Algoritmo para Redução do Número de Parâmetros de Modelos HMM Utilizados em Sistemas de Reconhecimento de Fala Contínua

Glauco F. G. Yared e Fábio Violaro

***Resumo*— Os Sistemas de Reconhecimento de Fala baseados em modelos HMM têm sido utilizados nos últimos anos em várias aplicações embarcadas que requerem processamento em tempo real, tais como telefones celulares e automóveis. Neste contexto, um importante aspecto que deve ser considerado é o tamanho dos modelos HMM, o qual está diretamente relacionado com a carga computacional do sistema e com a estimação confiável de parâmetros. Os trabalhos anteriores nesta área têm utilizado medidas de verossimilhança para a obtenção de modelos que apresentem um melhor compromisso entre resolução acústica e robustez. Este trabalho apresenta um novo método baseado em uma Medida de Importância da Gaussiana (GIM), utilizada em um Algoritmo de Eliminação de Gaussianas (GEA), para a determinação da complexidade mais apropriada do HMM. Os resultados serão comparados com o método clássico do Critério de Informação Bayesiana (BIC) e com um critério discriminativo.**

***Palavras-Chave*— Redução da complexidade de modelos HMM, Medida de Importância da Gaussiana, Algoritmo de Eliminação de Gaussianas.**

***Abstract*— Nowadays, HMM-based speech recognition systems are used in many real time processing applications, from cell phones to automobile automation. In this context, one important aspect to be considered is the HMM model size, which is directly related to the computational load and to the reliable parameter estimation. Previous works in this area have used likelihood measures in order to obtain models with a better compromise between acoustic resolution and robustness. This work presents a new approach based on a Gaussian Importance Measure (GIM) used in the Gaussian Elimination Algorithm (GEA) for determining the more suitable HMM complexity. The results are compared to the classical Bayesian Information Criterion.**

***Keywords*— HMM model complexity reduction, Gaussian Importance Measure, Gaussian Elimination Algorithm.**

I. Introdução

O número de aplicações que utilizam técnicas de reconhecimento de fala tem crescido consideravelmente nos últimos anos, variando desde sistemas independentes de locutor até sistemas adaptados, que operam em ambientes com baixo ou elevado ruído, etc. Além disso, os sistemas de reconhecimento de fala têm sido projetados visando apresentar um alto desempenho de processamento e uma alta taxa de reconhecimento. No intuito de atingir tais objetivos, deve-se analisar o tamanho de cada modelo HMM, uma vez que tal variável está diretamente relacionado com a carga computacional e com a capacidade de classificação do modelo. O tamanho do modelo HMM está relacionado com o número de componentes Gaussianas presentes em cada mistura. Este trabalho apresenta três abordagens utilizadas para a determinação do número de Gaussianas por estado em modelos HMM: o primeiro é o clássico Critério de Informação Bayesiano (BIC), o segundo é baseado em um critério discriminativo e o terceiro é o novo Algoritmo de Eliminação de Gaussianas proposto.

Resumidamente, três passos se sucedem até a obtenção dos resultados do reconhecimento. Inicialmente realiza-se o treinamento através do algoritmo de Baum-Welch, seguido pela seleção de topologia do modelo através das três abordagens mencionadas anteriormente, e finalmente a decodificação é realizada utilizando-se as ferramentas do HTK [1]. Os desempenhos dos sistemas obtidos são comparados em termos de taxa de reconhecimento de palavras.

II. Fundamentação Teórica e Prática

O problema de modelagem estatística possui alguns aspectos que independem da tarefa específica para a qual os modelos são obtidos. Um problema clássico que precisa ser contornado é o da sobre-parametrização [2], que pode ocorrer em modelos com grande grau de liberdade, ou seja, modelos com um número excessivo de parâmetros. Em geral, tais modelos apresentam baixa taxa de erro de treinamento, devido à alta flexibilidade, mas o desempenho do sistema utilizando dados de teste é quase sempre insatisfatório. Por outro lado, modelos com um número de parâmetros insuficiente não podem nem ao menos ser treinados. Neste ponto, observa-se que deve ser atingido um equilíbrio entre a treinabilidade e robustez do modelo, a fim de se obter um sistema com um desempenho elevado. No contexto de reconhecimento de fala, utiliza-se a taxa de reconhecimento de palavras como medida de desempenho e o número total de componentes Gaussianas como medida do tamanho do sistema.

Outro aspecto importante que deve ser considerado é que a estimação confiável de parâmetros é realizada somente quando existem dados suficientes disponíveis para tal tarefa, caso contrário os parâmetros são mal estimados [3]. Sabendo-se que a base de dados de treinamento normalmente apresenta um número diferente de amostras de cada fone, é razoável se esperar que o número de amostras disponíveis seja também um fator limitante para o aumento do número de clusters no modelo de uma determinada unidade acústica. Dessa forma, dependendo do número de amostras disponíveis, deve-se aumentar ou diminuir a resolução acústica dos modelos HMM no intuito de se realizar uma estimação de parâmetros confiável. Além disso, a complexidade das fronteiras das

distribuições dos parâmetros acústicos também determina o número de componentes necessário para modelar corretamente as diferentes classes.

Existem também argumentos de ordem prática [4] que sustentam a idéia de se determinar modelos HMM com um número variado de Gaussianas por estado. O custo computacional está diretamente relacionado com o número de componentes Gaussianas presentes no sistema. Como consequência imediata, o número de operações e a memória necessária para a realização das mesmas aumenta com o número de componentes.

Portanto, as razões de natureza teórica e prática apresentadas acima servem como base de sustentação para a idéia de se obter modelos acústicos com um número variado de componentes Gaussianas por estado.

Em linhas gerais, a seção III descreverá o sistema de reconhecimento e base de dados utilizados nos experimentos. Na seção IV, o método BIC será revisto. Na sequência, a seção V descreverá um método discriminativo para determinação da topologia do modelo. Na seção VI o novo Algoritmo de Eliminação de Gaussianas será introduzido. Finalmente, as seções VII, VIII e IX apresentarão os resultados experimentais, as discussões e as conclusões do trabalho respectivamente.

III. Sistema de Reconhecimento de Fala e Base de Dados

Nos experimentos realizados, utilizou-se uma base de dados de fala contínua com vocabulário reduzido em Português do Brasil (700 palavras diferentes) [5], contendo um conjunto de 200 sentenças diferentes, produzidas por 40 locutores (20 do sexo masculino e 20 do sexo feminino). Foram utilizadas 1200 sentenças para o treinamento e 400 sentenças para testar o sistema final.

Implementou-se um sistema de treinamento baseado na máxima verossimilhança (ML) a fim de se obter modelos HMM contínuos com 3 estados, do tipo *left-to-right*, para cada fone. São considerados 36 fones independentes de contexto (incluindo o silêncio), o que resulta um total de 108 misturas Gaussianas multidimensionais com um número fixo ou variável de componentes em cada mistura. Além disso, cada componente Gaussiana é representada num espaço de dimensão 39 (cada vetor amostra é constituído pela concatenação de 12 coeficientes mel-cepstrais, 1 parâmetro log-energia, e suas derivadas de primeira e segunda ordem), assumindo independência entre as componentes de cada amostra e dessa forma utilizando matriz de covariância diagonal.

A tarefa de decodificação é realizada através das ferramentas fornecidas pelo HTK [1], utilizando um modelo de linguagem do tipo *Back-off bigram*.

IV. Critério de Informação Bayesiano (BIC)

O critério BIC tem sido amplamente utilizado para a seleção de estruturas no processo de modelagem estatística em diversas áreas. O conceito fundamental que sustenta o critério BIC é o Princípio da Parcimônia, o qual determina que o modelo selecionado deve ser aquele que apresente a menor complexidade e ao mesmo tempo tenha uma elevada capacidade para modelar os dados de treinamento. Tal princípio pode ser observado claramente na equação (1)

$$BIC\left(M_l^j\right) = \sum_{t=1}^{N_j} \log P\left(x_t^j | M_l^j\right) - \lambda \frac{\nu_l^j}{2} \log N_j, \qquad (1)$$

onde M_l^j é o modelo candidato "*l*" do estado "*j*", N_j é o número de amostras associadas ao estado "*j*", x_t^j é a *t*-ésima amostra do estado "*j*", ν_l^j é o número de parâmetros livres (médias, variâncias e coeficientes de ponderação das Gaussianas) presentes em M_l^j e o parâmetro λ controla o termo de penalização.

De acordo com tal critério, o modelo selecionado deve ser aquele que apresente o maior valor de BIC dentre todos os modelos candidatos. Pode-se notar então que a topologia do modelo de cada estado é obtida sem levar em consideração os modelos dos demais estados existentes. Entretanto, algumas modificações no critério BIC [6] já foram propostas no intuito de levar em consideração todos os estados existentes durante a seleção de topologia.

V. Algoritmo Discriminativo para o Aumento da Resolução Acústica

O critério discriminativo proposto em [7], [8] tem como princípio a determinação de quais estados encontram-se modelados com baixa resolução acústica (número insuficiente de Gaussianas no modelo), de acordo com um limiar previamente estabelecido. A idéia consiste basicamente em decodificar os dados de treinamento, fazer um alinhamento de Viterbi usando as transcrições das sentenças decodificadas e comparar tal alinhamento com o alinhamento correto. Assim, é possível verificar quais estados estão sendo confundidos com outros e, desta forma, elaborar uma lista de confusão para cada estado, a qual é definida por $F\left(x_t\right)$.

Neste sentido, a Equação (2) mede o comportamento do modelo de um determinado estado em relação às amostras associadas ao próprio estado em questão

$$P_c^l = \frac{1}{N_{fr}^l} \sum_{t \ni C(x_t)=l} \frac{p\left(x_t | M_l\right)}{p\left(x_t | M_l\right) + \sum\limits_{j \in F(x_t)} p\left(x_t | M_j\right)}, \qquad (2)$$

onde x_t são as amostras associadas à classe $C\left(x_t\right)$ que corresponde ao estado "l", N_{fr}^l é o número de amostras associadas ao estado "l", $p\left(x_t | M_l\right)$ é o logaritmo da verossimilhança dada pelo modelo "M_l" do estado "l" e "j" são os estados da lista de confusão $F\left(x_t\right)$.

Uma vez calculado o P_c^l para cada estado, deve-se encontrar todos os estados que apresentem P_c^l inferior a um limiar pré-definido e substituir tais modelos por correspondentes que foram treinados em um sistema que utiliza uma maior resolução acústica (maior número de Gaussianas por estado).

É importante notar que, neste método, parte-se de um sistema menor que possui modelos inicialmente com "X" Gaussianas por estado e que, após a análise, terá modelos com "X" e "Y" Gaussianas por estado, onde "Y" é o número de componentes presentes em cada mistura do sistema de maior

complexidade. O aumento do número de Gaussianas tem como objetivo aumentar a resolução acústica dos modelos HMM, onde for necessário, de acordo com o critério discriminativo, e consequentemente a discriminabilidade do mesmos.

Trabalhos anteriores mostram que este critério pode fornecer resultados que superam os obtidos pelo clássico BIC, dependendo dos limiares escolhidos [8].

VI. Algoritmo de Eliminação de Gaussianas (GEA)

Os trabalhos anteriores nesta área têm utilizado medidas de verossimilhança nos critérios para seleção de topologia do modelo [6], [8], [9]. No entanto, este trabalho define uma medida de importância da Gaussiana (GIM) a qual é utilizada primeiramente em um novo algoritmo discriminativo e, na sequência, em um método baseado na distância Euclidiana modificada para a eliminação de Gaussianas do modelo.

A. Redução Discriminativa da Complexidade do Modelo

O método discriminativo proposto neste trabalho para a determinação do número de Gaussianas por estado difere dos anteriores nesta área no sentido de que o algoritmo parte inicialmente de sistemas previamente treinados e indica quais Gaussianas devem ser eliminadas dos modelos, utilizando o novo GIM, ao invés de medidas de verossimilhança.

Todas as Gaussianas multidimensionais $N(\mu, \mathbf{\Sigma})$ estão representadas num espaço acústico de dimensão 39, e a função densidade de probabilidade (pdf) é dada pela Equação (3)

$$f(\mathbf{O_t}) = \frac{1}{(2\pi)^{\dim/2}\left|\mathbf{\Sigma}\right|^{1/2}} e^{-(\mathbf{x}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)/2}, \tag{3}$$

em que $|\mathbf{\Sigma}|$ é o determinante da matriz de covariância. Se as componentes das amostras forem estatisticamente independentes (matriz de covariância diagonal), então a pdf pode ser escrita na forma da Equação (4)

$$f(\mathbf{O_t}) = \prod_{d=1}^{\dim} \frac{1}{\sqrt{2\pi\sigma_d^2}} e^{-\left[(x_d-\mu_d)^2/2\sigma_d^2\right]}. \tag{4}$$

Além disso, a contribuição de cada amostra para o GIM, ao longo de cada dimensão acústica, é dada pelas áreas indicadas nas Figuras 1(a) and 1(b).

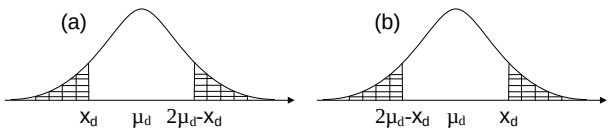

Fig. 1. *Contribuição de cada componente para o GIM. (a) para $x_d \leq \mu_d$. (b) para $x_d > \mu_d$.*

Assim, para cada Gaussiana, a contribuição ao longo de todas as dimensões para o GIM é calculada pela Equação (5)

$$GIM\,(O_t)^{(i;j;s)} = \prod_{d=1}^{\dim} \left(1 - \left[\frac{2}{\sqrt{\pi}} \int_0^{\|Z_d\|} e^{-(z_d)^2} dz_d\right]\right), \tag{5}$$

em que "dim" é a dimensão do vetor amostra, $z_d = \frac{x_d - \mu_{dij}}{\sqrt{2}\sigma_{dij}}$, e $\mathbf{O_t} = (x_1, x_2, \ldots, x_{\dim})$ é o vetor correspondente a uma amostra. Os valores μ_{dij} e σ_{dij} correspondem à média e ao desvio padrão respectivamente, ao longo da dimensão "*d*", da Gaussiana "*i*" que pertence ao estado "*j*".

O GIM da Gaussiana "*i*", que pertence ao estado "*j*", é calculado a partir de cada amostra associada ao estado "*s*", de tal forma que o valor médio do GIM pode ser obtido em relação a cada estado, conforme definido na Equação (6) abaixo

$$P_{GIM}^{(i;j;s)} = \frac{\sum_{t=1}^{N_s} GIM\,(O_t)^{(i;j;s)}}{N_s}, \tag{6}$$

onde N_s é o número de amostras associadas ao estado "*s*".

No intuito de se calcular o valor GIM, é necessário que se tenha uma base de dados segmentada, uma vez que a Equação (5) requer que as amostras tenham sido previamente associadas a cada estado do modelo. De forma alternativa, a segmentação pode ser obtida, por exemplo, pelo alinhamento de Viterbi realizado a partir das transcrições corretas de cada sentença, utilizando-se o melhor sistema HMM disponível.

Em um trabalho prévio [10], utilizou-se uma medida baseada na verossimilhança para o cálculo da contribuição de cada amostra para a medida de importância de cada Gaussiana. A nova medida proposta (GIM) se baseia na probabilidade das amostras se encontrarem fora do intervalo

$$\mu_d - \|x_d - \mu_d\| < x < \mu_d + \|x_d - \mu_d\|.$$

Pode-se notar que, quanto mais próxima a amostra se encontra da média da Gaussiana, maior é a contribuição para o GIM ao longo da dimensão analisada.

O $P_{GIM}^{(i;j;s)}$ pode ser então utilizado como medida de importância de cada Gaussiana em relação a cada estado. Assim, é possível implementar um método discriminativo de seleção de Gaussianas baseado em tal medida, em que o principal objetivo é maximizar a relação discriminativa, de tal forma que cada modelo obtido após a análise apresente o máximo $P_{GIM}^{(i;j;s)}$ para as amostras correspondentes ao estado "j" (s = j) e o mínimo $P_{GIM}^{(i;j;s)}$ para as demais amostras (s ≠ j) ao mesmo tempo. A relação discriminativa que deve ser maximizada é dada pela Equação (7)

$$DC^{(j)} = \frac{\left[\sum_{i=1}^{M_j} P_{GIM}^{(i;j;j)}\right]^K}{\left[\sum_{s\neq j}^{N}\sum_{i=1}^{M_j} P_{GIM}^{(i;j;s)}\right] \Big/ (N-1)} \tag{7}$$

onde K é o expoente de rigor, M_j é o número de Gaussianas do estado "*j*" e N é o número total de estados. Se o logaritmo da Constante Discriminativa (DC) for calculado, a expressão resultante é dada pela Equação (8), que é similar à Equação (1), no sentido que a primeira parcela mede a capacidade de modelar as amostras associadas ao estado "j" e a segunda parcela é um termo de penalização.

$$\log DC^{(j)} = K \log \sum_{i=1}^{M_j} P_{GIM}^{(i;j;j)} - \log \frac{\sum_{s \neq j}^{N} \sum_{i=1}^{M_j} P_{GIM}^{(i;j;s)}}{N-1}. \quad (8)$$

No entanto, as expressões diferem no sentido que o termo de penalização da Equação (1) somente considera aspectos inerentes ao modelo analisado, enquanto o termo de penalização na Equação (8) leva em consideração aspectos dos modelos de todos os estados presentes no sistema.

A principal idéia do método é a de eliminar Gaussianas de um sistema previamente treinado com um número fixo de componentes por estado e observar então o novo valor DC obtido. O valor da Constante Discriminativa (DC) pode aumentar ou diminuir dependendo da relevância da Gaussiana eliminada. Dessa forma, o expoente de rigor tem uma função importante na seleção de Gaussianas, uma vez que torna o critério discriminativo mais restritivo: quanto maior o valor do expoente de rigor, mais rigoroso se torna o critério e portanto menos Gaussianas são eliminadas.

O procedimento descrito acima é aplicado para cada estado dos modelos HMM. Uma vez concluído o processo de eliminação discriminativa, os modelos resultantes são treinados novamente pelo algoritmo de Baum-Welch, porém agora em uma condição bem menos flexível (menos parâmetros nos modelos).

É importante observar que o algoritmo discriminativo detecta apenas Gaussianas que pertencem a um dado estado, mas fornecem valores elevados de verossimilhança para dados pertencentes a outros estados. Entretanto, ainda podem restar Gaussianas excedentes no modelo após a aplicação do algoritmo discriminativo. Apesar de não serem detectadas pelo critério discriminativo, tais componentes precisam ser descartadas, uma vez que este procedimento pode ser realizado sem degradação da capacidade de classificação do sistema. Neste sentido, aplica-se em seguida um algoritmo baseado em distância a fim de se eliminar o excesso de Gaussianas que ainda pode existir nos modelos.

B. Eliminação de Gaussianas Baseada na Distância Euclidiana Modificada

Os modelos HMM obtidos após o treinamento com o algoritmo de Baum-Welch frequentemente apresentam componentes redundantes, ou seja, Gaussianas que convergiram para aproximadamente a mesma posição no espaço acústico e que apresentam praticamente a mesma contribuição para a classificação.

Neste sentido, torna-se interessante estabelecer um limiar de distância para a determinação da junção de duas ou mais Gaussianas numa única. Além disso, um limiar diferente deve ser utilizado para Gaussianas localizadas nas fronteiras da distribuição dos parâmetros acústicos e para aquelas localizadas nas partes centrais da distribuição. Dessa forma, é necessário determinar quais componentes se encontram nas fronteiras e quais as que se encontram nas partes centrais, e também descobrir qual limiar deve ser utilizado. Tal objetivo pode ser atingido através de uma medida indireta, a qual define-se como a distância Euclidiana modificada.

Inicialmente, a distância Euclidiana é calculada entre todas as Gaussianas do modelo de um determinado estado. Na sequência, o $P_{GIM}^{(i;j;s)}$ é utilizado para dar uma indicação da proximidade de cada Gaussiana em relação à fronteira da distribuição acústica e atribuir pesos diferentes para estas Gaussianas e para aquelas localizadas na parte central. Portanto, a medida de distância Euclidiana modificada é definida pela Equação (9)

$$M_{d_{xy}} = \frac{d_{xy}}{\frac{\sum_{i=x,y} P_{GIM}^{(i;j;j)}}{\sum_{i=x,y} P_{GIM}^{(i;j;j)} + \sum_{s \neq j} \sum_{i=x,y} P_{GIM}^{(i;j;s)}}}, \quad (9)$$

onde d_{xy} é dado por

$$d_{xy} = \sqrt{(\mu_x - \mu_y) \cdot (\mu_x - \mu_y)^T}, \quad (10)$$

sendo μ_x e μ_y os vetores de média das componentes Gaussianas *x* e *y* respectivamente.

As Gaussianas redundantes são então substituídas por aquela que apresentar o maior determinante da matriz de covariância.

VII. RESULTADOS

Os experimentos foram realizados utilizando o sistema de reconhecimento e a base de dados previamente descritos. Os primeiros resultados foram obtidos para sistemas com um número fixo de componentes Gaussianas por estado (sistemas de referência), os quais são apresentados na Figura 2.

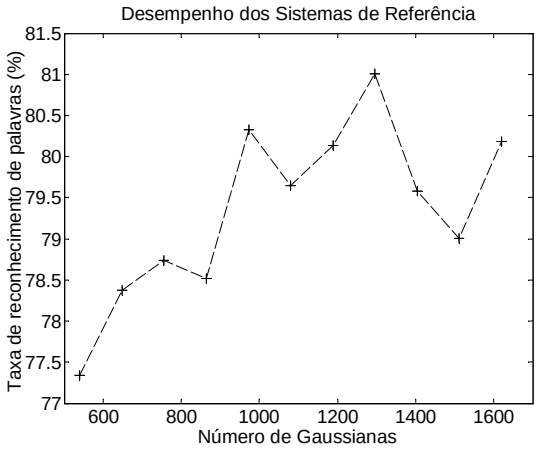

Fig. 2. *Cada ponto corresponde a um sistema de referência contendo um número fixo de componentes Gaussianas por estado (de 5 a 15 Gaussianas por estado). O HTK é utilizado na decodificação.*

O sistema contendo 1296 Gaussianas (12 Gaussianas por estado) apresenta o melhor desempenho na decodificação, como se pode notar na Figura 2. O procedimento adotado neste trabalho visa determinar sistemas com um número variado de Gaussianas por estado e que apresentem um melhor compromisso entre tamanho e desempenho em relação ao sistema de referência.

O primeiro método implementado foi o discriminativo para o aumento da resolução acústica dos modelos HMM. Uma vez que o sistema com 12 Gaussianas por estado resultou na melhor taxa de reconhecimento, utilizou-se tal sistema para o aumento da resolução acústica de sistemas menores. Assim, obteve-se modelos do tipo MYxM12-*threshold* (Y = 5, 7, 9 e 11; *threshold* = 0.2, 0.4, 0.6 e 0.8), em que Y é o número de Gaussianas por estado dos modelos iniciais, e *threshold* é o valor limiar de P_c^l abaixo do qual o modelo do estado é considerado com baixa resolução acústica e portanto precisa de um número maior de Gaussianas. Os resultados obtidos encontram-se na Figura 3.

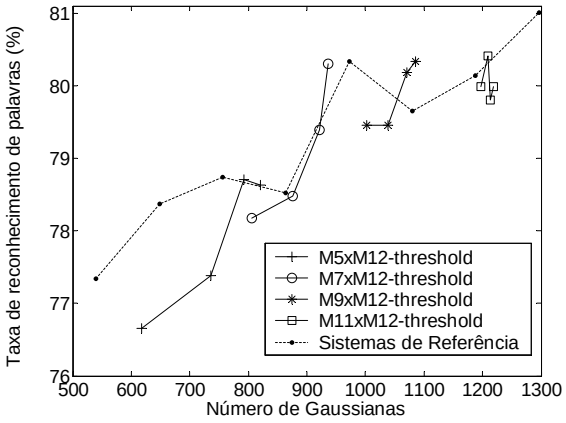

Fig. 3. *Sistemas obtidos com o critério discriminativo para o aumento da resolução acústica de modelos HMM. Para cada sistema inicial (MYxM12-threshold) com Y Gaussianas por estado, utilizam-se os seguintes valores de threshold: 0.2, 0.4, 0.6 e 0.8. Os sistemas de referência possuem de 5 até 12 Gaussianas por estado.*

Dessa forma pode-se obter sistemas que apresentem um desempenho superior ao dos sistemas de referência (com aproximadamente o mesmo tamanho) após o aumento da resolução acústica de alguns modelos HMM de acordo com o critério discriminativo. Além disso, tem-se também um melhor compromisso entre tamanho dos modelos e desempenho na decodificação. A comparação do resultado obtido pelo sistema M11xM12-0.4, que forneceu a maior taxa de reconhecimento utilizando o método discriminativo, com o fornecido pelo sistema contendo 12 Gaussianas por estado, que apresenta a maior taxa de reconhecimento dentre aqueles com número fixo de componentes por estado, está indicada na Tabela I.

TABELA I

COMPARAÇÃO ENTRE O SISTEMA M11XM12-0.4 E O DE REFERNÊCIA COM 12 GAUSSIANAS POR ESTADO.

M11xM12-0.4		12 Gaussianas por estado	
Número de Gaussianas	Taxa de Reconhec. (%)	Número de Gaussianas	Taxa de Reconhec. (%)
1209	80.41	1296	81.01

Conforme pode ser observado, o sistema M11xM12-0.4, comparado ao de referência contendo 12 Gaussianas por estado, apresenta uma economia de 6.71% no número de componentes Gaussianas, ao custo de uma degradação de 0.6% na taxa de reconhecimento.

Na sequência, o segundo método analisado é o clássico Critério de Informação Bayesiana. A idéia básica deste método é testar diversas topologias candidatas e selecionar aquela que apresentar o maior valor de BIC. Assim, diversos sistemas foram avaliados e alguns foram selecionados de acordo com o parâmetro λ, o qual assumiu os seguintes valores nos testes realizados: 0.3, 0.2, 0.15, 0.1, 0.07, 0.05, 0.03 e 0.01. A Figura 4 mostra os resultados obtidos.

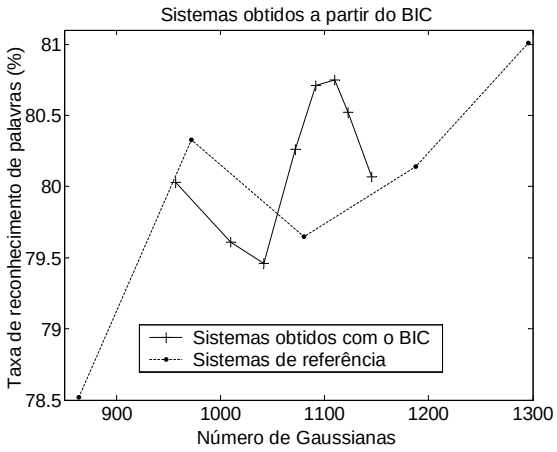

Fig. 4. *Sistemas obtidos com o BIC, utilizando-se os seguintes valores de λ: 0.3, 0.2, 0.15, 0.1, 0.07, 0.05, 0.03 e 0.01. Os sistemas de referência possuem de 8 até 12 Gaussianas por estado.*

De forma análoga, deseja-se comparar o sistema que apresenta o melhor desempenho obtido a partir do BIC (λ = 0.05) com o sistema de referência contendo 12 Gaussianas por estado, o que pode ser observado na Tabela II

TABELA II

COMPARAÇÃO ENTRE O SISTEMA OBTIDO COM O BIC (PARA λ = 0.05) E O DE REFERNÊCIA COM 12 GAUSSIANAS POR ESTADO.

BIC (λ = 0.05)		12 Gaussianas por estado	
Número de Gaussianas	Taxa de Reconhec. (%)	Número de Gaussianas	Taxa de Reconhec. (%)
1110	80.75	1296	81.01

Os resultados mostram que o sistema com maior desempenho obtido com o BIC apresenta uma economia de 14.3% e uma degradação de 0.26% na taxa de reconhecimento, quando comparado ao sistema com 12 Gaussianas por estado.

Finalmente, analisou-se o novo método proposto GEA, que parte do sistema previamente treinado com 12 Gaussianas por estado, e inicia um processo de eliminação de Gaussianas de forma discriminativa e de forma a evitar a presença de componentes redundantes. Neste sentido, o primeiro passo é determinar o expoente de rigor "K" e, na sequência, os limiares de distância Euclidiana modificada entre as Gaussianas dentro do modelo de cada estado. Assim, os resultados dos sistemas obtidos utilizando-se K = 10^5 e limiares de distância "d_{xy}" iguais a 0, 3.5, 4, 4.5, 5, 5.5, 6 e 7 estão mostrados na Figura 5.

O sistema que apresentou o melhor desempenho através do GEA foi com k = 10^5 e limiar de distância igual a 5.5. O resultado encontra-se na Tabela III.

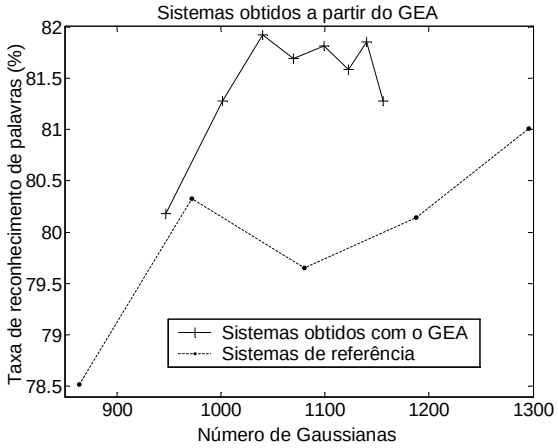

Fig. 5. *Sistemas obtidos com o GEA, utilizando-se $k = 10^5$ e os seguintes limiares de distância Euclidiana modificada: 0, 3.5, 4, 4.5, 5, 5.5, 6 e 7. Os sistemas de referência possuem de 8 até 12 Gaussianas por estado.*

TABELA III

COMPARAÇÃO ENTRE O SISTEMA OBTIDO COM O GEA (PARA K $= 10^5$ E LIMIAR DE DISTÂNCIA IGUAL A 5.5) E O DE REFERNÊCIA COM 12 GAUSSIANAS POR ESTADO.

GEA ($k = 10^5$ e $d_{xy} = 5.5$)		12 Gaussianas por estado	
Número de Gaussianas	Taxa de Reconhec. (%)	Número de Gaussianas	Taxa de Reconhec. (%)
1040	81.92	1296	81.01

Verifica-se então que o GEA permitiu a obtenção de um sistema que apresenta uma economia de 19.8% e ao mesmo tempo um ganho de 0.91% na taxa de reconhecimento em relação ao de referência contendo 12 Gaussianas por estado.

Portanto, dentre os métodos analisados e, de acordo com as condições descritas do experimento, a utilização do GEA fornece o sistema que apresenta o melhor compromisso entre o número de componentes Gaussianas presentes nos modelos HMM e o desempenho do sistema na decodificação.

VIII. DISCUSSÃO

O método discriminativo para o aumento da resolução acústica de modelos HMM mostrou ser eficiente para a obtenção de modelos menos complexos e com um maior desempenho na decodificação, em relação ao sistema original. Tal método parte de três condições iniciais (dois sistemas, cada um contendo um número fixo de Gaussianas por estado, e um valor de *threshold*) e resulta em modelos que apresentam um melhor compromisso entre tamanho e desempenho do sistema em relação aos de referência (com número fixo de Gaussianas por estado).

O BIC também permitiu a determinação de sistemas que superam o desempenho dos de referência. Além disso, necessita de duas condições iniciais: um conjunto de modelos candidatos e um valor para o parâmetro λ.

O GEA fornceu sistemas cujos desempenhos superam os obtidos pelos demais métodos, de acordo com as condições iniciais utilizadas no experimento. Porém, à medida que se aumenta demasiadamente o valor do limiar de distância Euclidiana modificada, pode-se obter sistemas com uma elevada degradação na taxa de reconhecimento em relação ao sistema original. Este método parte de três condições iniciais, as quais são o modelo inicial com número fixo de Gaussianas por estado (previamente treinado), o expoente de rigor K e o limiar de distância d_{xy}. Apesar de ter um equacionamento semelhante ao BIC, a etapa de eliminação discriminativa de Gaussianas permite evitar a presença de componentes que tenham convergido para distribuições erradas durante o treinamento. Além disso, a etapa de análise interna dos modelos HMM do processo de eliminação, baseada em distâncias Euclidianas modificadas, permite evitar a presença de Gaussianas redundantes, que podem eventualmente convergir para distribuições erradas, no caso de se retreinar o sistema. Dessa forma, a nova condição menos flexível (menos Gaussianas por estado) do modelo HMM obtido após a aplicação do GEA, evita problemas de sobre-parametrização e ao mesmo tempo contribui para a economia de recursos computacionais.

IX. CONCLUSÕES

O novo GEA proposto possibilitou, dentre os métodos analisados, a determinação das topologias dos modelos HMM que constituem o sistema de reconhecimento que apresenta o melhor compromisso entre número de Gaussianas por estado e desempenho na decodificação, com uma economia de 19.8% no tamanho do sistema e um aumento de 0.91% na taxa de reconhecimento, comparado ao sistema de referência com 12 Gaussianas por estado. Nos próximos trabalhos, pretende-se realizar os mesmos experimentos para uma base de dados de fala contínua com um vocabulário de maior tamanho e modelos de fones dependentes de contexto.

AGRADECIMENTOS

Os autores agradecem ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo apoio financeiro ao projeto.

REFERÊNCIAS

[1] *The HTK Book*, Cambridge University Engineering Departament, 2002.
[2] L. A. Aguirre, *An Introduction to Systems Identification - Linear and Non-linear Techniques Applied to Real Systems (in portuguese)*. Editora UFMG, 2000.
[3] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Prentice Hall, 1993.
[4] C. Lévy, G. Linarès, P. Nocera, and J.-G. Bonastre, "Reducing computational and memory cost for cellular phone embedded speech recognition system," in *IEEE International Conference on Acoustic, Speech and Signal Processing*, 2004.
[5] C. A. Ynoguti, "Continuous speech recognition using hidden markov models (in portuguese)," Ph.D. dissertation, State University of Campinas, 1999.
[6] A. Biem, "Model selection criterion for classification: Application to HMM topology optimization," in *7-th International Conference on Document Analysis and Recognition (ICDAR'03)*, 2003.
[7] Y. J. Chung and C. K. Un, "Use of different number of mixtures in continuous density hidden markov models," *Electronics Letters*, vol. 29, no. 9, pp. 824–825, 1993.
[8] M. Padmanabhan and L. R. Bahl, "Model complexity adaptation using a discriminant measure," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 2, pp. 205–208, 2000.
[9] Y. Gao, E.-E. Jan, M. Padmanabhan, and M. Picheny, "HMM training based on quality measurement," in *IEEE International Conference on Acoustic, Speech and Signal Processing*, 1999.
[10] G. F. G. Yared and F. Violaro, "Finding the more suitable HMM size in continuous speech recognition systems," in *International Information and Telecommunications Technologies Symposium*, 2004.