UNIVERSIDADE ESTADUAL DE CAMPINAS – UNICAMP
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO – FEEC
DEPARTAMENTO DE COMUNICAÇÕES – DECOM

NORMALIZAÇÃO DE LOCUTOR EM SISTEMA DE RECONHECIMENTO DE FALA

Raquel de Souza Ferreira Dias
Orientador: Prof. Dr. Fábio Violaro

Banca Examinadora:
Prof. Dr. Fábio Violaro – FEEC/UNICAMP – Presidente
Prof Dr. Ivandro Sanches – EPUSP
Prof. Dr. João Bosco Ribeiro de Val – FEEC/UNICAMP
Dr. José Antônio Martins – Motorola (Jaguariúna)

Tese submetida à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas – UNICAMP, como parte dos requisitos exigidos para obtenção do título de **Mestre em Engenharia Elétrica.**

- Novembro de 2000 -

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

D543n

Dias, Raquel de Souza Ferreira
Normalização de locutor em sistema de reconhecimento de fala / Raquel de Souza Ferreira Dias.--Campinas, SP: [s.n.], 2000.

Orientador: Fábio Violaro
Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Sistema de reconhecimento de voz. 2. Markov, Processos de. 3. Espectro de freqüência. 4. Filtros adaptativos. I. Violaro, Fábio. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

A Deus, por ter me dado forças e
perseverança para concluir meu trabalho.

Agradecimentos

Ao Prof. Dr. Fábio Violaro pelo apoio, incentivo e discussões de idéias, fundamentais para a conclusão deste trabalho.

À minha mãe e meu irmão pela compreensão e, principalmente, pela paciência nos momentos mais difíceis.

Aos colegas do LPDF, Fabrício, JanKees, Léo, Luís e Wilson, por proporcionarem um ambiente de trabalho alegre e descontraído.

Aos amigos Eurípides, Irene, Helder, Marta e Rodrigo, pela motivação e pelo auxílio nas discussões de idéias, e principalmente ao Dr. Carlos Ynoguti, pela paciência e orientação no desenvolvimento do trabalho.

Ao colega e amigo Prof. Dr. Antônio Marcos, e toda sua família, pela acolhida durante os primeiros tempos na cidade e pelas inúmeras discussões e idéias que me ajudaram na conclusão desta tese.

Aos professores e funcionários da FEEC.

Às pessoas que emprestaram suas vozes nas confecção da base de dados.

A todos os amigos que direta ou indiretamente participaram desta etapa da minha vida.

Ao CNPq pela concessão da bolsa de Mestrado, permitindo que este trabalho fosse realizado.

Obrigada a todos.

Índice

Resumo

Este trabalho tem por objetivo avaliar um sistema de reconhecimento de fala de vocabulário flexível, quando utilizada a técnica de normalização de locutor. A técnica de normalização adotada foi a de escalonamento ("warping") do eixo de freqüências. Este escalonamento foi realizado pela variação do banco de filtros, na escala Mel, na obtenção dos coeficientes Mel Cepstrais. Estes coeficientes e suas derivadas foram empregados nos Modelos Ocultos de Markov (HMMs) que modelam as sub-unidades da fala (fones). O banco de filtros foi então ajustado na tentativa de normalizar o comprimento do trato vocal dos diferentes locutores, sendo empregado um procedimento iterativo no retreinamento do sistema de reconhecimento. Na avaliação desta técnica, procuramos a melhor forma de realizar a normalização do sistema. Com a utilização deste método conseguiu-se reduzir a taxa de erro de um sistema básico, operando com um vocabulário de 400 palavras, de 19,25% para 11,25%.

Palavras Chave: Normalização de Locutor, Distorção do Eixo de Freqüências, Reconhecimento de Fala de Vocabulário Flexível, Modelos Ocultos de Markov Contínuos.

Abstract

This work was developed to evaluate a flexible vocabulary speech recognition system, when using speaker normalization. The normalization technique employed was the frequency warping. This warping was accomplished changing the Mel scale filter bank used in the mel-cepstral coefficients calculation. These mel-cepstral coefficients and its derivatives were employed in the Hidden Markov Models (HMMs) used to model the speech subunits (phones). The Mel scale filter bank was then adjusted in an effort to normalize the vocal tract length among different speakers and an iterative procedure was used to retrain the recognition system. In the evaluation of this technique, we searched the best way to accomplish the system normalization. With this method the error rate of a basic system, working with a 400 words vocabulary, decreased from 19.25% to 11.25 %.

Keywords: Speaker Normalization, Frequency Warping, Flexible Vocabulary Speech Recognition System, Continuous Hidden Markov Models.

1 Introdução

Com o avanço tecnológico, várias técnicas de análise e processamento de fala vêm sendo estudadas. Estas técnicas têm por objetivo tentar melhorar não só a qualidade de vida do ser humano, como também a oferta de serviços prestados à população.

A tecnologia em reconhecimento de fala, por exemplo, tem evoluído consideravelmente. E, sob condições controladas, vem permitindo a implementação de sistemas práticos que possibilitem interfaces amigáveis com o usuário. Estes sistemas, por sua vez, vem ocupando um espaço cada vez maior em diversas áreas, principalmente na área comercial.

Dentre as várias aplicações dos sistemas de reconhecimento de fala, podemos citar:

- ✓ Discagem através da fala, utilizada principalmente na telefonia móvel. A discagem dos números ou de palavras associadas a estes números é feita automaticamente, via voz, sem auxílio das mãos ou olhos.
- ✓ Na área de automação industrial, permitindo o controle de máquinas e dispositivos, sem o auxílio das mãos.
- ✓ No acesso a banco de dados, via voz, para verificação de saldo bancário, informações meteorológicas, etc.
- ✓ Na utilização de interfaces, via voz, para deficientes físicos, permitindo uma maior liberdade na realização de suas tarefas diárias.

Apesar do bom desempenho destes sistemas, ainda não se conseguiu um sistema de reconhecimento robusto, capaz de operar com vocabulário ilimitado e com independência de locutor. E um dos principais motivos desta limitação está relacionado à variabilidade dos sinais de fala.

A variabilidade da fala, encontrada em uma população, não deve ser associada apenas às diferenças de comprimento de trato vocal dos locutores, mas também às diferenças lingüísticas (sotaque, dialeto) e às condições físicas e emocionais destes locutores. Em geral, estas fontes de variabilidade não podem ser eliminadas, devendo, portanto, ser modeladas diretamente pela tecnologia de reconhecimento de fala adotada. Como exemplo de uma tecnologia muito utilizada atualmente, e que será usada neste trabalho, temos os HMMs (*Hidden Markov Models*), adotados pelo bom compromisso de seu custo computacional "versus" potência e flexibilidade, quando comparados a outras tecnologias.

Os efeitos causados pela variabilidade da fala são os maiores desafios para o estado-da-arte dos sistemas de reconhecimento de fala independente de locutor. Por este motivo, na tentativa de reduzir estes efeitos e conseqüentemente produzir uma melhoria no desempenho destes sistemas, têm-se investigado várias técnicas de transformação, que tentam normalizar (adaptar) as representações paramétricas do sinal de fala, através da reestimação de seus parâmetros acústicos [2,5,9,10,11 e 16].

Dentre as técnicas de transformação dos parâmetros acústicos, utilizadas na normalização de locutor, a que tem proporcionado os melhores resultados é a técnica de distorção ("warping") do eixo de freqüências. Esta técnica é proposta na tentativa de normalizar o comprimento do trato vocal, de diferentes locutores, com o objetivo de diminuir sua influência na obtenção dos parâmetros espectrais da fala.

Na técnica de distorção do eixo de freqüências os parâmetros acústicos são transformados de forma a escalonar, linearmente, o sinal de fala no domínio da freqüência. Este escalonamento pode ser realizado de duas maneiras. A primeira, proposta por [2], é feita pela compressão ou expansão do sinal da fala, no domínio da freqüência, sendo este sinal em seguida reamostrado, no domínio do tempo. A segunda maneira, proposta por [11], é realizada pela compressão ou expansão não mais do sinal de fala, e sim do banco de filtros, nas escalas Mel ou Bark, sem necessidade de mudança (reamostragem) do sinal original.

Neste trabalho adotaremos o método de distorção proposto por [11], por apresentar um mecanismo mais simples de implementação e por proporcionar uma eficiente melhora na robustez dos sistemas de reconhecimento de fala. Objetivando analisar o efeito causado pela variação do comprimento do trato vocal, entre diferentes locutores, ao longo deste trabalho

será discutida e avaliada a influência da distorção do banco de filtros na normalização das características espectrais destes locutores. A estrutura do trabalho é então desenvolvida da seguinte forma:

No capítulo 2 é apresentada uma introdução sobre os sistemas de reconhecimento de fala e suas restrições. Em seguida, é descrito o tipo de sistema abordado neste trabalho, que será denominado de *Sistema de Reconhecimento de Fala de Vocabulário Flexível*.

No capítulo 3 é apresentada a teoria sobre os HMMs, juntamente com aspectos relacionados à implementação do sistema de reconhecimento adotado.

No capítulo 4 descreve-se a base de dados e o ambiente de simulação usado tanto no treinamento quanto no reconhecimento do sistema adotado.

No capítulo 5 é apresentado o sistema de reconhecimento de fala flexível utilizado, descrevendo-se as etapas de extração de parâmetros, treinamento e reconhecimento. O capítulo é concluído avaliando-se o sistema implementado, sendo os resultados obtidos utilizados como referência para o sistema que será desenvolvido no capítulo 6.

No capítulo 6 são apresentadas a teoria e os procedimentos utilizados na normalização de locutor.

No capítulo 7 tem-se o treinamento e avaliação do sistema de reconhecimento de fala utilizando normalização de locutor.

No capítulo 8 apresentam-se as conclusões do trabalho, sugerindo-se ainda o desenvolvimento de trabalhos futuros a partir dos resultados obtidos.

2 Sistema de Reconhecimento de Fala Utilizando Vocabulário Flexível

2.1 - INTRODUÇÃO

Os sistemas capazes de reconhecer comandos de fala, pronunciados pausadamente ou de forma contínua, são genericamente chamados de sistemas de reconhecimento de fala.

As principais dificuldades relacionadas à tarefa de reconhecimento de fala devem-se principalmente à natureza aleatória do sinal (variabilidade da fala) e às distorções ocasionadas pelos equipamentos usados durante sua aquisição, além do próprio ambiente de aquisição que, quando sujeito a ruídos apreciáveis, pode comprometer o desempenho destes sistemas.

Embora diversas técnicas e estratégias tenham sido propostas e implementadas com o objetivo de se obter um sistema que permita a livre conversação entre o homem e a máquina [18], os sistemas atuais só conseguem atingir desempenho satisfatório quando submetidos a certas restrições.

2.2 - CLASSIFICAÇÃO DOS SISTEMAS DE RECONHECIMENTO DE FALA

Na prática, dependendo das restrições impostas, é possível obter uma melhora considerável no desempenho do sistema. Estas restrições levam em consideração o grau de dependência do locutor, estilo de fala (forma de pronúncia da sentença falada) e tamanho do vocabulário de reconhecimento.

2.2.1 - Dependência de Locutor

Quanto à dependência de locutor, os sistemas de reconhecimento de fala costumam ser classificados como dependentes de locutor e independentes de locutor.

A dependência de locutor é, essencialmente, uma característica resultante do modo como o reconhecedor será implementado. Pode-se notar que, à medida que locutores de diferentes aspectos lingüísticos (idade, sexo, nível sócio-cultural, entre outros fatores) são utilizados no treinamento do sistema, mais independente de locutor o sistema será, ou seja, quanto maior a variabilidade de pessoas participantes da base de treinamento, melhor poderá vir a ser o desempenho do sistema para um locutor genérico.

Os sistemas dependentes de locutor são capazes de reconhecer a fala de apenas um locutor, aquele para o qual foi treinado. Caso se queira mudar de locutor, os padrões de referência usados no treinamento também deverão ser modificados.

Os sistemas independentes de locutor, ao contrário dos sistemas dependentes de locutor, são capazes de reconhecer a fala de qualquer pessoa, mesmo que esta não tenha participado do treinamento, para obtenção dos padrões de referência.

A independência de locutor, a princípio, é difícil de ser atingida, pois as representações do sinal de fala, usados em sistemas práticos, são sensíveis às características de um locutor em particular. Tendo isso em mente, pesquisadores têm buscado encontrar parâmetros que possam representar o sinal de fala de maneira relativamente insensível à mudança de locutor.

Por outro lado, o reconhecimento dependente de locutor é mais simples de ser atingido, além de possuir a vantagem de apresentar baixas taxas de erros, quando comparado ao sistema independente de locutor. Entretanto, em aplicações via linha telefônica, por exemplo, muito utilizadas atualmente, há a necessidade de que os sistemas sejam independentes de locutor, uma vez que é impraticável a realização de um treinamento inicial para a fala de cada locutor que venha a utilizar a aplicação desejada.

Devido a esta necessidade, entre outras, têm-se investido em pesquisas no reconhecimento de fala irrestrita, independente de locutor e com vocabulário ilimitado, ou seja, sistemas capazes de reconhecer, independente de contexto, qualquer palavra, sem restrição quanto ao conteúdo ou ao estilo de fala empregado.

2.2.2 – Estilo de Fala

Com relação ao tipo de fala, os sistemas de reconhecimento costumam ser classificados como sistemas de reconhecimento de fala contínua e de palavras isoladas.

No reconhecimento de fala contínua o usuário fala naturalmente, ocorrendo a coarticulação (efeito que provoca alteração na forma de pronúncia dos sons devido à influência dos sons vizinhos) do final de uma palavra com o começo da próxima. Por causa deste fenômeno, a complexidade de um sistema para reconhecer fala contínua é maior que a de um sistema de reconhecimento de palavras isoladas.

No reconhecimento de palavras isoladas tem-se as palavras faladas isoladamente, ou seja, com uma pausa entre as mesmas. Um valor considerado razoável para esta pausa é de 200 ms, no mínimo. O objetivo desta pausa entre as palavras é de facilitar a detecção de início e fim das mesmas, além de permitir uma pronúncia clara destas palavras, evitando o efeito de coarticulação.

Apesar da dificuldade em se trabalhar com a fala contínua, devido principalmente à ausência de uma fronteira definida entre as palavras, tem-se como vantagem a velocidade na transmissão da informação, pois o número de palavras por unidade de tempo, quando pronunciadas continuamente, é maior que no caso das palavras isoladas.

2.2.3 - Tamanho do Vocabulário

Em relação ao tamanho do vocabulário, podemos classificá-lo como reduzido, médio e grande. Os sistemas de vocabulário reduzido são caracterizados por reconhecerem até 99 palavras, os sistemas de tamanho médio por reconhecerem de 100 a 999 palavras e os de tamanho grande por reconhecerem 1000 ou mais palavras.

O número de palavras a serem reconhecidas também influencia no desempenho dos sistemas de reconhecimento de fala. Um vocabulário grande, por exemplo, tem maior probabilidade de possuir palavras ambíguas, isto é, palavras que são aparentemente

semelhantes para o algoritmo de reconhecimento. Este fato, por sua vez, pode provocar um aumento na taxa de erros do sistema.

Em algumas aplicações, quando se tem vocabulários grandes, o tempo necessário para pesquisar a base de dados pode tornar o sistema proibitivo, sendo necessária a utilização de técnicas de busca que ignorem certos caminhos de pesquisa (proibição de certas seqüências de palavras, desconsideração de caminhos com probabilidades muito baixas, etc.). Entretanto, tais técnicas podem introduzir erros no sistema, pelo fato de desprezarem referências que poderiam estar corretas.

Em vocabulários pequenos pode-se fazer o modelamento individual para cada palavra. Porém, à medida que o vocabulário aumenta, a quantidade de dados para processamento e armazenamento torna-se inviável, caso se utilize um modelo para cada palavra. Nesse caso, é comum se modelar unidades menores, como os fones, por exemplo, onde o modelo das palavras é obtido através da concatenação dos modelos dos fones correspondentes. Entretanto, este procedimento produz uma queda de desempenho, pois os efeitos de coarticulação não são tão bem modelados como no modelo da palavra completa.

2.3 - VOCABULÁRIO FLEXÍVEL

Os sistemas que incluem vocabulário flexível oferecem grande flexibilidade e velocidade no desenvolvimento de aplicações, dispensando a geração de grandes bases de dados. Estes sistemas permitem a definição de um vocabulário de reconhecimento através da simples digitação de suas palavras constituintes.

O vocabulário é dito flexível quando formado por qualquer palavra do nosso idioma, não possuindo correlação com a base de dados utilizada no treinamento do sistema. Neste trabalho, por exemplo, optou-se por utilizar fala contínua (frases) para o treinamento dos HMMs e nomes de pessoas (palavras isoladas) para o reconhecimento do sistema.

O desenvolvimento de um sistema de reconhecimento de fala de vocabulário flexível consiste na construção de um conjunto de modelos de unidades de fala menores que a palavra (fones, difones, trifones), denominadas sub-unidades, capazes de gerar, quando concatenados,

qualquer palavra. Para isto, há a necessidade da elaboração de um amplo vocabulário de treinamento, contendo repetições de todas as sub-unidades de fala, escolhidas nos mais variados contextos.

A existência de um conjunto de modelos treinados de sub-unidades, possibilita a criação de modelos de palavras através da simples concatenação dos modelos das sub-unidades correspondentes. Assim, para se determinar a seqüência de sub-unidades correspondentes a uma determinada palavra, deve-se realizar a transcrição de seu texto ortográfico em uma seqüência de sub-unidades básicas (a mesma utilizada no treinamento). No caso da utilização de fones como sub-unidades básicas, esta transcrição nada mais é que uma conversão ortográfico-fonética [15].

Desenvolver um sistema prático de reconhecimento de vocabulário flexível é uma tarefa com grau de dificuldade elevado, pois exige o trabalho conjunto de lingüistas e engenheiros na elaboração da base de dados a ser utilizada. O primeiro grupo, os lingüistas, tem como objetivo obter um conjunto de palavras com o máximo balanceamento possível das sub-unidades adotadas. O segundo grupo, os engenheiros, são responsáveis pela criação dos modelos HMMs das sub-unidades adotadas, de forma clara e precisa, visando o melhor modelamento destas sub-unidades, com conseqüente melhora de desempenho do sistema.

3 Modelos Ocultos de Markov

3.1- INTRODUÇÃO

A teoria de HMM (Hidden Markov Models), aplicada ao reconhecimento de fala, assume que a fala pode ser modelada estatisticamente como um processo de Markov, resultando na representação dos sinais de fala por modelos compactos e com boa caracterização de suas propriedades temporais e espectrais.

O uso de HMM no reconhecimento de fala foi proposto inicialmente por Baker [3] e teve sua teoria fundamentada na década de 60 por Baum *et al* [4], sendo atualmente uma das principais ferramentas utilizadas nos sistemas de reconhecimento de fala.

3.2 - DESCRIÇÃO DO MODELO

Em um HMM, formado por uma cadeia de estados (cadeia de Markov), existem dois processos estocásticos associados, um envolvendo as transições entre os estados e outro envolvendo as observações de saída de cada estado. As observações de saída são manifestações do fenômeno sendo modelado, e são descritas por funções probabilísticas que podem ser obtidas de duas formas. A primeira delas, usualmente utilizada no modelamento acústico do sinal de fala, está associada à emissão de um símbolo no instante de chegada a um estado, e é conhecida como máquina de Moore. A segunda forma, geralmente utilizada no processamento de linguagem, está associada à emissão de um símbolo durante a transição entre estados, e é conhecida como máquina de Mealy.

A Figura 3.1 apresenta um exemplo de modelo HMM de 3 estados, correspondente às formas de Moore e Mealy respectivamente, sendo a primeira forma utilizada como modelo neste trabalho.

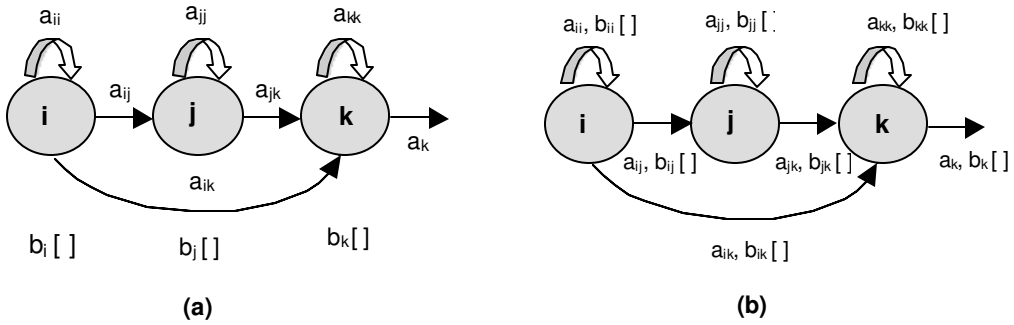


Figura 3.1: Exemplo de um HMM de 3 estados, associado à topologia Moore (a) e Mealy (b)

Na Figura 3.1,

i, j e k – representam os estados do modelo.

a_{ij} – representa a probabilidade de transição do estado *i* para o estado *j*.

b – representa a probabilidade de emissão de um símbolo, associado ou a um estado (b_i) ou a uma transição entre estados (b_{ij}).

Em geral, para o reconhecimento de fala, utiliza-se um modelo simplificado de HMM conhecido como modelo ***left-right***, ou modelo de Bakis [7]. Neste modelo, exemplificado na Figura 3.2, são permitidos apenas transições para o mesmo estado, ou transições de um estado *i* para um estado *j*, mais à direita, onde $a_{ij} = 0$ se $j > i + 2$.

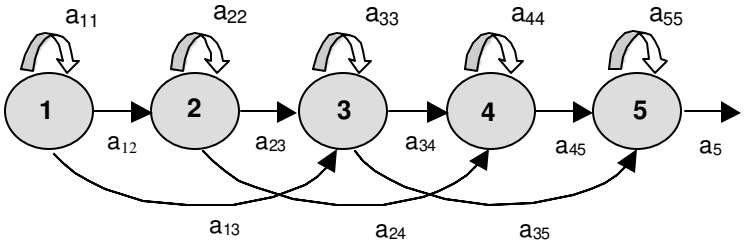


Figura 3.2: Exemplo de um HMM tipo *left-right* de cinco estados

Quando se aplica HMM a problemas de reconhecimento, as observações de saída são definidas por parâmetros representativos do sinal de fala, tais como os MFCC (Mel Frequency Cepstral Coefficients) [6 e 17], LFCC (Linear Frequency Cepstral Coefficients) [6], LPC (Linear Predictive Coding Coefficients) [13], entre outros.

Um HMM pode ser definido como sendo:

- Um conjunto de estados { S_j }, incluindo um estado inicial S_i e um estado final S_f .
- Uma matriz de transições $A=\{a_{ij}\}$, onde a_{ij} representa a probabilidade de se efetuar uma transição do estado i para o estado j.
- Uma matriz de probabilidades de saída $B=\{b_j(k)\}$, onde $b_j(k)$ define a probabilidade de emissão do símbolo k , ao se chegar ao estado *j* (modelo de Moore). O símbolo k pertence a um conjunto finito ou infinito de símbolos de saída.

Desde que a_{ij} e $b_j(k)$ sejam probabilísticos, as seguintes propriedades devem ser satisfeitas:

$$\sum_j a_{ij} = 1 \qquad \forall\, i,j, \text{ sendo} \quad a_{ij} \geq 0 \tag{3.1}$$

$$\sum_k b_j(k) = 1 \qquad \forall\, k, \text{ sendo } b_j(k) \geq 0 \tag{3.2}$$

Pode-se associar a seqüência de transições de estado a um processo estocástico X_t, e a seqüência de símbolos emitidos a outro processo Y_t, dependente do primeiro. Assim, a cada incremento em t existirá exatamente uma transição de estado (mesmo que seja para o mesmo estado – auto transição) e, conseqüentemente, a emissão de um símbolo. Utilizando esta notação temos:

$$a_{ij} = P\,(X_t = j \mid X_{t-1} = i) \tag{3.3}$$

$$b_j(k) = P(\, Y_t = k \mid X_t = j) \tag{3.4}$$

onde $X_t = j$ significa que a cadeia de Markov está no estado *j* no instante t, e $Y_t = k$ significa que o símbolo de saída emitido no instante t é k.

O conjunto de símbolos observáveis k corresponde à saída física do sistema que está sendo modelado. Desta forma, para sistemas de reconhecimento de fala, os símbolos devem corresponder a um conjunto de vetores de parâmetros extraídos do sinal sendo analisado.

3.3 - PARÂMETROS DO MODELO HMM

É comum o uso da notação observada em (3.5), utilizada para indicar um modelo HMM [20]:

$$\lambda=(A,B,\pi) \tag{3.5}$$

onde :

A é a matriz de elementos $a_{ij} = P [S_j \mid S_i]$, chamada matriz de transição de estados, de dimensão N^2, onde N é definido como o número de estados. S_i e S_j representam dois estados quaisquer do modelo.

B é a matriz de elementos $b_j (k) = P [V_k \mid S_j]$, que representa a probabilidade de se observar o resultado V_k (conjunto de resultados de saída esperados do modelo) estando no estado S_j. De acordo com a natureza dos elementos da matriz, os modelos HMMs podem ser classificados como discretos, contínuos ou semi-contínuos .

π : vetor de probabilidade inicial π (i). No caso do modelo left-right, π (i) = 1 para i = 1 e π (i) = 0 para i > 1.

3.4 - TIPOS DE HMM

Dependendo do tipo de função de probabilidade escolhida para os símbolos de saída, o HMM pode ser classificado como discreto, contínuo ou semi-contínuo.

- **Discreto:** O HMM é dito discreto quando o número de possíveis símbolos de saída, K, é finito e a probabilidade de se emitir o símbolo V_k , no estado S_j , é dada por $b_j(k)$.

A função de probabilidade B, mencionada em (3.5), tem as seguintes propriedades:

$$b_j(k) \geq 0 \qquad 1 \leq j \leq N \qquad 1 \leq k \leq K \tag{3.6}$$

$$\sum_{k=1}^{K} b_j(k) = 1 \qquad 1 \leq j \leq N \tag{3.7}$$

onde:

$b_j(k) = P[V_k \mid S_j]$

N - representa o número de estados do modelo.

K - representa o número de símbolos de saída.

No modelo HMM discreto, as seqüências de observações são formadas por índices de vetores de um dicionário. Este dicionário é formado por um conjunto de vetores chamados de palavras-código ou vetores-código. Em cada quadro do sinal de fala obtém-se um vetor de parâmetros que, após a quantização vetorial, é associado a um dos K possíveis vetores-código.

- **Contínuo:** O HMM é dito contínuo quando sua função densidade de probabilidade for contínua. Usualmente, utiliza-se a função densidade de probabilidade modelada como uma mistura finita de M gaussianas multidimensionais, dada por:

$$b_j(O_t) = \sum_{m=1}^{M} c_{jm} G(O_t, \mu_{jm}, U_{jm}) \qquad 1 \le j \le N \qquad (3.8)$$

onde:

O_t : é o vetor de parâmetros de entrada, suposto de dimensão D, no instante t.

c_{jm} : é o coeficiente da m-ésima componente da mistura no estado S_j .

G : é uma função densidade de probabilidade gaussiana multidimensional (dimensão D) com vetor média μ_{jm} e matriz covariância U_{jm}.

M - representa o número de gaussianas da mistura.

N - representa o número de estados.

Os coeficientes c_{jm} e a função densidade de probabilidade da mistura devem satisfazer as seguintes definições:

$$\sum_{m=1}^{M} c_{jm} = 1 \qquad 1 \le j \le N \qquad (3.9)$$

$$c_{jm} \ge 0 \qquad 1 \le j \le N \qquad (3.10)$$

$$\int_{-\infty}^{+\infty} b_j(\mathbf{x}) d\mathbf{x} = 1 \qquad (3.11)$$

sendo a integral da Equação (3.11) multidimensional (vetor **x** de dimensão D).

No modelo HMM contínuo a matriz B é representada pelo conjunto $B=\{b_j(.)\}$ de funções densidade de probabilidade, indexadas nos estados do HMM.

Quando as misturas (vetor média μ_{jm} e matriz covariância U_{jm}) são iguais para todos os estados do modelo, o HMM contínuo é denominado HMM "tied" contínuo [14]. Neste caso, de um estado para outro, haverá somente variação dos coeficientes de ponderação (c_{jm}) das componentes da mistura.

- **Semi-contínuo:** O HMM é dito semi-contínuo quando o modelo for um caso intermediário entre os HMMs discreto e contínuo. Neste caso, a densidade de probabilidade de emissão dos símbolos de saída é dada por:

$$b_j(O_t) = \sum_{V_k \in \eta(O_t)} c_j(k)\, f\,(O_t / V_k) \qquad 1 \le j \le N \tag{3.12}$$

onde:

O_t é o vetor de entrada.

K é o número de funções densidade de probabilidade (símbolos de saída).

$\eta(O_t)$ é o conjunto das funções densidade de probabilidade que apresentam os M maiores valores de $f(O_t \mid V_k)$, sendo $1 \le M \le K$.

V_k é o k-ésimo símbolo de saída.

$c_j(k)$ é a probabilidade de emissão do símbolo V_k no estado S_j.

$f(O_t \mid V_k)$ é o valor da k-ésima função densidade de probabilidade.

N - representa o número de estados utilizados.

No HMM semi-contínuo, o conjunto de funções densidade de probabilidade é o mesmo para todos os estados e todos os modelos, havendo modificação apenas dos coeficientes de ponderação $c_j(k)$.

Os coeficientes $c_j(k)$, os parâmetros das funções densidade de probabilidade e as probabilidades de transição são calculados e otimizados conjuntamente.

Quando o valor de M for igual a K, pode-se considerar o HMM semi-contínuo como um HMM contínuo, em que todas as misturas são iguais para todos os estados e todos os modelos, variando-se apenas os valores dos coeficientes das misturas de um estado para outro.

As matrizes B, para cada tipo de HMM, podem ser observadas mais detalhadamente em [14].

3.5 - HMMs EM RECONHECIMENTO DE FALA

Os sistemas de reconhecimento de fala geralmente são considerados como modelos de Markov de primeira ordem, para os quais são feitas duas hipóteses:

1) **Hipótese de Markov:** A probabilidade de uma cadeia estar em um dado estado, no instante t, depende apenas de seu estado no instante t-1.
2) **Hipótese de Independência de Emissão de Símbolos:** A probabilidade de emissão de um símbolo de saída, no instante t, depende apenas da transição realizada neste instante de tempo.

Os modelos de Markov, na prática, são usados para representar uma seqüência de observações. No reconhecimento de fala deseja-se determinar qual será o modelo λ_i (entre vários modelos disponíveis) que melhor representará a seqüência de observações desejada:

$$\max_i [P(\lambda_i | O)] = \max_i [P(O | \lambda_i)] \cdot P(\lambda_i) \quad (3.13)$$

Assumindo que os modelos λ_i são equiprováveis, a equação 3.13 resulta:

$$\max_i [P(\lambda_i | O)] = \max_i [P(O | \lambda_i)] \quad (3.14)$$

A estimação da Equação (3.14) consiste na própria identificação da palavra pronunciada, que é essencialmente um procedimento de classificação, onde deseja-se classificar uma determinada seqüência de observação O em alguma categoria i. Em outras palavras, deseja-se determinar o modelo λ_i que apresente a maior probabilidade de reproduzir a seqüência de observação desejada (O).

3.6 - MODELAMENTO DO HMM

Dado um HMM, tem-se 3 questões básicas que devem ser resolvidas: a avaliação, a decodificação e o treinamento.

Na avaliação deseja-se determinar qual o modelo, dentre os vários modelos, que mais provavelmente gerou uma dada seqüência de observações. Neste procedimento utiliza-se como solução o algoritmo Forward [8 e 14] ou o algoritmo de Viterbi [8 e 14].

Na decodificação utiliza-se como solução o algoritmo de Viterbi que, a partir de uma seqüência de observação, tem como função determinar a seqüência de estados que mais provavelmente produziu as observações.

No treinamento deseja-se determinar os parâmetros do modelo que maximizem a probabilidade de geração da observação. Neste procedimento utiliza-se como solução o algoritmo Forward-Backward, também conhecido como algoritmo de reestimação de Baum-Welch [8 e 14].

3.7 – HMM APLICADO EM RECONHECIMENTO DE FALA

O HMM pode ser utilizado na representação de quaisquer unidades ou sub-unidades da fala, tais como palavras, fones, difones, trifones, etc. Quando são criados modelos para as sub-unidades, a representação da palavra desejada é construída concatenando-se os modelos das sub-unidades correspondentes. Um exemplo de sub-unidade bastante utilizada são os fones, que podem ser classificados como dependentes de contexto ou independentes de contexto. Nos fones independentes de contexto os modelos HMMs são treinados independentemente, levando em consideração apenas a sua ocorrência e não o contexto em que ocorrem.

Uma das vantagens do uso da palavra como unidade está no fato da melhor representação dos efeitos da coarticulação. Porém, a escolha da unidade básica a ser modelada deve ser feita com cuidado, levando-se em consideração a viabilidade da obtenção dos dados para o treinamento desta unidade. Além disso, à medida que o tamanho da unidade básica aumenta, o número de unidades utilizadas na obtenção de qualquer palavra cresce

exponencialmente, juntamente com a quantidade de dados necessária para o treinamento do sistema.

Uma alternativa bastante razoável para escolha da unidade básica, é a utilização de sub-unidades fonéticas, que necessitam de vários exemplos de cada sub-unidades, e não mais de vários exemplos de cada palavra. Esta escolha pode ser considerada uma vantagem quando utilizam-se vocabulários grandes, onde a disponibilidade de um grande número de exemplos para cada palavra torna-se inviável.

Na escolha do número de estados ótimo, para representação do modelo HMM, não há regras. A maneira mais comum é fazer o número de estados igual ao número de fonemas da palavra. Neste caso, os modelos que representam palavras diferentes apresentarão números de estados diferentes.

Neste trabalho optou-se por utilizar os fones independentes de contexto como sub-unidades básicas, principalmente pela simplicidade de modelamento dos HMMs. Cada fone foi modelado através de um HMM de 3 estados. Desta forma, o modelo de cada palavra é gerado concatenando-se os modelos dos fones correspondentes a esta palavra.

Quanto ao número de símbolos de saída (HMM discreto), ou ao número de gaussianas por mistura (HMM contínuo) a ser utilizado, deve-se ter o compromisso entre o tamanho da seqüência de treinamento e o número de parâmetros a estimar, de modo que se consiga boas estimativas dos mesmos.

Quando um HMM é aplicado ao reconhecimento da fala, o sinal é representado por uma seqüência temporal de vetores $O=\{O_1, \ldots, O_T\}$. Para isto o sinal de fala deve ser dividido em segmentos, com um vetor de parâmetros (ou conjunto de vetores de parâmetros) caracterizando cada segmento. Assim, o vetor O_i corresponde a um vetor de parâmetros que representa o segmento i, e o número de vetores de observação O_i será igual ao número de segmentos T [14].

Nos sistemas utilizando HMM discreto, cada vetor de parâmetros $\hat{O}_i$ é representado por um símbolo pertencente a um conjunto de K símbolos, tornando-se necessário a utilização da quantização vetorial. Desta forma, tem-se uma densidade de probabilidade discreta associada a cada estado dos modelos.

Nos modelos utilizando HMM contínuo, por outro lado, não se emprega a quantização vetorial. A densidade de probabilidade associada a cada estado dos modelos é contínua, sendo representada normalmente por uma mistura de gaussianas multidimensionais. Neste caso evita-se o erro de quantização gerado na quantização vetorial dos vetores de parâmetros acústicos mas, em compensação, a estimação do modelo é bem mais complexa e requer uma quantidade maior de material de treinamento.

Como foram empregados vários parâmetros para modelamento do sinal de fala (coeficientes mel-cepstrais, delta-mel-cepstrais, delta-delta-mel-cepstrais, etc.), estes parâmetros foram considerados independentes entre si. Assim, a densidade de probabilidade de emissão de símbolos é dada pelo produto das densidades individuais de cada um dos parâmetros acima.

4 Base de Dados

4.1 - INTRODUÇÃO

A linguagem falada é a forma mais natural de comunicação humana, sendo sua estrutura moldada pelas estruturas fonológicas, sintáticas e prosódicas da língua, pelo ambiente acústico e pelo contexto em que a fala está sendo produzida [20].

O efeito causado pela variabilidade da fala é o principal responsável pela degradação dos sistemas de reconhecimento de fala. Por este motivo, tecnologias nesta área, tais como os HMMs, vêm tentando modelar estas variações utilizando métodos estatísticos que "aprendam" a partir de um conjunto de treinamento. Porém, para que estes métodos estatísticos sejam eficazes, é necessário que se utilize um conjunto de dados de treinamento considerável, tomando-se o cuidado para que as unidades básicas não deixem de ser modeladas ou sejam mal modeladas, causando degradação no desempenho do sistema.

O conjunto de dados utilizado no treinamento dos modelos deve ser escolhido de maneira que possa conter o máximo de locuções, com o maior número possível de unidades (palavras, fones, difones, etc.), nos mais diferentes contextos. Este procedimento, por sua vez, torna a base de dados custosa tanto em termos de trabalho, em sua confecção, quanto em termos financeiros, já que a base utilizada deve conter um conjunto considerável de locuções.

Para que os modelos estatísticos atinjam a eficácia desejada, a base de dados precisa ser extremamente grande. Estas bases de dados por sua vez, são geradas, na maioria das vezes, visando atender sistemas específicos, tornando-se difícil a avaliação de novas idéias de outros grupos de pesquisa, fazendo com que muitos pesquisadores tenham que dispender um tempo enorme no desenvolvimento de sua própria base. Este problema já não ocorre em países com maior tradição de pesquisa, pois possuem um maior incentivo para o desenvolvimento de grandes bases de dados, tanto de empresas quanto de instituições de pesquisa, incentivo este

que vem impulsionando de forma expressiva o desenvolvimento da tecnologia de fala nestes países.

4.2 – ESCOLHA DA BASE DE DADOS DE TREINAMENTO E DE TESTE

No presente trabalho, por motivo de economia de tempo, resolvemos treinar e normalizar o sistema de reconhecimento de fala de vocabulário flexível a partir da mesma base de dados elaborada por [20]. Essa base é constituída por frases foneticamente balanceadas e seqüências de dígitos conectados.

Para os teste de reconhecimento confeccionou-se outra base de dados, constituída de palavras isoladas, formada por nomes de pessoas e apelidos.

4.2.1 – Treinamento

A base de dados utilizada para o treinamento foi a mesma utilizada por [20], que é composta por 20 listas com cada lista contendo 10 frases foneticamente balanceadas, ou seja, frases geradas com uma distribuição fonética similar àquela encontrada na fala espontânea [1]. Adicionalmente foram empregadas mais 5 listas, também preparadas por [20], e compostas, cada uma, por 11 seqüências de 8 dígitos conectados, conforme listado no Apêndice A.

A base total é composta por 25 listas pronunciadas por diferentes locutores, sendo 23 homens (m01 a m25, com exceção de m19 e m22) e 19 mulheres (f01 a f21, com exceção de f14 e f16), resultando um total de 2000 locuções (1560 frases foneticamente balanceadas e 440 seqüências de 8 dígitos).

A gravação da base foi feita empregando-se uma placa de som SoundBlaster AWE64 de 16 bits, com freqüência de amostragem de 11,025 kHz, sendo os dados armazenados em formato Windows PCM (wav).

A maior parte dos locutores nasceu no interior do estado de São Paulo. Os outros são naturais de estados como Pernambuco, Ceará, Paraná e Amazonas. A maioria deles possui nível superior e todos tem pelo menos o 2º grau completo.

Os locutores foram divididos em grupos, onde cada grupo é composto por um determinado número de listas, repetidas tanto pelos locutores masculinos quanto pelos femininos, conforme a distribuição mostrada na Tabela 4.1.

Grupos	***Listas***	***Locutores***
1º grupo	1 2 3 4 **21**	f06 f09 f12 f13 m01 m07 m14 m20
2º grupo	5 6 7 8 **22**	f02 f07 f15 f20 (menos a lista **22**) m08 m16 m18 m23
3º grupo	9 10 11 12 **23**	f03 f08 f17 f18 m05 m09 m17 m21
4º grupo	13 14 15 16 **24**	f04 f10 f19 --- m03 m06 m11 m24
5º grupo	17 18 19 20 **25**	f01 f05 f11 f21 (menos a lista **25**) m04 m12 m13 m15
6º grupo	**21**	m25
7º grupo	**22**	m02
8º grupo	**23**	m10

Tabela 4.1: Tabela dos grupos de locutores da base de treinamento e suas respectivas listas. Os números em negrito representam as listas de dígitos conectados.

Conforme a Tabela 4.1, cada um dos locutores pronuncia um total de 5 listas, onde as 4 primeiras listas são compostas por 40 frases balanceadas e a última lista é formada por 11 seqüências de 8 dígitos conectados, representando um total de 51 locuções por locutor. Constituem exceção os locutores f20 e f21, que pronunciam apenas 40 locuções balanceadas (sem os dígitos conectados) e os locutores m02, m10 e m25 que pronunciam, cada um, apenas 11 seqüências de dígitos conectados .

4.2.2 – Teste

Para o teste do sistema de reconhecimento de fala de vocabulário flexível, foi preparado um vocabulário de 400 palavras, constituído por nomes de pessoas e apelidos. As palavras foram escolhidos de forma que contivessem a maior variação possível dos fones utilizados como unidades básicas (balanceamento fonético). Este vocabulário pode ser observado no Apêndice B.

O conjunto de 400 nomes e apelidos foi pronunciado por 20 locutores, sendo 10 homens e 10 mulheres, com cada um deles pronunciando 20 nomes distintos.

As gravações foram realizadas no laboratório de Processamento Digital da Fala (LPDF) do DECOM/ FEEC/ UNICAMP, utilizando uma placa de som SoundBlaster 16 (que apresenta desempenho inferior à placa utilizada no treinamento, AWE64). A freqüência de amostragem e o formato dos arquivos foram idênticos aos empregados na base de treinamento.

Da mesma forma que no treinamento, utilizamos locutores de diferentes naturalidades, como Pará, São Paulo, Minas Gerais, Espírito Santo e Pernambuco. A maioria dos locutores possui nível superior e todos tem pelo menos o 2º grau completo. Assim, quanto maior for a variabilidade de pessoas participantes do conjunto de dados, tanto de treinamento quanto de reconhecimento, mais confiável será a avaliação de desempenho do sistema implementado.

4.3 - TRANSCRIÇÃO FONÉTICA

A transcrição fonética tem como objetivo representar graficamente os sons existentes em uma língua. Esta representação pode variar com relação ao número de detalhes que se deseja representar, de modo que podemos ter uma transcrição larga, sem considerar muitos detalhes, ou uma transcrição estreita, procurando levar em consideração o maior número possível de detalhes acusticamente perceptíveis.

A transcrição é baseada no princípio de que um fonema é sempre representado por um símbolo e que este símbolo sempre representa um fonema. Dado um conjunto de símbolos para os fonemas, deve-se ser capaz de transcrever qualquer frase de uma língua numa

seqüência de fonemas, representando assim sua pronúncia de maneira inambígua para todos que sejam familiares a estes símbolos.

O conjunto de fones utilizado neste trabalho teve como referência o conjunto proposto por Alcaim et al [1] (35 fones), sendo estes fones correspondentes à realização prática dos fonemas. Entretanto, optou-se pela exclusão de alguns fones e inclusão de outros, com o objetivo de melhor modelar os HMMs para a base de dados utilizada neste trabalho. Desta forma, com base no novo conjunto de símbolos (39 fones), observado na Tabela 4.2, fez-se a transcrição fonética das frases e nomes que, posteriormente, serão utilizados no treinamento e no reconhecimento (teste) do sistema proposto.

É importante mencionar que no início e no final da transcrição, tanto para as frases e os dígitos quanto para os nomes, foi inserido um fone representando silêncio. Este fone foi inserido com o objetivo de fazer com que o sistema possa distinguir o início e o fim de cada palavra ou frase. O fone de silêncio é ainda utilizado sempre que há uma pausa mais longa, maior que 60 ms, entre uma palavra e outra, de forma que o HMM não confunda o silêncio de pausa para inspiração com qualquer outro fone, prejudicando assim o treinamento do sistema.

Consoantes

Símbolo	Exemplo
p	***p*** ato
b	***b*** ato
t	***t*** ato
d	***d*** ado
k	***c*** abide
g	***g*** ata
f	***f*** ato
v	***v*** aca
s	***s*** opa
z	***z*** orro
x	***ch*** ato
j	***j*** ato
l	***l*** ata
L	ca ***lh*** a
r	pê ***r*** a
rr	***r*** é
R	ca ***r*** ta (vibrante)
RR	ca ***r*** ta (retroflexo)
m	***m*** á
n	***n*** a
N	li ***nh*** a

Consoantes Fricativas

Símbolo	Exemplo
T	***t*** ia
D	***d*** ia

Semi Vogais

Símbolo	Exemplo
y	fu ***i***
w	cé ***u***

Vogais

Símbolo	Exemplo
i	***i*** da
e	***e*** le
E	***e*** la
a	***a*** ve
A	pat ***a***
o	t ***o*** do
O	***o*** va
u	***u*** va

Vogais Nasais

Símbolo	Exemplo
an	m ***ã*** e
en	p ***en*** te
in	p ***in*** ta
on	p ***õ*** e
un	f ***un*** do

Representação do Silêncio

Símbolo	Definição
#	Silêncio

Tabela 4.2: Sub-unidades acústicas utilizadas na transcrição fonética das locuções das bases de dados utilizadas no treinamento e teste do sistema proposto, incluindo-se um exemplo para cada fonema

5 Sistema Utilizado

5.1 - INTRODUÇÃO

Neste capítulo serão descritas as etapas utilizadas no desenvolvimento de um sistema de reconhecimento de fala utilizando vocabulário flexível.

O objetivo principal deste capítulo é avaliar o desempenho do sistema quanto à independência de locutor. Este sistema será tomado como referência para o modelamento do sistema desenvolvido para normalização de locutor.

O sistema de reconhecimento de fala utilizado é baseado no sistema implementado por [20], que recentemente foi ampliado para operar com HMMs contínuos. Este sistema é composto por 3 módulos, como mostrado na Figura 5.1, sendo estes módulos descritos a seguir.

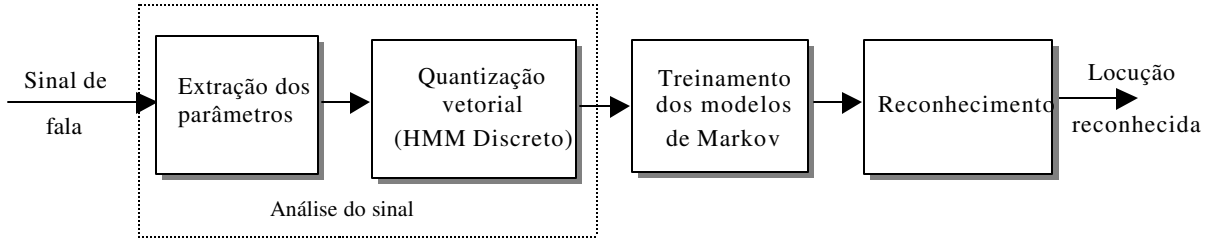


Figura 5.1: Diagrama de blocos do sistema de reconhecimento adotado

5.2 – ANÁLISE DO SINAL

O módulo de análise do sinal é dividido em 2 sub-módulos: o sub-módulo de extração dos parâmetros acústicos da fala e o sub-módulo de quantização vetorial (utilizado apenas para o HMM discreto). Estes módulos são responsáveis por transformar as locuções de entrada em parâmetros que possam ser interpretados pelos módulos seguintes.

5.2.1 - Extração de Parâmetros

O processo de obtenção dos parâmetros acústicos da fala envolve 3 etapas: a pré-ênfase, o janelamento e a análise espectral, representadas conforme a Figura 5.2.

Figura 5.2: Diagrama de blocos do processo de extração dos parâmetros mel-cepstrais

A pré-ênfase, realizada por um filtro passa altas ($1 - 0{,}95z^{-1}$), tem como função compensar a atenuação de 6dB/oitava nas altas freqüências. Esta atenuação é ocasionada pelo efeito combinado do espectro decrescente dos pulsos glotais (-12dB/oitava) e pelo efeito de radiação dos lábios (+6dB/oitava) [14].

Os parâmetros do sinal de fala são atualizados a cada 10 ms, sendo o janelamento do sinal calculado através da janela de Hamming de 20 ms. Este janelamento tem como função produzir suavização da amplitude do sinal amostrado, nos extremos do segmento de análise, dando maior ênfase às amostras localizadas no centro da janela. Desta forma, têm-se uma superposição entre os dados de análise de 2 janelas adjacentes. Este processo pode ser observado na Figura 5.3.

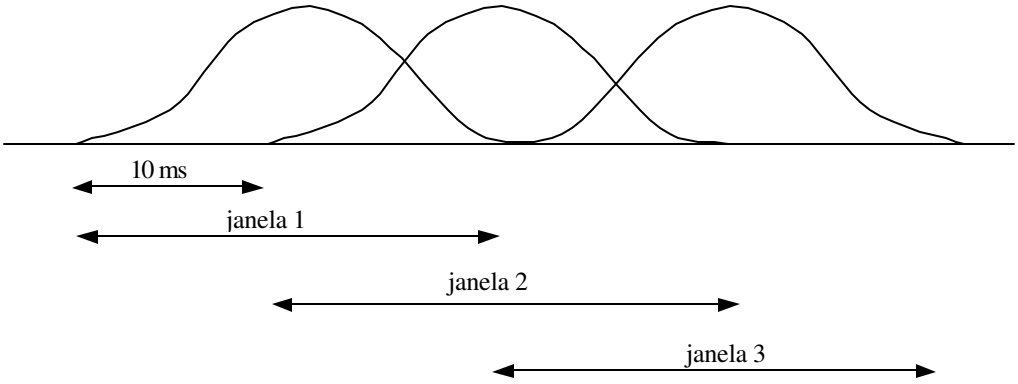


Figura 5.3: Processo de superposição de janelas para o cálculo dos parâmetros mel-cepstrais

Como finalização do processo, têm-se a análise espectral, onde é realizada a conversão da representação temporal do sinal analisado, para alguma forma de representação espectral.

Basicamente dois métodos de análise espectral predominam nos sistemas de reconhecimento de fala: o método de análise espectral LPC (Linear Predictive Coding) e o método de análise espectral por banco de filtros, obtido a partir da Transformada Rápida de Fourier (FFT). O segundo método será utilizado como referência neste trabalho, por ser mais eficiente na obtenção dos parâmetros mel-cepstrais.

Tanto o espectro resultante da FFT quanto o espectro resultante da predição linear são representações bem mais relacionadas ao processo de audição e percepção humana do que os métodos de representação temporal (taxa de cruzamentos por zero, perfil de energia, delta de energia, entre outros), utilizados na caracterização da fala. Isto justifica a ampla utilização de parâmetros extraídos a partir da representação espectral do sinal acústico, em relação aos parâmetros de representação temporal.

5.2.1.1 – Parâmetros Mel-Cepstrais Baseados em Banco de Filtros

Os parâmetros mel-cepstrais são obtidos, segundo [6], conforme ilustrado na Figura 5.4.

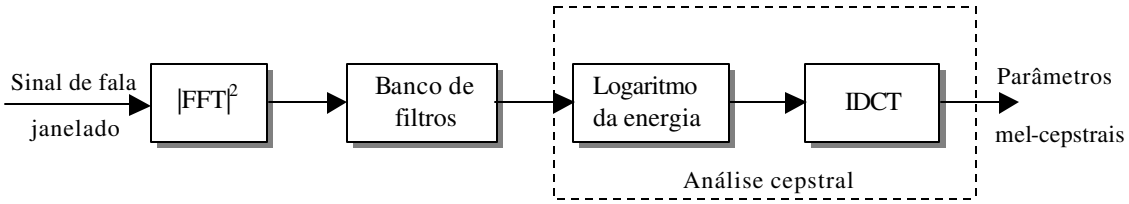


Figura 5.4: Diagrama de blocos utilizado no cálculo dos parâmetros mel-cepstrais utilizando banco de filtros

Inicialmente, calcula-se o quadrado do módulo da FFT das amostras pertencentes à janela de análise. Em seguida, filtra-se este sinal por um banco de filtros triangulares na escala Mel. Posteriormente, calcula-se o logaritmo da energia na saída dos filtros e, por fim, calcula-se a IDCT (Inverse Discrete Cosine Transform) sobre estes valores, obtendo-se os parâmetros desejados. No nosso caso são calculados 12 coeficientes mel-cepstrais por janela.

Além dos parâmetros cepstrais (mel-cepstrais), utilizou-se também a energia e os parâmetros diferenciais (delta-mel-cepstrais, delta-delta-mel-cepstrais, delta-energia), obtidos a partir do conjunto de seus respectivos parâmetros absolutos (mel-cepstrais e energia). Estes parâmetros foram utilizados no intuito de uma melhor caracterização das variações temporais do sinal de fala.

5.2.1.2 – Energia

A energia é calculada utilizando-se a seguinte equação.

$$E = \sum_{i=0}^{N-1} s^2(i) \qquad (5.1)$$

onde N representa o número de amostras da janela de análise e s(i) representa o sinal de fala janelado.

Neste trabalho utilizou-se a energia normalizada, dada por:

$$E_N = \log(E) - \log(E_{máx}) \tag{5.2}$$

onde $E_{máx}$ corresponde à janela de máxima energia para uma dada frase de treinamento.

5.2.1.3 – Parâmetros Diferenciais

Os parâmetros diferenciais adicionam informação sobre a derivada dos parâmetros empregados. Os parâmetros diferenciais utilizados foram calculados segundo a expressão:

$$\Delta_i(n) = \frac{1}{2K+1} \sum_{k=-K}^{K} k y_{i-k}(n) \tag{5.3}$$

onde :

K – está associado ao número de quadros adjacentes empregados no cálculo dos parâmetros diferenciais.

y_i – representa o vetor de parâmetros calculado no i-ésimo quadro.

Δ_i – representa o vetor delta correspondente ao vetor de parâmetros y_i calculado no i-ésimo quadro.

i – índice do quadro.

Neste trabalho são calculados, a cada 10 ms, 12 coeficientes delta-mel-cepstrais, 12 delta-delta-mel-cepstrais e mais o delta-energia, sendo utilizado para os parâmetros delta apenas um quadro de variação de cada lado (K = 1).

5.2.2 – Quantização Vetorial

Esta é a última etapa executada na fase de análise do sinal. Porém, só há necessidade desta etapa ser realizada se o sistema de reconhecimento de fala utilizado for baseado em HMMs discretos.

Nos HMMs discretos, os parâmetros que os individualizam são gerados a partir de seqüências de símbolos finitas. Entretanto, os parâmetros obtidos do sinal de fala, que serão utilizados como parâmetros de entrada deste modelo, apresentam comportamento e natureza contínua [8]. Daí a necessidade de uma conversão destes parâmetros, de natureza contínua, para parâmetros de natureza discreta. Esta conversão, ou codificação, realizada sobre os coeficientes empregados (mel, delta-mel, delta-delta-mel, energia e delta-energia), é chamada de *Quantização Vetorial*.

A quantização vetorial é realizada em 2 fases: uma responsável pela geração dos vetores códigos do quantizador, ou dicionário (codebook), e outra responsável pela quantização propriamente dita.

Para a geração dos vetores códigos do quantizador (treinamento), utilizou-se o algoritmo Lindo-Buzo-Gray (LBG) em sua versão "splitting" [12]. Este algoritmo tem por objetivo gerar um número N de vetores código, de forma a melhor representar uma grande amostra de vetores de parâmetros, obtidos de locuções utilizadas no treinamento do sistema. O critério utilizado para escolha destes vetores código é baseado na minimização da soma das distorções de cada vetor de treinamento, em relação ao seu vetor código mais próximo. No presente trabalho, para cada um dos parâmetros de entrada (mel, delta-mel, delta-delta-mel, energia e delta-energia) foi gerado um dicionário contendo 256 vetores.

O processo de quantização é então finalizado calculando-se a distorção entre cada vetor de entrada e todos os vetores código, escolhendo-se o vetor código que assegura a menor distorção. A medida de distorção utilizada neste procedimento foi a *Distância Euclidiana* e, como resultado final deste processo, tem-se um único valor inteiro (índice do vetor código escolhido) por quadro de fala. A seqüência destes vetores código, representando eventos de fala, será a seqüência de observação representativa dos fones.

5.3 – TREINAMENTO

Este módulo tem como função treinar os modelos HMMs das sub-unidades fonéticas a serem utilizados no módulo de reconhecimento.

O treinamento destas sub-unidades é realizado em 2 etapas: a etapa onde é feita a inicialização dos modelos e a etapa de treinamento.

5.3.1 – Inicialização

Na inicialização dos modelos HMM, as probabilidades de transição são consideradas equiprováveis, como mostrado na Figura 5.5.

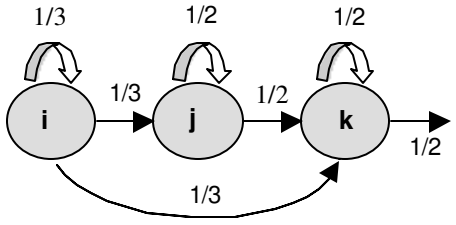


Figura 5.5: Valores iniciais utilizados como probabilidade de transição dos modelos dos fones

Quanto à densidade de probabilidade de emissão, no caso do HMM discreto, foi adotada uma distribuição uniforme. No caso do HMM contínuo, foi empregado o método via "Segmental K-Means".

Pelo método da distribuição uniforme (HMM discreto), assume-se que, inicialmente, todos os símbolos são equiprováveis, e que as probabilidades de emissão de saída destes símbolos recebem valor 1/num_vet, onde num_vet representa o número de vetores com o qual foi realizada a quantização vetorial dos parâmetros. No presente trabalho num_vet = 256.

No método via "Segmental K-Means", empregado para o HMM contínuo, a inicialização é realizada dividindo-se cada locução de treinamento em n partes iguais (de mesmo comprimento), sendo n definido como o número de sub-unidades acústicas da locução multiplicada pelo número de estados de cada modelo (que neste trabalho é igual a 3). Desta forma, cada estado do modelo é associado à n-ésima parte da locução. O modelo HMM é então criado, para cada locução, concatenando-se os modelos HMMs das sub-unidades acústicas referentes a sua transcrição fonética. Em seguida é feita uma contagem dos símbolos que ocorreram em cada uma das n partes da locução. Quando se tem repetição de sub-unidades numa mesma locução, a contagem dos símbolos é acumulada em uma mesma função densidade probabilidade.

Após a inicialização do modelo (discreto ou contínuo), é realizado novamente o procedimento "Segmental K-Means", que corresponde a um pré-treinamento dos modelos utilizando o algoritmo de Viterbi. Este procedimento é semelhante ao procedimento de inicialização descrito anteriormente, para o HMM contínuo, com a diferença de que agora a segmentação não é mais uniforme, ou seja, para cada um dos estados são associados mais ou menos quadros, dependentes do caminho escolhido pelo algoritmo de Viterbi. As probabilidades de emissão são atualizadas pela contagem dos símbolos emitidos em cada estado, e as probabilidades de transição são dadas pelo número de quadros obtidos em cada estado.

5.3.2 – Treinamento

Após a inicialização, vem o treinamento propriamente dito, no qual é utilizado o algoritmo de Baum-Welch [8]. Neste procedimento, para cada locução de treinamento é gerado um modelo HMM através da concatenação dos módulos referentes às sub-unidades de sua transcrição fonética. Em seguida, os modelos individuais das sub-unidades fonéticas são separados e as contagens (número de ocorrências de cada fone), geradas pelo algoritmo Baum-Welch, são acumuladas durante todo o processo de treinamento (uma época de treinamento), sendo em seguida transformadas em medidas de probabilidade.

Após cada época de treinamento, faz-se a verificação da convergência do processo de treinamento. Para cada locução de treinamento monta-se um modelo HMM correspondente, através da concatenação dos modelos das sub-unidades fonéticas, e calcula-se a probabilidade do modelo gerar a locução de entrada, empregando-se o algoritmo de Viterbi. Este procedimento é repetido para uma certa quantidade de locuções de treinamento (no nosso caso, 10% do material), obtendo-se a probabilidade média (verossimilhança: ML -"Maximum Likelihood") correspondente a estas locuções de treinamento. Na normalização do locutor, que será descrita posteriormente, a probabilidade média é calculada utilizando-se 10% do material de cada locutor. O treinamento é então finalizado quando a diferença relativa entre a verossimilhança média da época atual e a verossimilhança média da época anterior for menor que 0,001.

5.4 – RECONHECIMENTO

Este módulo é o responsável pelo mapeamento dos parâmetros acústicos correspondentes às locuções de entrada (locuções de teste), em sua transcrição ortográfica.

Dada uma locução de entrada, um sistema de reconhecimento de fala tem como objetivo gerar uma dada palavra, ou um conjunto de palavras, através de um processo de busca. Este processo de busca é realizado pela comparação entre a seqüência de vetores de características acústicas e os modelos das palavras que estão no vocabulário utilizado pelo sistema.

O "software" empregado no reconhecimento foi o mesmo desenvolvido por [20]. Este software foi desenvolvido para reconhecimento de fala contínua e admite duas opções, o algoritmo "Level Building" e o algoritmo "One-Step". Ambos são equivalentes sob o ponto de vista de resultados porém, quanto à implementação dos algoritmos, o "One-Step" proporciona facilidades como a redução do tempo de processamento das locuções. Deve-se ressaltar que não se está trabalhando com fala contínua, mas sim com palavras isoladas, sendo empregado o algoritmo de busca "Level Building" para o HMM discreto e o "One-Step" para o HMM contínuo, ambos com 3 níveis de busca. Estes níveis correspondem à palavra a ser reconhecida

e aos silêncios inicial e final desta palavra. Em ambos os casos, a verossimilhança é dada pelo algoritmo de Viterbi, que considera apenas a seqüência ótima de estados que produziu a seqüência de símbolos desejada.

O reconhecimento é então realizado, como observado na Figura 5.6, levando-se em consideração: os parâmetros acústicos das locuções a reconhecer (quantizados ou não), os modelos HMMs contendo as sub-unidades já treinadas e o vocabulário com o universo das palavras a serem reconhecidas. Os dois primeiros itens já foram abordados nas seções anteriores, de modo que nesta seção será dada ênfase apenas ao vocabulário de reconhecimento.

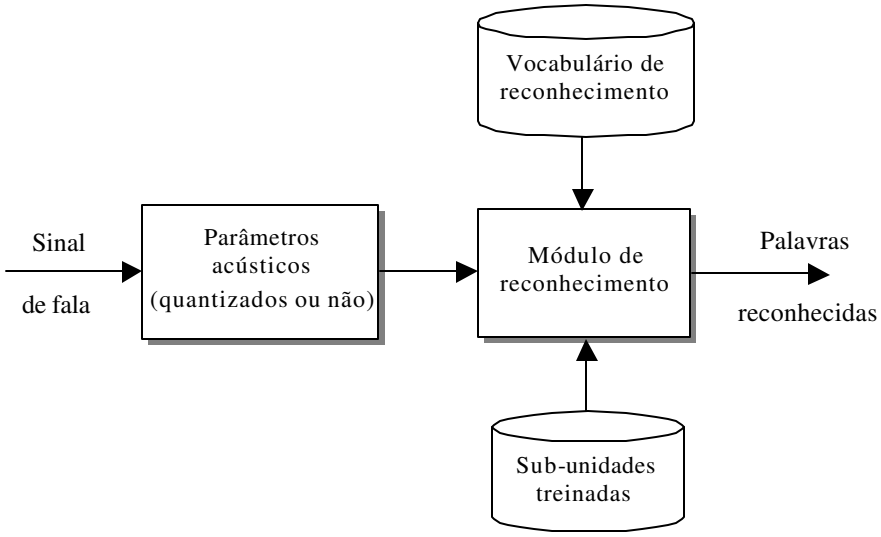


Figura 5.6: Figura representativa do módulo de reconhecimento, juntamente com os "parâmetros" de entrada utilizados neste módulo

5.4.1 - Vocabulário de Reconhecimento

O vocabulário de um sistema de reconhecimento de fala define o universo de palavras que serão reconhecidas, ou seja, toda e qualquer locução será mapeada em um conjunto de palavras deste universo. Em termos gerais, quanto maior e mais abrangente for o vocabulário,

mais flexível é o sistema, embora o reconhecimento torne-se cada vez mais difícil à medida que o vocabulário aumenta.

O primeiro passo para utilização de um sistema de reconhecimento de vocabulário flexível é a definição do vocabulário. Neste vocabulário, o modelo de cada palavra é obtido pela concatenção dos modelos das sub-unidades, obtidos após a transcrição ortográfica das palavras.

O arquivo de vocabulário requerido pelo sistema é dividido em duas partes. A primeira parte lista as sub-unidades fonéticas utilizadas na transcrição das palavras (as mesmas utilizadas no treinamento). A segunda parte lista a descrição das palavras do vocabulário. Esta descrição consta da transcrição fonética de cada palavra, da informação sobre sua duração (média e variância, respectivamente) e da classe à qual esta palavra pertence. Entretanto, como no caso de reconhecimento de palavras isoladas não há necessidade do uso do modelo de gramática, a classe no vocabulário adotado é definida como geral.

Nos testes com independência do locutor, sempre haverá casos em que as durações das palavras serão significativamente distantes daquelas armazenadas no modelo de duração. Isto pode fazer com que o reconhecimento seja prejudicado, sendo necessário algum procedimento de adaptação que minimize este problema. Por este motivo, optamos em não utilizar o modelo de duração de palavras, inserindo valor "zero" sobre a média e variância da duração de cada palavra do vocabulário. Um exemplo do arquivo de vocabulário utilizado é mostrado na Figura 5.7, e o arquivo contendo toda a listagem pode ser observado no Apêndice C.

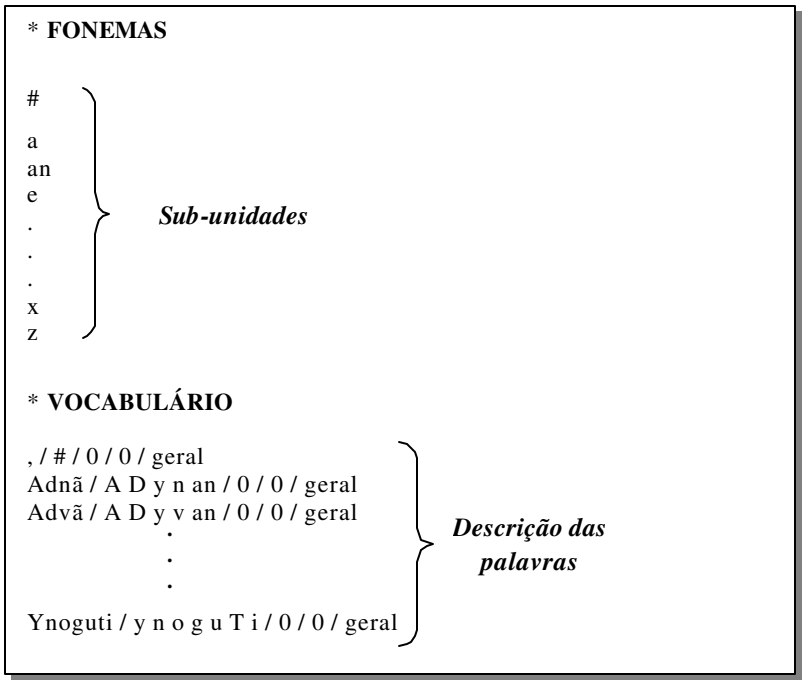


Figura 5.7 : Exemplo do arquivo de vocabulário utilizado no reconhecimento

5.5 – AVALIAÇÃO DO SISTEMA

Visando a obtenção de uma taxa de erros aceitável, realizamos vários testes com alguns parâmetros usualmente utilizados no reconhecimento (mel, dmel, ddmel, energ, denerg[1]), nas mais variadas possibilidades de utilização. Nestes testes foram utilizadas 300 locuções, pronunciadas por 9 locutores masculinos e 6 femininos, com cada um deles pronunciando 20 nomes distintos.

A seguir apresentaremos os resultados iniciais, que serão utilizados como referência na normalização de locutor, tanto para os HMMs discretos quanto para os HMMs contínuos.

[1] Os termos mel, dmel, ddmel, energ e denerg referem-se aos parâmetros mel-cepstrais, delta-mel-cepstrais, delta-delta-mel-cepstrais, energia e delta de energia, respectivamente.

5.5.1 – HMM Discreto

Após uma exaustiva análise das possibilidades de combinação dos parâmetros acústicos, obtivemos os seguintes resultados:

Parâmetro	% Erro
mel	33,81
mel, dmel	23,74
mel, dmel, ddmel	17,33
mel, dmel, ddmel, energ	21,58
mel, dmel, ddmel, energ, denerg	38,12

Tabela 5.1: Taxa de erros obtida para diferentes conjuntos de parâmetros (HMM Discreto)

Como observado na Tabela 5.1, o conjunto de parâmetros que apresentou o melhor resultado foi o mel, dmel e ddmel. No caso dos parâmetros dmel e ddmel utilizou-se K = 1. Foram feitos ainda testes adicionais com outros valores de K. No primeiro teste utilizou-se dois quadros de variação para o dmel (dm2) e um quadro para o ddmel (ddm1). No segundo teste, utilizou-se dois quadros para o dmel (dm2) e dois quadros para o ddmel (ddm2), obtendo-se os seguintes resultados:

HMM utilizando	% Erro
mdm1ddml	17,33
mdm2ddm1	18,06
mdm2ddm2	17,72

Tabela 5.2: Tabela representativa da variação de K para os parâmetros dmel e ddmel

Com os resultados observados nas Tabelas 5.1 e 5.2, e a partir das especificações dadas a seguir, obteve-se o modelamento final para o HMM discreto.

Especificações do sistema utilizado:

- Locuções para o treinamento: 2000 frases.
- Parâmetros utilizados: mel, dmel e ddmel (12 coeficientes cada).
- K: igual a 1 para o dmel e ddmel.
- Sub-unidades fonéticas utilizadas: fones independentes do contexto.
- Tipo de HMM: Discreto.
- Codebook : Foram obtidos 3 codebooks, um para cada parâmetro, contendo cada um 256 vetores. Cada um dos conjuntos de parâmetros foi quantizado, separadamente, com seu respectivo codebook e, a partir dos parâmetros quantizados, realizou-se o treinamento do sistema (utilizando-se todos os parâmetros).
- Algoritmo de Treinamento: Baum-Welch.
- Vocabulário de reconhecimento: 300 nomes (pronunciados por 9 homens e 6 mulheres).
- Algoritmo de reconhecimento: Level Building
- Avaliação do sistema: 17,33 % de erro.

5.5.2 – HMM Contínuo

Da mesma forma que no HMM discreto, definiu-se inicialmente os melhores parâmetros a serem utilizados pelo sistema, assim como as possibilidades de variação dos parâmetros dmel e ddmel.

Parâmetro utilizado (mistura de 3 gaussianas)	**% Erro**
mel	36,33
mel, dmel	24,04
mel, dmel, ddmel	17,00

Tabela 5.3: Taxa de erros para diferentes conjuntos de parâmetros (HMM Contínuo)

Parâmetro utilizado (mistura de 3 gaussianas)	**% Erro**
mdm1ddm1	17,00
mdm2ddm2	20,67

Tabela 5.4: Tabela representativa da variação de K para os parâmetros dmel e ddmel

Como podemos observar nas tabelas acima, o conjunto de parâmetros que produziu a menor taxa de erros foi o mel, dmel e ddmel, utilizando-se K = 1 tanto para o dmel quanto para o ddmel (como no HMM discreto). Entretanto, além da obtenção dos melhores parâmetros que caracterizam o sinal de fala, quando se utiliza HMM contínuo deve-se também ter a preocupação de se obter o número de misturas de gaussianas que melhor modele o sistema. Esta quantidade, por sua vez, dependerá da variedade de material utilizado no treinamento destes modelos, pois quanto maior for a base de dados de treinamento, maior poderá ser a quantidade de gaussianas por estado utilizadas no modelamento das unidades adotadas. Desta forma, variando-se o número de gaussinas, obtivemos os seguintes resultados:

Número de Gaussianas	% Erro
2 gaussianas	19,67
3 gaussianas	17,00
5 gaussianas	14,67
6 gaussianas	12,12
7 gaussianas	12,66
8 gaussianas	13,00

Tabela 5.5: Taxa de erros obtida variando-se o número de gaussianas

Neste momento, com todos os parâmetros definidos, e utilizando-se as especificações descritas a seguir, obtivemos o modelamento final para o HMM contínuo.

Especificações do sistema utilizado:

- Locuções para o treinamento: 2000 frases.
- Parâmetros utilizados: mel, dmel e ddmel (12 coeficientes cada).
- K: igual a 1 para o dmel e ddmel.
- Sub-unidades fonéticas utilizadas: fones independentes do contexto.
- Tipo de HMM: Contínuo.
- Número de gaussianas por estado e por parâmetro (densidades independentes): 3.
- Algoritmo de treinamento: Baum-Welch.
- Vocabulário de reconhecimento: 300 nomes (pronunciados por 9 homens e 6 mulheres).
- Algoritmo de reconhecimento: One-Step.
- Avaliação do sistema: 17,00 % de erro.

Deve-se mencionar, que apesar do desempenho do sistema ter sido melhor quando utilizou-se 6 gaussianas, optamos por utilizar 3 gaussinas nos primeiros testes exploratórios para normalização de locutor. Este número de gaussianas foi escolhido pois, como devem ser

realizados vários testes até que se encontre a melhor forma de normalizar o sistema, uma quantidade pequena de gaussianas requer menor tempo de treinamento. Entretanto, ao finalizarmos os testes para verificação da melhor forma de normalização de locutor, passaremos a utilizar 5 gaussianas.

5.5.3 – HMM Discreto com Separação dos Locutores por Sexo

Uma forma prática de verificarmos a importância da normalização de locutor nos sistemas de reconhecimento de fala, é analisarmos separadamente os locutores masculinos e femininos, de modo a diminuir a variabilidade do conjunto de treinamento [19]. Esta idéia decorre do fato do trato vocal dos locutores masculinos ser, em média, maior que o trato vocal dos locutores femininos, causando um deslocamento das formantes do sinal de fala para freqüências menores. Esta idéia foi testada utilizando-se HMM discreto.

As especificações utilizadas nos dois conjuntos de locutores foram as seguintes:

- Tipo de HMM: Discreto.
- Codebook : Inicialmente separou-se os locutores por sexo (feminino e masculino). Em seguida gerou-se um codebook para os locutores masculinos e outro para os locutores femininos (um codebook para o parâmetro mel, outro para o dmel e outro para o ddmel).
- Algoritmo de treinamento: Baum-Welch .
- Algoritmo de reconhecimento: Level Building
- Conjunto de treinamento:
    - Treinamento realizados somente com locutores masculinos: 1053 frases (23 locutores).
    - Treinamento realizados somente com locutores femininos: 947 frases (19 locutores).
- Vocabulário de reconhecimento:

- Somente locuções pronunciadas por locutores masculinos: 180 nomes (9 locutores)
- Somente locuções pronunciadas por locutores femininos: 120 nomes (6 locutores)

- Avaliação do sistema:

Ao treinarmos e reconhecermos cada conjunto de locutores, dentro de seu universo, obtivemos os seguintes resultados:

	Conj. de treinamento contendo locutores **Femininos**	Conj. de treinamento contendo locutores **Masculinos**
Reconhecimento de fala feminina	5,08 %	-------
Reconhecimento de fala masculina	-------	21 %

Tabela 5.6: Taxa de erros obtida utilizando-se diferentes conjuntos de treinamento

Como a taxa de erros do sistema inicial era de 17%, verificamos que a separação dos locutores quanto ao sexo causou uma diminuição considerável da taxa de erros para o grupo feminino (de 17% para 5,08%) e um aumento da taxa de erros para o grupo masculino (de 17% para 21%). O efeito positivo observado no grupo feminino reforça a idéia de que a variabilidade da fala entre locutores pode ser minimizada normalizando-se suas características espectrais. Já o efeito negativo observado no grupo masculino carece de uma justificativa convincente. Uma hipótese especulatória é a presença de alguns locutores masculinos com características espectrais próximas à feminina e que, devido à remoção dos locutores femininos da base de treinamento, tiveram suas características mal treinadas e assimiladas pelos HMMs, prejudicando assim o reconhecimento do sistema.

6 Normalização de Locutor

6.1 – INTRODUÇÃO

A Normalização de Comprimento do Trato Vocal ou ainda Normalização de Locutor, tem por objetivo tentar normalizar as representações paramétricas do sinal, de modo a reduzir os efeitos causados pela variabilidade da fala entre diferentes locutores.

O trato vocal, também denominado tubo acústico, é uma característica intrínseca do locutor. Ele é ajustado a todo momento para intensificar determinadas faixas de freqüências pela ressonância do som em seu interior. Esta ação de ajuste é denominada articulação e é responsável pela geração dos diferentes sinais acústicos que compõem o sinal de fala [8]. Assim, o trato vocal funciona como um filtro não estacionário, com diversas formas de ressonância que resultarão numa modificação do espectro do sinal fonte (ruído ou trem de pulsos periódicos), gerando o sinal de fala.

O tubo acústico possui diferentes formas e comprimentos para cada pessoa, resultando locuções com diferentes características acústicas. Na tentativa de minimizar esta variabilidade entre os locutores, uma das principais responsáveis pela degradação de desempenho dos sistemas de reconhecimento de fala, analisaremos ao longo deste capítulo o processo de normalização de locutor.

6.2 – DISTORÇÃO DO EIXO DE FREQÜÊNCIAS

A idéia de distorção na escala de freqüência, para normalização de diferentes tratos vocais, não é nova. Muitos trabalhos têm apresentado experimentos utilizando este método, como [5,11 e 21] que propõem uma extensão do trabalho de Andreou *et al* [2].

É intuitivo pensar que a distorção do eixo das freqüências deva ser realizada pela reamostragem do sinal no domínio do tempo, sendo este sinal em seguida passado através de um banco de filtros fixo. Entretanto, segundo [11], tem-se maior eficiência do método quando o processo de distorção é realizado diretamente no banco de filtros, na escala Mel, a partir do banco de filtros proposto por Davis-Mermelstein [6]. Este processo de variação do banco de filtros pode ser observado na Figura 6.1.

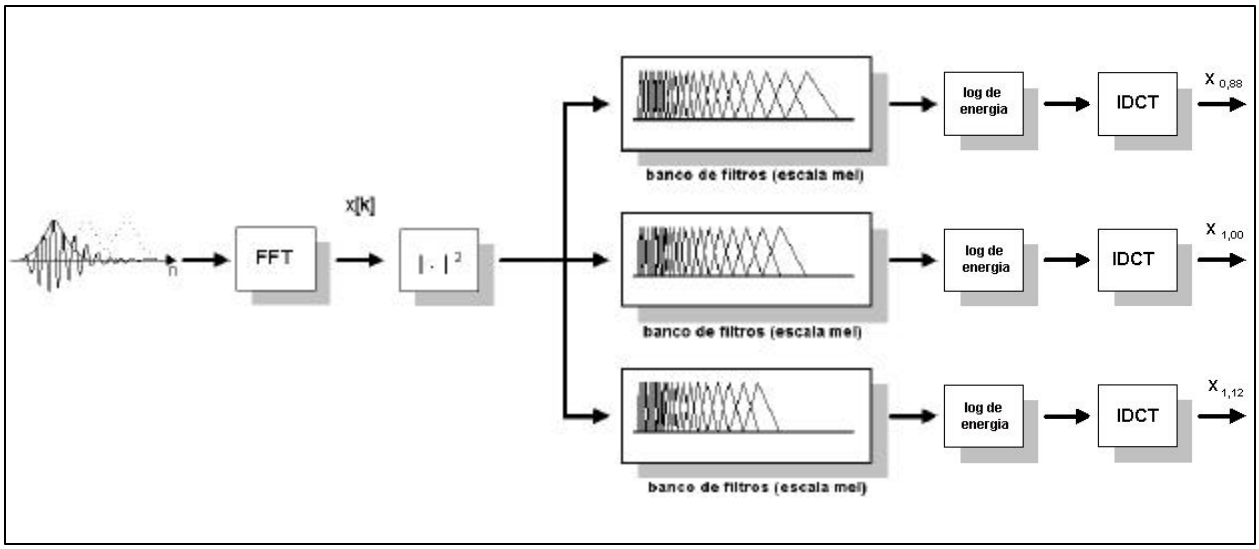


Figura 6.1: Sinal de fala com diferentes escalonamentos

No método proposto por Lee e Rose [11], observado na Figura 6.1, o banco de filtros é escalonado, no eixo das freqüências, por um fator de distorção apropriadamente calculado (estimado). Assim, com uma única FFT executada em cada segmento do sinal de fala analisado, pode-se empregar diferentes fatores de distorção (normalização) diretamente no banco de filtros, resultando em diferentes conjuntos de coeficientes mel-cepstrais, um para cada fator de distorção.

6.2.1 – Estimação do Fator de Distorção

O fator de distorção (α), utilizado na normalização de locutor, tem como função escalonar um determinado banco de filtros, a fim de compensar a variação do comprimento do trato vocal de um locutor em relação a um trato vocal de comprimento médio.

Este fator de distorção é conceitualmente representado pela razão entre o comprimento do trato vocal do locutor sendo analisado e o comprimento do trato vocal utilizado como referência, onde uma estimação segura deste comprimento deve ocorrer baseada em dados acústicos [11]. Entretanto, a estimação do comprimento do trato vocal baseada apenas em dados acústicos é difícil de ser obtida, uma vez que diferentes locutores possuem diferentes comprimentos de trato vocal, que podem ser modificados de acordo com o som produzido. Por este motivo, segundo [11], a melhor forma de se obter o fator responsável pelo escalonamento do banco de filtros é escolhê-lo de modo que a probabilidade de um conjunto de características acústicas, de um determinado locutor, seja maximizada em relação a um dado modelo acústico tomado como referência.

Embora o movimento dos lábios, entre outras variações, modifique o comprimento do trato vocal dos locutores, assume-se que este tipo de variação é similar entre os locutores e não tem efeito significativo na estimação do fator de distorção. Desta forma, o fator de distorção pode ser estimado, para cada locutor, usando-se todas as locuções por ele pronunciadas .

O fator de distorção ótimo, $\hat{\alpha}^{i}$, para cada locutor i, é então obtido pela máxima probabilidade de se obter um conjunto de características acústicas, dado um modelo λ e as transcrições de cada locução referente ao locutor i, ou seja :

$$\hat{\alpha}^{i} = \arg \max_{\alpha} P(X_i^{\alpha} / \lambda, W_i) \qquad (6.1)$$

onde:

$X_i^{á} = \{ X_{i,1}^{á}, X_{i,2}^{á}, ..., X_{i,N}^{á} \}$ representa o conjunto de características acústicas de todas as N locuções do locutor i, escalonadas de α .

$W_i = \{ W_{i,1}, W_{i,2},..., W_{i,N} \}$ representa o conjunto de transcrições de todas as N locuções associadas a um dado locutor i.

λ - representa o HMM treinado por uma grande população de locutores.

A estimação de $\hat{\alpha}$ é difícil de ser obtida pelo fato do fator de distorção corresponder a uma transformação não linear dos parâmetros característicos da fala. O fator de distorção ótimo ($\hat{\alpha}$) é então obtido pela procura entre 13 fatores, uniformemente espaçados entre 0,88 e 1,12 [11]. Esta faixa de valores foi escolhida com o objetivo de considerar uma variação de 25% do comprimento do trato vocal encontrado em um adulto e pelo fato de que uma variação maior destes valores poderia ocasionar uma variação muito grande na freqüência de corte do sinal analisado.

6.3 – NORMALIZAÇÃO DE LOCUTOR

O processo de normalização de locutor é representado pela transformação dos parâmetros acústicos da fala. Esta transformação é realizada por funções de distorção, que tentam resolver a dificuldade de se estimar a posição dos formantes que correspondam à "verdadeira" forma do trato vocal, para cada locutor, e então compensar estas diferenças.

A compensação do trato vocal é obtida pela estimação do melhor fator que escalonará o banco de filtros, de forma que se tenha um aumento na probabilidade do modelo acústico escalonado em relação ao modelo tomado como referência.

O banco de filtros na escala Mel, também chamado de banco de filtros de Davis-Mermelstein, que será escalonado apropriadamente, pode ser observado na Tabela 6.1.

Índice dos Filtros	Freq. Central (Hz)	Índice dos Filtros	Freq. Central (Hz)
1	100	13	1516
2	200	14	1741
3	300	15	2000
4	400	16	2297
5	500	17	2639
6	600	18	3031
7	700	19	3482
8	800	20	4000
9	900	21	4595
10	1000	22	5278
11	1149	23	6063
12	1320	24	6964

Tabela 6.1: Tabela contendo a freqüência central para o banco de filtros de Davis-Mermelstein

Na primeira coluna da Tabela 6.1 tem-se o índice dos filtros. O número de filtros utilizado varia com a freqüência de amostragem do sinal. Por exemplo, quando utiliza-se uma freqüência de amostragem de 8 kHz, tem-se uma freqüência máxima de fs/2 = 4 kHz e que corresponde, conforme a Tabela 6.1, à utilização de 19 filtros. Deve-se ressaltar que as freqüências centrais dos filtros variam linearmente até a freqüência de 1 kHz e, a partir daí, crescem exponencialmente com um fator de $2^{1/5}$. Neste trabalho utilizaremos uma freqüência de amostragem de 11,025 kHz (fs/2 = 5,512 kHz), impondo-se assim a utilização de 21 filtros.

Uma forma de implementação do banco de filtros, na escala Mel, é através de um conjunto de filtros passa-faixa triangulares, onde o k-ésimo filtro (k correspondente ao índice do filtro) é construído na freqüência central f (k), com sua faixa se estendendo da freqüência f (k –1) à freqüência f (k+ 1), de modo que f (k+1) não ultrapassasse fs/2.

Como exemplo de um banco de filtros triangulares, temos a Figura 6.2. Este conjunto de filtros será utilizado como referência na normalização dos locutores.

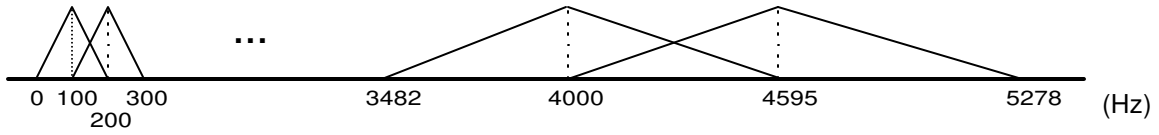


Figura 6.2: Representação de um banco de filtros triangulares na escala Mel

6.3.1 – Transformação dos Parâmetros

O processo de transformação dos parâmetros acústicos, ou ainda de normalização do trato vocal, é obtido pelo escalonamento das freqüências centrais do banco de filtros mel. O escalonamento destas freqüências é realizado, linearmente, por um fator de distorção (α) compreendido entre 0,88 e 1,12 (variando de 0,02), como mostrado na Equação 6.2. No final deste processo, tem-se um novo banco de filtros, com freqüências escalonadas. Dependendo do fator de escalonamento utilizado, estas freqüências ora serão expandidas ($\alpha < 1$), ora serão comprimidas ($\alpha > 1$):

$$f' = \beta . f \tag{6.2}$$

f – representa a freqüência original na escala Mel.

α - representa o fator de distorção.

$\beta = \frac{1}{\alpha}$ - representa o fator de escalonamento em freqüência (para α variando entre 0,88 e 1,12).

f' - representa a freqüência escalonada.

As tabelas 6.1a e b mostram os valores de $\dot{f}$, para os 13 valores de α, obtidos a partir da Equação 6.2, e que serão utilizados posteriormente na obtenção do melhor fator que caracterizará os parâmetros acústicos de cada locutor analisado.

	Freqüência Central Normalizada (f')						
	0,88	0,90	0,92	0,94	0,96	0,98	1,00
Índice dos Filtros							
1	113,64	111,11	108,70	106,38	104,17	102,04	**100**
2	227,27	222,22	217,39	212,77	208,33	204,08	**200**
3	340,91	333,33	326,09	319,15	312,50	306,12	**300**
4	454,55	444,44	434,78	425,53	416,67	408,16	**400**
5	568,18	555,55	543,48	531,91	520,83	510,20	**500**
6	681,82	666,67	652,17	638,30	625,00	612,24	**600**
7	795,45	777,78	760,87	744,68	729,17	714,29	**700**
8	909,09	888,89	869,57	851,06	833,33	816,33	**800**
9	1022,73	1000,00	978,26	957,45	937,50	918,37	**900**
10	1136,36	1111,11	1086,96	1063,87	1041,67	1020,41	**1000**
11	1305,68	1276,67	1248,91	1222,34	1196,88	1172,45	**1149**
12	1500,00	1466,67	1434,78	1404,26	1375,00	1346,94	**1320**
13	1722,73	1684,44	1647,83	1612,77	1579,17	1546,94	**1516**
14	1978,41	1934,44	1892,39	1852,13	1813,54	1776,53	**1741**
15	2272,73	2222,22	2173,91	2127,66	2083,33	2040,82	**2000**
16	2610,23	2552,22	2496,74	2443,62	2392,71	2343,88	**2297**
17	2998,86	2932,22	2868,48	2807,45	2748,96	2692,86	**2639**
18	3444,32	3367,78	3294,57	3224,47	3157,29	3092,86	**3031**
19	3956,82	3868,89	3784,78	3704,26	3627,08	3553,06	**3482**
20	4545,45	4444,44	4347,83	4255,32	4166,67	4081,63	**4000**
21	5221,59	5105,56	4994,57	4888,30	4786,46	4688,78	**4595**
22	5997,73	5864,44	5736,96	5614,89	5497,92	5385,71	**5278**
23	6889,77	6736,67	6590,22	6450,00	6315,62	6186,73	**6063**
24	7913,64	7737,78	7569,57	7408,51	7254,17	7106,12	**6964**

(a)

	Freqüência Central Normalizada (f')						
	1,00	1,02	1,04	1,06	1,08	1,10	1,12
Índice dos Filtros							
1	**100**	98,04	96,15	94,34	92,59	90,91	89,29
2	**200**	196,08	192,31	188,68	185,19	181,82	178,57
3	**300**	294,12	288,46	283,02	277,78	272,73	267,86
4	**400**	392,16	384,62	377,36	370,37	363,64	357,14
5	**500**	490,20	480,77	471,70	462,96	454,55	446,43
6	**600**	588,24	576,92	566,04	555,56	545,45	535,71
7	**700**	686,27	673,08	660,38	648,15	636,36	625,00
8	**800**	784,31	769,23	754,72	740,74	727,27	714,29
9	**900**	882,35	865,38	849,06	833,33	818,18	803,57
10	**1000**	980,39	961,54	943,40	925,93	909,09	892,86
11	**1149**	1126,47	1104,81	1083,96	1063,89	1044,55	1025,89
12	**1320**	1294,12	1269,23	1245,28	1222,22	1200,00	1178,57
13	**1516**	1486,27	1457,69	1430,19	1403,70	1378,18	1353,57
14	**1741**	1706,86	1674,04	1642,45	1612,04	1582,73	1554,46
15	**2000**	1960,78	1923,08	1886,79	1851,85	1818,18	1785,71
16	**2297**	2251,96	2208,65	2166,98	2126,85	2088,18	2050,89
17	**2639**	2587,25	2537,50	2489,62	2443,52	2399,09	2356,25
18	**3031**	2971,57	2914,42	2859,43	2806,48	2755,45	2706,25
19	**3482**	3413,73	3348,08	3284,91	3224,07	3165,45	3108,93
20	**4000**	3921,57	3846,15	3773,58	3703,70	3636,36	3571,43
21	**4595**	4504,90	4418,27	4334,91	4254,63	4177,27	4102,68
22	**5278**	5174,51	5075,00	4979,25	4887,04	4798,18	4712,50
23	**6063**	5944,12	5829,81	5719,81	5613,89	5511,82	5413,39
24	**6964**	6827,45	6696,15	6569,81	6330,91	6330,91	6217,86

(b)

Tabela 6.2 : Tabela contendo os valores da freqüência central, em Hz, para $0,88 \leq \alpha \leq 1,00$ (a) e para $1,00 \leq \alpha \leq 1,12$ (b)

Em testes preliminares, feitos por [11] e [21], verificou-se que com a normalização do locutor, na média, o fator 1,00 caracterizava a fala masculina, o fator 0,94 a fala feminina e o fator 0,975 caracterizava todos os locutores. A princípio, o fato de, na média, o fator de distorção para os homens ser maior do que para as mulheres é justificável, pois as mulheres tendem a ter comprimento do trato vocal mais curto que os homens, e conseqüentemente, freqüências formantes mais altas. Entretanto, o fato do fator médio obtido para todos os locutores ser diferente de 1,00, é algo inesperado, uma vez que o processo iterativo de treinamento inicial é constituído pelo modelo de todos os locutores não escalonados (α=1). Uma explicação para este fato, pode estar na diferença entre as larguras de banda dos filtros, entre os diferentes locutores (para diferentes valores de α), cujo eixo de freqüências ora é comprimido, ora é expandido.

6.4 – CONSIDERAÇÕES

Quando o eixo de freqüências é escalonado, a largura de faixa do sinal resultante difere da largura de faixa do sinal original. Esta variação, por sua vez, resulta em diferenças entre os parâmetros acústicos do sinal escalonado e do sinal original.

A diferença entre o sinal escalonado e o sinal original pode ser observada no momento em que as freqüências centrais $\dot{f}$ são obtidas, conforme as Tabelas 6.1 a e b. Neste instante, o limite de máxima largura de faixa do sinal, que era de fs/2 (metade da freqüência de amostragem), passa a assumir valores ora inferiores, ora superiores a este limite.

Para o experimento descrito neste trabalho, a taxa de amostragem utilizada é de 11,025 kHz, o que impõe um limite de faixa de 5,512 kHz e um número de filtros igual a 21. Porém, com o escalonamento da freqüência através do fator α, obtém-se valores entre 5,99773 kHz e 4,7125 kHz, quando utiliza-se α = 0,88 e α = 1,12, respectivamente.

Como podemos observar, estes valores ora serão maiores que fs/2 (freqüências expandidas), ora serão menores que fs/2 (freqüências comprimidas). Desta forma, ao

variarmos o fator de distorção α, estaremos provocando uma variação na freqüência de corte do sinal.

O efeito de compressão e expansão, ocasionado pela variação de α, por sua vez, provoca uma alteração na informação útil que será utilizada pelo sistema. Com o efeito de expansão, incluímos amostras adicionais do espectro de freqüências, ou seja, incluímos informação fora da banda normalmente utilizada. No caso da compressão, amostras do espectro de freqüências são descartadas, perdendo-se assim informação que poderia ser útil para o reconhecimento.

Uma solução para este problema, utilizada por [11], na tentativa de suavizar a variação entre a largura de faixa do sinal original e do sinal normalizado, é considerar uma função de escalonamento não linear, de tal forma que a largura de faixa do sinal escalonado seja a mais próxima do sinal original.

Como exemplo da função não linear, utilizada por [11], temos:

$$G(f) = \begin{cases} \beta . f & 0 \le f \le f_0 \\ \dfrac{f_{max} - \beta . f_0}{f_{max} - f_0}(f - f_0) + \beta . f_0 & f_0 \le f \le f_{máx} \end{cases} \qquad (6.3)$$

onde:

$\beta = 1/\alpha$ - para α compreendido entre 0,88 e 1,12.

f – representa a freqüência antes da normalização.

f_0 – representa a freqüência escolhida empiricamente, de valor acima da mais alta formante significativa [11].

$f_{máx}$ – representa a máxima largura de faixa (fs/2) do sinal original.

G(f) - representa a freqüência após a normalização.

Experimentos preliminares realizados por [11] usam tal função de escalonamento, para normalização do locutor, sugerindo que esta função possa ser mais robusta que um simples escalonamento linear de freqüências.

Neste trabalho, o valor de f_0 foi escolhido como sendo a máxima freqüência central utilizada (f_0 = 4,595 kHz). Conforme as Tabelas 6.1a e b, este é o valor de máxima freqüência, para o qual, após a normalização de freqüência (f'= β.f), não se terá valores de freqüência superiores a fs/2. O valor de $f_{máx}$ é obtido pela máxima freqüência que o sinal pode atingir (fs/2). Entretanto, ao utilizarmos como referência o banco de filtros Mel, a máxima freqüência que o sinal pode atingir é de 5,278 kHz.

Na normalização de locutor, realizada neste trabalho, os parâmetros mel-cepstrais são calculados a partir do banco de filtros escalonados (Tabelas 6.1a e b), levando-se em consideração a Equação 6.3. A Figura 6.3 mostra a função G(f) utilizada no escalonamento do banco de filtros.

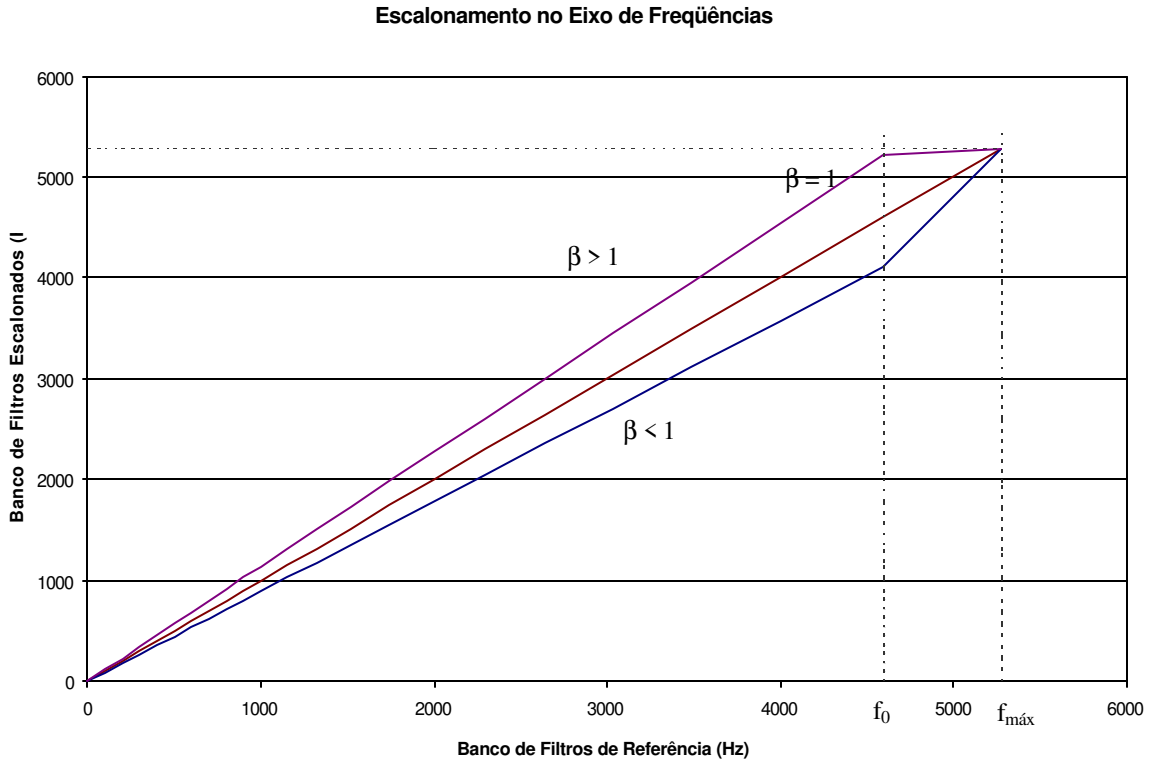


Figura 6.3 : Representação da função não linear usada no escalonamento do banco de filtros, na escala mel, segundo a função G(f)

Como podemos observar na Figura 6.3, para $\beta > 1$ tem-se a expansão do banco de filtros. Similarmente, para os valores de $\beta < 1$ tem-se a compressão do banco de filtros. Para uma melhor visualização de G(f) tem-se a Figura 6.4a que apresenta o banco de filtros de referência, sem escalonamento ($\beta = 1$), e as Figuras 6.4b e 6.4c, que mostram, respectivamente, a compressão e a expansão observada no banco de filtros.

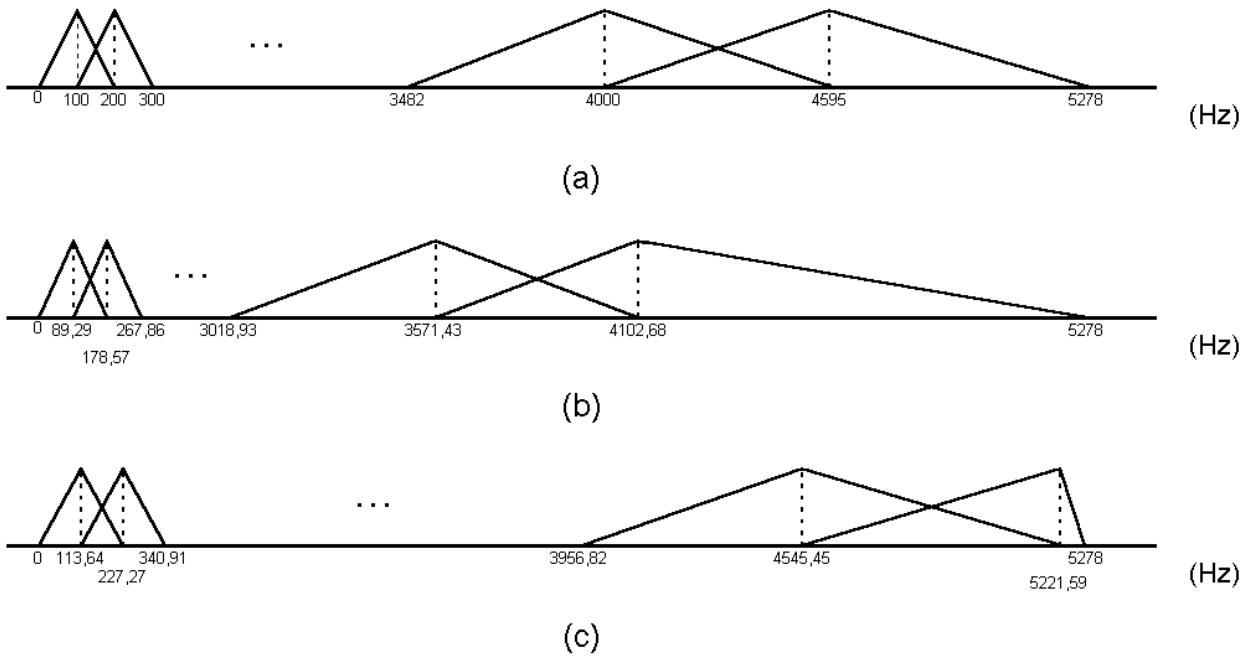


Figura 6.4: Figura representativa do banco de filtros referência (a), banco de filtros comprimido (b) e banco de filtros expandido (c)

7 Treinamento e Avaliação de HMM Utilizando Normalização de Locutor

7.1 – INTRODUÇÃO

Neste capítulo são apresentados os procedimentos utilizados na adaptação do sistema de reconhecimento de fala às características espectrais dos locutores.

O treinamento e avaliação deste sistema são realizados para o HMM contínuo. Inicialmente a idéia era trabalhar com HMM discreto, normalmente utilizado pela simplicidade de seu modelamento e por não necessitar de uma grande quantidade de dados no treinamento de seus modelos. Entretanto, como será discutido posteriormente, o resultado obtido com este modelo não foi o esperado, fazendo com que tivéssemos que redirecionar o trabalho para a utilização do HMM contínuo.

7.2 – NORMALIZAÇÃO DAS CARACTERÍSTICAS ESPECTRAIS DOS LOCUTORES

A normalização das características espectrais dos locutores nos sistemas de reconhecimento de fala é dividida em duas etapas, aplicadas iterativamente: escolha do fator de distorção (α) que melhor caracterize cada locutor analisado e o treinamento dos HMMs utilizando o $\alpha_{ótimo}$ obtido para cada locutor.

É importante ressaltar que, antes da escolha do melhor α e do treinamento dos HMMs, as locuções de treinamento devem ser separadas por locutor. Cada locutor será representado por um conjunto de locuções (frases ou palavras isoladas).

7.2.1 – Escolha do melhor α

Uma vez definidos os locutores, e suas respectivas locuções, é feita a escolha do melhor α, para cada locutor. O melhor α será aquele que, ao escalonarmos o banco de filtros, na escala Mel, como apresentado na seção 6.3, proporcionará a maior verossimilhança média nas suas locuções.

Na escolha do melhor fator de distorção são levados em consideração: todas as locuções do locutor sendo analisado, a transcrição fonética das locuções associadas a este locutor, as sub-unidades adotadas (fones) e o modelo HMM utilizado como referência, como ilustrado na Figura 7.1:

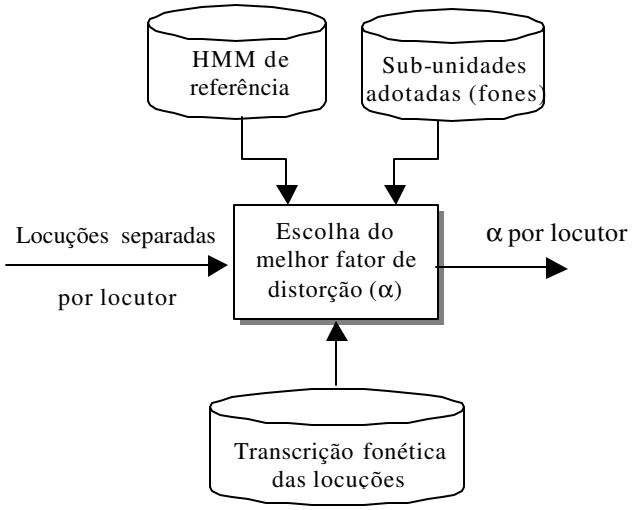


Figura 7.1: Figura representativa da escolha do melhor fator de distorção para cada locutor

O algoritmo utilizado no cálculo do melhor α é apresentado a seguir:

Inicialização:

As matrizes de transição e emissão, utilizadas no modelamento dos HMMs, são inicializadas com os valores obtidos no primeiro treinamento do sistema, sem normalização ($\alpha = 1$). O HMM obtido deste primeiro treinamento será chamado de HMM pré-treinado.

Recursão:

Inicialmente são definidas as seguintes variáveis:

α - fator de distorção ($0{,}88 \leq \alpha \leq 1{,}12$).

W_i – conjunto de transcrições referentes ao locutor i.

λ - modelo HMM pré-treinado.

X_i^{α} - conjunto de características espectrais, obtidas após escalonamento do banco de filtros por um fator α, para o locutor i.

max $[P(X_i^{\alpha} / \lambda, W_i)]$ – representa a máxima probabilidade de se obter um conjunto de observação X_i, escalonado de α, dado um modelo λ e um conjunto de transcrições W_i.

Para cada locutor i, faz-se:

- Variar α em intervalos de 0,02.
- Calcular a max $[P(X_i^{\alpha} / \lambda, W_i)]$ entre os 13 valores de α.
- Armazenar o valor de α que proporcionou a max $[P(X_i^{\alpha} / \lambda, W_i)]$, associando-o a seu respectivo locutor.

O algoritmo utilizado no cálculo da máxima verossimilhança foi o algoritmo de Viterbi.

Término:

O processo é finalizado quando se obtém, para cada locutor i, o seu respectivo fator de distorção α^i.

Uma melhor visualização do algoritmo pode ser observado na Figura 7.2:

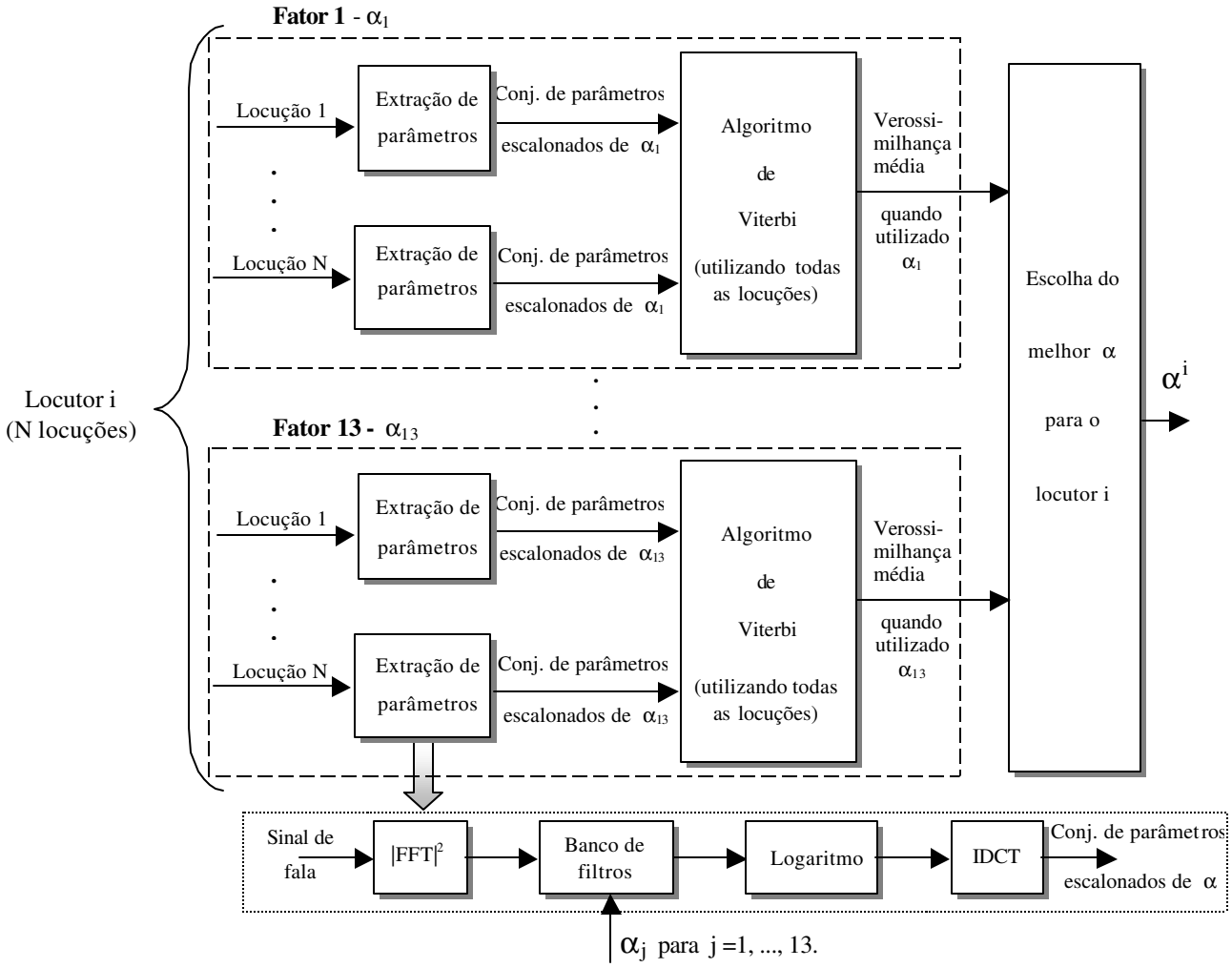


Figura 7.2: Diagrama de blocos utilizado na escolha do melhor α para cada locutor

É importante ressaltar que, quando utiliza-se HMM discreto, os parâmetros obtidos para os 13 valores de α, para cada locução, devem ser quantizados antes de se calcular a verossimilhança média entre as locuções. O codebook utilizado nesta quantização foi o mesmo utilizado no sistema sem normalização (um codebook para cada parâmetro utilizado).

7.2.2 – Treinamento

O processo de treinamento dos modelos HMMs utilizando normalização de locutor é semelhante ao treinamento descrito no Capítulo 5. Entretanto, é diferenciado na sua inicialização (realizada a partir do HMM pré-treinado) e na normalização das características espectrais dos locutores utilizados.

O treinamento dos modelos HMMs utilizando normalização de locutor é realizado a partir das sub-unidades adotadas (fones), das locuções de treinamento separadas por locutor, da transcrição fonética das locuções, dos α's obtidos para cada locutor e do modelo HMM (no primeiro treinamento o HMM corresponde ao HMM pré-treinado obtido quando $\alpha = 1$, sem normalização).

Inicialmente calculam-se os α's ótimos para cada locutor do conjunto de M locutores de treinamento. Em seguida estes α's, associados a seus respectivos locutores, são empregados para calcular um novo conjunto de parâmetros mel-cepstrais que serão utilizados no treinamento do sistema. O treinamento é então executado durante tantas épocas quantas foram necessárias para que a distorção desejada seja atingida (0,001 de diferença relativa entre a época anterior e a época atual). Depois de realizadas todas as épocas de treinamento, tem-se um novo modelo HMM (HMM_{NOVO}). Este novo HMM será utilizado na escolha de novos α's para cada locutor (α_{NOVO}). Em seguida é feito uma comparação entre os valores de α_{NOVO} e os valores de α inicialmente calculados. Caso pelo menos um dos α's, para um mesmo locutor, seja diferente, faz-se a atualização destes α's ($\alpha \Leftarrow \alpha_{NOVO}$), para o cálculo dos novos coeficientes mel-cepstrais, e do HMM ($HMM \Leftarrow HMM_{NOVO}$), para que em seguida seja executado um novo treinamento do sistema. A cada novo treinamento utilizando novos valores de α chamaremos de *Retreinamento*. Desta forma, segundo [11], o sistema deverá continuar sendo retreinado até que o fator de distorção de cada locutor (α^i) não seja mais alterado entre um retreinamento e outro. Obtém-se no final deste processo o $HMM_{NORMALIZADO}$ ($HMM_{NORMALIZADO} \Leftarrow HMM_{NOVO}$). A Figura 7.3 mostra uma representação esquemática deste treinamento.

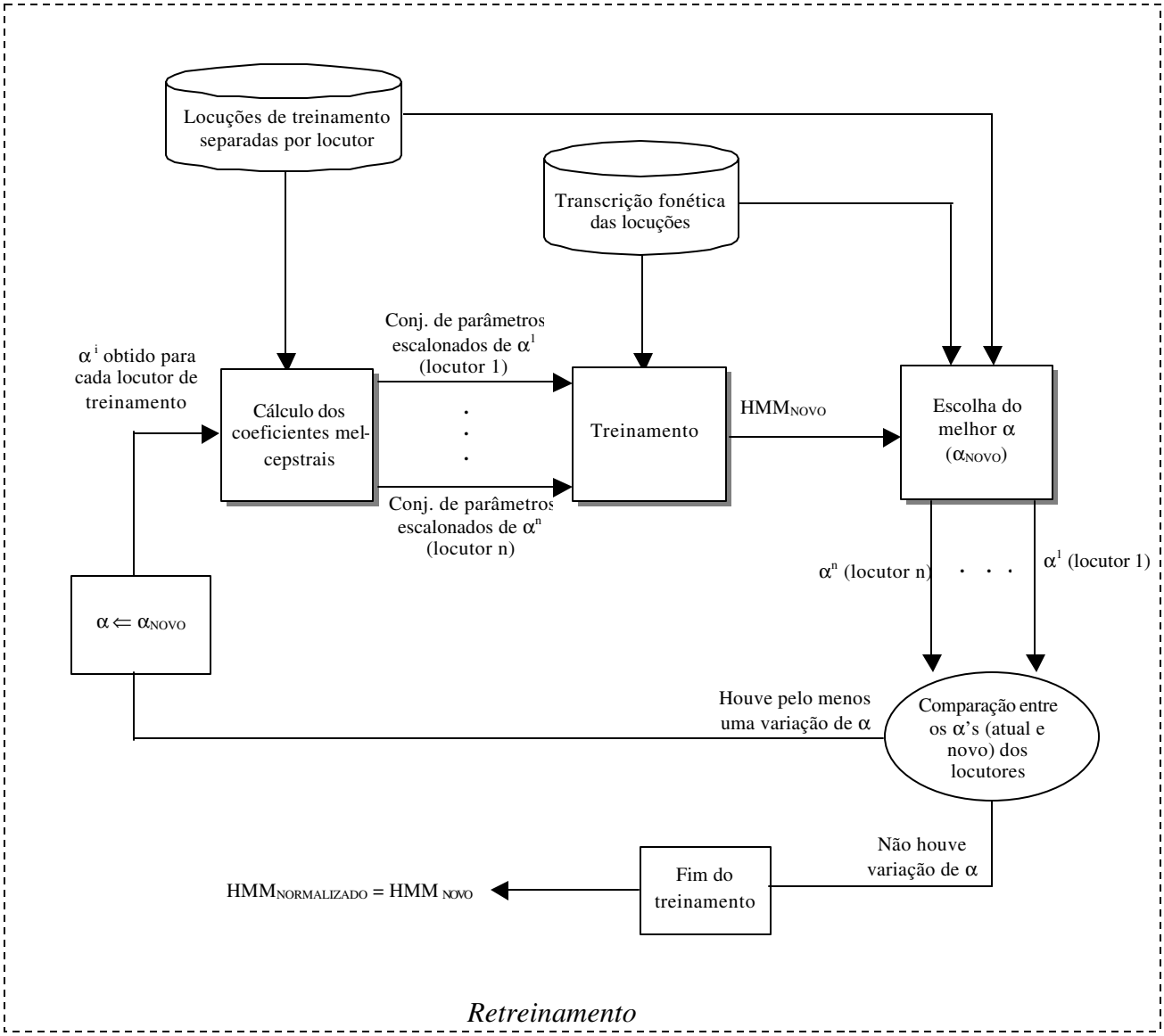


Figura 7.3: Figura representativa do treinamento dos HMMs utilizando normalização de locutor. O treinamento dos modelos é inicializado utilizando-se o HMM pré-treinado (sem normalização de locutor). A obtenção do α^i (para cada locutor i) é realizada antes do treinamento do sistema através do procedimento de escolha do melhor fator de distorção.

7.3 – RECONHECIMENTO

O reconhecimento das locuções é realizado da mesma forma que no sistema sem normalização de locutor. Entretanto, antes de se reconhecer cada locução, deve-se escolher o melhor fator de distorção para cada um dos locutores de teste (locutores utilizados no reconhecimento). A seguir, tem-se a representação deste sistema.

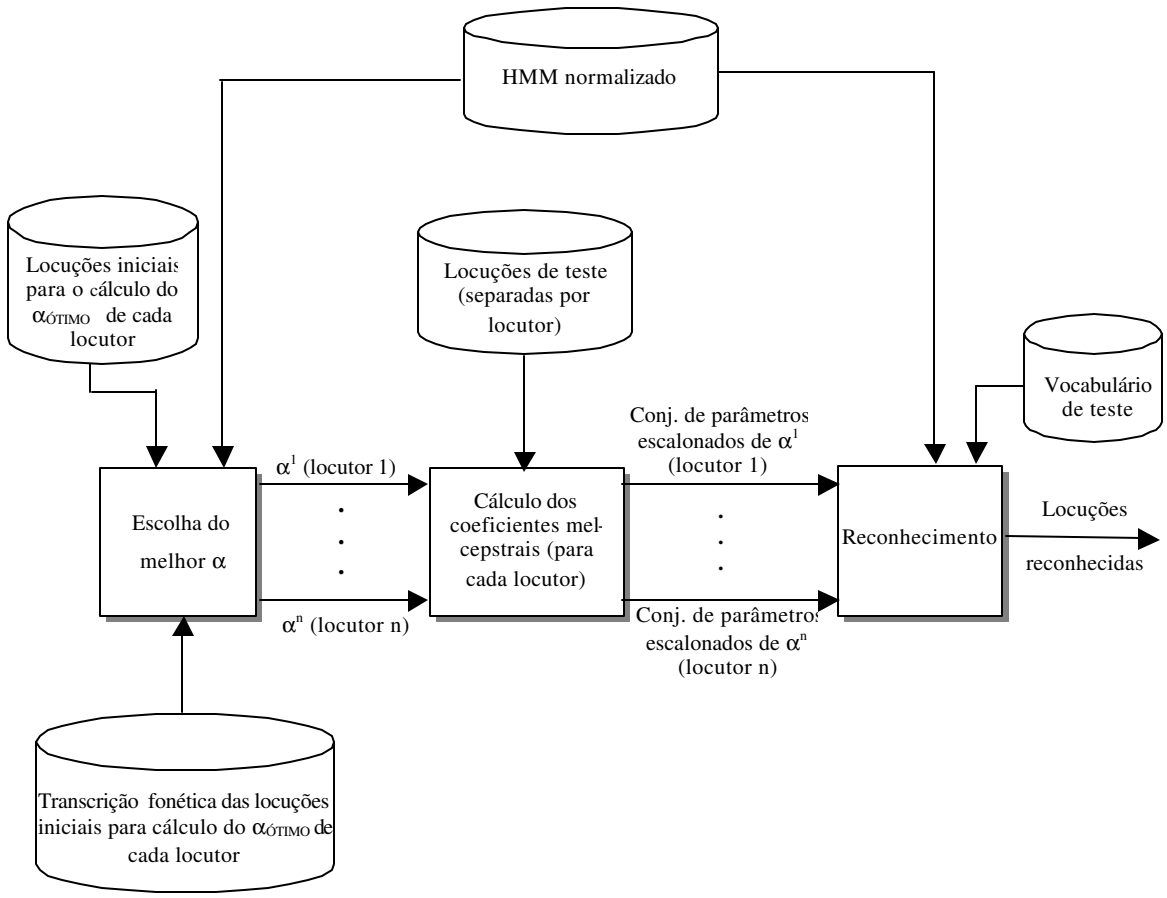


Figura 7.4: Representação do sistema de reconhecimento utilizando normalização de locutor

Como podemos observar na Figura 7.4, o HMM normalizado (HMM obtido após retreinamento do sistema) é utilizado tanto na escolha do melhor α quanto no reconhecimento das locuções.

Deve-se mencionar ainda que, tanto no treinamento quanto no reconhecimento, uma vez obtido o melhor fator de distorção, para cada locutor, este fator é utilizado para todas as demais locuções do respectivo locutor, na obtenção dos parâmetros acústicos de cada locução.

7.4 – AVALIAÇÃO DO PROCEDIMENTO DE ESCOLHA DO $\alpha_{ÓTIMO}$

Como visto anteriormente, para se normalizar as características espectrais das locuções, deve-se, antes do treinamento dos HMMs, obter o valor de α que, no sistema inicialmente treinado com $\alpha = 1$, melhor caracterize os parâmetros acústicos de cada locutor. Para demonstração dos resultados obtidos por este procedimento escolheu-se 4 locutores, dois femininos (F01 e F02) e dois masculinos (M01 e M11).

7.4.1 – Utilizando-se HMM Discreto

Na Tabela 7.1 são apresentadas as verossimilhanças obtidas para cada valor de α, de cada locutor, quando utilizado HMM discreto. A verossimilhança máxima definida para cada locutor é apresentada em negrito.

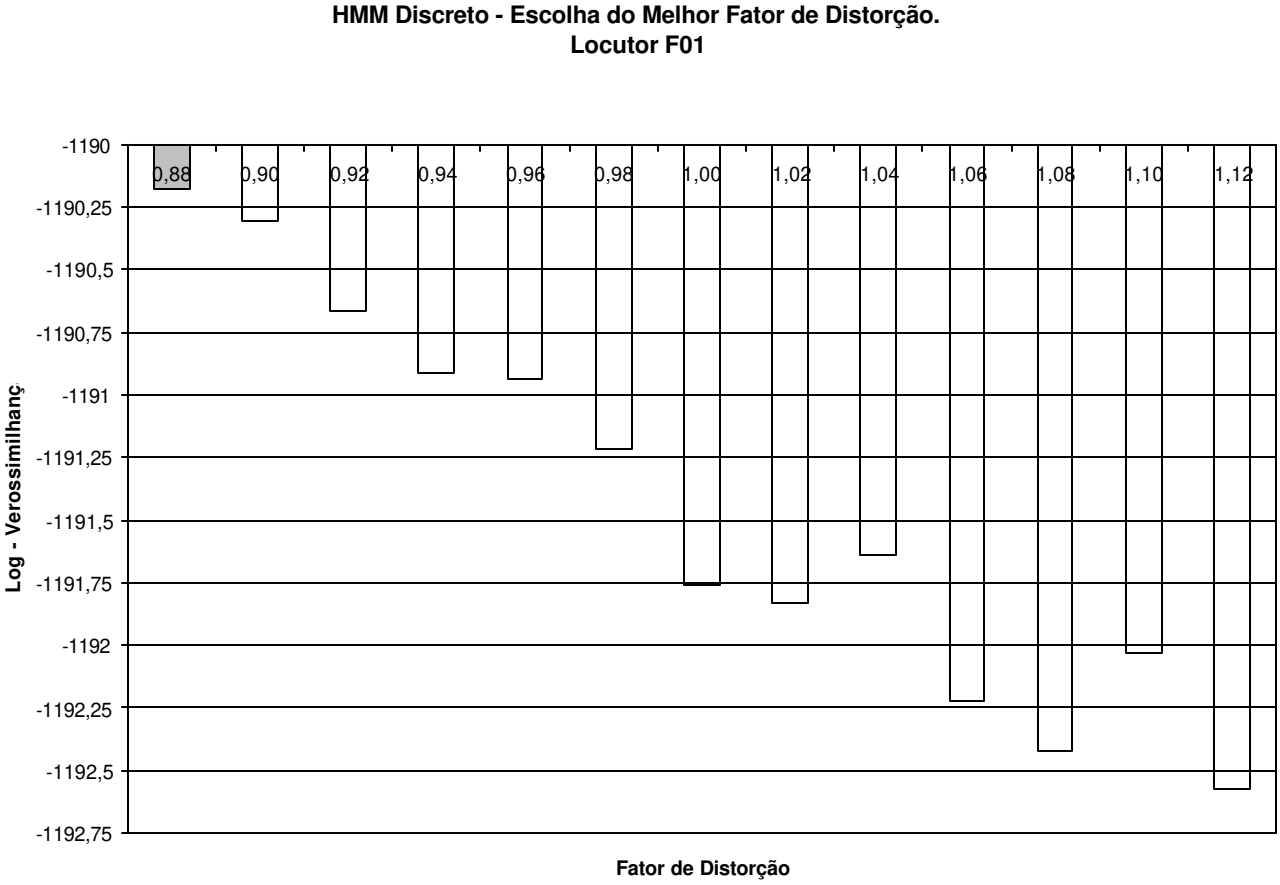
Fator de Distorção (α)	***F01***	***F02***	***M01***	***M11***
0.88	**-1190,18**	-1259,06	**-1054,31**	-1229,38
0.9	-1190,31	-1259,23	-1054,53	-1229,77
0.92	-1190,67	-1259,51	-1054,64	-1230,35
0.94	-1190,91	-1259,56	-1054,63	-1229,52
0.96	-1190,94	-1259,42	-1054,62	-1229,77
0.98	-1191,22	-1259,83	-1054,73	-1229,79
1.00	-1191,76	**-1256,42**	-1056,54	-1229,35
1.02	-1191,83	-1260,06	-1054,65	-1229,06
1.04	-1191,64	-1260,05	-1057,69	-1229,35
1.06	-1192,22	-1260,54	-1057,74	-1229,85
1.08	-1192,42	-1261,03	-1057,21	-1229,1
1.10	-1192,03	-1260,84	-1054,77	**-1227,19**
1.12	-1192,57	-1261,3	-1054,59	-1228,38

Tabela 7.1: Tabela representativa dos valores de verossimilhança, para cada valor de α, dos locutores (HMM Discreto)

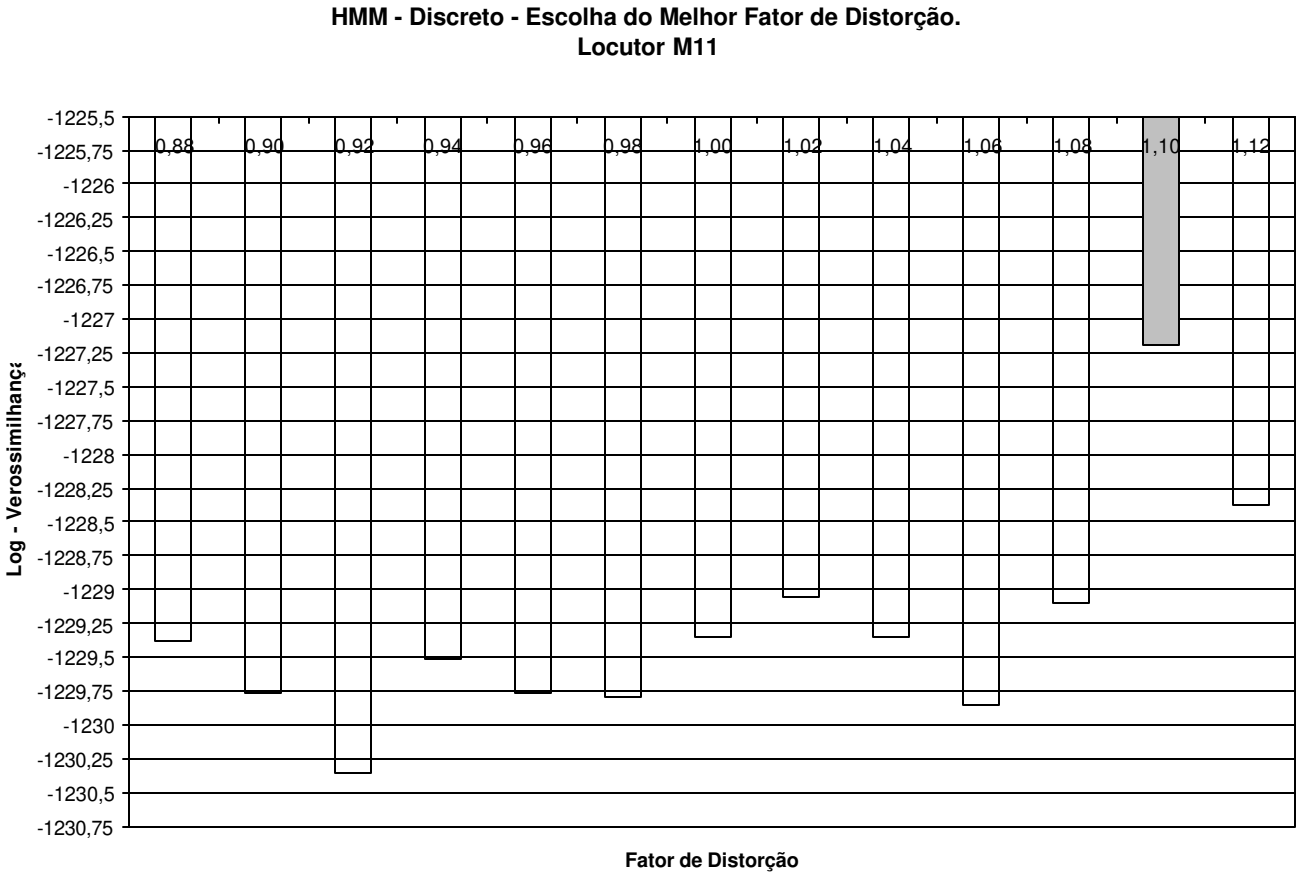
Como podemos observar na Tabela 7.1, os valores de verossimilhaça para cada fator de distorção (α) são muito próximos, sendo diferenciados apenas por sua parte decimal. Com uma proximidade tão grande entre estes valores, tem-se uma certa dificuldade em se afirmar qual o fator de distorção que melhor caracteriza cada locutor.

Podemos verificar ainda, nas Figuras 7.5a e b, traçadas para os locutores F01 e M11, que ocorreram vários máximos locais. Era de se esperar que ao variarmos α entre 0,88 e 1,12, a verossimilhança crescesse monotonicamente, atingisse um máximo, e depois passasse a decrescer monotonicamente. Entretanto, isto não ocorreu. Uma justificativa é a quantização vetorial utilizada nos vetores de parâmetros obtidos para cada valor de α. Estes vetores de

parâmetros podem estar tão próximos ao se variar α (de 0,02) que, ao serem quantizados vetorialmente, podem estar sendo confundidos, prejudicando assim a quantização e conseqüentemente a escolha do $\alpha_{ótimo}$. Nas Figuras 7.5a e b o $\alpha_{ótimo}$, definido para cada locutor, é representado em negrito.



(a)



(b)

Figura 7.5 : Histogramas da escolha do melhor α para os locutores F01 (a) e M11 (b), utilizando HMM discreto

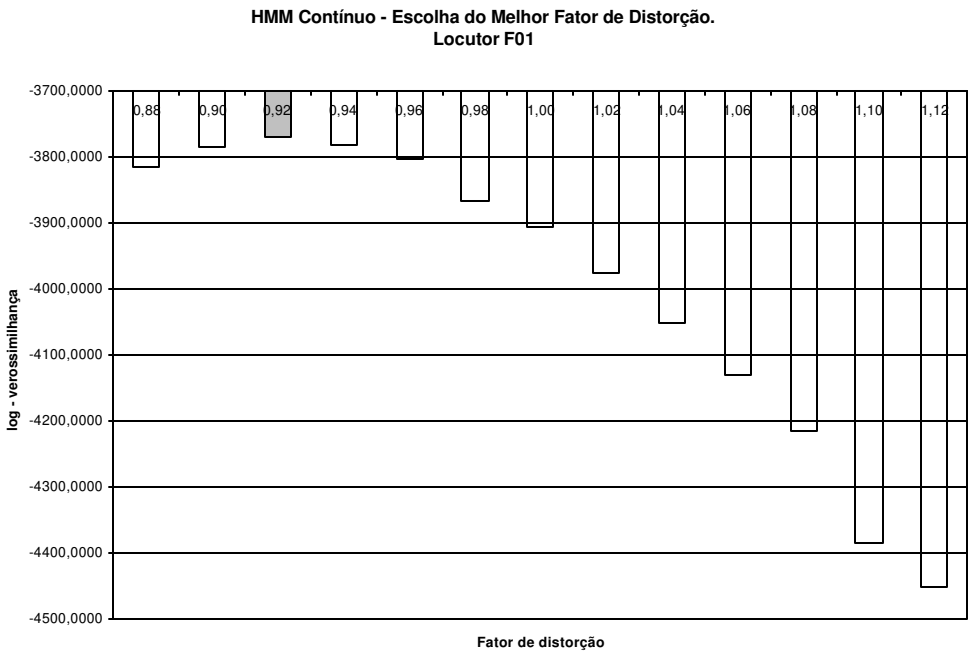
7.4.2 – Utilizando-se HMM Contínuo

Realizando-se para o HMM contínuo (3 gaussianas por estado) o mesmo teste utilizado na verificação do melhor α para o HMM discreto, temos:

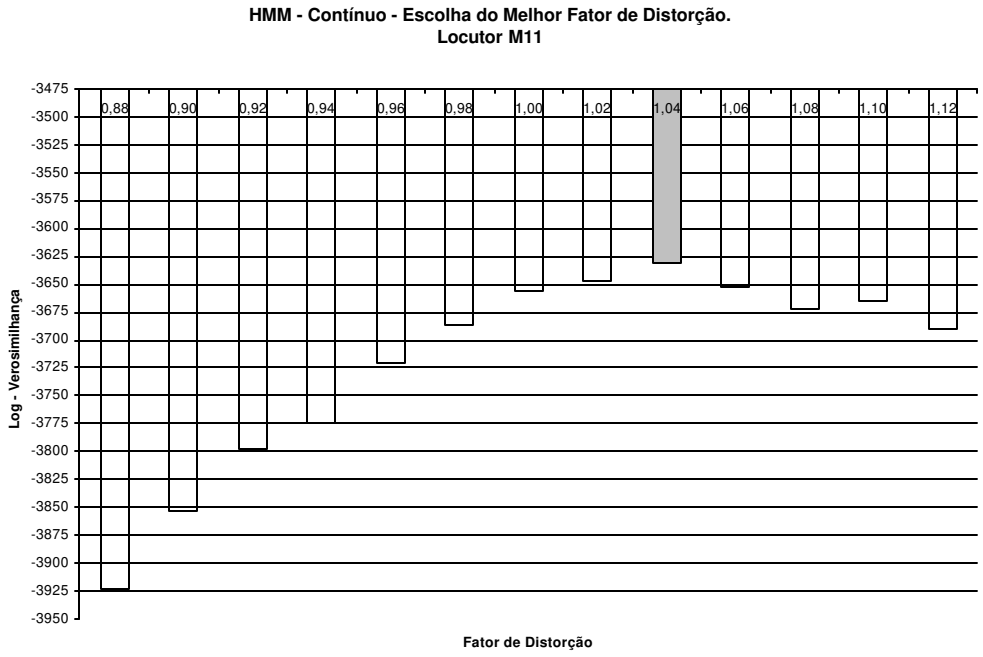
Fator de Distorção (α)	***F01***	***F02***	***M01***	***M11***
0.88	-2870,4541	-3597,8309	-2551,1280	-2889,36680
0.9	-2822,8541	-3622,6639	-2508,8116	-2834,49590
0.92	**-2810,8352**	-3571,6964	-2476,2912	-2792,87320
0.94	-2815,5426	-3537,1938	-2448,9662	-2767,02520
0.96	-2841,8261	-3522,8307	-2424,4140	-2734,97650
0.98	-2863,2061	-3537,8835	-2412,6645	-2716,06630
1.00	-2890,1879	**-3521,6385**	**-2397,5006**	-2690,61000
1.02	-2937,6140	-3547,0559	-2401,9416	-2692,70860
1.04	-2991,0265	-3574,1953	-2405,7590	**-2679,39680**
1.06	-3044,7666	-3610,7214	-2410,3009	-2688,19140
1.08	-3107,9111	-3669,4004	-2423,2746	-2707,76480
1.10	-3193,4661	-3746,3778	-2419,2502	-2695,44660
1.12	-3242,6063	-3796,8537	-2432,3044	-2712,58390

Tabela 7.2: Tabela representativa dos valores de verossimilhança, para cada valor de α dos locutores (HMM Contínuo)

Observando-se a Tabela 7.2 e comparando-a com a Tabela 7.1 verificamos que quando utilizado HMM contínuo a variabilidade entre as verossimilhanças, associadas a cada valor α, é bem maior do que no caso discreto. Isto nos permite uma melhor caracterização dos locutores, em relação a seu respectivo α. Quanto ao comportamento da curva de verossimilhança, podemos visualizar nas Figuras 7.6a e b que ela apresenta um máximo melhor definido, do que no caso discreto, como era de se esperar.



(a)



(b)

Figura 7.6 : Histogramas da escolha do melhor α para os locutores F01 (a) e M11 (b), utilizando HMM contínuo

Outro fato que deve ser ressaltado é a diferença entre o $\alpha_{ótimo}$ obtido, para cada locutor, quando utilizado HMM discreto e quando utilizado HMM contínuo. Este fato, por sua vez, reforça a idéia de que a quantização vetorial estaria prejudicando a escolha do $\alpha_{ótimo}$ entre os locutores.

Devido à dificuldade de se obter uma definição mais clara do $\alpha_{ótimo}$ ao se trabalhar com o HMM discreto, optamos por utilizar os HMMs contínuos no modelamento dos fones.

Na tentativa de analisar melhor a influência da escolha do $\alpha_{ótimo}$ na normalização dos locutores, podemos observar outros fatores bastante relevantes nesta escolha, como por exemplo o número de parâmetros acústicos e a quantidade de locuções necessárias na escolha deste α. A verificação da influência destes fatores na escolha do $\alpha_{ótimo}$ foi realizada utilizando-se o HMM contínuo com mistura de 3 gaussianas por estado.

- Quantidade de Parâmetros

A escolha do melhor fator de distorção, para um determinado locutor, pode ser realizada utilizando-se, por simplicidade, apenas 1 parâmetro acústico. Entretanto, ao variarmos a quantidade destes parâmetros, observamos os seguintes resultados.

	Locutores			
Quantidade de Parâmetros	**F01**	**F02**	**M01**	**M11**
1 parâmetro (mel)	$\alpha_{ótimo} = 0,92$	$\alpha_{ótimo} = 1,0$	$\alpha_{ótimo} = 1,0$	$\alpha_{ótimo} = 1,04$
3 parâmetros (mel, dmel, ddmel)	$\alpha_{ótimo} = 0,92$	$\alpha_{ótimo} = 0,96$	$\alpha_{ótimo} = 1,04$	$\alpha_{ótimo} = 1,04$

Tabela 7.3: Tabela ilustrativa da influência da quantidade de parâmetros utilizada na escolha do melhor α

Analisando-se os resultados da Tabelas 7.3, podemos perceber que com o aumento da quantidade de parâmetros na escolha do $\alpha_{ótimo}$, tem-se diferentes valores de α para um mesmo locutor. Este fato pode ser justificado uma vez que α depende do HMM utilizado como referência (HMM pré-treinado). Sendo o HMM pré-treinado obtido pela utilização de 3 parâmetros, consequentemente α deve ser obtido utilizando-se os 3 parâmetros também. A utilização de α obtido erroneamente pode comprometer o desempenho do sistema, uma vez que um ou mais locutores podem estar sendo treinados com α diferente do seu "ideal", na representação de suas características acústicas.

- Quantidade de Locuções

Assim como a quantidade de parâmetros, a quantidade de locuções empregada na escolha do $\alpha_{ótimo}$ pode influenciar o desempenho do sistema (empregando HMM já normalizado). Desta forma, quanto maior o número de locuções utilizadas por um locutor, maior será a probabilidade do α escolhido ser o ideal. Ao utilizarmos uma única frase, por exemplo, corremos o risco de que esta frase não possua um bom balanceamento fonético, influenciando assim na escolha do melhor α.

De modo a avaliar a influência do número de locuções, por locutor, na escolha do melhor α, fez-se uma análise preliminar utilizando-se todas as locuções dos locutores F01 e M01, por exemplo. Para cada conjunto de locuções é obtido o valor de α que melhor representa o locutor analisado. Estes resultados são mostrados a seguir:

Locutor F01

Quantidade de Locuções	α	Quantidade de Locuções	α	Quantidade de Locuções	α
1	0,92	18	0,92	35	0,92
2	0,92	19	0,92	36	0,92
3	0,92	20	0,92	37	0,92
4	0,92	21	0,92	38	0,92
5	0,92	22	0,92	39	0,92
6	0,92	23	0,92	40	0,92
7	0,92	24	0,92	41	0,92
8	0,92	25	0,92	42	0,92
9	0,92	26	0,92	43	0,92
10	0,92	27	0,92	44	0,92
11	0,92	28	0,92	45	0,92
12	0,92	29	0,92	46	0,92
13	0,92	30	0,92	47	0,92
14	0,92	31	0,92	48	0,92
15	0,92	32	0,92	49	0,92
16	0,92	33	0,92	50	0,92
17	0,92	34	0,92	51	0,92

Locutor M11

Quantidade de Locuções	α	Quantidade de Locuções	α	Quantidade de Locuções	α
1	1,04	18	1,04	35	1,04
2	1,04	19	1,04	36	1,04
3	1,04	20	1,04	37	1,04
4	1,04	21	1,04	38	1,04
5	1,04	22	1,04	39	1,04
6	1,04	23	1,04	40	1,04
7	1,04	24	1,06	41	1,04
8	1,04	25	1,06	42	1,04
9	1,04	26	1,04	43	1,04
10	1,04	27	1,04	44	1,04
11	1,04	28	1,04	45	1,04
12	1,04	29	1,04	46	1,04
13	1,04	30	1,04	47	1,04
14	1,04	31	1,04	48	1,04
15	1,04	32	1,04	49	1,04
16	1,04	33	1,04	50	1,04
17	1,04	34	1,04	51	1,04

Tabela 7.4: Tabelas representativas da variação do número de locuções para os locutores F01 (a) e M01 (b)

Nas Tabelas 7.4 a e b, percebemos que não há necessidade da utilização de todas as locuções associadas ao locutor, pois o valor de α não é modificado com o aumento do número de locuções, na escolha do melhor α. Assim, pode-se escolher uma quantidade de locuções menor, mas ao mesmo tempo suficiente para que não haja variação de α entre as locuções. Desta forma, por simplicidade, pelos resultados mostrados na Tabela 7.4a e b e principalmente pelo ganho de tempo na escolha do α, optamos por utilizar apenas 4 locuções por locutor, na escolha do $\alpha_{ótimo}$ (4 frases no treinamento e 4 nomes no reconhecimento, ou teste, do sistema).

7.5 – AVALIAÇÃO DO SISTEMA

Uma vez definida a melhor forma de escolha do α, iniciamos a avaliação do sistema, mostrada na Tabela 7.5, onde os HMMs foram implementados com misturas de 3 gaussianas por parâmetro. Nesta avaliação foram utilizados 42 locutores de treinamento (2000 frases) e 15 locutores de teste (300 nomes), definidos no Capítulo 4. Utilizou-se ainda 4 locuções por locutor na escolha do $\alpha_{ótimo}$. Deve-se lembrar que a cada novo retreinamento é escolhido um novo α para cada locutor.

Número de Retreinamentos		**Distorção (convergência)**	**α's Modificados**	**Taxa de Erros % (Teste)**
Treinamento inicial(α=1)		0,0009	-------	17,00
1º retreinamento	1º época	0,0089	-------	16,67
	2º época	0,0013		16,00
	3º época	0,0006		15,00
2º retreinamento	1º época	0,0021	22	15,00
	2º época	0,0006		15,33

Tabela 7.5: Resultado obtido para o sistema de reconhecimento com normalização do locutor, utilizando-se misturas de 3 gaussianas

Como podemos observar na Tabela 7.5, obteve-se uma melhora de desempenho quando comparado ao sistema sem normalização (17 % de erro), sendo necessário apenas 1 retreinamento antes que o desempenho do sistema fosse prejudicado. No entanto, o objetivo principal do sistema não foi atingido, isto é, o sistema deveria continuar sendo treinado até que não houvesse mais modificação entre os α's de um mesmo locutor entre um retreinamento e outro.

A partir dos resultados obtidos, surgiu a hipótese de que provavelmente os HMMs ainda não estariam treinados o suficiente para que pudessem passar para um próximo retreinamento. Seguindo esta idéia resolvemos aumentar o número de épocas de cada retreinamento, até que fossem atingidas, mais ou menos, 5 ou 6 épocas, ou enquanto houvesse melhora de desempenho do sistema. Assim, obtivemos os seguintes resultados:

		Distorção (convergência)	**α's Modificados**	**Taxa de Erros % (teste)**
Treinamento (α =1)		0,0009	-------	17,00
1º retreinamento	1º época	0,0089	-------	16,67
	2º época	0,0013		16,00
	3º época	0,0006		15,00
	4º época	0,0004		15,00
	5º época	0,0004		15,00
2º retreinamento	1º época	0,0019	22	14,00
	2º época	0,0005		14,00
	3º época	0,0003		
	4º época	0,0003		
	5º época	0,0002		
	6º época	0,0002		
	7º época	0,0002		
	8º época	0,0001		13,00
3º retreinamento	1º época	0,0005	10	13,33
	2º época	0,0002		13,33

Tabela 7.6: Tabela representativa do desempenho do sistema normalizado, utilizando-se misturas de 3 gaussianas, aumentando-se o número de épocas em cada retreinamento

Comparando-se as Tabelas 7.5 e 7.6, podemos notar que ao aumentarmos o número de épocas, em cada retreinamento, obtém-se uma melhora de desempenho em relação ao procedimento anterior. Entretanto, como pode ser visto na Tabela 7.6, não se pode afirmar, exatamente, quantas épocas devem ser realizadas para que haja melhora de desempenho no sistema. O único fato que pode ser constatado é que, com a normalização dos locutores, a faixa de distorção é reduzida de 0,001 para 0,0001.

Apesar de ter-se uma melhora de desempenho utilizando-se 2 retreinamentos, ainda há o fato de que os locutores continuam com valor de α variável entre um retreinamento e outro. Assim, tentando resolver este problema e o fato de não sabermos exatamente quantas épocas seriam necessárias para cada retreinamento, decidimos seguir 2 critérios de parada para o retreinamento do sistema. A primeira idéia seria de treinar os modelos HMMs com apenas 1 época por retreinamento, realizando-se uma nova escolha de α a cada novo retreinamento. Vários retreinamentos seriam executados até que não houvesse mais variação de α, de um mesmo locutor, entre um retreinamento e outro. A segunda idéia seria de que uma vez que os α's permanecessem constantes, para todos os locutores, seria verificada a distorção relativa obtida neste retreinamento. Caso esta distorção fosse maior que 0,0001, entre o retreinamento atual e retreinamento anterior, o retreinamento continuaria sendo realizado até que se obtivesse a distorção desejada.

Tomando como base o procedimento descrito anteriormente, obtivemos os resultados ilustrados na Tabela 7.7.

	Distorção	**α's Modificados**	**Taxa de Erros %**
Treinamento	0,0009	------	17,00
1º retreinamento	0,0089	------	16,67
2º retreinamento	0,0032	21	16,00
3º retreinamento	0,0013	8	16,00
4º retreinamento	0,0007	3	15,34
5º retreinamento	0,0005	3	15,00
6º retreinamento	0,0005	1	14,33
7º retreinamento	0,0004	1	14,33
8º retreinamento	0,0003	0	14,00
9º retreinamento	0,0002	0	14,33
10º retreinamento	0,0002	0	14,00
11º retreinamento	0,0002	0	13,67
12º retreinamento	0,0001	0	13,67

Tabela 7.7: Tabela ilustrativa do desempenho do sistema ao utilizarmos 1 época por retreinamento. A cada novo retreinamento é calculado um novo valor de α.

Como pode ser visualizado na Tabela 7.7, as duas condições adotadas foram bastante relevantes na escolha da melhor forma de normalização dos locutores. O limite de treinamento (número de épocas necessárias para cada retreinamento) é obtido de forma mais coerente que o procedimento anterior, obtendo-se ainda uma melhora de desempenho a cada retreinamento. Deve-se mencionar ainda que, com estas novas condições, diminuiu-se o tempo de treinamento do sistema. Onde antes utilizávamos 13 épocas para normalizar o sistema, agora utilizamos apenas 12 épocas.

7.6 – RESULTADOS FINAIS

Uma vez definida a melhor forma de retreinamento, reconhecimento e, principalmente, de escolha do melhor fator de distorção para normalização do sistema de reconhecimento de fala, apresentaremos os resultados finais deste processo ao utilizarmos misturas de 5 gaussianas por parâmetro.

No capítulo 5 foi mostrado que os HMMs eram melhor modelados quando utilizávamos misturas de 6 gaussianas. Entretanto, no decorrer dos testes pôde-se perceber que os locutores iam tendo suas características espectrais aproximadas (normalizadas), diminuindo assim a quantidade de gaussianas necessárias para sua representação. Desta forma, para melhor visualização do desempenho do sistema e para evitar que alguma gaussiana deixasse de ser corretamente modelada, devido à pequena quantidade de dados de treinamento utilizada, optamos por usar apenas 5 gaussianas na representação dos resultados finais do sistema.

Além das 5 gaussianas, utilizou-se também 20 locutores de teste, 5 a mais que os usados nos testes anteriores. Desta forma, tem-se um sistema com as seguintes especificações:

- Locutores de treinamento: 42 locutores, pronunciando um total de 2000 locuções.
- Parâmetros utilizados: mel, dmel e ddmel (12 coeficientes cada).
- K = 1 para o dmel e ddmel.
- Sub-unidades fonéticas utilizadas: fones independentes do contexto.
- Tipo de HMM: Contínuo.
- Número de gaussianas por estado e por parâmetro (densidades independentes): 5.
- Algoritmo de treinamento: Baum-Welch.
- Número de locuções utilizadas na escolha do melhor α: 4 frases para o treinamento e 4 nomes para o reconhecimento. Deve-se ressaltar que para o treinamento dos HMMs, o $\alpha_{ÓTIMO}$ obtido para cada locutor é utilizado para todas as locuções de seu respectivo locutor, na obtenção dos parâmetros cepstrais de cada locução.
- Critério de parada para o treinamento: distorção relativa = 0,0001.

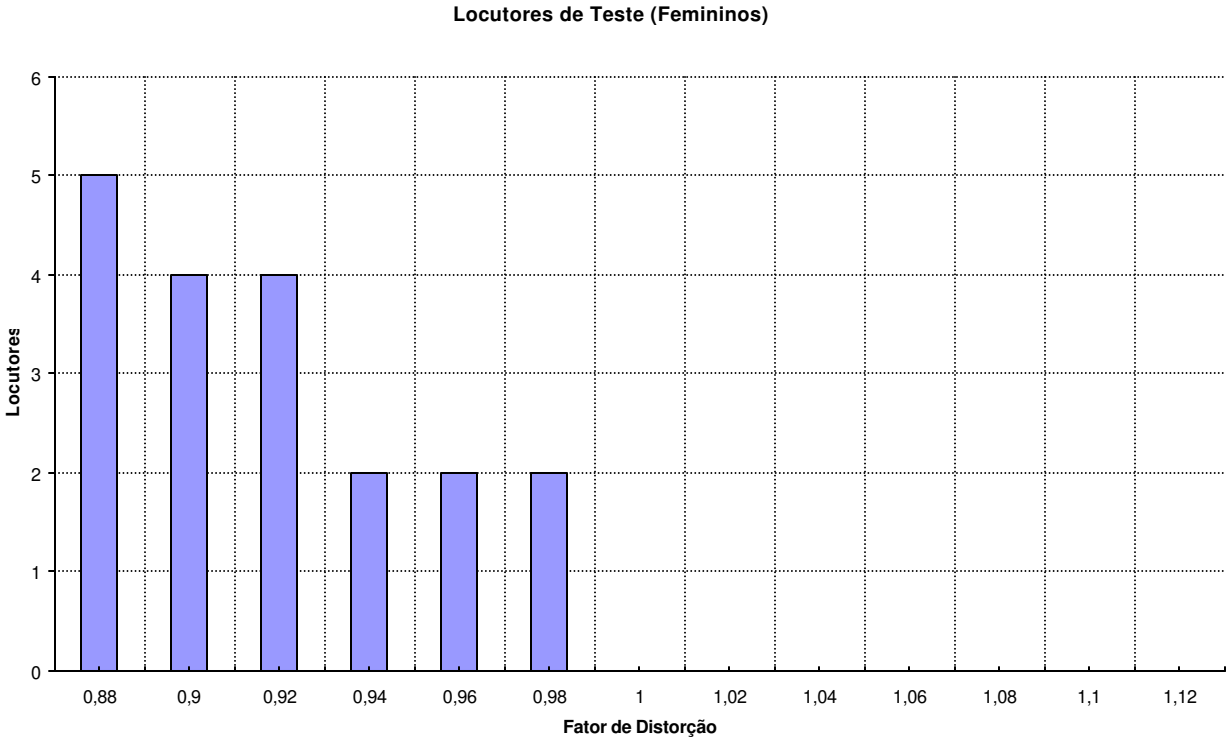
- Locutores de teste: 20 (10 homens e 10 mulheres) pronunciando um total de 400 nomes.
- Algoritmo de reconhecimento: One-Step.
- Avaliação do sistema:

	Distorção	**α's Modificados**	**Taxa de Erros %**	**Tempo de Treinamento (hora:minuto:segundo)**
Treinamento	0,001	------	19,25	27:58:51
1º retreinamento	0,0056	------	19,00	08:19:54
2º retreinamento	0,0035	26	17,50	08:32:24
3º retreinamento	0,0021	20	15,50	08:27:35
4º retreinamento	0,0012	9	15,00	07:47:45
5º retreinamento	0,0008	8	16,00	04:31:11
6º retreinamento	0,0007	5	15,75	08:59:24
7º retreinamento	0,0005	2	15,50	04:48:58
8º retreinamento	0,0004	2	14,00	07:49:08
9º retreinamento	0,0004	0	. . .	59:34:50
. . .	. . .			
19º retreinamento	0,0001		11,25	

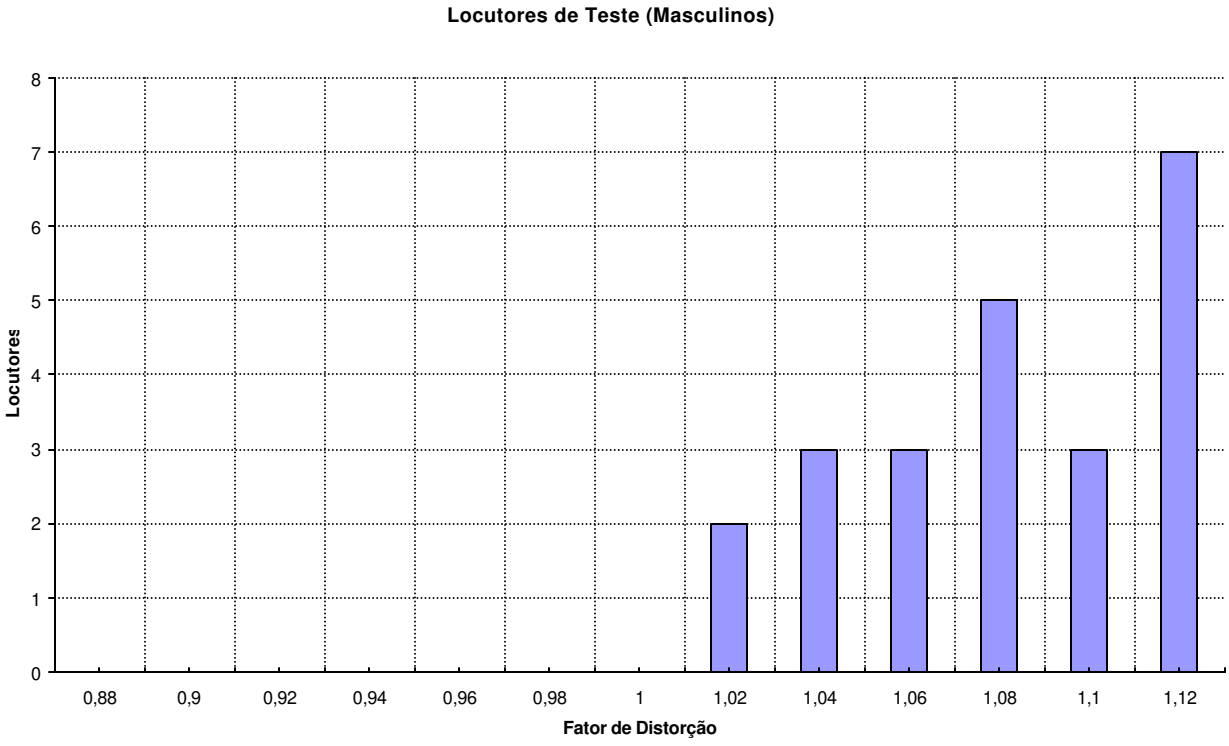
Tabela 7.8: Tabela ilustrativa do desempenho do sistema normalizado, utilizando 5 gaussianas. A duração de cada retreinamaneto leva em consideração o período de tempo utilizado para a escolha do α.

Na Tabela 7.8 o tempo apresentado do 9º ao 19º retreinamento, representa o valor total obtido durante 11 retreinamentos consecutivos, uma vez que não há variação do fator de distorção entre estes retreinamentos. O tempo de treinamento dos modelos foi verificado utilizando-se um Pentium II – 300 MHz.

Deve-se mencionar que o procedimento de escolha do $\alpha_{ÓTIMO}$ pode ser considerado um importante critério utilizado na caracterização de locutores (masculino e feminino), como ilustrado nas Figuras 7.7 a e b.



(a)



(b)

Figura 7.7: Histogramas representativos da faixa de valores de α escolhida para os locutores femininos (a) e masculinos (b), para o treinamento do sistema utilizando-se mistura de 5 gaussianas

Como pode ser observado na Figura 7.7, ao empregarmos o critério de normalização adotado neste trabalho, não se tem mais o problema dos locutores femininos serem caracterizados por apresentarem, na média, fator 0,94, os locutores masculinos por apresentarem, na média, fator 1,00 e para todos os locutores fator 0,975, como apresentado na seção 6.3.

Na Tabela 7.9 tem-se os valores de α definidos para os locutores de teste em cada novo retreinamento. Estes α's são obtidos utilizando o $HMM_{normalizado}$ para 5 misturas de gaussianas, conforme procedimento descrito na seção 7.3.

Locutor	**Número de Retreinamentos**								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**...**	**19**
F01	0,88	0,88	0,90	0,90	0,92	0,92	0,92	**...**	0,92
F02	0,88	0,88	0,88	0,88	0,88	0,88	0,88	**...**	0,88
F03	0,88	0,88	0,88	0,88	0,88	0,88	0,88	**...**	0,88
F04	0,96	0,92	0,92	0,92	0,92	0,92	0,92	**...**	0,92
F05	0,92	0,92	0,90	0,90	0,92	0,92	0,92	**...**	0,92
F06	0,92	0,90	0,90	0,88	0,88	0,88	0,88	**...**	0,88
F07	0,96	0,90	0,90	0,90	0,88	0,88	0,88	**...**	0,88
F08	0,88	0,88	0,88	0,88	0,88	0,88	0,88	**...**	0,88
F09	0,88	0,94	0,94	0,92	0,88	0,88	0,88	**...**	0,88
F10	0,92	0,90	0,90	0,90	0,92	0,92	0,92	**...**	0,92
M01	1,02	1,02	1,04	1,04	1,04	1,04	1,04	**...**	1,04
M02	1,04	1,06	1,06	1,06	1,08	1,08	1,08	**...**	1,08
M03	1,06	1,06	1,08	1,08	1,08	1,08	1,08	**...**	1,08
M04	0,96	0,98	0,98	0,98	1,00	1,00	1,00	**...**	1,00
M05	1,08	1,06	1,08	1,08	1,08	1,08	1,08	**...**	1,08
M06	1,04	1,02	1,02	1,02	1,04	1,04	1,04	**...**	1,04
M07	0,96	1,0	1,00	1,00	1,00	1,00	1,00	**...**	1,00
M08	0,96	0,96	0,98	0,98	0,96	0,96	0,96	**...**	0,96
M09	0,96	0,94	0,94	0,96	0,96	0,96	0,96	**...**	0,96
M10	1,04	1,04	1,04	1,04	1,04	1,04	1,04	**...**	1,04

Tabela 7.9: Tabela representativa dos valores de α obtidos para os locutores de teste (HMM normalizado com mistura de 5 gaussianas por estado)

Como pode ser observado na Tabela 7.9, depois de um certo número de retreinamentos os α's dos locutores de teste mantêm-se constantes, mas a verossimilhança média calculada com as locuções de treinamento continua aumentando. Uma justificativa para este fato é que o material utilizado no treinamento é bem maior que o material de teste.

Na Tabela 7.10 tem-se a taxa de erros do sistema ao reconhecer os locutores de teste, sem e com normalização de locutor. No caso de se empregar o sistema normalizado, foi também calculada a taxa de erros sem se providenciar a normalização destes locutores. Nota-se que, neste caso, a taxa de erros superou a do sistema com normalização. Isto era de se esperar uma vez que o HMM normalizado ficou especializado para um locutor médio.

Treinamento	**Reconhecimento**	
	Sem Normalização	Com Normalização
Sem Normalização	19,25 %	------
Com Normalização	24,25 %	11,25 %

Tabela 7.10: Taxa de desempenho do sistema sem e com normalização de locutor (HMM com 5 gaussianas/estado)

No Apêndice D é apresentada uma tabela contendo as locuções pronunciadas pelos locutores de teste e as locuções reconhecidas por estes locutores, utilizando-se normalização de locutor. As locuções reconhecidas erroneamente são mostradas em negrito, sendo apresentado ainda a colocação em que a locução correta estaria situada.

8 - Conclusões:

A normalização de locutor não deve ser confundida com a adaptação de locutor. Na primeira técnica todos os locutores utilizados no retreinamento do sistema são normalizados, iterativamente, em relação a um locutor médio. Na segunda técnica, o retreinamento é realizado para um locutor em particular, aquele para o qual o sistema será adaptado. Na adaptação ao locutor tem-se um período de retreinamento muito menor do que na normalização do locutor, uma vez que a adaptação é feita para um único locutor.

No presente trabalho avaliamos a técnica de normalização de comprimento do trato vocal entre diferentes locutores. Para avaliação desta técnica utilizou-se o *Sistema de Reconhecimento de Fala Independente do Locutor e de Vocabulário Flexível*.

A utilização de vocabulário flexível no sistema adotado foi de grande valia, principalmente por proporcionar uma maior flexibilidade quando da criação do vocabulário a ser reconhecido pelo sistema. Desta forma, pôde-se reconhecer locuções fora do universo com o qual o sistema foi treinado, daí a maior versatilidade do sistema.

Quanto à técnica utilizada, pôde-se comprovar que a normalização de comprimento do trato vocal realizada pelo escalonamento do banco de filtros, na escala Mel, é uma importante ferramenta a ser engajada nos sistemas de reconhecimento de fala. Pode-se ressaltar que, além de proporcionar uma melhora no desempenho do sistema, é de fácil implementação. Este tipo de normalização, segundo [11], tende a proporcionar melhores resultados do que as técnicas de *Separação das Características Acústicas (masculino e feminino)* e de *Normalização da Média Cepstral*.

Ao longo deste trabalho, verificamos a necessidade de uma estratégia que proporcionasse a melhor maneira de se normalizar o sistema, sendo escolhida uma estratégia de, iterativamente, calcular o α^i, para cada locutor, e retreinar o sistema a cada nova época.

Deve-se destacar ainda, o curto período de tempo necessário para o cálculo do $\alpha_{ótimo}$ de cada locutor, na fase de reconhecimento. Empregando-se 4 locuções (nomes) por locutor, obteve-se, avaliando-se um conjunto de 20 locutores de teste, um tempo médio de 19 minutos para o cálculo do $\alpha_{ótimo}$ de todos os locutores. Assim, para cada locutor, gastou-se aproximadamente 1 minuto no cálculo do $\alpha_{ótimo}$.

O procedimento de escolha do melhor fator de distorção (α) pode ser considerado um fator determinante no desempenho do sistema. Isto decorre do fato de que ao ser escolhido, para pelo menos um locutor, um valor de α diferente do seu "ideal", o treinamento para este locutor será realizado fora de suas características espectrais. Desta forma, os HMMs assumirão valores de verossimilhança bastante baixos, comprometendo assim o desempenho do sistema.

Como desvantagem da técnica de normalização tem-se a proximidade das gaussianas. A cada retreinamento do sistema, utilizando um novo conjunto de coeficientes α, mais normalizado torna-se o sistema. Esta normalização, por sua vez, faz com que algumas gaussianas utilizadas no modelamento dos HMMs deixem de ser relevantes, como eram no início do processo, passando a ter coeficientes (c_{jm}) bastante baixos. Este fato poderá gerar problemas de underflow no sistema, caso não tenha sido previsto. Outra desvantagem da técnica é o tempo gasto em cada retreinamento, além do tempo necessário para o treinamento inicial do sistema (sem normalização). Em contrapartida a estes problemas, tem-se o aumento considerável de desempenho do sistema.

A contribuição mais significativa deste trabalho, em relação a [11], foi a de proporcionar um método mais robusto para o retreinamento do sistema e para obtenção do $\alpha_{ótimo}$ de cada locutor.

Como sugestão para trabalhos futuros, podemos sugerir a avaliação da normalização do locutor, através do escalonamento do banco de filtros, somente para as baixas freqüências. Uma justificativa para esta proposta é o fato de que a variabilidade da fala, analisando os fones como unidades básicas, é maior nas vogais (concentradas nas baixas freqüências), quando pronunciadas por diferentes locutores. Desta forma, a influência das altas freqüências no cálculo dos coeficientes mel-cepstrais seria mantida constante, não havendo descontinuidade na banda de corte do sinal ao escalonarmos o banco de filtros.

Bibliografia

[1] ALCAIM, A., SOLEWICZ, J. A., MORAES, J. A. - "Freqüência de Ocorrência dos Fones e Lista de Frases Foneticamente Balanceadas no Português Falado no Rio de Janeiro. Revista Brasileira de Telecomunicações, vol 7, nº1, pp: 23 – 41. Dezembro 1992.

[2] ANDREOU, A., KAMM T. and COHEN, J. – "Experiments in Vocal Tract Normalization". Proceedings CAIP Workshop: Frontiers in Speech Recognition II, 1994.

[3] BAKER, J. K. – "The Dragon System- an Overview". IEEE Transactions on ASSP, vol. ASSP-23, nº 1, February 1975, pp 24-29.

[4] BAUM, L.E. *et al* – "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains". Ann. Math. Stat., Vol. 41, nº 1, 1970, pp 164-171.

[5] BURNETT, D. C. and FANTY, M. – "Rapid Unsupervised Adaptation to Children's Speech on a Connected-Digit Task. Proceedings of the ICSLP, Philadelphia, PA, October 1996 (publicação interna).

[6] DAVIS, S. B. and MERMELSTEIN, P. – "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-28, nº 4, August 1980, pp 357-368.

[7] DELLER Jr, J. R., PROAKIS, J. G., HANSEN, J. H. L. – "Discrete Time Processing of Speech Signals. MacMillan Publishing Companhy. New York, 1993.

[8] FAGUNDES, R. D. R. – "Reconhecimento de Voz, Linguagem Contínua, usando Modelos de Markov". Tese de Mestrado, Escola Politécnica da USP, 1993, pp 46-51, 81-98.

[9] HAO, Y. and FANG, D. – "Speech Recognition using Speaker Adaptation by System Parameter Transformation. IEEE Transactions on Speech and Audio Processing, Vol. 2, nº 1, Part 1, January, 1994, pp 63-67.

[10] HAZEN T. J. and GLASS, J. R. – "A Comparison of Novel Techniques for Instantaneous Speaker Adaptation". Spoken Language Systems Group. Laboratory for Computer Science MIT, Cambridge. Publicação obtida pela internet no endereço www.sls.lcs.mit.edu/sls/publications/1997/eurospeech97-hazen.pdf.

[11] LEE, L. and ROSE, R. – "A Frequency Warping Approach to Speaker Normalization". IEEE Transactions on Speech and Audio Processing, Vol. 6, nº 1, January 1998, pp 49-60.

[12] LINDO Y., BUZO, A., GRAY R. M. – "An Algorithm for Vector Quantizer Design". IEEE Transactions on Communications, COM-28(1), January 1980, pp 84-94.

[13] MARKEL, J. D. and Gray Jr., A. H. – "Linear Prediction of Speech". Berlim Heidelberg, New York, 1976, pp 10-32, 190-211.

[14] MARTINS, J. A. – "Avaliação de Diferentes Técnicas para Reconhecimento de Fala". Tese de Doutorado, FEEC/UNICAMP, Campinas, Dezembro 1997, pp 19-33, 37-40,72-73.

[15] NUNES, H. F. – "Reconhecimento de Fala Baseado em HMM". Tese de Mestrado, FEEC/UNICAMP, Campinas, Dezembro 1996, pp 61-70.

[16] PADMANABHAN M., LALIT R. B., NAHAMOO, D. and PICHENY, M. A. – "Speaker Clustering and Transformation for Speaker Adaptation in Speech Recognition Systems". IEEE Transactions on Speech and Audio Processing, Vol. 6, nº 1, January 1998, pp 71-77.

[17] PICONE, J. W. – "Signal Modeling Techniques in Speech Recognition". Proceedings of the IEEE, Vol. 81, nº 9, September 1993, pp 1215-1247.

[18] RABINER, L. R. – "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". Proceeding of the IEEE, Vol. 77, nº 2, February 1989, pp 257-286.

[19] RUNSTEIN, F. O. – "Sistema de Reconhecimento de Fala baseado em Redes Neurais Artificiais". Tese de Doutorado, FEEC/UNICAMP, Campinas, Setembro 1998, pp 129-139.

[20] YNOGUTI, C. A – "Reconhecimento de Fala Contínua usando Modelos Ocultos de Markov". Tese de Doutorado, UNICAMP, Campinas, Maio 1999, pp 24-32, 47-55, 58-82.

[21] ZHAN P. and WAIBEL, A. – "Vocal Tract Length Normalization for Large Vocabulary Continuous Speech Recognition". School of Computer Science, Carnegie Melon University, Pittsburgh, PA, May 1997. Publicação interna obtida pela internet no endereço www.lti.cs.cmu.edu/Research/cmt-tech-reports.html (CMU-LTI-97-150).

Apêndice A:

Lista dos Dígitos Utilizados no Treinamento do Sistema

Lista 21

UM	CINCO	DOIS	TRÊS	TRÊS	SEIS	CINCO	TRÊS
UM	UM	DOIS	CINCO	TRÊS	SETE	CINCO	CINCO
UM	NOVE	DOIS	QUATRO	TRÊS	OITO	CINCO	QUATRO
UM	DOIS	DOIS	OITO	TRÊS	NOVE	CINCO	OITO
UM	MEIA	DOIS	UM	TRÊS	ZERO	CINCO	UM
MEIA	TRÊS	UM	OITO	TRÊS	MEIA	QUATRO	TRÊS
DOIS	SETE	UM	TRÊS	TRÊS	TRÊS	QUATRO	SETE
DOIS	DOIS	UM	SEIS	TRÊS	DOIS	QUATRO	NOVE
DOIS	NOVE	UM	QUATRO	TRÊS	CINCO	QUATRO	QUATRO
DOIS	MEIA	UM	ZERO	TRÊS	QUATRO	QUATRO	UM
DOIS	DOIS	UM	SEIS	TRÊS	DOIS	QUATRO	NOVE

Lista 22

QUATRO	OITO	NOVE	SEIS	SETE	OITO	OITO	SEIS
QUATRO	ZERO	NOVE	SETE	SETE	ZERO	OITO	SETE
QUATRO	SEIS	NOVE	MEIA	SETE	SEIS	OITO	OITO
QUATRO	CINCO	NOVE	NOVE	SETE	CINCO	OITO	NOVE
QUATRO	DOIS	NOVE	ZERO	SETE	DOIS	OITO	ZERO
MEIA	CINCO	NOVE	UM	SETE	TRÊS	OITO	UM
CINCO	SETE	SEIS	TRÊS	SETE	SETE	OITO	MEIA
CINCO	DOIS	SEIS	SEIS	SETE	NOVE	OITO	DOIS
CINCO	NOVE	SEIS	QUATRO	SETE	QUATRO	OITO	CINCO
CINCO	SEIS	SEIS	ZERO	SETE	UM	OITO	QUATRO
CINCO	ZERO	SEIS	NOVE	SETE	MEIA	OITO	TRÊS

Lista 23

SEIS	CINCO	ZERO	OITO	UM	CINCO	DOIS	TRÊS
SEIS	UM	ZERO	ZERO	UM	UM	DOIS	CINCO
SEIS	MEIA	ZERO	SEIS	UM	NOVE	DOIS	MEIA
SEIS	DOIS	ZERO	MEIA	UM	DOIS	DOIS	OITO
SEIS	SETE	ZERO	DOIS	UM	MEIA	DOIS	UM
MEIA	QUATRO	ZERO	TRÊS	UM	OITO	DOIS	ZERO
NOVE	TRÊS	ZERO	SETE	UM	TRÊS	DOIS	SETE
NOVE	DOIS	ZERO	NOVE	UM	SEIS	DOIS	DOIS
NOVE	CINCO	ZERO	QUATRO	UM	QUATRO	DOIS	NOVE
NOVE	QUATRO	ZERO	UM	UM	ZERO	DOIS	SEIS
NOVE	OITO	ZERO	CINCO	UM	SETE	DOIS	QUATRO

Lista 24

ZERO	CINCO	MEIA	MEIA	CINCO	SEIS	SEIS	OITO
ZERO	UM	MEIA	UM	CINCO	SETE	SEIS	ZERO
ZERO	NOVE	MEIA	NOVE	CINCO	OITO	SEIS	SEIS
ZERO	DOIS	MEIA	DOIS	CINCO	MEIA	SEIS	CINCO
ZERO	SETE	MEIA	SETE	CINCO	ZERO	SEIS	DOIS
MEIA	ZERO	MEIA	OITO	CINCO	UM	SEIS	TRÊS
TRÊS	SETE	DOIS	ZERO	CINCO	TRÊS	SEIS	SETE
TRÊS	DOIS	CINCO	MEIA	CINCO	DOIS	SEIS	NOVE
TRÊS	NOVE	SEIS	OITO	CINCO	CINCO	SEIS	QUATRO
TRÊS	SEIS	ZERO	OITO	CINCO	QUATRO	SEIS	UM
TRÊS	MEIA	OITO	ZERO	CINCO	NOVE	SEIS	MEIA

Lista 25

OITO	TRÊS	SETE	SEIS	NOVE	OITO	TRÊS	TRÊS
OITO	CINCO	SETE	SETE	NOVE	ZERO	TRÊS	CINCO
OITO	MEIA	SETE	OITO	NOVE	SEIS	TRÊS	QUATRO
OITO	OITO	SETE	NOVE	NOVE	CINCO	TRÊS	OITO
OITO	UM	SETE	ZERO	NOVE	MEIA	TRÊS	UM
MEIA	SEIS	SETE	UM	NOVE	TRÊS	TRÊS	ZERO
SETE	TRÊS	OITO	SETE	NOVE	SETE	ZERO	TRÊS
SETE	DOIS	OITO	DOIS	NOVE	NOVE	ZERO	SEIS
SETE	CINCO	OITO	NOVE	NOVE	QUATRO	ZERO	QUATRO
SETE	QUATRO	OITO	SEIS	NOVE	UM	ZERO	ZERO
SETE	MEIA	OITO	QUATRO	NOVE	DOIS	ZERO	MEIA

Apêndice B:

Lista das Locuções de Teste Utilizadas no Sistema

Locutores Femininos

F01	***F02***	***F03***	***F04***	***F05***
Aderton	Albetã	Alda	Armelinda	Adnã
Atírson	Alessandra	Alex	Bárbara	Arantes
Bonifácio	Belém	Bianca	Barbie	Camões
Casagrande	Bruno	Clemildes	Beth	Carolina
Dunga	Camila	Creuza	Émerson	Coralina
Emília	Elaine	Didi	Gabriela	Cunha
Francine	Flávio	Dimitre	Gílce	Custódia
Gamarra	Giseli	Dráuzio	Igor	Dundun
Humberto	Luís	Ester	Janete	Fagundes
Líncon	Odilon	Irene	Joana	Feliciano
Linus	Patrícia	John	Júlio	Francisco
Maycol	Ricardo	Johny	Machado	Getúlio
Mengueli	Sabrina	Márcia	Natacha	Ivon
Plutarco	Sandra	Maurízio	Paulo	Keity
Raimundo	Tatiana	Osvaldo	Pereira	Liana
Rick	Tiradentes	Quinzinho	Susi	Mússio
Sasha	Wanessa	Sandro	Tales	Nenem
Ulaima	Wilson	Tiago	Valdeli	Pauline
Xanaia	Xavier	Tobias	Virgílio	Salviano
Zeus	Xuxa	Wilson	Xistos	Talita

F06	***F07***	***F08***	***F09***	***F10***
Adão	Anael	Afrodite	Abraão	Alcemir
Arley	Atos	Aramides	Artemis	Argel
Bruna	Bartolomeu	Borgonho	Benvindo	Bismarki
Daniele	Bedita	Camilo	Cassiel	Brígida
Eva	Contíjio	Cindi	Claudiomiro	Carvalho
Gertrudes	Diórgines	Cóssimo	Demóstines	Charles
Guarabira	Elena	Clodovil	Domingas	Domingos
Henrique	Faramildes	Djalminha	Edivanir	Dora
Jaílson	Guadalupe	Eleonor	Fragélico	Élsio
Josefina	Jesebel	Emanueli	Gílson	Frederico
Karina	Jesus	Eufrozina	Joelmir	Galvão
Kleiton	Katatau	Felícia	Júpiter	Juarez
Marina	Lucélia	Lanceloti	Kramer	Kekulê
Meloni	Miró	Messias	Lazaroni	Luana
Menezes	Narciso	Ozéias	Makino	Maira
Molina	Piaget	Policarpo	Menem	Policastro
Pitágoras	Ruth	Rincon	Paraíso	Renner
Plínio	Shakira	Ruan	Rubens	Selma
Sérgio	Vampeta	Selton	Serafim	Virginha
Tina	Xexeu	Vânia	Vanderley	Zulu

Locutores Masculinos

M01	***M02***	***M03***	***M04***	***M05***
Alexandra	Andrislane	Araken	Augusto	Andressa
Benedito	Antunes	Astrogildo	Bonfin	Antônia
Brigite	Brasilina	Bernardo	Bóris	Beatriz
César	Celso	Dartanhan	Catherine	Carlos
Clínton	Cláudia	Diana	Cumegundes	Charuto
Cuca	Cornélio	Dino	Eustáquio	Cristiana
Dalila	Dutra	Eliane	Feliz	Fábio
Duran	Edmílson	Eurípedes	Freud	Giovani
Eugênio	Guilherme	Felisberto	Hermegrobildo	João
Expedito	Haydee	Félix	Jupira	Joaquim
Glauter	Katiúcia	Ferreira	Kênia	José
Jezel	Maximíliam	Godofredo	Ludmila	Kátia
Leopoldo	Poliana	Guido	Marcela	Leovegildo
Merlin	Quaresma	Haroldo	Martinha	Manuel
Natali	Quitéria	Marcos	Miashita	Pelé
Petrúcio	Simoni	Matsuda	Natália	Pelópidas
Picolloto	Soraia	Ronaldo	Tatiane	Sílvia
Robson	Waldeci	Shirley	Ted	Simões
Suzano	Washington	Souza	Tinoco	Taís
Vítor	Watanabe	Wagner	Tito	Vitória

M06	***M07***	***M08***	***M09***	***M10***
Advã	Anabeli	Arimatéia	Arquimedes	Andréa
Apuã	Bastos	Baltazar	Borges	Bethoven
Cláudio	Beto	Crispim	Cibeli	Brizola
Dali	Cassiopéia	Danton	Cícero	Cleópatra
Erick	David	Denílson	Cruzoé	Dilermando
Glauber	Elisa	Ediméia	Dacildo	Duarte
Guaíra	Epaminondas	Gilberto	Damaceno	Euclídes
Karen	Ernin	Hernani	Edílson	Fabiel
Luciana	Fabiana	Ivete	Elivelton	Gabriel
Mussun	Fabrício	Josiane	Glória	Hélder
Pontes	Gilmar	Lucineide	Ieda	Ivo
Priscila	Ianka	Magno	Joshua	Juméia
Raquel	Josilene	Marcel	Kenedy	Lucrécia
Rodrigo	Luzia	Meireles	Letícia	Mara
Samanta	Minerva	Percivaldo	Maciel	Martinho
Talmo	Montovane	Rebeca	Micheli	Nálberti
Telma	Pietro	Rodrigues	Mijardela	Pitolomeu
Tuca	Resvala	Sizila	Noé	Reinaldo
Walter	Tieta	Veloso	Perlla	Sueli
Ynoguti	Zaratustra	Zurique	Ronivon	Vladimir

Apêndice C:

Dicionário de Pronúncias

Neste apêndice são mostrados os fonemas e as locuções de teste utilizadas no arquivo de vocabulário. As locuções de teste (nomes e apelidos) iniciadas com letra maiúscula representam locução pronunciada por locutor masculino e as iniciadas com letra minúscula representam locução pronunciada por locutor feminino.

A estrutura deste arquivo é a seguinte:

transc. gráfica / transc. fonética / média da duração (ms) / desv. padrão da dur. / classe das palavras

Como não utilizamos o modelo de duração no sistema, foi inserido valor "zero" nos valores da média e do desvio padrão. E, por não haver necessidade do uso de um modelo de gramática no sistema, a classe será definida como geral.

***fonemas**

#	f
a	g
A	j
an	k
e	l
E	L
en	m
i	n
y	N
in	p
o	r
O	rr
on	R
u	RR
w	s
un	t
b	T
d	v
D	x
	z

***fim**

***vocab**

, / # / 0 / 0 / geral
abraão / A b r A an w / 0 / 0 / geral
aderton / A d e rr t on / 0 / 0 / geral
adão / A d an w / 0 / 0 / geral
adnã / A D y n an / 0 / 0 / geral
Advã / A D y v an / 0 / 0 / geral
afrodite / A f r o D i T y / 0 / 0 / geral
albetã / A w b e t an / 0 / 0 / geral
alcemir / A w s e m i rr / 0 / 0 / geral
alda / a w d A / 0 / 0 / geral
alessandra / A l e s an d r A / 0 / 0 / geral
alex / A l E k y s y / 0 / 0 / geral
Alexandra / A l e x an d r A / 0 / 0 / geral
Anabeli / A n A b E l i / 0 / 0 / geral
anael / A n A E w / 0 / 0 / geral
Andréa / an d r E A / 0 / 0 / geral
Andressa / an d r e s A / 0 / 0 / geral
Andrislane / an d r i s l an n y / 0 / 0 / geral
Antônia / an t o n y A / 0 / 0 / geral
Antunes / an t u n y x / 0 / 0 / geral
Apuã / A p u an / 0 / 0 / geral
Araken / A r A k en / 0 / 0 / geral
aramides / A r A m i d y s / 0 / 0 / geral
arantes / A r an T y s / 0 / 0 / geral
argel / A rr j E w / 0 / 0 / geral
Arimatéia / A r i m A t E y A / 0 / 0 / geral
arley / A rr l e y / 0 / 0 / geral
armelinda / A R m e l in d A / 0 / 0 / geral
Arquimedes / A rr k i m E D y s / 0 / 0 / geral
artemis / A rr t e m i s / 0 / 0 / geral
Astrogildo / A s t r o j y w d w / 0 / 0 / geral
atírson / A t i r s on / 0 / 0 / geral
atos / a t w s / 0 / 0 / geral
Augusto / A w g u s t w / 0 / 0 / geral
Baltazar / b A w t A z A rr / 0 / 0 / geral
bárbara / b a RR b A r A / 0 / 0 / geral
barbie / b a RR b y / 0 / 0 / geral
bartolomeu / b A RR t o l o m e w / 0 / 0 / geral
Bastos / b a s t u s / 0 / 0 / geral
Beatriz / b e A t r i s / 0 / 0 / geral
bedita / b e D i t A / 0 / 0 / geral
belém / b e l en / 0 / 0 / geral
Benedito / b e n e D i t w / 0 / 0 / geral
benvindo / b en v in d w / 0 / 0 / geral
Bernardo / b E rr n A rr d w / 0 / 0 / geral
beth / b E T y / 0 / 0 / geral
Bethoven / b e t o v en / 0 / 0 / geral
Beto / b E t o / 0 / 0 / geral
bianca / b y an k A / 0 / 0 / geral
bismarki / b i s m A rr k i / 0 / 0 / geral
Bonfin / b on f in / 0 / 0 / geral
bonifácio / b o n y f a s i w / 0 / 0 / geral
Borges / b O RR j y x / 0 / 0 / geral
borgonho / b O R g o N w / 0 / 0 / geral
Bóris / b O r i s / 0 / 0 / geral
Brasilina / b r A z i l i n A / 0 / 0 / geral
brígida / b r i j i d a / 0 / 0 / geral
Brigite / b r i j i T y / 0 / 0 / geral
Brizola / b r i z O l A / 0 / 0 / geral
bruna / b r u n A / 0 / 0 / geral
bruno / b r u n w / 0 / 0 / geral
camila / k an m i l A / 0 / 0 / geral
camilo / k an m i l w / 0 / 0 / geral
camões / k A m on y s / 0 / 0 / geral
Carlos / k a R l w s / 0 / 0 / geral
carolina / k A r o l i n A / 0 / 0 / geral
carvalho / k A rr v A L w / 0 / 0 / geral
casagrande / k A z A g r an D y / 0 / 0 / geral
cassiel / k A s i E w / 0 / 0 / geral
Cassiopéia / k A s i o p E y A / 0 / 0 / geral
Catherine / k A t e r i n y / 0 / 0 / geral
Celso / s E w s w / 0 / 0 / geral
César / s E z A RR / 0 / 0 / geral
charles / x a rr l e s / 0 / 0 / geral
Charuto / x A r u t w / 0 / 0 / geral
Cibeli / s i b E l i / 0 / 0 / geral
Cícero / s i s e r w / 0 / 0 / geral
cindi / s in D i / 0 / 0 / geral
Cláudia / k l a w d y A / 0 / 0 / geral
Cláudio / k l a w d y w / 0 / 0 / geral
claudiomiro / k l A w D y o m i r w / 0 / 0 / geral
Cleópatra / k l E O p A t r A / 0 / 0 / geral
clemildes / k l e m y w d y x / 0 / 0 / geral
Clínton / k l in t un / 0 / 0 / geral
contíjio / k on T i j y w / 0 / 0 / geral
coralina / k o r A l i n A / 0 / 0 / geral
Cornélio / k O rr n E l y w / 0 / 0 / geral
cóssimo / k O s i m w / 0 / 0 / geral
creuza / k r e w z A / 0 / 0 / geral
Crispin / k r i s p in / 0 / 0 / geral
Cristiana / k r i x t i an n A / 0 / 0 / geral
Cruzoé / k r u z o E / 0 / 0 / geral
Cuca / k u k A / 0 / 0 / geral
Cumegundes / k u m e g un d y s / 0 / 0 / geral
cunha / k u N A / 0 / 0 / geral
custódia / k u s t O D y A / 0 / 0 / geral
Dacildo / d A s i w d w / 0 / 0 / geral
Dali / d A l y / 0 / 0 / geral
Dalila / d A l i l A / 0 / 0 / geral
Damaceno / d an m A s e n w / 0 / 0 / geral
daniele / d an n y E l y / 0 / 0 / geral
Danton / d an t on / 0 / 0 / geral

darci / d a rr s i / 0 / 0 / geral
Dartanhan / d A rr t an N an / 0 / 0 / geral
David / d e y v i D y / 0 / 0 / geral
demóstines / d e m O s T i n i x / 0 / 0 / geral
Denílson / d e n i w s on / 0 / 0 / geral
Diana / D y an n A / 0 / 0 / geral
didi / D i D i / 0 / 0 / geral
Dilermando / D i l e rr m an d w / 0 / 0 / geral
dimitre / D i m i t r y / 0 / 0 / geral
Dino / D i n w / 0 / 0 / geral
diorgines / D y O RR j i n y s / 0 / 0 / geral
djalminha / D j a w m i N a / 0 / 0 / geral
domingas / d o m in g A x / 0 / 0 / geral
domingos / d o m in g u s / 0 / 0 / geral
dora / d O r A / 0 / 0 / geral
dráuzio / d r a w z y w / 0 / 0 / geral
Duarte / d u A rr T y / 0 / 0 / geral
dundum / d un d un / 0 / 0 / geral
dunga / d un g A / 0 / 0 / geral
Duran / d u r an / 0 / 0 / geral
Dutra / d u t r A / 0 / 0 / geral
Edílson / e d y w s on / 0 / 0 / geral
Ediméia / e D i m E y A / 0 / 0 / geral
edivanir / e D i v a n i rr / 0 / 0 / geral
Edmílson / e d y m y w s on / 0 / 0 / geral
elaine / e l an y n y / 0 / 0 / geral
elena / e l e n A / 0 / 0 / geral
eleonor / e l e o n o R / 0 / 0 / geral
Eliane / e l i an n y / 0 / 0 / geral
Elisa / e l i z A / 0 / 0 / geral
Elivelton / e l i v E w t on / 0 / 0 / geral
élsio / E w s y w / 0 / 0 / geral
emanueli / e m a n u E l y / 0 / 0 / geral
émerson / E m e RR s on / 0 / 0 / geral
emília / e m i l y A / 0 / 0 / geral
Epaminondas / e p A m i n on d A s / 0 / 0 / geral
Erick / E r i k y / 0 / 0 / geral
Ernim / e R n in / 0 / 0 / geral
ester / y x t E rr / 0 / 0 / geral
Euclídes / e w k l i D y s / 0 / 0 / geral
eufrozina / e w f r o z i n a / 0 / 0 / geral
Eugênio / e w j e n y w / 0 / 0 / geral
Eurípedes / e w r i p e d y s / 0 / 0 / geral
Eustáquio / e w x t a k y w / 0 / 0 / geral
eva / E v A / 0 / 0 / geral
Expedito / e x p e d i t w / 0 / 0 / geral
Fabiana / f A b y an n A / 0 / 0 / geral
Fabiel / f A b y E w / 0 / 0 / geral
Fábio / f a b y w / 0 / 0 / geral
Fabrício / f A b r i s y o / 0 / 0 / geral
fagundes / f A g un d y s / 0 / 0 / geral
faramildes / f A r A m y w d i s / 0 / 0 / geral
felícia / f e l i s y A / 0 / 0 / geral
feliciano / f e l i s y an n w / 0 / 0 / geral
Felisberto / f e l i s b E r t w / 0 / 0 / geral
Félix / f E l i k i s / 0 / 0 / geral
Feliz / f e l i s / 0 / 0 / geral
Ferreira / f e rr e y r A / 0 / 0 / geral
flávio / f l a v y w / 0 / 0 / geral
fragélico / f r A j E l i k u / 0 / 0 / geral
francine / f r an s i n y / 0 / 0 / geral
francisco / f r an s i s k w / 0 / 0 / geral
frederico / f r e d e r i k w / 0 / 0 / geral
Freud / f r O y D y / 0 / 0 / geral
Gabriel / g A b r y E w / 0 / 0 / geral
gabriela / g A b r y E l A / 0 / 0 / geral
galvão / g A w v an w / 0 / 0 / geral
gamarra / g A m a rr A / 0 / 0 / geral
gertrudes / j e r t r u d y s / 0 / 0 / geral
getúlio / j e t u L w / 0 / 0 / geral
Gilberto / j i w b E rr t w / 0 / 0 / geral
gilce / j y w s y / 0 / 0 / geral
Gilmar / j y w m a RR / 0 / 0 / geral
gílson / j y w s on / 0 / 0 / geral
Giovani / j y o v an n i / 0 / 0 / geral
giseli / j i z E l y / 0 / 0 / geral
Glauber / g l a w b e R / 0 / 0 / geral
Glauter / g l a w t e RR / 0 / 0 / geral
Glória / g l O r y A / 0 / 0 / geral
Godofredo / g o d o f r e d w / 0 / 0 / geral
goreti / g o r E T y / 0 / 0 / geral
guadalupe / g w A d A l u p y / 0 / 0 / geral
Guaíra / g w A y r A / 0 / 0 / geral
guarabira / g w A r A b i r A / 0 / 0 / geral
Guido / g i d w / 0 / 0 / geral
Guilherme / g i L E r m y / 0 / 0 / geral
Haroldo / A r o w d w / 0 / 0 / geral
Haydee / A y d e / 0 / 0 / geral
Hélder / E w d e rr / 0 / 0 / geral
henrique / en rr i k y / 0 / 0 / geral
Hermegrobildo / e rr m e g r o b y w d w / 0 / 0 / geral
Hernani / e rr n an n i / 0 / 0 / geral
humberto / un b E rr t w / 0 / 0 / geral
Ianka / i an k a / 0 / 0 / geral
Ieda / i e d a / 0 / 0 / geral
igor / i g o RR / 0 / 0 / geral
irene / i r e n y / 0 / 0 / geral
Ivete / i v E T / 0 / 0 / geral
Ivo / i v w / 0 / 0 / geral
ivon / i v on / 0 / 0 / geral
jaílson / j A y w s on / 0 / 0 / geral
janete / j an n E T / 0 / 0 / geral
jesebel / j e z e b E w / 0 / 0 / geral
jesus / j e z u s / 0 / 0 / geral
Jezel / j E z E w / 0 / 0 / geral
joana / j o an n A / 0 / 0 / geral
João / j w an w / 0 / 0 / geral

Joaquim / j o A k in / 0 / 0 / geral
joelmir / j o e w m i rr / 0 / 0 / geral
john / D y on / 0 / 0 / geral
johny / D y on n i / 0 / 0 / geral
José / j w z E / 0 / 0 / geral
josefina / j o z e f i n A / 0 / 0 / geral
Joshua / j O x w A / 0 / 0 / geral
Josiane / j o z i an n y / 0 / 0 / geral
Josilene / j o z i l e n e / 0 / 0 / geral
juarez / j w A r e s / 0 / 0 / geral
júlio / j u L w / 0 / 0 / geral
Juméia / j u m E y A / 0 / 0 / geral
Jupira / j u p i r A / 0 / 0 / geral
júpiter / j u p i t e rr / 0 / 0 / geral
Karen / k a r en / 0 / 0 / geral
karina / k a r i n A / 0 / 0 / geral
katatau / k A t A t A w / 0 / 0 / geral
Kátia / k a T y A / 0 / 0 / geral
Katiúcia / k A T y w s y A / 0 / 0 / geral
keity / k e y T i / 0 / 0 / geral
kekulê / k e k u l e / 0 / 0 / geral
Kenedy / k e n e D i / 0 / 0 / geral
Kênia / k e N y A / 0 / 0 / geral
kleiton / k l e y t on / 0 / 0 / geral
kramer / k r an m e / 0 / 0 / geral
lanceloti / l an s e l O T y / 0 / 0 / geral
lazaroni / l A z A r on n i / 0 / 0 / geral
Leopoldo / l e o p o w d w / 0 / 0 / geral
Leovegildo / l e o v e g y w d w / 0 / 0 / geral
Letícia / l e T i s y A / 0 / 0 / geral
liana / l i an n A / 0 / 0 / geral
líncon / l in k on / 0 / 0 / geral
linus / l i n u x / 0 / 0 / geral
luana / l u an n A / 0 / 0 / geral
lucélia / l u s E l y A / 0 / 0 / geral
Luciana / l u s y an n A / 0 / 0 / geral
Lucineide / l u s i n e y D i / 0 / 0 / geral
Lucrécia / l u k r E s y A / 0 / 0 / geral
Ludmila / l u d y m i l A / 0 / 0 / geral
luís / l u y s / 0 / 0 / geral
Luzia / l u z y A / 0 / 0 / geral
machado / m A x A d w / 0 / 0 / geral
Maciel / m A s y E w / 0 / 0 / geral
Magno / m a g i n w / 0 / 0 / geral
maira / m a y r A / 0 / 0 / geral
makino / m a k i n u / 0 / 0 / geral
Manuel / m A n u E w / 0 / 0 / geral
Mara / m a r A / 0 / 0 / geral
Marcel / m A rr s E w / 0 / 0 / geral
Marcela / m A rr s E l A / 0 / 0 / geral
márcia / m a rr s y A / 0 / 0 / geral
Marcos / m a rr k w s / 0 / 0 / geral
marina / m a r i n A / 0 / 0 / geral
Martinha / m A rr t i n y A / 0 / 0 / geral
Martinho / m a rr T i N w / 0 / 0 / geral
Matsuda / m A T y s u d A / 0 / 0 / geral
maurízio / m A w r i z y w / 0 / 0 / geral
Maximíliam / m A x i m i l y an / 0 / 0 / geral
maycol / m a y k o / 0 / 0 / geral
Meireles / m e y r E l y s / 0 / 0 / geral
meloni / m E l o n i / 0 / 0 / geral
menem / m e n en / 0 / 0 / geral
menezes / m e n e z y s / 0 / 0 / geral
mengueli / m en g E l y / 0 / 0 / geral
Merlin / m E rr l in / 0 / 0 / geral
messias / m e s y A s / 0 / 0 / geral
Miashita / m y A x i t A / 0 / 0 / geral
Micheli / m i x E l i / 0 / 0 / geral
Mijardela / m i j A rr d E l A / 0 / 0 / geral
Minerva / m i n e RR v A / 0 / 0 / geral
miró / m i r O / 0 / 0 / geral
molina / m O l y n A / 0 / 0 / geral
Montovane / m on t o v an n y / 0 / 0 / geral
mússio / m u s y w / 0 / 0 / geral
Mussum / m u s un / 0 / 0 / geral
Nálberti / n a w b e rr T i / 0 / 0 / geral
narciso / n A rr s i z w / 0 / 0 / geral
natacha / n A t A x A / 0 / 0 / geral
Natali / n A t A l i / 0 / 0 / geral
Natália / n A t a L A / 0 / 0 / geral
nenem / n e n en / 0 / 0 / geral
Noé / n O E / 0 / 0 / geral
odilon / o D i l on / 0 / 0 / geral
osvaldo / O s v A w d w / 0 / 0 / geral
ozéias / o z E y A s / 0 / 0 / geral
paraíso / p A r A i z u / 0 / 0 / geral
patrícia / p A t r i s y A / 0 / 0 / geral
pauline / p A w l i n y / 0 / 0 / geral
paulo / p a w l w / 0 / 0 / geral
Pelé / p e l E / 0 / 0 / geral
Pelópidas / p E l o p i d A s / 0 / 0 / geral
Percivaldo / p e rr s i v a w d u / 0 / 0 / geral
pereira / p e r e y r A / 0 / 0 / geral
Perlla / p E rr l A / 0 / 0 / geral
Petrúcio / p e t r u s y w / 0 / 0 / geral
piaget / p y A j e / 0 / 0 / geral
Picolloto / p i k o l o t w / 0 / 0 / geral
Pietro / p y e t r o / 0 / 0 / geral
pitágoras / p i t a g o r A s / 0 / 0 / geral
Pitolomeu / p i t o l o m e w / 0 / 0 / geral
plínio / p l i n y w / 0 / 0 / geral
plutarco / p l u t a rr k o / 0 / 0 / geral
Poliana / p o l y an n A / 0 / 0 / geral
policarpo / p o l i k a R p w / 0 / 0 / geral
policastro / p o l y k A s t r w / 0 / 0 / geral
Pontes / p on T y s / 0 / 0 / geral
Priscila / p r i s i l A / 0 / 0 / geral
Quaresma / k w A r e x m A / 0 / 0 / geral

quinzinho / k in z i N w / 0 / 0 / geral
Quitéria / k i t E r y a / 0 / 0 / geral
raimundo / rr a y m un d o / 0 / 0 / geral
Raquel / rr A k E w / 0 / 0 / geral
Rebeca / rr e b E k A / 0 / 0 / geral
Reinaldo / rr e y n a w d w / 0 / 0 / geral
renner / rr e n E rr / 0 / 0 / geral
Resvala / rr e s v A l A / 0 / 0 / geral
ricardo / rr i k a d w / 0 / 0 / geral
rick / rr i k i / 0 / 0 / geral
rincon / rr in k on / 0 / 0 / geral
Robson / rr O b y s un / 0 / 0 / g eral
Rodrigo / rr o d r i g w / 0 / 0 / geral
Rodrigues / rr o d r i g i s / 0 / 0 / geral
Ronaldo / rr o n a w d w / 0 / 0 / geral
Ronivon / rr o n i v on / 0 / 0 / geral
ruan / rr u an / 0 / 0 / geral
rubens / rr u b en x / 0 / 0 / geral
ruth / rr u T y / 0 / 0 / geral
sabrina / s A b r i n A / 0 / 0 / geral
salviano / s A w v y an n w / 0 / 0 / geral
Samanta / s A m an t A / 0 / 0 / geral
sandra / s an d r A / 0 / 0 / geral
sandro / s an d r w / 0 / 0 / geral
sasha / s a x A / 0 / 0 / geral
selma / s E w m A / 0 / 0 / geral
selton / s E w t on / 0 / 0 / geral
serafim / s E r A f in / 0 / 0 / geral
sérgio / s E r j y w / 0 / 0 / geral
shakira / x a k i r A / 0 / 0 / geral
Shirley / x i r l e y / 0 / 0 / geral
Sílvia / s y w v y A / 0 / 0 / geral
Simões / s i m on y s / 0 / 0 / geral
Simoni / s i m o n i / 0 / 0 / geral
Sizila / s i z i l A / 0 / 0 / geral
Soraia / s O r A y A / 0 / 0 / geral
Souza / s o w z A / 0 / 0 / geral
Sueli / s w e l i / 0 / 0 / geral
susi / s u z i / 0 / 0 / geral
Suzano / s u z A n w / 0 / 0 / geral
Taís / t A y s / 0 / 0 / geral
tales / t a l y s / 0 / 0 / geral
talita / t A l i t A / 0 / 0 / geral
Talmo / t a w m w / 0 / 0 / geral
tatiana / t A T y an n A / 0 / 0 / geral
Tatiane / t A t y an n y / 0 / 0 / geral
Ted / t E d y / 0 / 0 / geral
Telma / t E w m A / 0 / 0 / geral
tiago / T y a g w / 0 / 0 / geral
Tieta / T y e t A / 0 / 0 / geral
tina / t i n A / 0 / 0 / geral
Tinoco / t i n o k w / 0 / 0 / geral
tiradentes / T i r A d en t y s / 0 / 0 / geral
Tito / t i t w / 0 / 0 / geral
tobias / t o b y A x / 0 / 0 / geral
Tuca / t u k A / 0 / 0 / geral
ulaima / u l a y m A / 0 / 0 / geral
valdeli / v A w d e l i / 0 / 0 / geral
vampeta / v an p e t A / 0 / 0 / geral
vanderley / v an d e rr l e y / 0 / 0 / geral
vânia / v an n y A / 0 / 0 / geral
Veloso / v e l o z w / 0 / 0 / geral
virgílio / v i R j i l y w / 0 / 0 / geral
virginha / v i rr j i n y A / 0 / 0 / geral
Vítor / v i t o rr / 0 / 0 / geral
Vitória / v i t O r y A / 0 / 0 / geral
Vladimir / v l A D i m i rr / 0 / 0 / geral
Wagner / v a g i n e rr / 0 / 0 / geral
Waldeci / v A w d e s i / 0 / 0 / geral
Walter / v a w t e rr / 0 / 0 / geral
wanessa / v A n e s A / 0 / 0 / geral
Washington / w O x i t on / 0 / 0 / geral
Watanabe / w A t A n A b y / 0 / 0 / geral
wilson / w y w s on / 0 / 0 / geral
wilton / w y w t on / 0 / 0 / geral
xanaia / x A n a y A / 0 / 0 / geral
xavier / x A v y E R / 0 / 0 / geral
xexeu / x E x E w / 0 / 0 / geral
xistos / x i x t w s / 0 / 0 / geral
xuxa / x u x A / 0 / 0 / geral
Ynoguti / y n o g u T i / 0 / 0 / geral
Zaratustra / z A r A t u s t r A / 0 / 0 / geral
zeus / z e w x / 0 / 0 / geral
zulu / z u l u / 0 / 0 / geral
Zurique / z u r i k i / 0 / 0 / geral

***fim**

Apêndice D:

Locuções Reconhecidas.

Neste apêndice é mostrada uma tabela contendo as locuções reconhecidas utilizando-se normalização de locutor. A estrutura desta tabela é a seguinte:

locução pronunciada pelo locutor; locução reconhecida após normalização; verossimilhança obtida; duração da decodificação; classificação da locução correta.

Na primeira coluna desta tabela, as locuções iniciadas com letra minúscula representam a fala feminina (F01 a F10) e as iniciadas com letra maiúscula representam a fala masculina (M01 a M10).

Na coluna em que é apresentada a duração da busca pelas locuções (decodificação), utilizou-se como base para o cálculo deste tempo um microcomputador Pentium II (MMX – 300 MHz).

No caso de erro de decodificação, a palavra reconhecida erroneamente, a sua verossimilhança e o tempo de decodificação são representados em negrito. A classificação da locução correta, no caso de erro de decodificação, é também representada em negrito. Essa classificação é feita da seguinte maneira: a maior verrossimilhança é considerada de ordem 0 e define a palavra reconhecida; a verossimilhança seguinte é ordenada como 1 e corresponde ao candidato mais próximo, e assim por diante. Por exemplo, a palavra "Armelinda" foi reconhecida erroneamente como "Karina". Isto significa que a palavra "Karina" apresentou a maior verossimilhança (ordem 0). A classificação 7 mostrada na última coluna indica que a verossimilhança apresentada pela palavra correta, "Armelinda", está 7 posições abaixo (ordem 7).

A vírgula, antes e após a palavra reconhecida, representa a verificação do silêncio inicial e final na locução.

LOCUTOR F01				
aderton	, aderton ,	-1967,3660	00:00:05	-----
atírson	, atírson ,	-2110,3753	00:00:08	-----
bonifácio	, bonifácio ,	-2088,7574	00:00:05	-----
casagrande	, casagrande ,	-1900,0169	00:00:05	-----
dunga	, dunga ,	-1256,1243	00:00:03	-----
emília	, emília ,	-1770,9317	00:00:04	-----
francine	, francine ,	-2096,6262	00:00:05	-----
gamarra	, gamarra ,	-1604,9176	00:00:05	-----
humberto	, humberto ,	-1955,2226	00:00:04	-----
líncon	, líncon ,	-1696,3878	00:00:07	-----
linus	, linus ,	-1840,4182	00:00:05	-----
maycol	, maycol ,	-1445,6241	00:00:06	-----
mengueli	, mengueli ,	-1766,9488	00:00:04	-----
plutarco	, plutarco ,	-2055,3218	00:00:04	-----
raimundo	, raimundo ,	-1970,5349	00:00:05	-----
rick	, rick ,	-1676,9733	00:00:03	-----
sasha	, sasha ,	-1734,8170	00:00:04	-----
ulaima	, ulaima ,	-1802,0919	00:00:07	-----
xanaia	, xanaia ,	-1856,4811	00:00:05	-----
zeus	, zeus ,	-1788,4195	00:00:09	-----

LOCUTOR F02				
albetã	, albetã ,	-1590,6467	00:00:09	-----
alessandra	, alessandra ,	-1975,4712	00:00:06	-----
belém	, belém ,	-1474,1214	00:00:05	-----
bruno	, bruno ,	-1502,3280	00:00:05	-----
camila	**, shakira ,**	**-1196,9349**	**00:00:04**	**8ª**
elaine	, elaine ,	-1574,5978	00:00:05	-----
flávio	, flávio ,	-1501,9500	00:00:04	-----
giseli	, giseli ,	-1539,2491	00:00:05	-----
luís	, luís ,	-1213,4682	00:00:04	-----
odilon	, odilon ,	-1571,8531	00:00:05	-----
patrícia	, patrícia ,	-1700,3597	00:00:05	-----
ricardo	, ricardo ,	-1522,0052	00:00:04	-----
sabrina	, sabrina ,	-1894,8462	00:00:05	-----
sandra	**, Andréa ,**	**-1411,3355**	**00:00:05**	**1ª**
tatiana	, tatiana ,	-1822,0787	00:00:05	-----
tiradentes	, tiradentes ,	-2020,9449	00:00:07	-----
wanessa	, wanessa ,	-1565,3309	00:00:06	-----
wilson	, wilton ,	-1204,2058	00:00:05	-----
xavier	, xavier ,	-1148,2601	00:00:07	-----
xuxa	, xuxa ,	-1570,9596	00:00:09	-----

LOCUTOR F03				
alda	, alda ,	-1609,4874	00:00:08	-----
alex	, alex ,	-2193,0193	00:00:18	-----
bianca	**, Ianka ,**	**-1835,7599**	**00:00:21**	**14ª**
clemildes	, clemildes ,	-2756,2387	00:00:13	-----
creuza	, creuza ,	-2562,8274	00:00:11	-----
didi	, didi ,	-1768,1339	00:00:08	-----
dimitre	, dimitre ,	-2143,3518	00:00:14	-----
dráuzio	, dráuzio ,	-2096,7843	00:00:06	-----
ester	, ester ,	-2206,2746	00:00:07	-----
irene	**, Cibeli ,**	**-1605,6088**	**00:00:06**	**1ª**
john	**, Diana ,**	**-1666,5989**	**00:00:06**	**3ª**
johny	, johny ,	-2095,1709	00:00:05	-----
márcia	, márcia ,	-1957,0838	00:00:06	-----
maurízio	, maurízio ,	-2081,2528	00:00:17	-----
osvaldo	, osvaldo ,	-2425,7647	00:00:06	-----
quinzinho	, quinzinho ,	-2179,4063	00:00:07	-----
sandro	, sandro ,	-1707,9505	00:00:09	-----
tiago	, tiago ,	-2029,9271	00:00:09	-----
tobias	**, Arquimedes ,**	**-1891,4720**	**00:00:05**	**5ª**
wilson	, wilson ,	-1570,5854	00:00:04	-----

LOCUTOR F04				
armelinda	**, karina ,**	**-1807,2623**	**00:00:08**	**7ª**
bárbara	, bárbara ,	-1679,7510	00:00:06	-----
barbie	, barbie ,	-1675,7791	00:00:08	-----
beth	, beth ,	-2412,5454	00:00:09	-----
émerson	, émerson ,	-1924,1243	00:00:06	-----
gabriela	, gabriela ,	-2375,4488	00:00:09	-----
gilce	, gilce ,	-2008,1173	00:00:06	-----
igor	, igor ,	-1636,9754	00:00:05	-----
janete	, janete ,	-2049,3867	00:00:07	-----
joana	, joana ,	-1740,7949	00:00:06	-----
júlio	, júlio ,	-2059,3326	00:00:06	-----
machado	, machado ,	-2095,1701	00:00:07	-----
natacha	**, sasha ,**	**-2237,6271**	**00:00:07**	**1ª**
paulo	**, Talmo ,**	**-1548,1401**	**00:00:06**	**1ª**
pereira	, pereira ,	-1225,1450	00:00:07	-----
susi	, susi ,	-1830,8429	00:00:06	-----
tales	, tales ,	-1512,7436	00:00:06	-----
valdeli	**, Nálberti ,**	**-2020,0654**	**00:00:07**	**1ª**
virgílio	, virgílio ,	-1848,7444	00:00:06	-----
xistos	, xistos ,	-2064,4173	00:00:07	-----

LOCUTOR F05				
adnã	, adnã ,	-2167,3847	00:00:16	-----
arantes	**, Karen ,**	**-1659,4646**	**00:00:06**	**8ª**
camões	, camões ,	-1927,2851	00:00:06	-----
carolina	, carolina ,	-2718,2779	00:00:08	-----
coralina	, coralina ,	-2490,3801	00:00:07	-----
cunha	, cunha ,	-1768,6525	00:00:04	-----
custódia	, custódia ,	-2546,3931	00:00:07	-----
dundun	, dundum ,	-2242,7817	00:00:07	-----
fagundes	, fagundes	-1688,7375	00:00:06	-----
feliciano	, feliciano ,	-3148,9244	00:00:09	-----
francisco	, francisco ,	-2986,5352	00:00:08	-----
getúlio	, getúlio ,	-2741,6849	00:00:07	-----
ivon	, ivon ,	-1824,5278	00:00:07	-----
keity	, keity ,	-2014,2295	00:00:06	-----
liana	, liana ,	-1724,4307	00:00:06	-----
mússio	, mússio ,	-2118,0360	00:00:06	-----
nenem	, nenem ,	-2430,2223	00:00:07	-----
pauline	**, rubens ,**	**-2158,3309**	**00:00:06**	**1ª**
salviano	, salviano ,	-2950,8754	00:00:10	-----
talita	, talita ,	-2409,9119	00:00:08	-----

LOCUTOR F06				
adão	**, gamarra ,**	**-1689,8067**	**00:00:08**	**30ª**
arley	**, charles ,**	**-1802,7857**	**00:00:06**	**1ª**
bruna	, bruna ,	-1363,1174	00:00:04	-----
daniele	, daniele ,	-1886,7065	00:00:06	-----
eva	, eva ,	-1582,7301	00:00:06	-----
gertrudes	, gertrudes ,	-2541,6403	00:00:07	-----
guarabira	, guarabira ,	-2166,7478	00:00:08	-----
henrique	, henrique ,	-2361,3970	00:00:08	-----
jaílson	, jaílson ,	-2910,5838	00:00:15	-----
josefina	, josefina ,	-2463,1218	00:00:08	-----
karina	, karina ,	-2232,4045	00:00:08	-----
kleiton	, kleiton ,	-1619,8679	00:00:05	-----
marina	, marina ,	-1379,4182	00:00:07	-----
meloni	, meloni ,	-1988,8061	00:00:08	-----
menezes	, menezes ,	-1521,6563	00:00:06	-----
molina	, molina ,	-2133,7114	00:00:09	-----
pitágoras	, pitágoras ,	-1730,5620	00:00:07	-----
plínio	, plínio ,	-1736,7885	00:00:05	-----
sérgio	, sérgio ,	-1344,1032	00:00:05	-----
tina	, tina ,	-1197,4774	00:00:09	-----

LOCUTOR F07				
anael	, anael ,	-1962,8897	00:00:06	-----
atos	, atos ,	-1965,9122	00:00:05	-----
bartolomeu	, bartolomeu ,	-2548,8970	00:00:07	-----
bedita	, bedita ,	-2113,7702	00:00:06	-----
contíjio	, contíjio ,	-1975,9070	00:00:05	-----
diorgines	, diorgines ,	-2039,4203	00:00:07	-----
elena	, elena ,	-1772,8166	00:00:04	-----
faramildes	, faramildes ,	-2203,9999	00:00:07	-----
guadalupe	, guadalupe ,	-2360,5531	00:00:05	-----
jesebel	, jesebel ,	-2053,8722	00:00:06	-----
jesus	, jesus ,	-2155,7826	00:00:06	-----
katatau	, katatau ,	-2152,3470	00:00:06	-----
lucélia	, lucélia ,	-1958,6394	00:00:05	-----
miró	**, feliciano ,**	**-1803,4804**	**00:00:05**	**5ª**
narciso	, narciso ,	-2286,0730	00:00:06	-----
piaget	**, Eugênio ,**	**-1672,1928**	**00:00:06**	**40ª**
ruth	, ruth ,	-1936,9750	00:00:06	-----
shakira	, shakira ,	-2250,0327	00:00:05	-----
vampeta	, vampeta ,	-2145,5945	00:00:05	-----
xexeu	, xexeu ,	-1938,7769	00:00:06	-----

LOCUTOR F08				
afrodite	, afrodite ,	-2134,0872	00:00:05	-----
aramides	, aramides ,	-1770,3923	00:00:05	-----
borgonho	, borgonho ,	-1659,2171	00:00:04	-----
camilo	, camilo ,	-1394,6292	00:00:03	-----
cindi	, cindi,	-1031,9735	00:00:02	-----
clodovil	, clodovil,	-1730,8442	00:00:07	-----
cóssimo	, cóssimo,	-1361,1299	00:00:02	-----
djalminha	, djalminha ,	-2264,6107	00:00:04	-----
eleonor	, eleonor,	-1543,3869	00:00:03	-----
emanueli	, emanueli,	-1473,8340	00:00:04	-----
eufrozina	, eufrozina,	-2451,4632	00:00:04	-----
felícia	, felícia,	-1558,9788	00:00:03	-----
lanceloti	, lanceloti,	-2043,5241	00:00:04	-----
messias	**, Miashita ,**	**-1631,8189**	**00:00:03**	**14ª**
ozéias	, ozéias,	-1119,9095	00:00:06	-----
policarpo	, policarpo,	-1974,5641	00:00:04	-----
rincon	**, líncon ,**	**-1740,8451**	**00:00:03**	**1ª**
ruan	, ruan ,	-1149,9361	00:00:02	-----
selton	, selton,	-1380,3982	00:00:04	-----
vânia	**, Dalila,**	**-1349,4691**	**00:00:03**	**10ª**

LOCUTOR F09				
abraão	, abraão ,	-1568,2995	00:00:04	**-----**
artemis	**, afrodite ,**	**-2029,2006**	**00:00:05**	**2ª**
benvindo	, benvindo ,	-2025,0430	00:00:05	**-----**
cassiel	, cassiel ,	-1807,2537	00:00:05	**-----**
claudiomiro	, claudiomiro ,	-2300,9746	00:00:06	**-----**
demóstines	, demóstines ,	-2204,5799	00:00:06	**-----**
domingas	, domingas ,	-1869,4009	00:00:05	**-----**
edivanir	, edivanir ,	-1766,3285	00:00:04	**-----**
fragélico	, fragélico ,	-2138,4748	00:00:06	**-----**
gílson	, gílson ,	-2081,1791	00:00:04	**-----**
joelmir	, joelmir ,	-1757,7629	00:00:05	**-----**
júpiter	, júpiter ,	-2081,9068	00:00:05	**-----**
kramer	, kramer ,	-1934,6166	00:00:05	**-----**
lazaroni	, lazaroni ,	-2091,1672	00:00:06	**-----**
makino	, makino ,	-2372,9310	00:00:09	**-----**
menen	**, benvindo ,**	**-1730,7110**	**00:00:04**	**1ª**
paraíso	, paraíso ,	-2094,3307	00:00:06	**-----**
rubens	, rubens ,	-1749,8214	00:00:05	**-----**
serafim	, serafim ,	-1943,6985	00:00:05	**-----**
vanderley	, vanderley ,	-1821,4325	00:00:12	**-----**

LOCUTOR F10				
alcemir	, alcemir ,	-1918,0375	00:00:08	**-----**
argel	**, Marcel ,**	**-1331,5312**	**00:00:03**	**2ª**
bismarki	, bismarki ,	-1892,9320	00:00:04	**-----**
brígida	**, bedita ,**	**-1389,3678**	**00:00:03**	**5ª**
carvalho	, carvalho ,	-1412,7940	00:00:06	**-----**
charles	, charles ,	-1352,5777	00:00:03	**-----**
domingos	, domingos	-1847,0570	00:00:04	**-----**
dora	, dora ,	-931,2043	00:00:02	**-----**
élsio	, élsio ,	-1438,2601	00:00:03	**-----**
frederico	, frederico ,	-1826,1169	00:00:04	**-----**
galvão	, galvão ,	-1289,0382	00:00:03	**-----**
juarez	, juarez ,	-1611,0838	00:00:04	**-----**
kekulê	, kekulê ,	-1578,0711	00:00:03	**-----**
luana	, luana ,	-1417,3168	00:00:03	**-----**
maira	, maira ,	-1078,0847	00:00:03	**-----**
policastro	, policastro ,	-2473,9610	00:00:04	**-----**
renner	, renner ,	-1156,6860	00:00:02	**-----**
selma	, selma ,	-1311,1629	00:00:02	**-----**
virginha	**, Cícero ,**	**-1826,2399**	**00:00:04**	**1ª**
zulu	, zulu ,	-1351,1124	00:00:03	**-----**

LOCUTOR M01				
Alexandra	, Alexandra ,	-3087,5488	00:00:07	-----
Benedito	, Benedito ,	-2948,9478	00:00:07	-----
Brigite	, Brigite ,	-3033,5153	00:00:10	-----
César	, César ,	-3187,6568	00:00:10	-----
Clínton	, Clínton ,	-2622,0247	00:00:06	-----
Cuca	, Cuca ,	-2763,1157	00:00:06	-----
Dalila	, Dalila ,	-1907,4047	00:00:05	-----
Duran	, Duran ,	-2006,8430	00:00:06	-----
Eugênio	, Eugênio ,	-2829,3073	00:00:06	-----
Expedito	, Expedito ,	-2625,4539	00:00:07	-----
Glauter	**, juarez**	**-2191,6115**	**00:00:05**	**4ª**
Jezel	, Jezel ,	-2767,7280	00:00:07	-----
Leopoldo	, Leopoldo ,	-2783,6317	00:00:07	-----
Merlin	, Merlin ,	-3038,6633	00:00:07	-----
Natali	, Natali ,	-2700,7061	00:00:11	-----
Petrúcio	, Petrúcio ,	-3694,6434	00:00:07	-----
Picolloto	, Picolloto ,	-3588,8823	00:00:08	-----
Robson	, Robson ,	-1883,9998	00:00:03	-----
Suzano	, Suzano ,	-1902,8463	00:00:04	-----
Vítor	**, frederico ,**	**-2562,7578**	**00:00:06**	**1ª**

LOCUTOR M02				
Andrislane	, Andrislane ,	-2857,3243	00:00:07	-----
Antunes	, Antunes ,	-1656,4885	00:00:05	-----
Brasilina	, Brasilina ,	-2064,2571	00:00:12	-----
Celso	, Celso ,	-1330,9542	00:00:04	-----
Cláudia	, Cláudia ,	-2452,0179	00:00:07	-----
Cornélio	, Cornélio ,	-1884,3604	00:00:06	-----
Dutra	, Dutra ,	-1263,6402	00:00:03	-----
Edmílson	, Edmílson ,	-2421,3589	00:00:06	-----
Guilherme	, Guilherme ,	-2358,7354	00:00:08	-----
Haydee	, Haydee ,	-1909,3097	00:00:08	-----
Katiúcia	, Katiúcia ,	-2872,2473	00:00:09	-----
Maximíliam	, Maximíliam ,	-2573,2138	00:00:09	-----
Poliana	, Poliana ,	-2437,0646	00:00:08	-----
Quaresma	, Quaresma ,	-2271,9573	00:00:06	-----
Quitéria	, Quitéria ,	-2188,4048	00:00:07	-----
Simoni	**, Simões ,**	**-1619,6700**	**00:00:04**	**1ª**
Soraia	, Soraia ,	-2349,3765	00:00:07	-----
Waldeci	, Waldeci ,	-2596,5453	00:00:08	-----
Washington	, Washington ,	-2715,6213	00:00:10	-----
Watanabe	, Watanabe ,	-3050,7115	00:00:15	-----

LOCUTOR M03				
Araken	, Araken ,	-1643,2917	00:00:05	-----
Astrogildo	, Astrogildo ,	-2065,7641	00:00:06	-----
Bernardo	, Bernardo ,	-2145,4901	00:00:06	-----
Dartanhan	, Dartanhan ,	-1913,7656	00:00:08	-----
Diana	, Diana ,	-1732,8985	00:00:08	-----
Dino	, Dino ,	-1985,9460	00:00:06	-----
Eliane	, Eliane ,	-1881,5707	00:00:06	-----
Eurípedes	, Eurípedes ,	-2457,9271	00:00:07	-----
Felisberto	, Felisberto ,	-1911,2457	00:00:06	-----
Félix	, Félix ,	-1966,3529	00:00:06	-----
Ferreira	, Ferreira ,	-1647,6428	00:00:06	-----
Godofredo	, Godofredo ,	-1874,6291	00:00:06	-----
Guido	, Guido ,	-1927,8495	00:00:06	-----
Haroldo	, Haroldo ,	-1758,3824	00:00:04	-----
Marcos	, Marcos ,	-2130,1082	00:00:06	-----
Matsuda	, Matsuda ,	-2834,3533	00:00:08	-----
Ronaldo	, Ronaldo ,	-2271,4702	00:00:07	-----
Shirley	, Shirley ,	-2406,1742	00:00:08	-----
Souza	, Souza ,	-2225,2039	00:00:07	-----
Wagner	, Wagner ,	-2921,0186	00:00:08	-----

LOCUTOR M04				
Augusto	, Augusto ,	-2455,3872	00:00:05	-----
Bonfin	, Bonfin ,	-2369,6091	00:00:13	-----
Bóris	, Bóris ,	-2107,4803	00:00:05	-----
Catherine	**, armelinda ,**	**-2231,4704**	**00:00:06**	**1ª**
Cumegundes	, Cumegundes ,	-2637,2931	00:00:06	-----
Eustáquio	, Eustáquio ,	-2263,4620	00:00:06	-----
Feliz	, Feliz ,	-2278,9828	00:00:06	-----
Freud	, Freud ,	-2054,9726	00:00:06	-----
Hermegrobildo	, Hermegrobildo ,	-2797,5513	00:00:08	-----
Jupira	, Jupira ,	-2416,2744	00:00:07	-----
Kênia	**, David ,**	**-1681,2272**	**00:00:05**	**59ª**
Ludmila	, Ludmila ,	-2442,8531	00:00:05	-----
Marcela	, Marcela ,	-2036,2749	00:00:06	-----
Martinha	**, Martinho ,**	**-2451,1693**	**00:00:07**	**1ª**
Miashita	, Miashita ,	-2212,0158	00:00:06	-----
Natália	, Natália ,	-2056,0930	00:00:05	-----
Tatiane	, Tatiane ,	-2026,0878	00:00:06	-----
Ted	, Ted ,	-1411,2831	00:00:04	-----
Tinoco	, Tinoco ,	-2407,5407	00:00:08	-----
Tito	, Tito ,	-1625,3755	00:00:04	-----

LOCUTOR M05				
Andressa	, Andressa ,	-1951,5085	00:00:05	-----
Antônia	, Antônia ,	-2005,2339	00:00:05	-----
Beatriz	, Beatriz ,	-2356,5908	00:00:07	-----
Carlos	, Carlos ,	-1638,1499	00:00:10	-----
Charuto	, Charuto ,	-2079,4507	00:00:06	-----
Cristiana	, Cristiana ,	-2102,2052	00:00:05	-----
Fábio	**, Fabiel ,**	**-1809,7609**	**00:00:05**	**1ª**
Giovani	**, João ,**	**-2048,2500**	**00:00:06**	**6ª**
João	, João ,	-1661,2774	00:00:05	-----
Joaquim	, Joaquim ,	-2068,1070	00:00:06	-----
José	, José ,	-1639,8297	00:00:04	-----
Kátia	, Kátia ,	-1781,5753	00:00:05	-----
Leovegildo	, Leovegildo ,	-2205,9005	00:00:07	-----
Manuel	, Manuel ,	-1818,8734	00:00:06	-----
Pelé	, Pelé ,	-1575,8928	00:00:05	-----
Pelópidas	**, Eurípedes ,**	**-2317,7056**	**00:00:06**	**1ª**
Sílvia	, Sílvia ,	-1997,8197	00:00:06	-----
Simões	, Simões ,	-2142,4899	00:00:05	-----
Taís	**, tales ,**	**-1740,9986**	**00:00:08**	**1ª**
Vitória	, Vitória ,	-1879,1190	00:00:05	-----

LOCUTOR M06				
Advã	, Advã ,	-1674,5792	00:00:05	-----
Apuã	, Apuã ,	-1812,2136	00:00:05	-----
Cláudio	**, contíjio ,**	**-1881,7396**	**00:00:04**	**2ª**
Dali	**, darci ,**	**-1708,1536**	**00:00:05**	**32ª**
Erick	, Erick ,	-1672,9250	00:00:05	-----
Glauber	, Glauber ,	-1808,7929	00:00:04	-----
Guaíra	, Guaíra ,	-2007,6698	00:00:11	-----
Karen	, Karen ,	-1324,7956	00:00:06	-----
Luciana	, Luciana ,	-1708,8837	00:00:04	-----
Mussun	, Mussum ,	-1829,1671	00:00:04	-----
Pontes	, Pontes ,	-1780,3992	00:00:05	-----
Priscila	, Priscila ,	-1920,5863	00:00:06	-----
Raquel	, Raquel ,	-1736,1375	00:00:08	-----
Rodrigo	, Rodrigo ,	-1987,6740	00:00:11	-----
Samanta	, Samanta ,	-2179,5558	00:00:05	-----
Talmo	**, Augusto ,**	**-1300,9352**	**00:00:04**	**1ª**
Telma	, Telma ,	-1658,3687	00:00:05	-----
Tuca	, Tuca ,	-1392,6710	00:00:04	-----
Walter	**, Guaíra ,**	**-1883,6296**	**00:00:06**	**1ª**
Ynoguti	, Ynoguti ,	-2016,8275	00:00:06	-----

LOCUTOR M07				
Anabeli	, Anabeli ,	-1386,2789	00:00:05	-----
Bastos	, Bastos ,	-1238,0106	00:00:04	-----
Beto	, Beto ,	-1311,6277	00:00:04	-----
Cassiopéia	, Cassiopéia ,	-1603,0840	00:00:06	-----
David	**, arley ,**	**-1099,2968**	**00:00:05**	**1ª**
Elisa	, Elisa ,	-1195,6213	00:00:04	-----
Epaminondas	, Epaminondas ,	-1555,2237	00:00:07	-----
Ernin	, Ernim ,	-1083,3009	00:00:05	-----
Fabiana	, Fabiana ,	-1945,0002	00:00:08	-----
Fabrício	, Fabrício ,	-1782,1562	00:00:06	-----
Gilmar	, Gilmar ,	-1207,8769	00:00:04	-----
Ianka	, Ianka ,	-1093,2296	00:00:05	-----
Josilene	, Josilene ,	-1332,8878	00:00:05	-----
Luzia	, Luzia ,	-1100,1374	00:00:05	-----
Minerva	, Minerva ,	-1003,5411	00:00:10	-----
Montovane	, Montovane ,	-1444,5361	00:00:05	-----
Pietro	**, pereira ,**	**-1174,4370**	**00:00:05**	**3ª**
Resvala	, Resvala ,	-1273,3904	00:00:05	-----
Tieta	, Tieta ,	-1262,5113	00:00:07	-----
Zaratustra	, Zaratustra ,	-1711,0155	00:00:07	-----

LOCUTOR M08				
Arimatéia	, Arimatéia ,	-1604,9024	00:00:05	-----
Baltazar	, Baltazar ,	-1501,6620	00:00:05	-----
Crispin	, Crispin ,	-1509,2826	00:00:04	-----
Danton	, Danton ,	-1216,4916	00:00:03	-----
Denílson	, Denílson ,	-1643,0716	00:00:09	-----
Ediméia	, Ediméia ,	-1261,0404	00:00:06	-----
Gilberto	, Gilberto ,	-1746,5503	00:00:04	-----
Hernani	, Hernani ,	-1551,3697	00:00:03	-----
Ivete	, Ivete ,	-1247,6328	00:00:03	-----
Josiane	, Josiane ,	-1543,4909	00:00:04	-----
Lucineide	, Lucineide ,	-1715,3122	00:00:04	-----
Magno	, Magno ,	-1180,8716	00:00:04	-----
Marcel	**, Maciel ,**	**-1328,8542**	**00:00:05**	**1ª**
Meireles	, Meireles	-1668,2599	00:00:09	-----
Percivaldo	, Percivaldo ,	-1714,5966	00:00:06	-----
Rebeca	, Rebeca ,	-1391,5119	00:00:05	-----
Rodrigues	, Rodrigues	-1623,9725	00:00:05	-----
Sizila	, Sizila ,	-1520,5147	00:00:05	-----
Veloso	, Veloso ,	-1659,1408	00:00:05	-----
Zurique	**, afrodite ,**	**-1494,6831**	**00:00:04**	**4ª**

LOCUTOR M09				
Arquimedes	, Arquimedes ,	-2027,8525	00:00:06	-----
Borges	, Borges ,	-1366,9302	00:00:04	-----
Cibeli	, Cibeli ,	-2137,0669	00:00:08	-----
Cícero	, Cícero ,	-2115,4691	00:00:08	-----
Cruzoé	, Cruzoé ,	-1183,3756	00:00:03	-----
Dacildo	, Dacildo ,	-1523,8457	00:00:05	-----
Damaceno	, Damaceno ,	-1609,7663	00:00:05	-----
Edílson	**, gílson ,**	**-1075,5766**	**00:00:02**	**4ª**
Elivelton	, Elivelton ,	-1770,0000	00:00:05	-----
Glória	, Glória ,	-1271,1690	00:00:03	-----
Ieda	, Ieda ,	-1396,1488	00:00:03	-----
Joshua	, Joshua ,	-2551,1751	00:00:05	-----
Kenedy	, Kenedy ,	-1973,6199	00:00:06	-----
Letícia	, Letícia ,	-1535,4664	00:00:05	-----
Maciel	, Maciel ,	-1743,7390	00:00:05	-----
Micheli	**, ester ,**	**-1397,1366**	**00:00:03**	**1ª**
Mijardela	, Mijardela ,	-1776,5635	00:00:04	-----
Noé	**, goreti ,**	**-1565,5017**	**00:00:04**	**9ª**
Perlla	**, Telma ,**	**-1156,3293**	**00:00:04**	**1ª**
Ronivon	, Ronivon ,	-2487,6920	00:00:07	-----

LOCUTOR M10				
Andréa	, Andréa ,	-1360,2707	00:00:06	-----
Bethoven	, Bethoven ,	-1683,9760	00:00:06	-----
Brizola	, Brizola ,	-1469,3299	00:00:06	-----
Cleópatra	, Cleópatra ,	-1647,5146	00:00:06	-----
Dilermando	, Dilermando ,	-1773,3420	00:00:13	-----
Duarte	, Duarte ,	-1636,2107	00:00:05	-----
Euclídes	, Euclídes ,	-1937,0265	00:00:06	-----
Fabiel	, Fabiel ,	-1587,8028	00:00:10	-----
Gabriel	, Gabriel ,	-1510,0716	00:00:08	-----
Hélder	, Hélder ,	-1346,0422	00:00:05	-----
Ivo	, Ivo ,	-1316,6747	00:00:05	-----
Juméia	, Juméia ,	-1674,5265	00:00:06	-----
Lucrécia	, Lucrécia ,	-1873,2982	00:00:06	-----
Mara	, Mara ,	-1364,3738	00:00:11	-----
Martinho	, Martinho ,	-1758,2740	00:00:05	-----
Nálberti	, Nálberti ,	-1768,1028	00:00:05	-----
Pitolomeu	, Pitolomeu ,	-1643,4476	00:00:05	-----
Reinaldo	, Reinaldo ,	-1644,7367	00:00:05	-----
Sueli	, Sueli ,	-1821,5017	00:00:05	-----
Vladimir	, Vladimir ,	-1643,1419	00:00:05	-----