X Congresso Brasileiro de Inteligência Computacional (CBIC'2011), 8 a 11 de Novembro de 2011, Fortaleza, Ceará

SISTEMA DE INFERÊNCIA *FUZZY* BASEADO NA TRANSFORMADA COSSENO DISCRETA PARA RECONHECIMENTO DE VOZ

Washington L. S. Santos
Universidade Federal do Maranhão
Departamento de Engenharia Elétrica
Programa de Pós-graduação em Engenharia de Eletricidade
washington.wlss@ifma.edu.br

Ginalber L. O. Serra
Instituto Federal de Educação, Ciência e Tecnologia do Maranhão
Departamento de Eletro-Eletrônica
Laboratório de Inteligência Computacional Aplicada à Tecnologia
Av. Getúlio Vargas, 04, Monte Castelo, CEP: 65030-005, São Luís, Maranhão, Brasil
ginalber@ifma.edu.br

Resumo – A utilização da transformada cosseno discreta (TCD) na compressão de dados e na classificação de padrões aumentou muito nos últimos anos, e isso deve-se principalmente ao fato do seu desempenho aproximar-se muito dos resultados obtidos com a transformada de Karhunen-Loève que é considerada ótima. Neste trabalho procura-se demonstrar o potencial da Transformada Cosseno Discreta, bem como Sistemas Fuzzy no reconhecimento de voz. Essas duas ferramentas mostraram bons resultados no modelamento temporal do sinal de voz. Após uma exposição do modelamento matemático da voz utilizada neste artigo, aborda-se de forma sucinta a extração das características temporais do sinal de voz e define-se um sistema de reconhecimento automático de voz, como classificador, que extrai as características das locuções, coeficientes mel cepstrais de duas dimensões, e através da transformada discreta cosseno são apresentados os padrões para o classificador fuzzy.

Palavras-chave – Transformada cosseno discreta, reconhecimento de voz, sistema fuzzy, mel- cepstral.

Abstract – The use Discrete Cosine Transform (DCT) on data compression and pattern classification has increased in recent years, and this is mainly due to the fact that their performance much closer to the results obtained with the Karhunen-Loève transform that is considered optimal. In this paper we aimed to demonstrate the potential of Discrete Cosine Transform and Fuzzy Systems in speech recognition. These two tools showed good temporal modeling of voice signal. After discussing the mathematical modeling of the voice used in this article, we discuss briefly the extraction of temporal characteristics of the voice signal and sets up a system for recognizing an automated voice, as a classifier, which extracts the characteristics of the phrase , mel-cepstrals coefficients in two dimensions and through Discrete Cosine Transform are presented the patterns for the fuzzy classifier.

Keywords – Discrete cosine transform, speech recongnition, fuzzy systems, mel-cepstral.

1. INTRODUÇÃO

A base para a maioria dos algoritmos de processamento digital de voz é um modelo de sistema no tempo discreto para a produção de amostras do sinal de voz. A parametrização, isto é, codificação de um sinal analógico de voz, é um dos primeiros passos no processo de reconhecimento de voz. Várias técnicas de análise de sinal têm sido sugeridas na literatura especializada. Essas técnicas, normalmente, pretendem produzir representações paramétricas com algum significado perceptual da voz, onde se procura destacar as características mais importantes da voz para maximizar o desempenho no processo de reconhecimento [1]. A seleção das melhores representações paramétricas do sinal de voz é uma tarefa muito importante no desenvolvimento de qualquer sistema de reconhecimento de voz. O objetivo da seleção da melhor forma de codificar o sinal é comprimir os dados de voz eliminando informações não pertencentes à análise fonética do sinal e melhorar aqueles aspectos do sinal que contribuem significativamente às detecções das diferenças fonéticas dos sons de voz [2]. O problema de reconhecimento de padrões pode ser formulado como segue: sejam S_k classes, onde $k = 1, 2, 3...K$, contidas num espaço de padrões com dimensão $\Re^n$. Tomando-se um espaço qualquer de padrões com dimensão $\Re^x$, onde $x \leq n$, pode-se tranformar este espaço em um novo espaço de padrões com dimensão $\Re^a$, onde $a < x \leq n$. Então, supondo-se uma estatística de segunda ordem medida ou modelada para cada S_k, através de uma função de covariância representada por $\left[\Phi^x_{(k)}\right]$, a matriz de covariância generalizada descritiva do problema de reconhecimento de padrões torna-se:

$$[\Phi_x] = \sum_{k=1}^{K} P(S_k) \left[\Phi_x^{(k)}\right] \tag{1}$$

onde $P(S_k)$ é uma função de distribuição da classe S_k, *a priori*, com $0 \leq P(S_k) \leq 1$. Um operador de transformação linear através da matriz $\mathbf{A}$ irá mapear o espaço de padrões dentro de um espaço transformado onde os vetores bases serão colunas ortogonais dessa matriz. Os padrões do novo espaço são combinações lineares dos eixos originais conforme a estrutura da matriz $\mathbf{A}$. A estatística de segunda ordem no espaço transformado é dada por:

$$\Phi_{\mathbf{A}} = \mathbf{A}^T [\Phi_x] \mathbf{A} \tag{2}$$

onde $\Phi_{\mathbf{A}}$ corresponde à matriz de covariância no espaço gerado pela matriz $\mathbf{A}$ e o operador T corresponde à transposta de uma matriz. A partir de então, pode-se extrair características que forneçam maior poder discriminatório para a classificação a partir da dimensão do espaço gerado [3]. Uma das mais difundidas técnicas para reconhecimento dos padrões de voz é o "Hidden Markov Model (HMM)" [4], [5]. Apesar de sua capacidade de reconhecimento, é bem conhecido que uma das principais deficiências do HMM clássico está relacionada com o modelamento inadequado da duração do evento acústico associado com cada estado. Desde que a probabilidade de recorrência para o mesmo estado é constante, a probabilidade de duração do evento acústico associado com o estado tem uma probabilidade exponencial decrescente com o tempo. A hipótese básica é que a voz é um sinal quase estacionário e a sua parte estacionária pode ser representada por um simples estado do HMM. Este tipo de duração não representa a estrutura temporal da voz. Outra fragilidade do HMM é a hipótese de que dentro de cada estado os vetores observações são não correlacionados, enquanto na realidade o que acontece é o oposto da hipótese admitida. Freqüentemente erros ocorrem porque uma sequência de observação é decodificada por poucos estados, tipicamente absorvendo segmentos de baixa energia e com alta probabilidade de duração. Os outros estados, em vez disso, são rapidamente atravessados devido a sua distribuição não se adaptar bem ao restante da observação. Esses erros, portanto, não dependem da confusão intrínseca de palavras de acústica semelhantes, mas principalmente pela falta de boa modelagem da duração do evento acústico o que produz hipótese fracamente relacionada à acústica da palavra correta [6]. Para justificar a estrutura dinâmica dos vetores de observação, incluindo as variações locais e globais, este artigo, propõe um sistema de reconhecimento de voz de dígitos isolados que não se baseia diretamente no modelamento da duração estado/palavra; em vez disso, baseia-se nas variações globais das características espectrais de cada palavra e sua correlação no tempo, duas importantes características que são exploradas parcialmente pelo HMM clássico. Este artigo propõe um sistema de parametrização e reconhecimento do sinal de voz, utilizando-se a Transformada Cosseno Discreta (TCD) [7] e sistema de inferência fuzzy. A utilização da TCD na compressão de dados e na classificação de padrões aumentou muito nos últimos anos, e isso deve-se principalmente ao fato do seu desempenho aproximar-se muito dos resultados obtidos com a transformada de Karhunen-Loève que é considerada ótima. Neste artigo procura-se demonstrar o potencial da TCD, bem como sistema de inferência fuzzy, no reconhecimento de voz. Essas duas ferramentas mostraram bons resultados no modelamento temporal do sinal de voz. Após uma exposição do modelamento matemático da voz utilizada neste artigo, aborda-se de forma sucinta a extração das características temporais do sinal de voz e define-se um sistema de reconhecimento automático de voz, como classificador, que extrai as características das locuções, coeficientes mel cepstrais de duas dimensões, e através da transformada discreta cosseno são apresentados os padrões para o classificador fuzzy.

2 SISTEMA DE RECONHECIMENTO DE VOZ TCD-FUZZY

Na figura 1 mostra-se o diagrama de bloco do sistema proposto para o reconhecimento do sinal de voz.

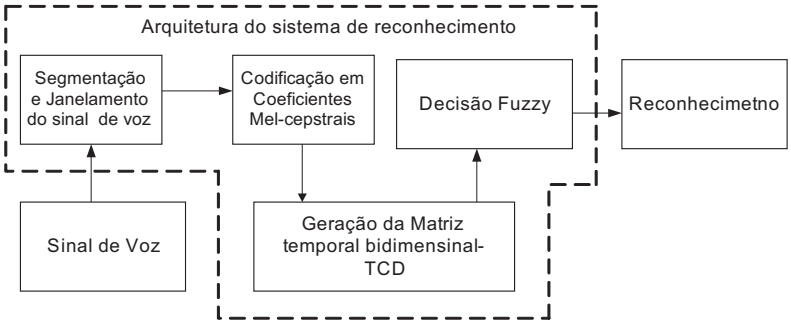

Figura 1: Diagrama de bloco do sistema de reconhecimento

Inicialmente o sinal de voz é dividido em segmentos, os quais são janelados e em seguida codificado em uma quantidade de parâmetros definidos pela ordem dos coeficientes mel-cepstrais. Os coeficientes TCD são calculados e, finalmente, as funções de pertinências dos padrões são geradas para inferências no reconhecimento final do dígito.

2.1 Segmentação e janelamento do sinal de voz

Quando uma janela retangular é aplicada a um determinado sinal, ela seleciona uma pequena parcela deste sinal, a qual será analisada, denominada segmento. A análise de Fourier de curto-prazo efetuada sobre esses segmentos, é chamada análise de sinal segmento por segmento. A duração do segmento T_f é definida como a extensão de tempo sobre o qual um conjunto de parâmetros é considerado válido. O período do segmento é utilizado para determinar a extensão de tempo entre os cálculos de sucessivos parâmetros. Para processamento de voz, tipicamente, o período de segmento está entre $10ms$ e $30ms$. Valores nesta faixa represetam um compromisso entre a razão de mudança do espectro e a complexidade do sistema [1]. Devido ao fato de nas extremidades das janelas o sinal analisado sofrer um amortecimento excessivo em suas amostras, faz-se necessário à utilização de um processo denominado sobreposição para controlar quão rapidamente os parâmetros do sinal podem mudar de segmento para segmento. Em processamento de voz a janela mais utilizada é a de Hamming, que é um caso particular da janela de Hanning dada por

$$\omega(n) = \frac{\alpha_\omega - (1-\alpha_\omega)cos(2n\pi)/(N_S-1)}{\beta_\omega} \tag{3}$$

onde $\alpha_\omega = 0.54$, com 0≤n≤N_S e $\omega(n)$=0 para n fora do intervalo; α_ω é definida como uma constante no intervalo [0,1], N_S é o tempo de duração da janela e β_ω é uma constante de normalização definida tal que o valor da raiz média quadrática (rms) da janela é igual a unidade, como segue:

$$\beta_\omega = \sqrt{\frac{1}{N}\sum_{n=0}^{N_S-1}\omega^2(n)}. \tag{4}$$

Assim, a cada novo segmento apenas uma fração do sinal irá mudar. Na figura 2 é ilustrado um processo de segmentação e janelamento onde são tomados N segmentos de K amostras do sinal.

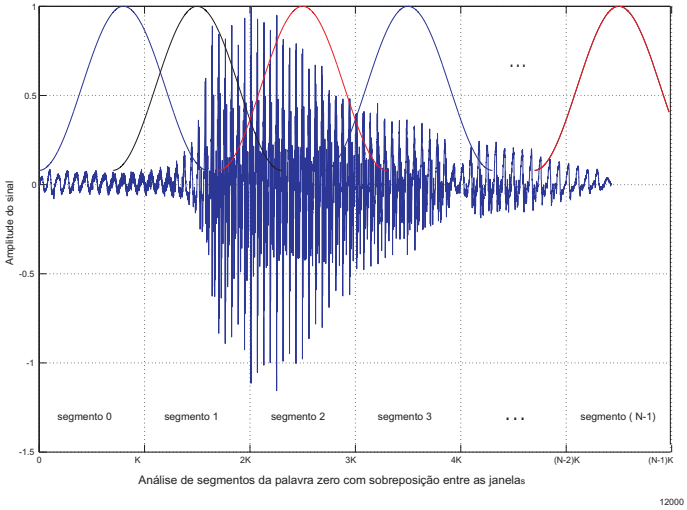

Figura 2: Análise de segmentos da palavra com sobreposição entre as janelas.

A sobreposição entre as janelas é dada por:

$$\text{sobreposição}(\%) = \frac{T_w - T_f}{T_w} \times 100 \tag{5}$$

onde T_w é o tempo de duração da janela e T_f é o tempo de duração do segmento. Assim, por exemplo, a combinação do período do segmento de $20ms$ e duração de janela de $30ms$ corresponde a aproximadamente 33% de sobreposição. O objetivo da sobreposição é reduzir o ruído introduzido pelo janelamento e o ruído de canal não estacionário.

2.2 Codificação em coeficientes mel-cepstrais

Experimentos com a percepção humana tem mostrado que frequências de um som complexo dentro de uma certa largura de banda de alguma frequência nominal não pode ser individualmente identificada. Quando um dos componentes deste som está fora da largura de banda considerada, essa componente não pode ser distinguida. Normalmente, considera-se uma largura de banda crítica para voz como sendo de 10% a 20% da frequência central do som considerado. Uma das formas mais populares de se mapear a frequência de um dado sinal de som para valores de freqências perceptuais, isto , capaz de excitar a audição humana é através da escala mel [1]. Esta escala tenta mapear as frequências perceptíveis de um tom ou de uma frequência de *pitch* em uma escala linear. Neste artigo utilizou-se uma frequência limite para segmentação uniforme $F_u = 1kHz$, uma distribuição em 10 intervalos uniformes, uma frequência de amostragem mínima de 8kHz e a escala mel [2] dada por:

$$mel = 2595log\left(1 + \frac{f}{700}\right) \tag{6}$$

O banco de filtros utilizado abrange a faixa de 0 a 4600Hz sendo distribuído em 20 filtros, e, através da Transformada rápida de Fourier (FFT), gera-se a saída log-energia já devidamente espaçada na escala mel denominada de em.

2.3 Geração da matriz temporal bidimensional-TCD

Os coefecientes mel-cepstrais são calculados através da seguinte equação:

$$mfcc(n+1) = mfcc(n) + \sum_{k=1}^{NF} em(k)cos\left[\frac{n\,(k-0.5)}{NF.\pi}\right] \tag{7}$$

onde em são os coeficientes mel-cepstrais e NF é o número de filtros. A matriz de coeficientes mel-cepstrais de duas dimensões, que é resultado da TDC realizada em uma seqüência de T vetores de observação de coeficientes mel-cepstrais no eixo do tempo, é obtida pela equação:

$$C_k(n,T) = \frac{1}{N}\sum_{t=1}^{T} mfcc_k(t)cos\frac{(2t-1)n\pi}{2T} \tag{8}$$

onde $k, 1 \leq k \leq K$, refere-se a k-ésima (linha) componente do $t-$ésimo segmento da matriz e $n, 1 \leq n \leq N$(coluna), refere-se a ordem da TCD. Dessa forma, obtém-se a matriz de duas dimensões, onde o interesse está nos coeficientes de baixa ordem de k e n que codificam as variações de longo prazo do envelope espectral do sinal de voz [6]. Este procedimento é realizado para cada palavra falada. Assim, tem-se uma matriz bidimensional $C_k(n,T)$ para cada sinal de entrada. Os elementos da matriz são obtidos da seguinte forma:

1. Para uma dada palavra ***P*** são tomados dez exemplos de pronúncias dessa palavra. Esses exemplos são devidamente codificados em T segmentos distribuídos ao longo do eixo do tempo;

2. Cada segmento de um dado exemplo da palavra ***P*** gera uma quantidade K de coeficientes mel-cepstrais, dessa forma são retiradas às características significantes para cada segmento ao longo do tempo. Calcula-se a TCD de ordem N para cada coeficiente mel-cepstral de mesma ordem dentro dos segmentos distribuídos ao longo do eixo do tempo, isto é, a TCD de ordem N será calculada para os coeficientes c_1 do segmento $t = 1$, c_1 do segmento $t = 2, ...,$ c_1 do segmento $t = T$, e assim por diante gerando os elementos $c_{11}, c_{12}, c_{13}, ..., c_{1N}$ da matriz dada na equação (8), até mapear todos os coeficientes em todos os segmentos. Assim, é gerada uma matriz para cada exemplo da palavra ***P***;

3. São calculadas uma matriz de média e uma de variância para representar o modelo da palavra ***P***.

A seguir tem-se as matrizes formadas:

$$\mathbf{C^0} = \begin{pmatrix} c_{11}^0 & c_{12}^0 \\ c_{21}^0 & c_{22}^0 \end{pmatrix}$$

$$\mathbf{C^1} = \begin{pmatrix} c_{11}^1 & c_{12}^1 \\ c_{21}^1 & c_{22}^1 \end{pmatrix}$$

$$\vdots$$

$$\mathbf{C^9} = \begin{pmatrix} c_{11}^9 & c_{12}^9 \\ c_{21}^9 & c_{22}^9 \end{pmatrix}$$

2.4 Sistema de inferência fuzzy para decisão

A etapa de decisão é realizada por um sistema de inferência fuzzy baseado no conjunto de regras obtidas a partir das médias e das variâncias das matrizes temporais de duas dimensões de cada dígito falado. Para este artigo optou-se por utilizar uma matriz com o número mínimo possível de parâmetros (2×2) e que ainda permita um desempenho satisfatório quando comparado com reconhecedores de padrões disponíveis na literatura. Os elementos das matrizes $\mathbf{C}^j$, com $j = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$, onde j representa o padrão usado no treinamento, são utilizados pelo sistema de inferência fuzzy para gerar quatro funções de pertinência gaussianas, correspondente a cada elemento $c_{kn}^j\,\big|^{k=1,2;n=1,2}$ da matriz. O modelo do sistema de inferência fuzzy para o reconhecimento é dado na figura 3. O desenvolvimento do sistema de inferência TCD-Fuzzy utiliza funções gaussianas para a geração das funções de pertinência, onde foram tomadas as médias e variâncias dos elementos da matriz $\mathbf{C}^j$ correspondente a cada padrão. Deste modo, para cada função de pertinência, tem-se o grau de pertinência $\mu\left(c_{kn}^j\right)$ dado conforme segue:

$$\mu\left(c_{11}^{j}\right) = \left[\mu_{c_{11}}^{0}\mu_{c_{11}}^{1}\mu_{c_{11}}^{2}\mu_{c_{11}}^{3}\ldots\mu_{c_{11}}^{9}\right] \tag{9}$$

$$\mu\left(c_{12}^{j}\right) = \left[\mu_{c_{12}}^{0}\mu_{c_{12}}^{1}\mu_{c_{12}}^{2}\mu_{c_{12}}^{3}\ldots\mu_{c_{12}}^{9}\right] \tag{10}$$

$$\mu\left(c_{21}^{j}\right) = \left[\mu_{c_{21}}^{0}\mu_{c_{21}}^{1}\mu_{c_{21}}^{2}\mu_{c_{21}}^{3}\ldots\mu_{c_{21}}^{9}\right] \tag{11}$$

$$\mu\left(c_{22}^{j}\right) = \left[\mu_{c_{22}}^{0}\mu_{c_{22}}^{1}\mu_{c_{22}}^{2}\mu_{c_{22}}^{3}\ldots\mu_{c_{22}}^{9}\right] \tag{12}$$

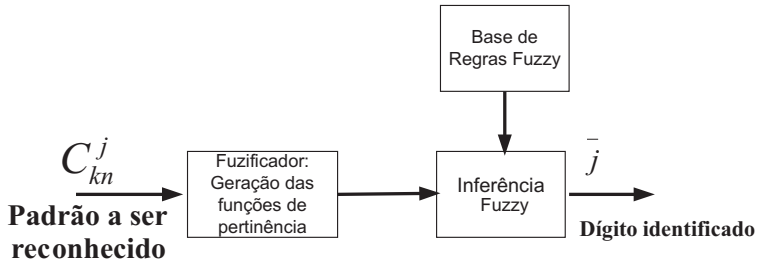

Figura 3: Sistema Fuzzy utilizado no reconhecimento.

O grau de ativação da regra é dado por:

$$h_i^j\left(c_{kn}^{j}\right) = \mu_{c_{11}}^{j} \times \mu_{c_{12}}^{j} \times \mu_{c_{21}}^{j} \times \mu_{c_{22}}^{j} \tag{13}$$

onde $j = 0, 1, 2, ..., 9$ representa o padrão e $i = 1, 2, 3, ..., 10$ representa o índice da regra. O vetor do grau de ativação de cada regra é dado por:

$$\overline{h_i} = \left[h_1^0\ h_2^1\ h_3^2\ h_4^3\ h_5^4\ h_6^5\ h_7^6\ h_8^7\ h_9^8\ h_{10}^9\right] \tag{14}$$

O grau de ativação normalizado da regra é dado por:

$$y_i^j = \frac{h_i^j}{\sum_{l=1}^{L} h_l^j} \tag{15}$$

e

$$\sum_{l=1}^{L} y_i^j = 1 \tag{16}$$

onde $L = 10$. O sistema de inferência fuzzy para reconhecimento toma a decisão através das bases de regras, ecolhendo o maior valor do vetor dado na equação (15).

3 RESULTADOS EXPERIMENTAIS

Os parâmetros da matriz C_{kn}^{j} e as variâncias dos seus elementos foram utilizados para gerar as funções de pertinências, mostradas nas figuras 4 a 7. Assim, para cada padrão j treinado tem-se um conjunto de centros correspondente aos elementos c_{kn}^{j} da matriz temporal $\mathbf{C}^{j}$, utilizados na geração das funções de pertinências, respectivamente.

3.1 Treinamento do Sistema

Na fase de treinamento, para a parametrização dos modelos, foram utilizados doze locutores, sendo sete masculinos (locutores de 1 a 6 e locutor 11) e cinco femininos (locutores de 7 a 10 e locutor 12) distribuídos como segue:

1. Os locutores de 1 a 10 falaram, em duas séries, os dígitos de 0 a 9 num total de 200 locuções pronunciadas em ambiente com baixos níveis de ruído (laboratório), das quais as 100 primeiras foram utilizadas para treinamento. Assim, por exemplo, para o dígito zero foram pronunciadas dez locuções por locutores diferentes e, sucessivamente, para os demais dígitos. Para cada dez exemplos de um dígito (padrão) foi gerada uma matriz $\mathbf{C}^{j}$. Para finalizar o cômputo das matrizes de treinamento realizou-se o cálculo da média e da variância dos elementos da matriz $\mathbf{C}^{j}$, gerando-se assim duas matrizes de ordem (2×2), uma de média e outra de variância que representam os parâmetros do padrão de cada dígito. As outras 100 locuções foram utilizadas no procedimento de teste.

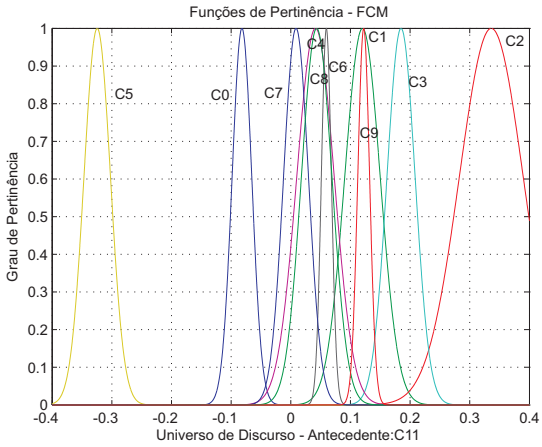

Figura 4: Função de pertinência do parâmetro C_{11}^{j}

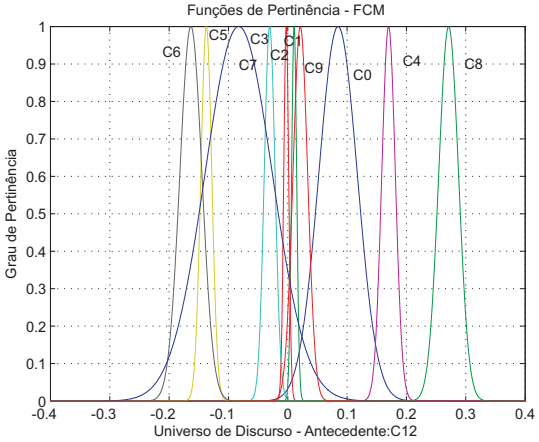

Figura 5: Função de pertinência do parâmetro C_{12}^{j}

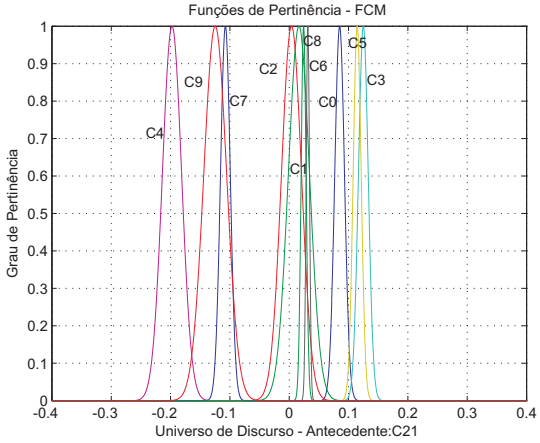

Figura 6: Função de pertinência do parâmetro C_{21}^{j}

2. Os locutores 11 e 12 falaram, também, em duas séries tomadas em dias diferentes e horários diferentes, em condições diferentes, dez vezes os dígitos de 0 a 9 num total de cem locuções por série, visando, também, os procedimentos de teste.

3.2 Modo teste

Neste modo, utilizou-se, 100 locuções pronunciadas em ambiente com controle de nível de ruído e 400 locuções pronunciadas em ambiente sem nenhum tipo de controle de ruído. Para cada dez exemplos de cada dígito falado, foi gerada uma matriz temporal de coeficientes cepstrais bidimensional $\mathbf{C}^{j}$, utilizada no procedimento de teste. Efetivamente, foram realizados cinco tipos de testes:

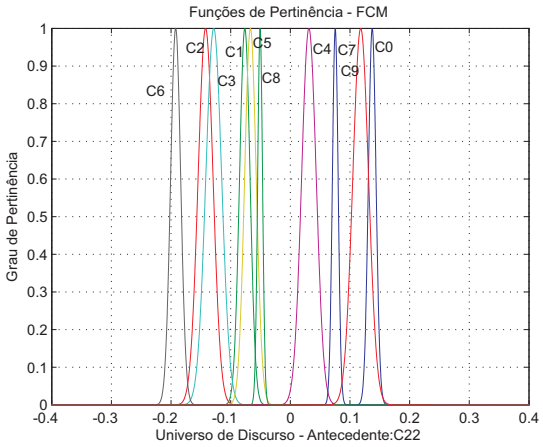

Figura 7: Função de pertinência do parâmetro C_{22}^{j}

1. TESTE 1: Reconhecimento estritamente dependente do locutor, onde as palavras utilizadas para o treinamento e testes foram pronunciadas por um mesmo grupo de 10 locutores.

2. TESTE 2: Reconhecimento com dependência parcial do locutor, onde o locutor submetido ao reconhecimento participou do processo de treinamento com dois exemplos para cada dez exemplos de cada dígito (Locutor masculino).

3. TESTE 3: Reconhecimento com dependência parcial do locutor, onde o locutor submetido ao reconhecimento participou do processo de treinamento com dois exemplos para cada dez exemplos de cada dígito (Locutor feminino).

4. TESTE 4: Reconhecimento independente do locutor, onde o locutor submetido aos testes não teve nenhuma participação no processo de treinamento dos modelos (Locutor masculino).

5. TESTE 5: Reconhecimento independente do locutor, onde o locutor submetido aos testes não teve nenhuma participação no processo de treinamento dos modelos (Locutor feminino).

A tabela 1 mostra a análise do desempenho do sistema TCD-*Fuzzy* para o reconhecimento de voz, considerando-se a ordem mínima da matriz temporal com coeficientes mel-cepstrais e, submetido aos testes supracitados. Observa-se, claramente, a eficiência da metodologia proposta comparada com o método HMM largamente usado na literatura.

	TCD-*Fuzzy*	HMM
TESTE 1	89%	84%
TESTE 2	84%	50%
TESTE 3	79%	66%
TESTE 4	75%	71%
TESTE 5	82%	50%

Tabela 1: Resultados dos testes com o TCD-*Fuzzy* e HMM.

4 CONCLUSÕES

Observa-se pelos resultados que a proposta de Reconhecedor de voz baseado em um classificador TCD-Fuzzy, mesmo com uma quantidade mínima de parâmetros nos padrões gerados foi capaz de extrair mais fielmente as características temporais do sinal de voz e apresentar bons resultados de reconhecimento, quando comparado com o HMM. No desenvolvimento deste trabalho não foi utilizada nenhuma técnica de específica de redução de ruído, tais como os utilizados normalmente nos reconhecedores baseados em HMM. Acredita-se que com o tratamento adequado da relação sinal-ruído nos processos de treinamento e teste, poderá acarretar em um melhor desenvolvimento do Reconhecedor TCD-*Fuzzy*. Um aumento nos exemplos utilizados no banco de geração dos padrões poderá aumentar o grau de confiabiidade melhorando também o desempenho do TCD-*Fuzzy*.

AGRADECIMENTOS

Os autores agradecem ao CNPq pelo apoio financeiro e ao grupo de inteligência computacional aplicada à tecnologia do Instituto Federal do Maranhão pela infra-estrutura disponível para realização desta pesquisa e obtenção dos resultados experimentais.

REFERÊNCIAS

[1] J. W. Picone. *Signal modeling techiniques in speech recognition*. IEEE Transactions on Computer, vol.79, n.4, p.1214-1247, vol.2 edition, April 1991.

[2] L. Rabiner and J. Biing-Hwang. *Fundamentals of Speech Recognition.* Prentice Hall, New Jersey, 1993.

[3] H. C. Andrews. *Multidimensional Rotations in Feature Selection*. IEEE Transaction on Computers, September 1971.

[4] D. F. W. Z. Shenouda, S.D. and D. A. Goneid. *Hybrid Fuzzy HMM System for Arabic Connectionist Speech Recognition*. The 23rd National U.Jio Science Conference (NRSC 2006), Egypt, March 2006.

[5] Yong-Qian and Y. P.-Y. Woo. *Speech Recognition Using Fuzzy Logic*. IEEE, Northern Illinois University, Dekalb, 1999.

[6] P. L. L. Fissore and E. Rivera. *Using word temporal structure in HMM Speech recongnition*. ICASSP 97, vol.2, p.975-978, Munich-Germany, April 1997.

[7] T. N. N. Ahmed and K. Rao. *Discrete Cosine Trasnform*. IEEE Transaction on Computers, vol.c-24 edition, January 1974.