sentenças reconhecidas para os experimentos 1 e 3 encontra-se no Apêndice C. Nesta tabela tem-se que :

	M. de duração	M. de linguagem	S (%)	D (%)	I (%)	E (%)
Pré-REMAP	Sim	$P(M\|\Theta) = 1$	11,07	8,50	13,46	33,03
REMAP-1	Sim	$P(M\|\Theta) = 1$	5,67	5,14	10,27	21,08
REMAP-2	Sim	$P(M\|\Theta) = 1$	4,69	4,87	8,24	17,80
REMAP-3	Sim	$P(M\|\Theta) = 1$	4,51	4,96	7,26	16,74
REMAP-3	Não	$P(M\|\Theta) = 1$	27,54	2,13	12,75	42,42
REMAP-3	Sim	Gramática $P - Gram.$	0	0,35	0,18	0,53

Tabela 7.2: Resultados da avaliação do sistema híbrido ANN-HMM implementado, no reconhecimento das 100 sentenças do Apêndice C.

	S (%)	D (%)	I (%)	E (%)
HMM-1-MCC	24,0	3,10	27,10	54,21
HMM-1-MCC-D	16,82	3,89	20,72	41,45
HMM-3-MCC	10,89	3,80	14,70	29,40
HMM-3-MCC-D	6,64	2,48	9,12	18,2

Tabela 7.3: Resultados da avaliação de um sistema em desenvolvimento no LPDF-UNICAMP, baseado apenas em HMM discreto, no reconhecimento das 100 sentenças apresentadas no Apêndice C.

- $S = 100 \cdot \frac{N_S}{I_W}$ - Taxa de erro de substituição (N_S - número total de erros de substituição e I_W - número total de palavras presentes no léxico).
- $D = 100 \cdot \frac{N_E}{I_W}$ - Taxa de erro de deleção (N_E - número total de erros de deleção).
- $I = 100 \cdot \frac{N_I}{I_W}$ - Taxa de erro de inserção (N_I - número total de erros de inserção).
- $E = 100 \cdot \frac{N_S+N_E+N_I}{I_W}$ - Taxa total de erro de palavras.

Neste momento, está sendo desenvolvido no LPDF - UNICAMP, um sistema para reconhecimento de fala contínua empregando apenas modelos ocultos de Markov discretos com codebook de 256 posições. Este sistema foi avaliado com as 100 sentenças do Apêndice C e os resultados do sistema encontram-se na Tabela 7.3. Nesta tabela tem-se que :

- **HMM-1-MCC** - HMM com um estado por fone e utilizando somente coeficientes Mel Cepstrais (12 coeficientes).

- **HMM-1-MCC-D** - HMM com um estado por fone e utilizando coeficientes Mel Cepstrais e coeficientes Delta Mel Cepstrais.
- **HMM-3-MCC** - HMM com três estados por fone e utilizando somente coeficientes Mel Cepstrais.
- **HMM-3-MCC-D** - HMM com três estados por fone e utilizando coeficientes Mel Cepstrais e coeficientes Delta Mel Cepstrais.

Durante a obtenção dos resultados da Tabela 7.3 fez-se uso do mesmo modelo de duração utilizado no sistema híbrido, e não foi utilizado qualquer tipo de restrição gramatical $P(M|\Theta) = 1$.

Capítulo 8

Considerações Finais

8.1 Análise de Desempenho

8.1.1 Reconhecimento de Fones

- **Bases de Dados Desbalanceada** O desempenho da ANN como classificadora de unidades sub-lexicais poderia ter sido melhor se a base de dados utilizada não fosse tão desbalanceada. Os fones /u/ e /i/ possuem 482 e 411 exemplares, respectivamente, enquanto os fones /η/ e /ʃ/ possuem, respectivamente, apenas 6 e 2 exemplares.

- **Parâmetros de Entrada** No sistema implementado não houve uma preocupação em otimizar o módulo de Pré-processamento. Por este motivo foram utilizados apenas 12 coeficientes Mel Cepstrais [5]. Porém acredita-se que o desempenho do sistema pode ser melhorado se forem utilizados além dos coeficientes Mel Cepstrais, outros parâmetros tais como : o Log de Energia Normalizada [26], coeficientes Delta Mel Cepstrais[1] [6] e Delta Log de Energia Normalizada [26].

- **Dimensões da Rede** As dimensões da ANN foram ajustadas para melhor otimizar o processo de estimação (Pré-REMAP). A rede MLP foi dimensionada com 70 neurônios na camada escondida porque a base de dados inicial (fones segmentados manualmente) possuía poucos exemplares. Mas durante a reestimação a quantidade de exemplos aumentou significativamente uma vez que estes passaram a ser tomados a cada quadro de análise e não apenas centrados nas marcas de segmentação. Neste caso acredita-se que a quantidade de 70 neurônios tenha sido insuficiente durante as etapas de reestimação (REMAP) dos parâmetros da ANN. Duas possíveis soluções

[1] Com relação aos coeficientes Delta ainda existem dúvidas quanto a sua eficiência, uma vez que os modelos híbridos já são sensíveis a contexto. A influência destes coeficientes talvez limite-se às bordas do vetor de entrada $\mathbf{X}_{n-c}^{n+d}$.

para este problema são : (1) dimensionar a ANN de forma a melhor se adequar tanto aos exemplos utilizados durante a etapa de estimação (Pré-REMAP) como aos exemplos utilizados durante as etapas de reestimação (REMAP), (2) Utilizar uma ANN com poucos neurônios durante o Pré-REMAP e a partir desta ANN gerar os alvos suaves a serem utilizados no REMAP-1 e então utilizar uma outra ANN com um número maior de neurônios para as etapas de reestimação. O problema neste caso é que a ANN não poderá utilizar os pesos sinápticos da etapa Pré_REMAP.

8.1.2 Reconhecimento de Sentenças com o uso de Modelos de Sentenças

- **Eficiência do modelamento** Os resultados para o reconhecimento de sentenças mostram a eficiência, tanto da ANN no modelamento das *variabilidades acústicas* como dos HMMs na absorção das *variabilidades temporais* da fala. Os resultados mostram que o sistema foi capaz de reconhecer sentenças de tamanhos significativamente diferentes, frases com apenas 1 palavra e frase com até 20 palavras. Além disto mostra a capacidade do sistema em distinguir frases que diferem entre si em apenas uma única palavra como por exemplo : "*A cotação do dólar aumentou, e as bolsas fecharam em baixa*" e "*A cotação do dólar aumentou, mas as bolsas fecharam em baixa*".

- **Flexibilidade** O reconhecimento com o uso de modelos de sentenças somente faz sentido para uma aplicação específica em que o número de sentenças permitidas pelo sistema seja bastante reduzido (porque devem ser montados explicitamente o modelo de cada uma das sentenças). Apesar desta restrição os modelos híbridos ANN-HMM, conforme definidos neste trabalho, apresentam a grande flexibilidade de permitir ao usuário expandir o número de sentenças permitidas pelo sistema, bastando somente, que ele forneça a conversão ortográfico-fonética desta nova sentença. Isto é possível porque para montar o HMM de uma nova sentença o sistema necessita saber apenas quais HMMs de fones devem ser concatenados.

8.1.3 Reconhecimento de Sentenças com o uso de Modelos de Palavras

- **Modelos dos fones com apenas um único estado** No sistema implementado foi utilizado um único estado para modelar um fone. Em geral, os sistemas baseados em HMM que representam o estado-da-arte em reconhecimento de fala, utilizam 3 estados por fone com o objetivo de modelar melhor as variações estatísticas existentes ao longo de um fone.

- **Modelos de duração de palavras** Os resultados apresentados no Capítulo 7 mostram que o uso da informação sobre as durações médias das palavras durante o procedimento de decodificação realizado pelo algoritmo Level Building é capaz de melhorar, significativamente, a taxa de acertos de palavras do sistema.

- **Algoritmos de busca** O algoritmo de busca implementado, algoritmo "level building", não é muito eficiente do ponto de vista computacional. Uma versão mais otimizada do "Level Building" é o "Frame Sinchronous Level Building", que realiza a mesma tarefa do "Level Building", porém, permitindo a realização dos cálculos de todos os L níveis de procura de forma paralela e não seqüencial como é realizado no "Level Building".

8.2 Contribuições

- **Inovação** Primeiro trabalho na área de sistemas híbridos ANN-HMM aplicados ao reconhecimento de fala contínua realizado no Brasil.

- **Levantamento Bibliográfico** Foi realizado um vasto levantamento bibliográfico sobre reconhecimento de fala contínua, que será utilizado como base, no LPDF (Laboratório de Processamento Digital de Sinais de Fala da UNICAMP), para outros trabalhos sobre modelos híbridos ANN-HMM, técnicas de otimização conjunto de ANNs e HMMs, algoritmos de busca, modelamento de unidades sub-lexicais, independência de locutor e adaptabilidade de locutor.

- **Definição de metas de pesquisa na área de reconhecimento de fala contínua no LPDF - UNICAMP** Um dos grandes méritos deste trabalho foi o estabelecimento de bases para a realização de uma nova linha de pesquisa dentro do LPDF, o reconhecimento de fala contínua. Nessa linha deve-se destacar a utilização de modelos híbridos ANN-HMM para o modelamento acústico de unidades sub-lexicais e o domínio de algoritmos básicos para a decodificação acústica.

8.3 Sugestões para Trabalhos Futuros

- **Pré-Processamento** Realizar um melhor pré-processamento do sinal, utilizando Log de Energia Normalizada, coeficientes Delta e possivelmente análises em multitaxas e segmentações não uniformes [42].

- **Construção de um segmentador automático** Para continuação dos trabalhos em reconhecimento de fala contínua é necessário a aquisição de uma nova base de dados com um número maior de sentenças e de palavras, e preferivelmente independente de locutor. No caso de sistemas baseados em modelos híbridos esta nova base deve ser pré-segmentada em termos de unidades sub-lexicais, necessitando, portanto, de um segmentador automático, possivelmente baseado em HMMs e alinhamentos de Viterbi.

- **Explorar melhor as não estacionariedades da fala** Melhorar o modelamento das unidades sub-lexicais utilizando um número maior de estados por fone (possivelmente 3 estados por fone) para modelar melhor as variabilidades estatísticas do sinal ao longo de uma unidade sub-lexical.

- **Unidades sub-lexicais dependentes de contexto** As unidades sub-lexicais utilizadas no sistema implementado (fones) foram considerados independentes de contexto, isto é, o modelo de cada fone utilizado não levava em consideração qual o fone que o precedia ou sucedia. O modelamento de fones dependentes de contexto não é uma tarefa difícil de ser implementada com o uso de um sistema híbrido ANN-HMM [33].

- **Métodos para otimização das ANNs** Desenvolver novos algoritmos para otimização das ANNs com o objetivo de melhor estimar as probabilidades $P(q_k|\mathbf{x}_n)$. Como sugestões podem ser desenvolvidos algoritmos híbridos baseados em algoritmos genéticos e métodos de gradiente [25].

- **Algoritmos de busca mais eficientes computacionalmente** Implementação de algoritmos de busca mais eficientes computacionalmente como é o caso do "Frame Synchronous Level Building" [17].

- **Restrições gramaticais** Realizar um estudo detalhado sobre restrições gramaticais a serem utilizadas durante o procedimento de busca e também como uma etapa de pós-processamento.

- **Uso de bases de dados padrões** Com o objetivo de avaliar melhor o sistema e poder comparar os resultados obtidos com os resultados apresentados em periódicos internacionais, utilizar bases de dados padrão em Inglês, como por exemplo o TIMIT (6300 sentenças, 6299 palavras, previamente rotulada em unidades sub-lexicais e gravada em ambiente de estúdio) e o NTIMIT (semelhante ao TIMIT, porém com a simulação de ruído telefônico).

- **Modelos híbridos ANN-HMM discriminativos** Implementar o sistema híbrido ANN-HMM discriminativo proposto por Konig e H. Bourlard [14].

Apêndice A

Critério MMI

O critério de maximização da informação mútua - MMI ("Maximum Mutual Information"), procura ressaltar a discriminação entre os modelos que competem entre si, na tentativa de utilizar da melhor forma possível as informações diponíveis no limitado conjunto de treinamento. Para uma dada seqüência de vetores acústicos $\mathbf{X}_{M_j}^{(s)}$ o critério MMI treina o modelo correto M_j (modelo associado a $\mathbf{X}_{M_j}^{(s)}$) positivamente e treina negativamente todos os outros modelos M_i, com $i \neq j$, ajudando a diferenciar melhor os modelos que competem entre si, e também melhorando a capacidade de discriminação durante a etapa de avaliação. A informação mútua entre uma seqüência de vetores acústicos $\mathbf{X}_{M_j}^{(s)}$ e o seu modelo associado M_j é definida segundo [27] como :

$$I\left(\mathbf{X}_{M_j}^{(s)}, M_j|\Theta\right) = \log \frac{p\left(\mathbf{X}_{M_j}^{(s)}, M_j|\Theta\right)}{p\left(\mathbf{X}_{M_j}^{(s)}|\Theta\right) \cdot P\left(M_j|\Theta\right)} \tag{A.1}$$

$$= \log \frac{p\left(\mathbf{X}_{M_j}^{(s)}|M_j, \Theta\right)}{p\left(\mathbf{X}_{M_j}^{(s)}|\Theta\right)} \tag{A.2}$$

$$= \log p\left(\mathbf{X}_{M_j}^{(s)}|M_j, \Theta\right) - \log p\left(\mathbf{X}_{M_j}^{(s)}|\Theta\right) \tag{A.3}$$

utilizando o conceito de probabilidade marginal e realizando algumas manipulações simples, a Equação $(A.3)$ pode ser reescrita, como:

$$I\left(\mathbf{X}_{M_j}^{(s)}, M_j|\Theta\right) = \underbrace{\log p\left(\mathbf{X}_{M_j}^{(s)}|M_j, \Theta\right)}_{\mathbf{I}} - \underbrace{\sum_i \log p\left(\mathbf{X}_{M_j}^{(s)}|M_i, \Theta\right) \cdot P\left(M_i|\Theta\right)}_{\mathbf{II}} \tag{A.4}$$

sendo que :

I - Representa o treinamento positivo do modelo correto M_j (justamente como no critério ML).

II - Representa o treinamento negativo de todos os outros modelos M_i com $i \neq j$.

O treinamento com o critério MMI consiste na determinação dos parâmetros Θ que maximizem a informação mútua

$$\Theta_{MMI} = \arg\max I\left(\mathbf{X}_{M_j}^{(s)}, M_j|\Theta\right) \tag{A.5}$$

Infelizmente , esta equação não pode ser solucionada por análise direta ou por reestimação [33]. Uma possível solução baseia-se em métodos de gradiente [33]. Entretanto tais procedimento de otimização estão sujeitos a problemas de implementação númérica [27].

Também pode ser verificado através da Equação ($A.2$) que o critério MMI é equivalente ao critério de maximização da probabilidade a posteriori MAP ("Maximum A Posteriori"), o qual é expresso por $P\left(M_j|\mathbf{X}_{M_j}^{(s)}, \Theta\right)$ conforme Equação (2.9). Para verificar esta equivalência basta reescrever $P\left(M_j|\mathbf{X}_{M_j}^{(s)}, \Theta\right)$ em termos da regras de Bayes :

$$P\left(M_j|\mathbf{X}_{M_j}^{(s)}, \Theta\right) = \frac{p\left(\mathbf{X}_{M_j}^{(s)}|M_j, \Theta\right) \cdot P\left(M_j|\Theta\right)}{p\left(\mathbf{X}_{M_j}^{(s)}|\Theta\right)} \tag{A.6}$$

Como o termo $P\left(M_j|\Theta\right)$ que representa o modelo da língua em geral é considerado independente de Θ, $P\left(M_j|\Theta\right) = P\left(M_j\right)$ e como a função logaritmo é uma função monotônica, então :

$$\arg\max_{\Theta} P\left(M_j|\mathbf{X}_{M_j}^{(s)}, \Theta\right) \Leftrightarrow \arg\max_{\Theta} \left\{\log \frac{p\left(\mathbf{X}_{M_j}^{(s)}|M_j, \Theta\right)}{p\left(\mathbf{X}_{M_j}^{(s)}|\Theta\right)}\right\} \tag{A.7}$$

Apêndice B

Dicionário de Pronúncias

A seguir são apresentadas todas as palavras que compõem o léxico utilizado, seguidas de suas respectivas pronúncias (representação em termos de fones). É importante observar que existem palavras com mais de uma pronúncia.

1.	(SILÊNCIO)	#
2a.	A	*a*
2b.	A	*á*
3.	ACEITARÃO	*á s e t á ɾ̈ ã u*
4.	ACORDO	*a k o ɾ̈ d u*
5.	ADELAIDE	*a d e l á i D i*
6.	AFIRMAR	*a f i r m á r*
7.	AGUARDAREMOS	*á g u á ɾ̈ d a ɾ̈ ẽ m u z*
8.	AINDA	*á ĩ d*
9.	AJUSTES	*a j u s T i s*
10.	ALUNOS	*a l u n o z*
11.	AMANHÃ	*á m ã ɳ ã*
12a.	ANALISTAS	*á n á l i s t a z*
12b.	ANALISTAS	*á n á l i s t a s*
13.	ANTECIPAR	*ã t e s i p á ɾ̈*
14.	ANTERIOR	*ã t e ɾ̈ i o r*
15a.	AO	*a u*
15b	AO	*á u*
16a.	AOS	*á u s*
16b.	AOS	*a u s*
17.	APLICAÇÕES	*á p l i k á s õ i z*
18.	APROVEITAR	*á p ɾ̈ o v e i t á ɾ̈*
19.	AQUI	*á k i*
20a.	AS	*a s*
20b.	AS	*á s*
20c.	AS	*á z*
21.	ASSIM	*á s ĩ*
22.	ASSINADO	*á s i n á d u*
23.	ATÉ	*á t é*
24a.	ATRASADAS	*á t ɾ̈ á z á d a z*
24b.	ATRASADAS	*á t ɾ̈ á z á d a s*
25.	ATRÁS	*á t ɾ̈ á z*
26.	ATRATIVIDADE	*á t ɾ̈ á T i v i d á D i*
27.	ATUALIZAÇÃO	*á t u á l i z á s ã u*

28.	AUMENTO	*á u m ẽ t u*
29a.	AUMENTOU	*á m ẽ t o u*
29b.	AUMENTOU	*á u m ẽ t o u*
30.	BAIXA	*b á i ʃ a*
31.	BAIXO	*b á i ʃ u*
32a.	BANCO	*b ã k u*
32b.	BANCO	*b ã k*
33.	BANCOS	*b ã k u z*
34.	BARROSO	*b á R o z u*
35.	BASTANTE	*b á s t ã T i*
36.	BENEFICIANDO	*b e n e f i s i ã d u*
37a.	BOLSA	*b o u s a*
37b.	BOLSA	*b o s a*
38a.	BOLSAS	*b o u s a s*
38b.	BOLSAS	*b o s a s*
39.	BRASIL	*b ɾ̈ a z i u*
40.	BRASILEIRA	*b ɾ̈ a z i l e ɾ̈ a*
41.	CADASTRAL	*k á d á s t ɾ̈ á u*
42.	CADERNETA	*k á d e ɾ̈ n e t a*
43.	CAFÉ	*k á f é*
44.	CAIXAS	*k á i ʃ a z*
45.	CARTÃO	*k á r t ã u*
46a.	CENTO	*s ẽ t*
46b.	CENTO	*s ẽ t u*
47.	CENTRAL	*s ẽ t ɾ̈ á u*
48.	CENTRO	*s ẽ t ɾ̈ u*
49a.	CHAMADA	*ʃ ã m á d a*
49b.	CHAMADA	*ʃ á m á d a*
50.	CHEGARAM	*ʃ e g á ɾ̈ ã u*
51.	CHEQUE	*ʃ é k i*
52.	CIDADE	*s i d á D i*
53.	CINCO	*s ĩ k u*
54.	CINQUENTA	*s ĩ k u ẽ t a*
55.	CLIENTE	*k l i ẽ T i*
56a.	CLIENTES	*k l i ẽ T i z*
56b.	CLIENTES	*k l i ẽ T z*
57.	CÓDIGO	*k ó D i g u*
58.	COLOCARÁ	*k o l o k á ɾ̈ á*
59a.	COM	*k õ*
59b.	COM	*k o*
60.	COMPARECIMENTO	*k õ p á ɾ̈ e s i m ẽ t u*
61.	CONDOMÍNIO	*k õ d o m i n i u*
62.	CONFORTO	*k õ f o r t u*
63.	CONSELHO	*k õ s e λ* u
64a.	CONSIDERADO	*k õ s i d e ɾ̈ á d*
64b.	CONSIDERADO	*k õ s i d e ɾ̈ á d u*
65a.	CONSIDERAVELMENTE	*k õ s i d e ɾ̈ á v e u m ẽ T i*
65b.	CONSIDERAVELMENTE	*k õ s d e ɾ̈ á v e u m ẽ T i*
65c.	CONSIDERAVELMENTE	*k õ s i d e ɾ̈ á v i u m ẽ T i*
66a.	CONSUMO	*k õ s u m u*
66b.	CONSUMO	*k õ s ũ m u*
67.	CONTA	*k õ t a*
68.	CONTAS	*k õ t á s*
69a.	CONTINUAM	*k õ T i n u ã*
69b.	CONTINUAM	*k õ T i n u ã u*
70a.	CONTRIBUINTES	*k õ t ɾ̈ i b u ĩ T i z*
70b.	CONTRIBUINTES	*k õ t ɾ̈ i b u ĩ T i s*
71.	CONVENCEMOS	*k õ v ẽ s ẽ m u z*
72a.	CONVÊNIO	*k õ v ẽ n i u*
72b.	CONVÊNIO	*k õ v e n i u*

73.	CORRIGIDAS	*k o R i j i d a s*
74.	COTAÇÃO	*k o t á s ã u*
75a.	CRÉDITO	*k r̈ é D i t u*
75b.	CRÉDITO	*k r̈ é D t u*
76.	CRUZ	*k r̈ u s*
77.	CRUZEIROS	*k r̈ u z e r̈ u z*
78.	CULTURAS	*k u t u r̈ a z*
79.	CUMPRIMENTO	*k ũ p r̈ i m ẽ t u*
80.	CURVA	*k u r̈ v a*
81a.	DA	*d*
81b.	DA	*d a*
82.	DADOS	*d a d u z*
83a.	DAS	*d á z*
83b.	DAS	*d a s*
83c.	DAS	*d a z*
84a.	DE	*d e*
84b.	DE	*D ĩ*
84c.	DE	*D*
85a.	DEPÓSITOS	*d e p ó z i t u z*
85c.	DEPÓSITOS	*d e p ó z t u z*
86.	DESCONTOS	*d e s k õ t u z*
87.	DESENVOLVER	*d e z ẽ v o v e r̈*
88.	DESTE	*d e s T i*
89.	DESTINO	*d e s T ĩ n*
90.	DETECTADO	*d e t e k t á d u*
91.	DETERMINAÇÃO	*d e t e r̈ m i n á s ã u*
92.	DEVE	*d é v i*
93a.	DEVEM	*d é v ẽ*
93b.	DEVEM	*d é v ẽ i*
94.	DEVEMOS	*d e v ẽ m u z*
95a.	DEVIDO	*d e v i d u*
95b.	DEVIDO	*d e v i d*
96.	DEZEMBRO	*d e z ẽ b r̈ u*
97.	DEZENOVE	*d e z ẽ n ó v i*
98.	DEZESSEIS	*d e z e s e z*
99.	DEZESSETE	*d e z e s é T i*
100.	DIA	*D i a*
101.	DIARIAMENTE	*D i á r̈ i á m ẽ T i*
102.	DIFERENTES	*D i f e r̈ ẽ T i s*
103.	DISPONÍVEIS	*D i s p o n i v e s*
104a.	DISPONÍVEL	*D s p o n i v i u*
104b.	DISPONÍVEL	*D i s p o n i v i*
104c.	DISPONÍVEL	*D i s p o n i v i u*
105.	DISPOSIÇÃO	*D s p o z i s ã u ũ*
106.	DO	*d u*
107.	DOCUMENTO	*d o k u m ẽ t u*
108.	DOIS	*d o i z*
109a.	DÓLAR	*d ó l a r̈*
109b.	DÓLAR	*d ó l á*
110.	DUZENTOS	*d u z ẽ t u z*
111a.	E	*e*
111b.	E	*i*
112.	É	*é*
113.	EFICÁCIA	*e f i k á s i a*
114.	EFICIÊNCIA	*e f i s i ẽ s i a*
115.	ELE	*e l i*
116a.	ELETRÔNICOS	*e l e t r̈ õ n i k u z*
116b.	ELETRÔNICOS	*e l e t r̈ o n i k u s*
117a.	EM	*ẽ i*
117b.	EM	*ẽ*

118.	EMBARQUE	*ê b á ɾ̈ k*
119a.	EMPRESA	*ẽ p ɾ̈ e z*
119b.	EMPRESA	*ẽ i p ɾ̈ e z a*
120.	EMPRESÁRIO	*ẽ p ɾ̈ e z á ɾ̈ i u*
121.	ENTRE	*ẽ t ɾ̈ i*
122.	ENTREGOU	*ẽ t ɾ̈ e g o u*
123.	ESTA	*é s t*
124a.	ESTÁ	*í s t á*
124b.	ESTÁ	*s t á*
125.	ESTAÇÃO	*e s t á s ã u*
126.	ESTARÃO	*i s t á ɾ̈ ã u*
127.	ESTATAL	*e s t á t á u*
128.	ESTÃO	*i s t ã u*
129a.	ESTÁVEL	*e s t á v e u*
129b.	ESTÁVEL	*e s t á v i u*
130.	EXPLÍCITA	*e s p l i s i t a*
131.	EXPRESSO	*s p ɾ̈ é s u*
132.	FAZER	*f a z e ɾ̈*
133.	FECHARAM	*f e ʃ á ɾ̈ ã*
134.	FICARÁ	*f i k á ɾ̈ á*
135.	FINANCIAMENTO	*f i n ã s i á m ẽ t u*
136.	FOI	*f o i*
137a.	FORMULÁRIOS	*f o ɾ̈ m u l á ɾ̈ i u s*
137b.	FORMULÁRIOS	*f o ɾ̈ m u l á ɾ̈ i u z*
138.	FORTALEZA	*f o ɾ̈ t á l e z a*
139.	FUNCIONÁRIO	*f ũ s õ n á ɾ̈ i u*
140.	FUTURO	*f u t u ɾ̈ u*
141a.	GOVERNO	*g o v e R n u*
141b.	GOVERNO	*g o v e r n u*
142.	HAVERÁ	*á v e ɾ̈ á*
143a.	HORAS	*ó ɾ̈ a s*
143b.	HORAS	*ó ɾ̈ a z*
144.	IBGE	*i b e j e é*
145.	IMEDIATO	*i m e D i á t u*
146.	IMPORTAÇÃO	*ĩ p o r t á s ã u*
147.	IMPOSTO	*ĩ p o s t u*
148.	IMPULSIONADO	*ĩ p u s i õ n á d u*
149.	IMPULSIONOU	*ĩ p u s õ n o u*
150a.	INADEQUADO	*i n á d e k u á d u*
150b.	INADEQUADO	*ĩ n á d e k u á d u*
151.	INCOMPLETO	*ĩ k õ p l é t u*
152.	INDEXADORES	*ĩ d e k s á d o ɾ̈ i s*
153.	INFORMA	*ĩ f ó r m a*
154a.	INÍCIO	*i n i s i u*
154b.	INÍCIO	*i n i s u*
155.	INSTALADOS	*ĩ s t á l á d u z*
156.	INSTITUIÇÕES	*ĩ s T i t u i s õ i z*
157.	INSUFICIENTE	*ĩ s u f i s i ẽ T i*
158.	INTEGRAÇÃO	*ĩ t e g ɾ̈ á s ã u*
159a.	INTERCÂMBIO	*ĩ t e ɾ̈ k ã b i u*
159b.	INTERCÂMBIO	*ĩ t e ɾ̈ k ã b i*
159c.	INTERCÂMBIO	*ĩ t e r k ã b i u*
160.	INTERESSANTE	*ĩ t e ɾ̈ e s ã T i*
161.	INTERNO	*ĩ t é ɾ̈ n u*
162a.	INVESTINDO	*ĩ v e s T ĩ d u*
162b.	INVESTINDO	*ĩ v e s T ĩ n*
163.	INVESTIR	*ĩ v e s T i r*
164.	ISTO	*i s t u*
165.	JUNHO	*j ũ ɳ o*
166.	JUROS	*j u ɾ̈ u z*

167.	-LHES	*λ i s*
168.	LOCALIZADO	*l o k á l i z á d o*
169.	MAIORIA	*m á i o ɾ̈ i a*
170.	MAIS	*m á i z*
171a.	MAS	*m á s*
171b.	MAS	*m á z*
172.	MEDIDA	*m e D i d a*
173.	MELHORES	*m e λ ó ɾ̈ i z*
174a.	MERCADO	*m e ɾ̈ k á d u*
174b.	MERCADO	*m e R k á d u*
175.	MÊS	*m e s*
176a.	MIL	*m i u*
176b.	MIL	*m i*
177.	MILHÕES	*m i λ õ i s*
178.	MINUTO	*m ĩ n u t u s*
179.	MODIFICAÇÕES	*m o d i f i k á s õ i z*
180.	MOMENTO	*m o m ẽ t u*
181.	MONETÁRIO	*m õ n e t á ɾ̈ i u*
182.	MUITO	*m ũ i t*
183.	NA	*n a*
184.	NACIONAL	*n á s õ n á u*
185.	NÃO	*n ã u*
186.	NAQUELE	*n á k e l i*
187.	NAS	*n a z*
188.	NECESSÁRIO	*n e s e s á ɾ̈ i u*
189.	NEM	*n ẽ i*
190.	NO	*n u*
191.	NOME	*n õ m i*
192.	NORMALMENTE	*n o ɾ̈ m á u m ẽ T*
193.	NOS	*n u s*
194.	NOTÍCIA	*n o T i s i a*
195.	NOVE	*n ó v i*
196.	NOVO	*n o v u*
196.	NÚMERO	*n ũ m i ɾ̈* u
198a.	O	*o*
198b.	O	*u*
199.	OITENTA	*o i t ẽ t a*
200.	ONTEM	*õ t ẽ ĩ*
201a.	OPERAÇÕES	*o p e ɾ̈ á s õ i s*
201b.	OPERAÇÕES	*o p e ɾ̈ á s õ i z*
202.	OPINIÃO	*o p i n i ã u*
203.	OPORTUNIDADE	*o p o r t u n i d á D i*
204a.	OS	*u s*
204b.	OS	*u z*
205a.	OU	*o u*
205b.	OU	*o u*
206.	OUTROS	*o t ɾ̈ u z*
207.	PÁGINAS	*p á j i n a s*
208.	PAÍS	*p á i s*
209a.	PARA	*p a ɾ̈*
209b.	PARA	*p á ɾ̈* a
210.	PARECE	*p á ɾ̈ é s*
211a.	PARTIR	*p á r T i r*
211b.	PARTIR	*p á ɾ̈ T i ɾ̈*
212.	PASSA	*p á s a*
213.	PASSADA	*p á s á d a*
214.	PASSADO	*p á s á d u*
215.	PASSAGEIROS	*p á s á j e ɾ̈ u s*
216.	PASSARÁ	*p á s á r á*
217.	PAULO	*p á u l u*

218.	PELO	*p e l u*
219.	PELOS	*p e l u z*
220.	PEQUENA	*p e k e n a*
221.	PERDA	*p e r d a*
222.	PERIGOSA	*p e ɾ̈ i g ó z a*
223.	PERMITA	*p e ɾ̈ m i t á*
224a.	PERMITE	*p e r m i T*
224b.	PERMITE	*p e ɾ̈ m i T*
224c.	PERMITE	*p e ɾ̈ m ĩ T i*
225.	PERMITIDO	*p e ɾ̈ m í T i d u*
226.	PESQUISA	*p e s k i z a*
227.	PLANO	*p l ã n u*
228.	POR	*p u ɾ̈*
229.	PORCENTO	*p u ɾ̈ s ẽ t u*
230.	PORQUE	*p u ɾ̈ k e*
231.	PORTÃO	*p o ɾ̈ t ã u*
232.	POSSO	*p ó s o*
233.	POUPANÇA	*p o u p ã s a*
234.	PRECISO	*p ɾ̈ e s i z*
235.	PREÇO	*p ɾ̈ e s u*
236.	PREÇOS	*p ɾ̈ e s u s*
237.	PRESTAÇÃO	*p ɾ̈ e s t á s ã u*
238.	PREZADO	*p ɾ̈ e z á d u*
239.	PROBLEMA	*p ɾ̈ o b l ẽ m á*
240.	PROJETO	*p ɾ̈ o j é t u*
241.	PROVOCANDO	*p ɾ̈ o v o k ã n*
242.	PRÓXIMO	*p ɾ̈ ó s i m u*
243.	PÚBLICA	*p u b l i k a*
244.	QUADRA	*k u á d ɾ̈ a*
245.	QUANDO	*k u ã d u*
246.	QUARENTA	*k u á ɾ̈ ẽ t*
247.	QUATRO	*k u á t ɾ̈ u*
248.	QUATROCENTOS	*k u á t ɾ̈ u s ẽ t u z*
249.	QUE	*k i*
250.	QUEDA	*k é d a*
251.	QUINHENTOS	*k ĩ ẽ t u s*
252.	QUINZE	*k ĩ z i*
253.	RADICAIS	*R á d i k á i z*
254.	REAIS	*R e á i s*
255a.	REAL	*R e á u*
255b.	REAL	*R é á u*
256a.	RECENTEMENTE	*R e s ẽ T i m ẽ T i*
256b.	RECENTEMENTE	*R e s ẽ T i m ẽ T*
257.	RECIFE	*R e s i f i*
258.	REGISTRADO	*R e j i s t ɾ̈ á d u*
259a.	REGRAS	*R é g ɾ̈ a s*
259b.	REGRAS	*R é g ɾ̈ a z*
260.	RESOLUÇÃO	*R e z o l u s ã u*
261.	REUNIÃO	*R e u n i ã u*
262.	RIO	*R i u*
263a.	SALDO	*s á u d u*
263b.	SALDO	*s á u d*
264.	SANTA	*s ã t a*
265.	SÃO	*s ã u*
266.	SAQUES	*s á k i s*
267.	SE	*s i*
268.	SEGUINTE	*s i g ĩ T i*
269.	SEGUIR	*s e g i ɾ̈*
270.	SEGUNDO	*s e g ũ d u*
271.	SEMANA	*s e m ã n a*

272.	SEMANAS	*s e m ã n a s*
273a.	SEMPRE	*s ẽ p r̈* i
273b.	SEMPRE	*s ẽ p r̈ e*
274.	SER	*s e r*
275.	SERÁ	*s e r̈ á*
276.	SESSENTA	*s e s ẽ t a*
277.	SETE	*s é T i*
278.	SETENTA	*s e t ẽ t*
279.	SEU	*s e u*
280.	SEUS	*s e u s*
281.	SIM	*s ĩ*
282.	SOBRE	*s o b r̈*
283.	SOFRERÁ	*s o f r̈ e r̈* á
284.	SUA	*s u a*
285.	SUBINDO	*s u b ĩ d u*
286.	SUBSTITUÍDO	*s u b i s T i t u i d u*
287.	SUFICIENTE	*s u f i s i ẽ T i*
288.	SUL	*s u u*
289.	TAXAS	*t á ʃ a z*
290.	TELEBRAS	*t e l e b r̈ á s*
291.	TELECOMUNICAÇÕES	*t é l i k o m u n i k á s õ i z*
292a.	TELEFÔNICA	*t e l e f õ n i k a*
292b.	TELEFÔNICA	*t e l e f õ n i k*
293.	TELEFÔNICAS	*t e l e f õ n i k a s*
294.	TELESP	*t e l é s p*
295.	TEM	*t ẽ i*
296.	TEMPO	*t ẽ p u*
207.	TERÁ	*t e r̈ á*
298.	TIVEMOS	*T i v ẽ m u z*
299.	TODAS	*t o d a z*
300a.	TODOS	*t o d u z*
300b.	TODOS	*t o d u s*
301.	TRABALHO	*t r̈ á b á λ u*
302.	TRATA	*t r̈ á t a*
303.	TRÊS	*t r̈ e s*
304.	TREZENTOS	*t r̈ e z ẽ t u z*
305a.	TRINTA	*t r̈ ĩ t a*
305b.	TRINTA	*t r̈ ĩ t*
306.	ÚLTIMA	*u T i m a*
307.	ÚLTIMAS	*u T i m a s*
308.	UM	*ũ*
309a.	UMA	*ũ m a*
309b.	UMA	*u m a*
310.	UNA	*u n a*
311.	UNIVERSIDADES	*u n i v e r s i d á D i z*
312a.	VALOR	*v á l o r*
312b.	VALOR	*v á l o r̈*
313a.	VENCIMENTO	*v ẽ s i m ẽ t*
313b.	VENCIMENTO	*v ẽ s i m ẽ t u*
314.	VIGOR	*v i g o r*
315.	VINCULADAS	*v ĩ k u l á d a z*
316a.	VINTE	*v ĩ T*
316b.	VINTE	*v ĩ T i*
317.	VISA	*v í z a*
318.	VOCÊ	*v o s e*
319.	VÔO	*v o*

Apêndice C

Lista com as Sentenças Reconhecidas

Este Apêndice apresenta uma lista com as 100 sentenças utilizadas para a avaliação do sistema híbrido ANN-HMM implementado. Também são apresentadas as sentenças reconhecidas pelo sistema para os casos : (1) Pré-REMAP (sem reestimação), (2) REMAP-1, (3) REMAP-2 e (4) REMAP-3, estes quatro casos sem o uso de restrições gramaticais e (5) utilizando uma Gramática *P-Gram.* Em todos estes cinco casos foram utilizados o modelo de duração de palavras discutido na Seção 5.2.3.

---

1 Original A cotação do dólar aumentou, e as bolsas fecharam em baixa
Pré-REMAP *para* cotação ** ** aumentou *mas* bolsas *antecipar* em baixa *tempo*
REMAP-1 A cotação ** dólar aumentou *mas* bolsas fecharam em baixa
REMAP-2 A cotação ** dólar aumentou, e as bolsas fecharam em baixa
REMAP-3 A cotação ** dólar aumentou, e as bolsas fecharam em baixa
P-Gram. A cotação do dólar aumentou, e as bolsas fecharam em baixa

2 Original A cotação do dólar aumentou, mas as bolsas fecharam em baixa
Pré-REMAP *para* cotação ** dólar aumentou mas as bolsas fecharam *o* em baixa *cento*
REMAP-1 A cotação ** dólar aumentou *mais* as bolsas fecharam *uma* baixa
REMAP-2 A cotação do dólar aumentou, mas as bolsas fecharam *o* em baixa
REMAP-3 A cotação ** dólar aumentou *mais* as bolsas fecharam *o* em baixa
P-Gram. A cotação do dólar aumentou, mas as bolsas fecharam em baixa

3 Original A bolsa ficará estável ou sofrerá uma pequena queda
Pré-REMAP *passa* bolsa ficará *e* estável ** sofrerá uma pequena queda *por*
REMAP-1 A bolsa ficará *é* estável ** sofrerá uma pequena queda
REMAP-2 A bolsa ficará *estarão* sofrerá uma pequena queda
REMAP-3 A bolsa ficará *está Rio* sofrerá uma pequena queda
P-Gram A bolsa ficará estável ou sofrerá uma pequena queda

4 Original Não haverá ajustes, nem modificações radicais no plano
Pré-REMAP *plano* haverá ajustes nem modificações *a dia mais* no plano tempo
REMAP-1 Não haverá *as* ajustes nem modificações *a de queda* no plano
REMAP-2 Não haverá *as* ajustes nem modificações *a de queda* no plano
REMAP-2 Não haverá ajustes nem modificações *a de queda* no plano
P-Gram Não haverá ajustes, nem modificações radicais no plano

5 Original Foi detectado um problema em seu cartão; ele deve ser substituído
Pré-REMAP *para* foi detectado ** problema *São* cartão
REMAP-1 Foi detectado ** problema *vinte São* cartão *para , ele de é ele ser substituído*
REMAP-2 Foi detectado ** problema *vinte são* cartão, ele de é ele ser substituído
REMAP-3 Foi detectado ** problema *vinte* seu cartão, ele ele substituído
P-Gram Foi detectado um problema em seu cartão, ele deve ser substituído

6 Original É necessário que o convênio permita o intercâmbio
Pré-REMAP *até* necessário ** ** convênio permita o intercâmbio *país*
REMAP-1 É necessário *do* convênio permita *ao* intercâmbio
REMAP-2 É necessário *do* convênio permita o intercâmbio
REMAP-3 É necessário ** ** convênio permita o intercâmbio
P-Gram É necessário que o convênio permita o intercâmbio

7 Original Posso afirmar-lhes que o convênio permite o intercâmbio
Pré-REMAP *passa* afirmar *início* convênio permite ** intercâmbio *tempo*
REMAP-1 Posso afirmar *início* convênio permite ** intercâmbio
REMAP-2 Posso afirmar *ele se* o convênio permite ** intercâmbio
REMAP-3 Posso afirmar *ele se* o convênio permite ** intercâmbio
P-Gram Posso afirmar-lhes que o convênio permite o intercâmbio

8 Original O convênio que foi assinado recentemente permite o intercâmbio
Pré-REMAP *posso* convênio ** foi *e* assinado recentemente permite ** intercâmbio *país*
REMAP-1 o convênio ** foi *é* assinado recentemente permite o intercâmbio
REMAP-2 O convênio que foi *é* assinado recentemente permite o intercâmbio
REMAP-3 O convênio que foi *é* assinado recentemente permite o intercâmbio
P-Gram O convênio que foi assinado recentemente permite o intercâmbio

9 Original O convênio permite o intercâmbio porque visa a integração entre alunos de culturas diferentes
Pré-REMAP *tempo* convênio permite ** intercâmbio porque visa a integração entre *dados* de culturas diferentes *cento*
REMAP-1 O convênio permite o intercâmbio porque visa *é* integração *em terá* alunos de culturas diferentes
REMAP-2 O convênio permite o intercâmbio porque visa a *é* integração entre *é* alunos de culturas diferentes
REMAP-3 O convênio permite o intercâmbio porque visa a integração entre *é* alunos de culturas diferentes
P-Gram O convênio permite o intercâmbio porque visa a integração entre alunos de culturas diferentes

10 Original O convênio permite o intercâmbio quando se trata de universidades vinculadas ao projeto de integração
Pré-REMAP o convênio permite ** intercâmbio quando se trata de universidades vinculadas ao projeto de integração
REMAP-1 o convênio permite ** intercâmbio quando se trata de universidades vinculadas ao projeto de integração
REMAP-2 ** convênio permite ** intercâmbio quando se trata de universidades vinculadas ao projeto de integração
REMAP-3 ** convênio permite ** intercâmbio *com no* se trata de universidades vinculadas ao projeto de integração
P-Gram. O convênio permite o intercâmbio quando se trata de universidades vinculadas ao projeto de integração

11 Original O convênio, que foi assinado recentemente, permite o intercâmbio
Pré-REMAP *tempo* convênio que foi *e* assinado recentemente *permitido* intercâmbio *para*
REMAP-1 O convênio que foi *e* assinado *a* recentemente *permitido* intercâmbio
REMAP-2 O convênio que foi *e* assinado *a* recentemente *permitido* intercâmbio
REMAP-3 O convênio que foi *e* assinado *a* recentemente, permite *e* intercâmbio
P-Gram O convênio, que foi assinado recentemente, permite o intercâmbio

12 Original O convênio assinado na última reunião é mais interessante do que o anterior
Pré-REMAP *pelo* convênio assinado *no* última reunião ** mais interessante do que ** anterior *muito*
REMAP-1 O convênio assinado *no* última reunião ** mais interessante do *Rio* anterior *horas*
REMAP-2 O convênio assinado *não* última reunião ** mais interessante *do Rio* anterior *horas*
REMAP-3 O convênio assinado *não* última reunião ** mais interessante do *Rio* anterior
P-Gram O convênio assinado na última reunião é mais interessante do que o anterior

13 Original A medida que o tempo passa, mais nos convencemos da eficácia do convênio
Pré-REMAP *país dia pelo* tempo passa mais nos convencemos da eficácia do convênio *país*
REMAP-1 A *mil devido* tempo passa , mais nos convencemos da eficácia do convênio
REMAP-2 A *mil devido* tempo passa, mais nos convencemos da eficácia do convênio
REMAP-3 A *mil devido* tempo passa, mais nos convencemos da eficácia do convênio
P-Gram A medida que o tempo passa, mais nos convencemos da eficácia do convênio

14 Original Se o convênio permite o intercâmbio, devemos aproveitar a oportunidade para desenvolver o projeto de integração
Pré-REMAP *próximo* convênio permite ** intercâmbio devemos aproveitar *ao* oportunidade para desenvolver o projeto de integração
REMAP-1 Se o convênio permite ** intercâmbio devemos aproveitar *ao* oportunidade para *dezembro convênio* projeto de integraçã
REMAP-2 Se o convênio permite ** intercâmbio devemos aproveitar *ao* oportunidade para desenvolver o projeto de integração
REMAP-3 Se o convênio permite ** intercâmbio devemos aproveitar *ao* oportunidade para desenvolver o projeto de integração
P-Gram. Se o convênio permite o intercâmbio, devemos aproveitar a oportunidade para desenvolver o projeto de integração

15 Original Localizado a uma quadra do centro da cidade, o condomínio permite que você una trabalho e conforto
Pré-REMAP *o* localizado ** uma quadra do cento ** cidade *por* condomínio permite que você *uma* trabalho e conforto
REMAP-1 Localizado ** uma quadra do cento da cidade *por* condomínio permite que você *uma* trabalho e conforto
REMAP-2 Localizado ** uma quadra do *cento* da cidade *por* condomínio permite que você *um na* trabalho e conforto
REMAP-3 Localizado ** uma *a* quadra do centro da cidade *por* condomínio permite ** você *um na* trabalho e conforto
P-Gram Localizado a uma quadra do centro da cidade, o condomínio permite que você una trabalho e conforto

16 Original É suficiente
Pré-REMAP *até* suficiente *por*
REMAP-1 É suficiente
REMAP-2 É suficiente
REMAP-3 É suficiente
P-Gram É suficiente

17 Original Isto é suficiente
Pré-REMAP *país com* é suficiente *para*
REMAP-1 Isto *o* é suficiente
REMAP-2 Isto *o* é suficiente
REMAP-3 Isto *o* é suficiente
P-Gram Isto é suficiente

18 Original O saldo é suficiente
Pré-REMAP *tempo São projeto* suficiente *para*
REMAP-1 O saldo *o* é suficiente
REMAP-2 O saldo *o* é suficiente
REMAP-3 O saldo *o* é suficiente
P-Gram O saldo é suficiente

19 Original O saldo de sua conta é suficiente
Pré-REMAP *posso* saldo de sua conta *até* suficiente *tempo*
REMAP-1 O saldo de sua conta *até* suficiente
REMAP-2 O saldo de sua conta *até* suficiente *em*
REMAP-3 O saldo de sua conta *até* suficiente *em*
P-Gram O saldo de sua conta é suficiente

20 Original O saldo disponível é insuficiente
Pré-REMAP *pelo* saldo disponível *o* é insuficiente *cento*
REMAP-1 O saldo disponível *com* é insuficiente *em*
REMAP-2 O saldo disponível *com* é insuficiente *em*
REMAP-3 O saldo disponível *com* é insuficiente *em*
P-Gram O saldo disponível é insuficiente

21 Original O saldo disponível em sua conta é insuficiente
Pré-REMAP *cotação do* disponível ** sua conta *até* insuficiente *até*
REMAP-1 O saldo disponível em sua *o com parece* insuficiente
REMAP-2 O saldo disponível em sua *o com parece* insuficiente
REMAP-3 O saldo disponível em sua *o com parece* insuficiente
P-Gram O saldo disponível em sua conta é insuficiente

22 Original Isto parece insuficiente
Pré-REMAP *país* parece insuficiente *tempo*
REMAP-1 *-lhes* parece insuficiente
REMAP-2 Isto parece insuficiente
REMAP-3 Isto parece insuficiente
P-Gram Isto parece insuficiente

23 Original O saldo parece ser insuficiente
Pré-REMAP *passado* parece *e* ser insuficiente *tempo*
REMAP-1 *do* saldo parece *ele* insuficiente
REMAP-2 O saldo parece *ele* insuficiente
REMAP-3 O saldo parece *ele* insuficiente
P-Gram O saldo parece ser insuficiente

24 Original O saldo sempre está disponível
Pré-REMAP *tempo* saldo sempre está disponível *para*
REMAP-1 O saldo sempre está disponível
REMAP-2 O saldo sempre está disponível
REMAP-3 O saldo sempre está disponível
P-Gram O saldo sempre está disponível

25 Original O saldo está sempre disponível no início do mês
Pré-REMAP *tempo* saldo está sempre disponível ** início *mil em isto*
REMAP-1 O saldo está sempre *de se número mil* início *mil em -lhes*
REMAP-2 O saldo está sempre *de se número mil vencimento as*
REMAP-3 O saldo está sempre *se número mil vencimento as*
P-Gram O saldo está sempre disponível no início do mês

26 Original No início do mês, o saldo está disponível
Pré-REMAP *tempo* início *mais* o saldo está disponível *para*
REMAP-1 *de* início *nem os* o saldo está disponível
REMAP-2 *de* início do *melhores* o saldo está disponível
REMAP-3 *de* início *melhores* o saldo está disponível
P-Gram ** início do mês, o saldo está disponível

27 Original Esta é a última chamada para o vôo 737 da Rio-Sul
Pré-REMAP *até as terá últimas* chamada *por os o se esta por e e* sete da Rio *São tempo*
REMAP-1 Esta é ** última chamada *por os o* sete três sete da Rio Sul *um*
REMAP-2 Esta é ** última chamada *por os* o sete três sete da Rio Sul *um*
REMAP-3 Esta é ** última chamada *por os* o *os esta três* sete da Rio Sul *um*
P-Gram Esta é a última chamada para o vôo sete três sete da Rio-Sul

28 Original Isto é uma pesquisa de opinião pública
Pré-REMAP *país a* uma pesquisa de opinião pública *tempo*
REMAP-1 *estão com* pesquisa de opinião pública
REMAP-2 *estão* uma pesquisa de opinião pública
REMAP-3 *estão* uma pesquisa de opinião pública
P-Gram Isto é uma pesquisa de opinião pública

29 Original O valor de sua conta telefônica é baixo
Pré-REMAP *portão dois* de sua conta telefônica é baixa *tempo*
REMAP-1 O valor de sua conta telefônica é baixo
REMAP-2 O valor de sua conta telefônica é baixo
REMAP-3 O valor de sua conta telefônica é baixo
P-Gram O valor de sua conta telefônica é baixo

30 Original É de trinta mil cruzeiros, o valor de sua conta telefônica
Pré-REMAP *parece* de trinta mil cruzeiros o valor de sua conta telefônica *empresa*
REMAP-1 É de trinta mil cruzeiros o valor de sua conta telefônica
REMAP-2 É de trinta mil cruzeiros, o valor de sua conta telefônica
REMAP-3 É de trinta mil cruzeiros o valor de sua conta telefônica
P-Gram É de trinta mil cruzeiros, o valor de sua conta telefônica

31 Original O vencimento de sua prestação será no dia quatro de junho
Pré-REMAP *todos* vencimento de sua prestação será *última* quatro de *número país*
REMAP-1 O vencimento de *Sul o* prestação será no dia quatro *visa nome*
REMAP-2 O vencimento de *Sul o* prestação será no dia quatro *visa nome*
REMAP-3 O vencimento de *Sul o* prestação será no dia quatro *visa uma um*
P-Gram O vencimento de sua prestação será no dia quatro de *consumo*

32 Original O preço aumentou
Pré-REMAP *todos* preço *o* aumentou *para*
REMAP-1 O preço *o* aumentou *para*
REMAP-2 O preço *o* aumentou *para*
REMAP-3 O preço *o* aumentou
P-Gram O preço aumentou

33 Original O preço do café aumentou
Pré-REMAP ** preço do café aumentou *para*
REMAP-1 *do* preço do café aumento *o*
REMAP-2 O preço do café aumentou
REMAP-3 O preço do café aumentou
P-Gram O preço do café aumentou

34 Original O preço do café expresso aumentou
Pré-REMAP ** preço ** café expresso aumentou *para*
REMAP-1 O preço *do* café expresso *o* aumentou
REMAP-2 O preço ** café expresso *o* aumentou
REMAP-3 O preço ** café expresso *o* aumentou
P-Gram O preço do café expresso aumentou

35 Original O preço do café aumentou consideravelmente
Pré-REMAP *plano* preço ** café ** aumentou consideravelmente *tempo*
REMAP-1 O preço do café ** aumentou consideravelmente
REMAP-2 O preço do café ** aumentou consideravelmente
REMAP-3 O preço do café ** aumentou consideravelmente
P-Gram O preço do café aumentou consideravelmente

36 Original O preço do café aumentou consideravelmente na semana passada
Pré-REMAP *porque São* café aumentou consideravelmente na semana passada *tempo*
REMAP-1 *do* preço do café aumentou consideravelmente na semana passada
REMAP-2 *do* preço do café aumentou consideravelmente na semana passada
REMAP-3 O preço do café aumentou consideravelmente na semana passada
P-Gram O preço do café aumentou consideravelmente na semana passada

37 Original Aumentou o preço do café
Pré-REMAP *para* aumentou ** preço do café *cento*
REMAP-1 Aumentou ** preço do café
REMAP-2 Aumentou ** preço do café
REMAP-3 Aumentou ** preço do café
P-Gram Aumentou o preço do café

38 Original As taxas de juros no mercado interno estão subindo bastante
Pré-REMAP *país* taxas *início* mercado interno estão subindo bastante *tempo*
REMAP-1 As taxas de juros** mercado interno estão subindo bastante
REMAP-2 As taxas de *Sul* no mercado interno estão subindo bastante *em vinte*
REMAP-3 As taxas de juros no mercado interno estão subindo bastante *em vinte*
P-Gram As taxas de juros no mercado interno estão subindo bastante

39 Original As contas chegaram atrasadas
Pré-REMAP *passa* contas chegaram atrasadas *imposto*
REMAP-1 As contas chegaram atrasadas
REMAP-2 As contas chegaram atrasadas
REMAP-3 As contas chegaram atrasadas
P-Gram As contas chegaram atrasadas

40 Original As contas chegaram atrasadas ontem
Pré-REMAP *passa* contas chegaram *muito* atrasadas *o TELESP*
REMAP-1 As contas chegaram *muito* atrasadas *com o TELESP*
REMAP-2 As contas chegaram *muito* atrasadas *com o TELESP*
REMAP-3 As contas chegaram *muito* atrasadas *com o TELESP*
P-Gram As contas chegaram atrasadas ontem

41 Original As contas telefônicas deste mês chegaram muito atrasadas ao banco
Pré-REMAP *país* contas telefônicas deste *mil aceitarão* muito atrasadas ao banco *tempo*
REMAP-1 As contas telefônicas *recentemente* chegaram muito atrasadas ao banco
REMAP-2 As contas telefônicas *recentemente* chegaram muito atrasadas ao banco
REMAP-3 As contas telefônicas *recentemente* chegaram muito atrasadas ao banco
P-Gram As contas telefônicas deste mês chegaram muito atrasadas ao banco

42 Original Ontem, as contas chegaram aqui muito atrasadas
Pré-REMAP *Paulo ele* as contas chegaram *bastante* muito atrasadas *para*
REMAP-1 *posso ele* as contas chegaram *bastante* muito *a* atrasadas
REMAP-2 *posso ele* as contas chegaram *partir* muito *a* atrasadas
REMAP-3 *Com o ele* as contas chegaram *partir* muito *a* atrasadas
P-Gram Ontem, as contas chegaram aqui muito atrasadas

43 Original Chegaram atrasadas
Pré-REMAP Chegaram atrasadas *trinta*
REMAP-1 Chegaram atrasadas
REMAP-2 Chegaram atrasadas
REMAP-3 Chegaram atrasadas
P-Gram Chegaram atrasadas

44 Original Chegaram atrasadas todas as contas telefônicas deste mês
Pré-REMAP *até* chegaram atrasadas todas as contas telefônicas deste *nem cento*
REMAP-1 Chegaram *com* atrasadas todas as contas telefônicas deste *nem -lhes*
REMAP-2 Chegaram *com* atrasadas todas as contas telefônicas deste mês
REMAP-3 Chegaram *com* atrasadas todas as contas telefônicas deste mês
P-Gram Chegaram atrasadas todas as contas telefônicas deste mês

45 Original O governo aumentou o imposto no mês passado
Pré-REMAP *tempo* governo aumentou ** imposto ** mês passado *conta*
REMAP-1 *ao* governo aumentou ** imposto no mês passado
REMAP-2 *ao* governo aumentou ** imposto no mês passado
REMAP-3 *ao* governo aumentou ** imposto no mês passado
P-Gram O governo aumentou o imposto no mês passado

46 Original O governo aumentou o imposto sobre importação
Pré-REMAP *tempo* governo aumentou imposto sobre importação *tempo*
REMAP-1 O governo aumentou ** imposto sobre importação
REMAP-2 O governo aumentou ** imposto sobre importação
REMAP-3 O governo aumentou ** imposto sobre importação
P-Gram O governo aumentou o imposto sobre importação

47 Original O governo entregou os formulários aos contribuintes
Pré-REMAP *Paulo* governo entregou ** formulários aos contribuintes *para*
REMAP-1 O governo entregou ** formulários aos contribuintes
REMAP-2 O governo entregou ** formulários aos contribuintes
REMAP-3 O governo entregou ** formulários aos contribuintes
P-Gram O governo entregou os formulários aos contribuintes

48 Original O governo entregou aos contribuintes os formulários
Pré-REMAP *Paulo* governo entregou aos contribuintes os formulários *para*
REMAP-1 O governo entregou *horas* contribuintes *nos* formulários
REMAP-2 O governo entregou *horas* contribuintes *nos* formulários
REMAP-3 O governo entregou aos contribuintes *nos* formulários
P-Gram O governo entregou aos contribuintes os formulários

49 Original O banco colocará a sua disposição o novo cheque
Pré-REMAP *tempo* banco colocará ** sua disposição ** novo *sete para*
REMAP-1 *do* banco colocará ** sua disposição ** novo cheque
REMAP-2 *do* banco colocará ** sua disposição ** novo cheque
REMAP-3 O banco colocará ** sua disposição ** novo cheque
P-Gram O banco colocará a sua disposição o novo cheque

50 Original A conta telefônica em nome de Adelaide Barroso terá vencimento amanhã
Pré-REMAP *terá* conta telefônica *e* nome ** Adelaide Barroso *da* , terá vencimento *aumento*
REMAP-1 A conta telefônica *e* nome *dia* Adelaide Barroso *da* , terá vencimento *aumento da*
REMAP-2 A conta telefônica *e* nome *dia* Adelaide Barroso *da,* terá vencimento *ao não em a da*
REMAP-3 A conta telefônica *e* nome *dia* Adelaide Barroso *da,* terá vencimento *a uma sim em a*
P-Gram A conta telefônica em nome de Adelaide Barroso terá vencimento amanhã

51 Original A perda da atratividade das aplicações em caderneta de poupança está provocando um aumento de consumo no país
Pré-REMAP A *prezado* atratividade das aplicações em caderneta de poupança *estarão quando* aumento de consumo ** país
REMAP-1 A perda da atratividade das aplicações em caderneta de poupança está provocando *nome entre* de consumo ** país
REMAP-2 A perda da atratividade das aplicações em caderneta de poupança *de* está provocando *nome entre* de consumo ** país
REMAP-3 A perda da atratividade das aplicações em caderneta de poupança está provocando *o* aumento de consumo ** país
P-Gram. A perda da atratividade das aplicações em caderneta de poupança está provocando um aumento de consumo no país

52 Original O mercado foi considerado inadequado
Pré-REMAP *tempo* mercado foi considerado inadequado *tempo*
REMAP-1 *do* mercado foi considerado inadequado
REMAP-2 *do* mercado foi considerado inadequado
REMAP-3 O mercado foi considerado inadequado
P-Gram O mercado foi considerado inadequado

53 Original O mercado foi considerado inadequado pelos analistas
Pré-REMAP *tempo* mercado foi considerado inadequado pelos analistas *tempo*
REMAP-1 O mercado foi considerado inadequado pelos analistas
REMAP-2 O mercado foi considerado inadequado pelos analistas
REMAP-3 O mercado foi considerado inadequado pelos analistas
P-Gram O mercado foi considerado inadequado pelos analistas

54 Original O mercado foi considerado inadequado naquele momento
Pré-REMAP *tempo* mercado foi considerado inadequado naquele momento *trinta*
REMAP-1 *uma é a do* foi considerado inadequado naquele momento
REMAP-2 *uma é a do* foi considerado inadequado naquele momento
REMAP-3 *uma é a do* foi considerado inadequado naquele momento
P-Gram O mercado foi considerado inadequado naquele momento

55 Original Naquele momento, o mercado foi considerado inadequado pelos analistas das melhores instituições de pesquisa
Pré-REMAP Naquele momento *por* mercado foi considerado inadequado pelos analistas das melhores instituições *e* pesquisa
REMAP-1 Naquele momento *por* mercado foi considerado inadequado pelos analistas das *ele* melhores instituições de pesquisa
REMAP-2 Naquele momeno *por* mercado foi considerado inadequado pelos analistas das melhores instituições de pesquisa
REMAP-3 Naquele momeno *por* mercado foi considerado inadequado pelos analistas das melhores instituições de pesquisa
P-Gram. Naquele momento, ** mercado foi considerado inadequado pelos analistas das melhores instituições de pesquisa

56 Original Diariamente
Pré-REMAP *perda* diariamente *para*
REMAP-1 *dia é cliente*
REMAP-2 *dia é cliente*
REMAP-3 Diariamente
P-Gram Diariamente

57 Original Curva perigosa
Pré-REMAP Curva *preço das ao*
REMAP-1 Curva perigosa
REMAP-2 Curva perigosa
REMAP-3 Curva perigosa
P-Gram Curva perigosa

58 Original Dia vinte do sete
Pré-REMAP *pública* vinte do sete *para*
REMAP-1 Dia vinte do sete *para*
REMAP-2 Dia vinte do sete
REMAP-3 Dia vinte do sete
P-Gram Dia vinte do sete

59 Original Sim
Pré-REMAP sim
REMAP-1 sim
REMAP-2 Sim
REMAP-3 Sim
P-Gram Sim

60 Original Não
Pré-REMAP *por* não *tempo*
REMAP-1 Não
REMAP-2 Não
REMAP-3 Não
P-Gram Não

61 Original Saldo: vinte e cinco reais
Pré-REMAP *passado país* vinte ** cinco reais *tempo*
REMAP-1 Saldo , *de* vinte ** cinco *regras*
REMAP-2 Saldo, *de* vinte ** cinco *regras*
REMAP-3 Saldo, *de* vinte ** cinco real -lhes
P-Gram Saldo: vinte e cinco reais

62 Original Estação Santa Cruz
Pré-REMAP *prestação* Santa Cruz *para*
REMAP-1 Estação Santa Cruz
REMAP-2 Estação Santa Cruz
REMAP-3 Estação Santa Cruz
P-Gram Estação Santa Cruz

63 Original Passageiros com destino a São Paulo, Recife e Fortaleza, embarque imediato
Pré-REMAP Passageiros com deste *na* São Paulo Recife ** Fortaleza *ainda aqui* imediato *com é*
REMAP-1 *as fazer as* com deste *na* São *posso* Recife ** Fortaleza , *vinte Barroso* imediato *com*
REMAP-2 *as fazer as* com *deste na* São *posso* Recife ** Fortaleza, *sim banco aqui* imediato
REMAP-3 *as fazer as* com *deste na* São *posso* Recife ** Fortaleza, embarque imediato
P-Gram Passageiros com destino a São Paulo, Recife e Fortaleza, embarque imediato

64 Original Os bancos atrás de mais eficiência
Pré-REMAP *pelos* bancos atrás *e* mais eficiência *tempo*
REMAP-1 *O* bancos atrás ** mais eficiência
REMAP-2 Os bancos atrás ** mais eficiência
REMAP-3 Os bancos atrás ** mais eficiência
P-Gram ** bancos atrás de mais eficiência

65 Original Descontos de até 50%
Pré-REMAP *tempo* descontos ** até cinqüenta porcento *para*
REMAP-1 Descontos *e* até cinqüenta porcento
REMAP-2 Descontos ** até cinqüenta porcento
REMAP-3 Descontos *a* até cinqüenta porcento
P-Gram Descontos de até cinqüenta porcento

66 Original Número incompleto
Pré-REMAP *por* número incompleto *conta*
REMAP-1 *da* número incompleto
REMAP-2 Número incompleto
REMAP-3 Número incompleto
P-Gram Número incompleto

67 Original Vinte e cinco
Pré-REMAP *o* vinte ** cinco
REMAP-1 *o* vinte ** cinco
REMAP-2 Vinte ** cinco
REMAP-3 Vinte ** cinco
P-Gram Vinte e cinco

68 Original Vinte cinco reais
Pré-REMAP *tempo* vinte ** cinco *real TELESP*
REMAP-1 *de* vinte cinco *regras*
REMAP-2 *de* vinte cinco *regras*
REMAP-3 Vinte cinco *regras*
P-Gram Vinte cinco reais

69 Original Cento e vinte cinco
Pré-REMAP *até* cento ** vinte *cinquenta*
REMAP-1 Cento *ele* vinte ** cinco
REMAP-2 Cento *ele* vinte ** cinco
REMAP-3 Cento *ele* vinte ** cinco
P-Gram Cento e vinte cinco

70 Original Cento e vinte e cinco reais
Pré-REMAP *posso e entre* vinte ** cinco *da TELESP*
REMAP-1 *setenta e* vinte ** cinco *regras*
REMAP-2 Cento de vinte ** cinco *regras*
REMAP-3 Cento e vinte ** cinco *regras*
P-Gram Cento e vinte e cinco reais

71 Original Quatrocentos e quarenta e nove
Pré-REMAP *para* quatrocentos e quarenta *estável tempo*
REMAP-1 Quatrocentos e quarenta e nove
REMAP-2 Quatrocentos e quarenta e nove
REMAP-3 Quatrocentos e quarenta e nove
P-Gram Quatrocentos e quarenta e nove

72 Original Dois mil, cento e vinte e cinco
Pré-REMAP *todos Rio centro* vinte ** cinco *tempo*
REMAP-1 Dois mil cento e vinte ** cinco
REMAP-2 Dois mil cento e vinte ** cinco
REMAP- Dois mil, cento e vinte ** cinco
P-Gram Dois mil, cento e vinte e cinco

73 Original Dezesseis mil e quinhentos
Pré-REMAP *até* dezesseis mil e quinhentos *país*
REMAP-1 Dezesseis mil e quinhentos
REMAP-2 Dezesseis mil e quinhentos
REMAP-3 Dezesseis mil e quinhentos
P-Gram Dezesseis mil e quinhentos

74 Original Oitenta milhões, trezentos e sessenta mil e duzentos e setenta e um
Pré-REMAP *aproveitar* milhões trezentos ** sessenta mil e duzentos ** setenta e *uma para*
REMAP-1 Oitenta milhões trezentos ** sessenta mil *mil do cento os* setenta e *uma*
REMAP-2 Oitenta milhões trezentos ** sessenta mil *mil do cento os* setenta e *uma*
REMAP-3 *com setenta* milhões trezentos ** sessenta mil *mil* duzentos ** setenta e *uma*
P-Gram Oitenta milhões, trezentos e sessenta mil e duzentos e setenta e um

75 Original A TELEBRÁS, a empresa de telecomunicações brasileira, está investindo em pesquisa
Pré-REMAP *mas* TELEBRÁS empresa *os* telecomunicações brasileira , está investindo *e* pesquisa *até*
REMAP-1 *até* TELEBRÁS *está* empresa *das* telecomunicações brasileira , está investindo *e* pesquisa
REMAP-2 *até* TELEBRÁS *está* empresa *das* telecomunicações brasileira, está investindo *de* pesquisa
REMAP-3 *até* TELEBRÁS, *assim* empresa *das* telecomunicações brasileira, está investindo *de* pesquisa
P-Gram A TELEBRÁS, a empresa de telecomunicações brasileira, está investindo em pesquisa

76 Original A TELEBRÁS, uma empresa estatal, está investindo em pesquisa
Pré-REMAP *para* TELEBRÁS *com* uma empresa estatal , está investindo ** pesquisa *tempo*
REMAP-1 *até* TELEBRÁS , uma empresa estatal , está investindo em pesquisa
REMAP-2 *até* TELEBRÁS, uma empresa estatal, está investindo em pesquisa
REMAP-3 *até* TELEBRÁS, uma empresa estatal, está investindo em pesquisa
P-Gram A TELEBRÁS, uma empresa estatal, está investindo em pesquisa

77 Original Tivemos recentemente a seguinte notícia : a TELEBRÁS passará a investir mais em pesquisa
Pré-REMAP *até* tivemos recentemente ** seguinte notícia *país* TELEBRÁS passará ** investir mais em pesquisa
REMAP-1 Tivemos recentemente ** seguinte notícia , a TELEBRÁS passará ** investir mais em pesquisa
REMAP-2 Tivemos recentemente ** seguinte notícia, a TELEBRÁS passará a investir mais *investir visa*
REMAP-3 Tivemos recentemente ** seguinte notícia, a TELEBRÁS passará ** investir mais *vinte* pesquisa
P-Gram Tivemos recentemente a seguinte notícia, a TELEBRÁS passará a investir mais em pesquisa

78 Original TELESP informa : dezenove horas e trinta minutos
Pré-REMAP TELESP informa , dezenove horas e trinta *mil outros sempre*
REMAP-1 TELESP informa , dezenove horas e trinta minutos
REMAP-2 TELESP informa, dezenove horas e trinta minutos
REMAP-3 TELESP informa, dezenove horas e trinta minutos
P-Gram TELESP informa : dezenove horas e trinta minutos

79 Original Empresário, é preciso antecipar o futuro
Pré-REMAP *para* empresário ** preciso antecipar *os* futuro *pelo*
REMAP-1 *é* empresário , é preciso antecipar o futuro
REMAP-2 *é* empresário, é preciso antecipar o futuro
REMAP-3 *é* Empresário, é preciso antecipar o futuro
P-Gram Empresário, é preciso antecipar o futuro

80 Original Prezado cliente, aguardaremos o seu comparecimento
Pré-REMAP Prezado cliente *para* aguardaremos o seu comparecimento *para*
REMAP-1 Prezado cliente , aguardaremos o seu comparecimento
REMAP-2 Prezado cliente, aguardaremos o seu comparecimento
REMAP-3 Prezado cliente, aguardaremos o seu comparecimento
P-Gram Prezado cliente, aguardaremos o seu comparecimento

81 Original O código foi registrado pelo funcionário
Pré-REMAP *Paulo* código foi registrado *de os impulsionado tempo*
REMAP-1 *no* código foi registrado pelo funcionário
REMAP-2 *no* código foi registrado pelo funcionário
REMAP-3 O código foi registrado pelo funcionário
P-Gram O código foi registrado pelo funcionário

82 Original O convênio, um documento de trinta páginas, tem permitido o intercâmbio
Pré-REMAP *pelo* convênio ** documento de trinta páginas ** permitido ** intercâmbio *com tempo*
REMAP-1 O convênio ** documento de trinta páginas , tem *na* permitido ** intercâmbio *com*
REMAP-2 O convênio ** documento de trinta páginas, *trinta* permitido ** intercâmbio *com*
REMAP-3 O convênio ** documento de trinta páginas, *trinta* permitido ** intercâmbio
P-Gram O convênio, um documento de trinta páginas, tem permitido o intercâmbio

83 Original Os caixas eletrônicos não aceitarão mais depósitos
Pré-REMAP *poupança caixas* eletrônicos não aceitarão ** depósitos *para*
REMAP-1 Os *a* caixas eletrônicos não aceitarão ** depósitos
REMAP-2 Os caixas eletrônicos não aceitarão ** depósitos
REMAP-3 Os caixas eletrônicos não aceitarão ** depósitos
P-Gram Os caixas eletrônicos não aceitarão mais depósitos

84 Original Os caixas eletrônicos não aceitarão mais depósitos a partir das 15 horas
Pré-REMAP *bolsas os* eletrônicos não aceitarão mais depósitos a partir das quinze *saques*
REMAP-1 *bolsa* caixas eletrônicos não aceitarão mais depósitos a partir *da sete vinte* horas
REMAP-2 *bolsa* caixas eletrônicos não aceitarão mais depósitos a partir das *sete vinte* horas
REMAP-3 Os caixas eletrônicos não aceitarão mais depósitos a partir *da sete vinte* horas
P-Gram Os caixas eletrônicos não aceitarão mais depósitos a partir das quinze horas

85 Original Os caixas eletrônicos não aceitarão mais depósitos a partir das 15 horas do próximo dia vinte e nove
Pré-REMAP Os *esta os* eletrônicos não aceitarão mais depósitos a partir das quinze horas , do próximo dia vinte ** nove
REMAP-1 Os *café se os* eletrônicos não aceitarão mais depósitos a partir das quinze horas , do próximo dia vinte ** nove
REMAP-2 Os caixas eletrônicos não aceitarão mais depósitos a partir das quinze horas, do próximo dia vinte ** nove
REMAP-3 Os caixas eletrônicos não aceitarão mais depósitos a partir das quinze horas, do próximo dia vinte ** nove
P-Gram Os caixas eletrônicos não aceitarão mais depósitos a partir das quinze horas do próximo dia vinte e nove

86 Original E não estarão disponíveis para saques a partir das 17 horas do dia trinta
Pré-REMAP *opinião* estarão disponíveis para *a* saques , a partir das dezessete horas do dia trinta *esta*
REMAP-1 *de* não estarão disponíveis passa saques , a partir das dezessete horas do dia trinta
REMAP-2 *de* não estarão disponíveis para saques, a partir das dezessete horas do dia trinta
REMAP-3 E não estarão disponíveis para saques, a partir das dezessete horas do dia trinta
P-Gram E não estarão disponíveis para saques a partir das dezessete horas do dia trinta

87 Original Todos os bancos devem fazer a atualização cadastral até o dia 31 de dezembro
Pré-REMAP Todos os *o* bancos devem fazer ** atualização cadastral até ** ** trinta *Rio* de dezembro
REMAP-1 Todos os o bancos devem fazer ** atualização cadastral até ** ** trinta *é* e um de dezembro
REMAP-2 Todos os bancos devem fazer ** atualização cadastral até ** ** trinta *aqui* um de dezembro
REMAP-3 Todos os bancos devem fazer ** atualização cadastral até ** ** trinta *aqui* um de dezembro
P-Gram Todos os bancos devem fazer a atualização cadastral até o dia trinta e um de dezembro

88 Original Todos os bancos devem fazer a atualização cadastral de seus clientes até 31 de dezembro
Pré-REMAP Todos os bancos devem fazer ** atualização cadastral de seus clientes até trinta *mil mil* dezembro *com*
REMAP-1 Todos os bancos devem fazer ** atualização cadastral de seus clientes até trinta *é* e um de dezembro
REMAP-2 Todos os bancos devem fazer ** atualização cadastral de seus clientes até trinta *é* e um de dezembro
REMAP-3 Todos os bancos devem fazer ** atualização cadastral de seus clientes até trinta *é* e um de dezembro
P-Gram Todos os bancos devem fazer a atualização cadastral de seus clientes até trinta e um de dezembro

89 Original De acordo com a determinação do Banco Central
Pré-REMAP De acordo *quadra* determinação ** banco central *conta*
REMAP-1 De acordo com a determinação ** banco central
REMAP-2 De acordo com a determinação ** Banco Central
REMAP-3 De acordo com a determinação ** Banco Central
P-Gram De acordo com a determinação do Banco Central

90 Original Em cumprimento a Resolução 2025 do Conselho Monetário Nacional
Pré-REMAP Em cumprimento *a* resolução dois mil ** vinte ** cinco , *o* conselho monetário nacional *com*
REMAP-1 Em cumprimento *a* resolução dois mil ** vinte ** cinco *para* , do *com ser* monetário nacional
REMAP-2 Em cumprimento *nove nas o São* dois mil ** vinte ** cinco, do *com ser* Monetário Nacional
REMAP-3 Em cumprimento *nove* resolução dois mil ** vinte ** cinco, do *com ser* Monetário Nacional
P-Gram Em cumprimento a Resolução 2025 do Conselho Monetário Nacional

91 Original De acordo com a determinação explícita do Banco Central
Pré-REMAP De acordo com ** determinação explícita do banco central *para*
REMAP-1 *a* , de acordo com ** determinação explícita do banco central
REMAP-2 *a,* de acordo com ** determinação explícita do Banco Central
REMAP-3 De acordo com a determinação explícita do Banco Central
P-Gram De acordo com a determinação explícita do Banco Central

92 Original As operações continuam
Pré-REMAP *atrás* operações continuam *país*
REMAP-1 As operações continuam
REMAP-2 As operações continuam
REMAP-3 As operações continuam
P-Gram As operações continuam

93 Original As operações de crédito continuam
Pré-REMAP *atrás* operações de crédito continuam *tempo*
REMAP-1 As operações de crédito continuam
REMAP-2 As operações de crédito continuam
REMAP-3 As operações de crédito continuam
P-Gram As operações de crédito continuam

94 Original As operações de crédito e financiamento continuam a seguir as regras
Pré-REMAP *para* as operações de crédito ** financiamento continuam ** seguir as regras *tempo*
REMAP-1 As operações de crédito e financiamento continuam ** seguir as regras
REMAP-2 As operações de crédito e financiamento continuam ** seguir as regras
REMAP-3 As operações de crédito e financiamento continuam ** seguir as regras
P-Gram As operações de crédito e financiamento continuam a seguir as regras

95 Original As operações de crédito e financiamento continuam a seguir as regras do banco central
Pré-REMAP *mas* operações de crédito ** financiamento continuam ** seguir as regras do banco central *com*
REMAP-1 As operações *dia* crédito *mil* financiamento continuam ** seguir *nas* regras do banco central
REMAP-2 As operações de crédito *mil* financiamento continuam ** seguir as regras do banco central
REMAP-3 As operações de crédito *mil* financiamento continuam ** seguir *nas* regras do banco central
P-Gram As operações de crédito e financiamento continuam a seguir as regras do banco central

96 Original As operações de crédito e financiamento corrigidas por outros indexadores
Pré-REMAP *mas* operações de crédito ** financiamento corrigidas por outros indexadores *que*
REMAP-1 As operações de crédito e financiamento corrigidas por outros indexadores *que*
REMAP-2 As operações de crédito e financiamento corrigidas por outros indexadores *com*
REMAP-3 As operações de crédito e financiamento corrigidas por outros indexadores
P-Gram As operações de crédito e financiamento corrigidas por outros indexadores

97 Original Continuam normalmente a seguir as regras do banco central ainda em vigor
Pré-REMAP Continuam ** ** seguir as regras do banco central , ainda ** vigor
REMAP-1 Continuam ** ** seguir as regras do banco central , ainda *e e* vigor
REMAP-2 Continuam ** ** seguir as regras do banco central, ainda *e e* vigor
REMAP-3 Continuam ** ** seguir as regras do banco central, ainda *e e* vigor
P-Gram Continuam ** a seguir as regras do banco central ainda em vigor

98 Original O aumento no consumo devido ao Plano Real impulsionou os preços
Pré-REMAP *momento* consumo devido ** plano real impulsionou os preços *o*
REMAP-1 O aumento no consumo devido ao plano real impulsionou os preços
REMAP-2 O aumento ** consumo devido ao Plano Real impulsionou os preços
REMAP-3 O aumento ** consumo devido ao Plano Real impulsionou os preços
P-Gram O aumento no consumo devido ao Plano Real impulsionou os preços

99 Original Segundo dados do IBGE, o aumento no consumo devido ao plano real impulsionou os preços
Pré-REMAP *para seu* dados do IBGE , ** *documento* consumo devido ** plano real impulsionou os preços
REMAP-1 Segundo dados do IBGE , documento no consumo devido plano real impulsionou os preço cheque
REMAP-2 Segundo dados do IBGE, *o documento* no consumo devido ** plano real impulsionou os preços *cheque*
REMAP-3 Segundo dados do IBGE, o *documento* no consumo devido ** plano real impulsionou os preços *cheque*
P-Gram Segundo dados do IBGE, o aumento no consumo devido ao plano real impulsionou os preços

100 Original O aumento no consumo devido ao Plano Real tem impulsionado consideravelmente os preços nas últimas semanas
Pré-REMAP
REMAP-1 * aumento no consumo devido * Plano Real *para em* impulsionado consideravelmente os preço *bancos última* semanas.
REMAP-2 * aumento no consumo devido * Plano Real *para em* impulsionado consideravelmente os *preço não os última* semanas.
REMAP-3 * aumento no consumo devido * Plano Real *por em* impulsionado consideravelmente os *preço não os última* semanas.
P-Gram. * aumento no consumo devido ao Plano Real tem impulsionado consideravelmente os preços nas últimas semanas

Bibliografia

[1] BENGIO, Y., MORI, R. D., FLAMMIA, G. AND KOMPE, R., "Global Optimization of a Neural Network-Hidden Markov Model Hybrid", IEEE Trans. on Neural Networks., vol. 3, pp. 252-259, **1992**.

[2] BISHOP, C, M., *Neural Networks for Patterns Recognition*, Oxford University Press, **1996**.

[3] BOURLARD, H., HERMANSKY, H. AND MORGAN, N., "Towards increasing speech recognition error rates", Speech Communication., vol. 12, pp. 205-231, **1996**.

[4] BOURLARD, H. AND WELLEKENS, C. J., "Links Between Markov Models and Multilayer Perceptrons", IEEE Trans. Patt. Anal. and Mach. Intell., vol. 12, pp. 1167-1178, **1990**.

[5] DAVIS, S. B. AND MELMELSTEIN, P., "Comparision of Parametric Representation for Monosyllabic Word Recognition in Continuously Spoken Sentences", IEEE Transactions on Acoustic, Speech and Signal Processing, ASSP - 28, No. 4, August **1980**.

[6] DELLER, J., PROAKIS, J. AND HANSEN, J., *Discrete-Time Processing of Speech Signals*, Macmillan, New York, **1993**.

[7] DEVILLERS, L. AND DUGAST, C., "Hybrid System Combining Expert-TDNN and HMMs for Continuous Speech Recognition" , IEEE ICASSP94, , pp 165-168, **1994**.

[8] DUGAST, C., DEVILLERS, L. AND AUBERT, X., "Combining TDNN and HMM in a Hybrid System for Improved Continuous-Speech Recognition" , IEEE Trans. Acoustic, Speech and Audio Processing, vol. 2, pp 217-223, jan. **1994**.

[9] FRITSCH, J., "Modular Neural Networks for Speech Recognition," Diploma Thesis, Iteractive Systems Laboratories, Carnegie Mellon University, **1996**.

[10] HAYKIN, S., *Neural Networks : A Comprehensive Foundation*, Prentice Hall, New Jersey, **1994**.

[11] HUANG, X. C., HON, H. W., HWANG H. W. AND LEE K. F., "A Comparative Study of Discrete, Semicontinuous, and Continuous Hidden Markov Models", Computer Speech and Language, vol(7), pp. 359-368, **1993**.

[12] HUANG, X. D., ARIKI, Y. AND JACK, M. A., "Hidden Markov Models for Speech Recognition", Edinburgh University Press, Edinburgh, **1990**.

[13] JELINEK, F., "A Fast Sequential Decoding Algorithm Using A Stack," IBM Journal, Res. Develop., Vol. 13, pp. 675-685, Nov. **1969**.

[14] KONIG, Y., "REMAP : Recursive Estimation and Maximization of A Posteriori Probabilities in Transition-Based Speech Recognition", PhD Thesis, University of California at Berkeley, **1996**.

[15] KONIG, Y., BOURLARD, H. AND MORGAN, N., "REMAP : Experiments with Speech Recognition", IEEE ICASSP96, Atlanta, pp. 7-10, **1996**.

[16] LEE, K. F., *Automatic Speech Recognition*, Kluwer Academic Publishers, 2ª edition, **1992**.

[17] LEE, C. H. AND RABINER, L. R., "A Frame-Synchronous Network Search Algorithm for Connected Word Recognition", IEEE Transaction on Acoustics, Speech, and Signal Processing, Vol. 37, No. 11, November **1989**.

[18] MORAIS E. S. E VIOLARO, F., "Sistema Híbrido ANN-HMM para Reconhecimento de Fala Contínua", Anais do XV SBT - Simpósio Brasilerio de Telecomunicações, Recife, pp. 117-120, **1997**.

[19] MORAIS E. S., VIOLARO, F., YNOGUTI, C. A. E NETTO, M. A., "Sistemas Híbridos ANN-HMM Baseados nos Critérios ML e MAP para Reconhecimento de Séries Temporais", Anais do 3º SBAi - Simpósio Brasilerio de Automação Inteligente, Vitória, pp. 406-411, **1997**.

[20] MORAIS E. S., VIOLARO, F. E NETTO, M. A., "Modelo Oculto de Markov Parametrizado através de uma ANN e sua Aplicação no Reconhecimento de Fala Contínua", Anais do III CBRN - Congresso Brasileiro de Redes Neurais, Florianópolis, pp. 324-329, **1997**.

[21] MORGAN, N. AND BOURLARD, H., "Continuous Speech Recogniton : An introduction to the Hybrid HMM/Connectionist Approach", IEEE Signal Processing Magazine, pp. 25-49, may **1995**.

[22] MORGAN, N. AND BOURLARD, H., "Neural Networks for Statistical Recognition of Continuous Speech", Proceedings of the IEEE., vol. 83, pp. 741-770, may **1995**.

[23] NEY, H., "The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, pp 253-271, April **1984**.

[24] NUNES, H. F., "Reconhecimento de Fala Baseado em HMM", Tese de Mestrado, UNICAMP, Campinas, **1996**.

[25] PACHECO, S. S. AND THOMÉ, A. G., "Turbo-Shake", Anais do III Congresso Brasileiro de Redes Neurais, pp. 51-55, Julho **1997**.

[26] PICONE, J. W., "Signal Modeling Techiques in Speech Recognition", Procedings of the IEEE, Vol. 81, No 9, 1215-1247, September **1993**.

[27] RABINER, L. AND JUANG, B. H., *Fundamentals of Speech Recognition*, Prentice Hall **1993**.

[28] RABINER, L. R., "A Tutorial on Hidden Markov Models an Selected Applications in Speech Recognition", Procedings of the IEEE, vol. 77, No. 2, pp. 257-286, **1989**.

[29] RABINER, L. R. AND LEVINSON, S. E., "A Speaker-Independent, Syntax-Directed, Connected Word Recognition System Based on Hidden Markov Models and Level Building," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP - 33, No. 3, pp. 561-573, June **1985**.

[30] RICHARD, M. D. AND LIPPMANN, R., "Neural Network Classifers Estimate Bayesian A-Posteriori Probabilities", Neural Computation 3(4), pp. 461-483, **1991**.

[31] ROBINSON, A. J., "An Application of Recurrent Nets to Phone Probability Estimation", IEEE Trans. on Neural Networks., vol. 5, pp. 298-305, **1994**.

[32] TATMAN, G. AND JANNARONE, R., "Real Time Neural Networks, III: Alternative Neural Networkds For Speech Aplications", IEEE Computer Society Press., pp. 591-596, **1991**.

[33] TEBELSKIS, J., "Speech Recognition using Neural Networks", PhD Thesis, Carnegie Mellon University, **1995**.

[34] TEBELSKIS, J., "Performance Through Consistency : Connectionist Large Vocabulary Contiunuous Speech Recognition", IEICE on Neuro-Computing, pp. 391-398, **1990**.

[35] VINTSYUK, T. K., "Element-Wise Recognition of Continuous Speech Composed of Words from a Specified Dictionary", Kibernetika, Vol 7, pp. 133-143, March-April. **1971**.

[36] YAN, Y., FANTY, M. AND COLE, R., "Speech Recognition Using Neural Networks with Forward-Backward Probability Generated Targets", IEEE ICASSP97, Munich, April **1997**.

[37] YOMA, N. B., "Reconhecimento Automático de Palavras Isoladas : Estudo e Aplicação dos Métodos Determinístico e Estocástico", Tese de Mestrado, UNICAMP, Campinas, **1993**.

[38] YOUNG, S., "A Review of Large-Vocabulary Continuous-Speech Recognition", IEEE Signal Processing Magazine, pp. 45-57, september **1996**.

[39] WAIBEL, A., ZEPPENFELD, T. AND HOUGHTON, R., "Improving The MS-TDNN for Word Spotting", International Acoustic Speech and Signal Processing, pp. 475-478, **1993**.

[40] WAIBEL, A., HANAZAWA, T., HINTON, G., SHIKANO, K. AND LANG, K. J., "Phoneme Recognition Using Time-Delay Neural Networks", IEEE Trans. on Acoustics Speech and Signal Processing., vol. 37, pp. 328-338, march **1989**.

[41] WAN, E. A., "Times Series Prediction by Using a Connectionist Network with Internal Delay Lines", *Forecasting the Future and Undertanding the Past*, A. S. Weigend and N. A. Gershenfeld, pp. 195-217, Addison-Wesley, **1993**.

[42] OSTENDORF, M., AND S. ROUKOS., "A Stochastic Segment Model for Phoneme-Based Contiunuous Speech Recognition", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-37, pp 1857-1869, **1989**.