

Algoritmo Competitivo Aplicado ao Reconhecimento Automático da Identidade Vocal de Locutores

J. M. Fechine¹ F. Madeiro R. M. Vilar B. G. Aguiar Neto M. S. de Alencar *
LAPS, DEE, COPELE – Universidade Federal da Paraíba
58.109-970 - Campina Grande, PB, Brasil

¹Departamento de Estatística e Informática – Universidade Católica de Pernambuco
50.050-900 - Recife, PE, Brasil
{joseana,madeiro,rosa,bganeto,malencar}@dee.ufpb.br

Abstract

In recent works, two unsupervised algorithms (referred to as SSC and KMVVT) were successfully applied for vector quantization codebook design, leading to good results in signal processing applications. The present paper is concerned with a comparative study of SSC, KMVVT and the traditional LBG algorithm for designing codebooks of acoustic parameters in a speaker identification system based upon vector quantization. It is shown that SSC is a promising technique since it leads to good recognition rates (up to 97.8%), significantly higher than the ones obtained by using KMVVT or LBG. The authors also show that the speaker identity seems to be suitably represented by cepstral coefficients. In fact, these acoustic parameters lead to higher recognition rates when compared to those provided by Δ -cepstral coefficients.

1. Introdução

O processo de reconhecimento da identidade vocal de locutores consiste na extração de parâmetros de voz de um dado locutor, de forma a definir um modelo que preserve as características vocais deste locutor, possibilitando diferenciá-lo de outros locutores.

Os sistemas de entrada vocal, por não exigirem as mãos e os olhos do usuário para sua operação, podem ser utilizados em diversas aplicações, tais como: auxílio a deficientes físicos, controle de qualidade e inspeção e controle de acesso a ambientes restritos. Esses sistemas também se aplicam em situações nas quais: é inconveniente o uso de teclado; as mãos do usuário estão ocupadas; mobilidade é exigida durante o processo de entrada de dados e os olhos do operador devem permanecer fixos sobre um *display* [1].

Dado um sinal de voz de entrada, o objetivo do reconhecimento de locutor é identificar a pessoa mais provável de ser o locutor (dentro uma população conhecida) - Identificação de Locutor, ou verificar se o locutor é quem ele alega ser - Verificação de Locutor [2].

Em um sistema de reconhecimento de locutor a identificação é levada a efeito a partir da comparação dos padrões de teste do locutor a ser reconhecido com os padrões previamente armazenados (padrões de referência) de todos os locutores do sistema. Nesse sistema, o locutor é identificado como aquele que proporcionar o melhor “casamento” entre os padrões de teste e os padrões de referência [1]. Este tipo de identificação pode ocorrer de duas formas: conjunto-aberto (o locutor pode não fazer parte da população) e conjunto-fechado (sabe-se *a priori* que o locutor é um membro da população). Além disso, a identificação pode ser dependente do texto (o locutor pronuncia uma senha ou sentença pré-definida) ou independente do texto (não é feita a exigência anterior).

Com o objetivo de obter sistemas eficientes de reconhecimento automático de locutor, diversas técnicas têm sido utilizadas, dentre as quais destacam-se: Modelos de Markov Escondidos (HMMs - *Hidden Markov Models*) [3], Redes Neurais Artificiais [4], Quantização Vetorial (VQ - *Vector Quantization*) [5], Análise por Predição Linear [2] e Alinhamento Dinâmico no Tempo (DTW - *Dynamic Time Warping*) [2].

A tarefa de reconhecimento, apresentada neste trabalho, corresponde à identificação automática da identidade vocal de locutores, em um conjunto fechado e dependente do texto, que utiliza a Quantização Vetorial (QV) paramétrica para construção dos padrões representativos dos locutores. Esses parâmetros são obtidos a partir da análise por predição linear. Neste trabalho são avaliados dois tipos de coeficientes obtidos a partir dessa análise: coeficientes Cepstrais e Delta Cepstrais. Para o projeto dos dicionários do quantizador vetorial é realizada uma análise comparativa de três técnicas. A primeira utiliza o algoritmo tradicional LBG [6]. As demais técnicas utilizam redes neurais não supervisionadas, denominadas KMVVT (Kohonen Modificado com Vizinhança Centralizada em Torno do Vetor de Treino) [7, 8] e SSC (Competitivo no Espaço Sináptico) [9].

2. Descrição do Sistema

O sistema de reconhecimento (identificação) automático da identidade vocal de locutores desenvolvido é composto das seguintes etapas:

* Os autores gostariam de expressar os agradecimentos à CAPES e ao CNPq pelo apoio financeiro ao trabalho.

1. Processamento do sinal;
2. Extração de características;
3. Quantização vetorial;
4. Regra de decisão.

Como toda tarefa de reconhecimento de padrões, o sistema se divide em duas fases: treinamento e reconhecimento (identificação), descritas a seguir.

- Fase de Treinamento (Figura 1):

1. Elocução da sentença de treinamento;
2. Processamento do sinal;
3. Extração dos parâmetros da fala, a partir da elocução da sentença de treinamento, de forma a obter uma seqüência de vetores de características, utilizando a análise por predição linear (coeficientes Cepstrais e Delta Cepstrais). Para cada tipo de coeficiente é obtida uma seqüência de vetores de características;
4. Definição e armazenamento (para cada um dos L locutores cadastrados) de um conjunto de padrões de referência. Para cada tipo de coeficiente é obtido um dicionário de padrões vocais.

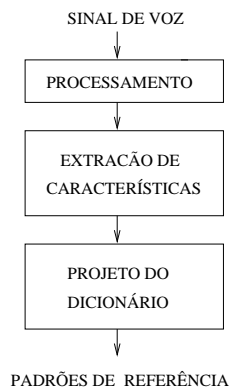


Figura 1: Fase de treinamento do Sistema de Reconhecimento (Identificação) de Locutor.

- Fase de Identificação (Figura 2):

1. Elocução da sentença de teste;
2. Processamento do sinal;
3. Extração dos parâmetros da fala, proporcionando a obtenção da seqüência de vetores de características representativa do locutor (coeficientes Cepstrais e Delta Cepstrais), uma seqüência de vetores para cada tipo de coeficiente;
4. Comparação dos padrões vocais de teste com o conjunto de padrões vocais de referência associados aos L locutores do sistema (obtidos durante a fase de treinamento);

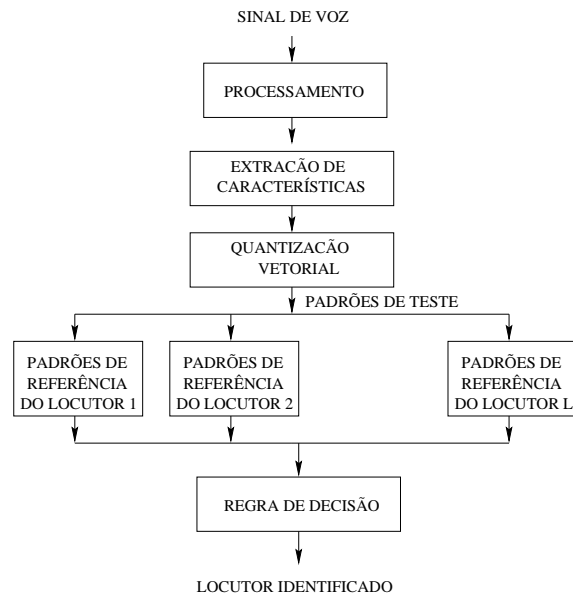


Figura 2: Fase de identificação do Sistema de Reconhecimento (Identificação) de Locutor.

5. Utilização de uma regra de decisão que permitirá a identificação do locutor.

A seguir serão descritas as etapas necessárias à implementação do sistema.

2.1. Processamento do sinal de voz

Esta etapa inclui a aquisição do sinal, pré-ênfase e janelamento em blocos, cada bloco contendo N_A amostras.

A aquisição e tratamento dos sinal de voz foram realizados da seguinte forma:

1. Aquisição: Placa Sound Blaster disponível no LAPS/DEE/UFPB, em ambiente relativamente silencioso, com um microfone comum (modelo Leadership);
2. Formato: WAV (sem cabeçalho);
3. Taxa de amostragem: 11 kHz;
4. Resolução: 16 bits (mono);
5. Pré-ênfase do sinal (fator de pré-ênfase: 0,95) [1, 10];
6. Tamanho dos segmentos de voz (N_A): 220 amostras (20 ms), garantindo as condições de estacionariedade do sinal [1].
7. Janelamento do sinal: Janela de Hamming, com superposição de 50% [1, 10].

2.2. Extração de características

O vetor de características é a interface entre as fases de treinamento e identificação e deve conter toda a

informação relevante à fase de classificação. Este vetor deve ser insensível às variações irrelevantes devidas às alterações das características acústicas quando da elocução de uma sentença e, ao mesmo tempo, ter uma baixa dimensionalidade visando minimizar a demanda de tempo computacional na etapa de classificação.

Os sistemas de reconhecimento de locutor têm usado diversas características [11], como formantes, intensidade, coeficientes obtidos a partir da análise por predição linear (Coeficientes LPC, Cepstrais, Cepstrais Ponderados, Delta Cepstrais, Delta Cepstrais Ponderados, etc.) [12], entre outras. Dentre estas últimas características, segundo Reynolds [12, 13], uma das mais utilizadas são os coeficientes Cepstrais.

A análise por predição linear é uma das mais importantes técnicas para análise de voz. Este método tem sido a técnica predominante para estimar os parâmetros da voz, utilizados para representação da voz em transmissão a baixa taxa de bits ou armazenamento. A importância desse método reside tanto na habilidade de fornecer estimativas extremamente corretas dos parâmetros da voz, quanto na relativa velocidade computacional [2].

No contexto desse trabalho é utilizada a técnica da análise por predição linear para obtenção dos parâmetros que irão compor os vetores de características de cada locutor, seguida de uma análise comparativa do desempenho de dois tipos de coeficientes: Cepstrais e Delta Cepstrais, de forma a obter um conjunto de parâmetros que proporcione a separabilidade máxima de um locutor, em relação aos demais locutores do sistema. Para cada segmento de fala janelado é extraído um conjunto de 12 coeficientes, tanto para a fase de treinamento quanto de identificação.

2.2.1 Coeficientes Cepstrais

Os coeficientes Cepstrais são usados para descrever a envoltória espectral do sinal de voz a curtos intervalos de tempo. O cepstrum é a transformada inversa de Fourier do logaritmo do espectro do sinal a curtos intervalos de tempo. Através da operação logaritmo, as funções de transferência do trato vocal e da fonte de voz são separadas [14].

Os coeficientes Cepstrais, ce_i , podem ser calculados recursivamente a partir dos coeficientes LPC, c_i [12]:

$$ce_i(n) = c_i(n) + \sum_{j=1}^{n-1} \left(\frac{j}{n}\right) ce_i(j) c_i(n-j), \quad (1)$$

com:

n - índice do coeficiente ($1 \leq n \leq K$);

i - índice do bloco de amostras.

Este método vem sendo utilizado comumente para definir as características do locutor, pois modela o trato vocal, o qual é a peça chave para distinguir locutores [12].

2.2.2 Coeficientes Delta Cepstrais

Os coeficientes Cepstrais representam as propriedades espectrais de um dado bloco de amostras de voz. Entretanto, estes não caracterizam a informação temporal ou de transição de uma seqüência de blocos de amostras de voz. Para aplicações envolvendo o reconhecimento de fala dependente do texto, um melhor desempenho pode ser obtido com a introdução da derivada cepstral no espaço de características, em virtude de esta derivada ser capaz de capturar a informação de transição da voz.

Os coeficientes Delta Cepstrais, Δce_i , podem ser obtidos da seguinte forma [12]:

$$\Delta ce_i(n) = \left[\sum_{q=-Q}^Q q ce_{i-q}(n) \right] G, \quad (2)$$

em que:

n - índice do coeficiente ($1 \leq n \leq K$);

i - índice do bloco de amostras;

G - termo de ganho (=0,375);

K - número de coeficientes Delta Cepstrais;

$2Q + 1$ - número de blocos de amostras sobre os quais o cálculo é realizado ($Q = 2$).

2.3. Quantização Vetorial

A Quantização Vetorial [15] é uma técnica de codificação usada tipicamente para transmissão a baixa taxa de bits. A eficiente taxa de redução de dados da QV, na parametrização de voz, é útil em Reconhecimento de Locutor para minimizar a memória utilizada, sendo sua principal vantagem a produção do dicionário para determinação da similaridade entre elocuições para um mesmo locutor e para discriminação entre locutores [16].

Um quantizador vetorial K -dimensional de M -níveis, é um mapeamento, q , que assume para cada vetor de entrada, $\vec{x} = \{x_0, \dots, x_{K-1}\}$, um vetor de reprodução, $\vec{\tilde{x}} = q(\vec{x})$, extraído de um alfabeto de reprodução finito $W = \{\vec{w}_i; i = 1, \dots, M\}$. O quantizador q é completamente descrito pelo alfabeto de reprodução (ou dicionário) W junto com a partição, $S = \{S_i; i = 1, \dots, M\}$, do espaço vetorial de entrada nos conjuntos $S_i = \{\vec{x} : q(\vec{x}) = \vec{w}_i\}$.

O conjunto W é denominado dicionário de reconstrução, M é o tamanho (número de níveis) do dicionário e \vec{w}_i são os vetores-código, ou vetores de reconstrução, de dimensão K [15].

O Algoritmo de Linde-Buzo-Gray (LBG) [6, 15], apesar de ser a técnica predominante para o projeto de dicionários, apresenta alguns problemas e limitações comumente reportados [17, 18], como por exemplo: a velocidade de convergência e o desempenho do dicionário final dependem do dicionário inicial; alguns vetores-código podem ser sub-utilizados e, em casos extremos, até mesmo

nunca serem acessados. Outras abordagens têm sido utilizadas para projeto de quantizadores vetoriais, tais como: algoritmo auto organizativo de Kohonen [19], e mais especificamente, os algoritmos KMVVT (Kohonen Modificado com Vizinhança Centrada em Torno do Vetor de Treino) [7, 8] e SSC (Competitivo no Espaço Sináptico) [9]. Devido as suas características, a aplicação das Redes Neurais no projeto dos dicionários para a quantização vetorial vem se tornando cada vez mais popular em aplicações envolvendo compressão de voz/imagem e reconhecimento de voz e de locutor.

2.3.1 O Algoritmo de Kohonen

Ao se utilizar o mapa auto-organizativo de Kohonen (SOM, *self-organizing map*) [19] para projetos de dicionários, os vetores-código (neurônios) K -dimensionais são geralmente associados a nós em um arranjo unidimensional ou em uma grade bidimensional. Os neurônios \vec{w}_i são aleatoriamente inicializados e iterativamente atualizados de acordo com a utilização de uma seqüência de treino. Para cada vetor de treino (vetor de entrada) \vec{x} , o neurônio vencedor \vec{w}_{i^*} (isto é, o neurônio mais “semelhante” a \vec{x} , segundo um critério de distorção) é determinado. É gerada uma vizinhança $\mathcal{N}_{\vec{w}_{i^*}} = \{\vec{w}_i | d_g(\vec{w}_i, \vec{w}_{i^*}) \leq r_g(n)\}$ em torno do neurônio vencedor, em que $r_g(n)$ denota o raio de vizinhança, medido na grade bidimensional e $d_g(\cdot)$ é uma distância medida na grade. Os neurônios pertencentes a $\mathcal{N}_{\vec{w}_{i^*}}$ são atualizados na direção do vetor de treino \vec{x} , de acordo com a regra de aprendizagem:

$$\Delta w_{ij}(n+1) = \eta(n) \cdot \mathcal{O}_i(i^*, n) \cdot (x_j - w_{ij}(n)), \quad (3)$$

em que

n - passo do algoritmo;

Δw_{ij} - modificação introduzida na j -ésima componente (peso sináptico) do neurônio \vec{w}_i ;

η - taxa de aprendizagem, com $0 < \eta < 1$;

\mathcal{O}_i - função que define a vizinhança em torno do neurônio vencedor \vec{w}_{i^*} ($\mathcal{O}_i = 1$ para $\vec{w}_i \in \mathcal{N}_{\vec{w}_{i^*}}$; $\mathcal{O}_i = 0$, caso contrário);

x_j - j -ésima componente do vetor de treino \vec{x} ;

w_{ij} - j -ésima componente do neurônio \vec{w}_i , $1 \leq i \leq M$ e $1 \leq j \leq K$.

Assim, dado o vetor de pesos sinápticos $\vec{w}_i(n)$ associado ao i -ésimo nó no passo n , a versão atualizada $\vec{w}_i(n+1)$ no passo $n+1$ é definida como

$$\vec{w}_i(n+1) = \vec{w}_i(n) + \eta(n) \cdot \mathcal{O}_i(i^*, n) \cdot (\vec{x} - \vec{w}_i(n)). \quad (4)$$

A cada passo n , a função de vizinhança $\mathcal{O}_i(i^*, n)$ é usada para alterar a taxa de aprendizagem do i -ésimo vetor-código (neurônio). Trata-se de uma função da

“distância física” do nó (na grade ou mapa bidimensional), correspondente ao i -ésimo neurônio, ao nó i^* , correspondente ao neurônio vencedor [18]. Tipicamente, $\mathcal{O}_i(i^*, n)$ é não-nula para nós próximos a i^* na grade, sendo nula para nós fora dessa vizinhança. A taxa de aprendizagem e o raio de vizinhança decrescem com n .

Podem ser identificadas duas fases no processo de aprendizagem que define o algoritmo não-supervisionado de Kohonen: fase de auto-organização ou ordenamento e fase de convergência. Na primeira, é obtido um ordenamento topológico dos neurônios na grade de nós; para tanto, a função raio de vizinhança deve incluir, no início do processo de treinamento, quase todos os neurônios cujos nós (na grade) estejam centrados no nó que identifica o neurônio vencedor; a função raio de vizinhança é então gradualmente reduzida de modo a incluir poucos neurônios ou, eventualmente, apenas o neurônio vencedor. A segunda fase é necessária para proporcionar uma melhor “sintonia” dos neurônios com a distribuição estatística da seqüência de treinamento. Nessa fase, tanto o raio de vizinhança como a taxa de aprendizagem são mantidos pequenos.

2.3.2 O Algoritmo Modificado de Kohonen com Vizinhança Centrada em Torno do Vetor de Treino

O algoritmo KMVVT, embora inspirado no processo de treinamento proposto por Kohonen, utiliza uma abordagem diferente para definição da vizinhança de atualização dos neurônios. Em virtude de os pesos sinápticos poderem ser vistos como as coordenadas dos neurônios (vetores-código, nesse contexto) no espaço R^K , a vizinhança pode ser adequada e satisfatoriamente definida no próprio espaço de padrões R^K , como uma hiper-esfera centrada em torno do vetor de treino \vec{x} . No algoritmo KMMVT a vizinhança é definida como:

$$\mathcal{N}_{\vec{x}} = \{\vec{w}_i | d_e(\vec{x}, \vec{w}_i) \leq r_e(n)\}, \quad (5)$$

em que $r_e(n)$ denota o raio de hiper-esfera na n -ésima iteração e $d_e(\cdot)$ denota a distância euclideana, medida no espaço R^K .

São definidas duas fases no algoritmo KMVVT. Na primeira, todos os neurônios pertencentes à vizinhança $\mathcal{N}_{\vec{x}}$ têm seus pesos atualizados. Na segunda, apenas o neurônio vencedor tem seus pesos ajustados.

Podem ser apontadas, portanto, diversas diferenças entre o algoritmo KMVVT e o algoritmo de Kohonen (SOM): no primeiro, a vizinhança é definida no próprio espaço R^K (precisamente, como uma hiper-esfera), enquanto que no segundo é definida em um arranjo topológico de nós (em geral, um mapa bidimensional, que pode ser retangular, hexagonal, etc.); no algoritmo KMVVT, a vizinhança é centrada no vetor de treino, enquanto que no algoritmo SOM é centrada em torno do nó (em uma grade) do neurônio vencedor. Por este motivo, em sua primeira fase, ao contrário do algoritmo SOM, o algoritmo KMVVT dispensa a necessidade de

determinação do neurônio vencedor. Além disso, a segunda fase do algoritmo de Kohonen contempla a possibilidade de serem atualizados os neurônios pertencentes a uma pequena vizinhança; no algoritmo KMVVT, por sua vez, não se utiliza vizinhança na segunda fase de treinamento: apenas o neurônio vencedor é atualizado.

2.3.3 Algoritmo Competitivo no Espaço Sináptico

O algoritmo SSC pode ser visto como uma versão simplificada do algoritmo KMVVT, em virtude de não ser necessário definir uma estrutura de vizinhança. No algoritmo SSC, apenas o neurônio vencedor tem seus pesos sinápticos atualizados. Neste sentido, o algoritmo SSC corresponde à dinâmica da segunda fase de treinamento do algoritmo KMVVT. Em termos de complexidade computacional, portanto, apresenta-se como uma alternativa mais adequada que o KMVVT.

2.4. Regra de Decisão

A regra de decisão utilizada pode ser descrita da seguinte forma: comparação dos padrões de teste com o conjunto de padrões de referência. Os padrões de referência que proporcionarem o menor erro médio de quantização identificam o locutor, desde que este erro seja maior que um dado limiar, cuja definição visa evitar o acesso de locutores não cadastrados pelo sistema, para contemplar o reconhecimento em um conjunto-aberto.

3. Apresentação e Análise dos Resultados

O sistema de reconhecimento de locutor foi implementado utilizando linguagem C. A base de dados foi formada por 20 locutores (10 do sexo masculino e 10 do sexo feminino), sendo estes estudantes universitários com idades próximas. Todos os locutores foram utilizados para treinamento e teste. A sentença utilizada para treinamento e teste foi a mesma (*Quero usar a máquina*), com duração de aproximadamente 1,5 segundos. Os dados de treinamento e de teste foram obtidos em sessões distintas e cada locutor pronunciou a sentença 5 vezes na fase de treinamento e 20 vezes na fase de teste.

Foram implementados três tipos de sistema, o primeiro utilizando quantização vetorial com dicionários projetados a partir do algoritmo LBG (QV-LBG), o segundo e o terceiro utilizando a quantização vetorial com dicionários projetados a partir de Redes Neurais (QV-RN) – o segundo sistema usa o algoritmo KMVVT e o terceiro utiliza o algoritmo SSC. Para cada sistema, os parâmetros acústicos representativos dos locutores foram obtidos, utilizando, inicialmente, os coeficientes Cepstrais e, em seguida, os coeficientes Delta Cepstrais.

Os resultados obtidos para o sistema de reconhecimento automático da identidade vocal de locutores (taxas médias de identificação, falsa aceitação e falsa rejeição e confiabilidade) são apresentados nas Tabelas 1 e 2.

A partir da Tabela 1, pode-se observar que para os coeficientes Cepstrais o sistema obteve um melhor desempenho com a utilização do algoritmo QV-SSC, visto

Tabela 1: Desempenho do sistema para os três algoritmos: QV-LBG, QV-KMVVT e QV-SSC, utilizando os coeficientes Cepstrais.

TAXAS MÉDIAS	QV-LBG	QV-KMVVT	QV-SSC
Identificação	92,0%	96,5%	97,8%
Falsa Aceitação	1,5%	1,3%	0,8%
Falsa Rejeição	6,5%	2,3%	1,5%
Confiabilidade	98,4%	98,7%	99,2%

Tabela 2: Desempenho do sistema para os três algoritmos: QV-LBG, QV-KMVVT e QV-SSC, utilizando os coeficientes Delta Cepstrais.

TAXAS MÉDIAS	QV-LBG	QV-KMVVT	QV-SSC
Identificação	87,8%	88,0%	88,5%
Falsa Aceitação	6,0%	7,3%	3,3%
Falsa Rejeição	6,3%	4,8%	8,3%
Confiabilidade	93,6%	92,4%	96,5%

que este proporciona a maior taxa média de identificação (97,8%) e as menores taxas médias de falsa aceitação (0,8%) e rejeição (1,5%), bem como a maior confiabilidade (99,2%). Observa-se que o algoritmo QV-KMVVT, mesmo com medidas de desempenho inferiores às obtidas com o algoritmo QV-SSC, apresenta desempenho superior ao apresentado pelo algoritmo QV-LBG.

A Tabela 2 mostra que, para os coeficientes Delta Cepstrais, o melhor desempenho é obtido com o algoritmo QV-SSC, à exceção da taxa de falsa rejeição. Entretanto, é importante destacar que a utilização desses coeficientes proporciona um desempenho inferior ao obtido com o uso dos coeficientes Cepstrais.

Conforme apresentado nas Tabelas 1 e 2, para quaisquer dos algoritmos e coeficientes, as taxas médias de falsa rejeição são, de uma forma geral, maiores que as de falsa aceitação. Este fato representa uma característica importante do sistema, pois é melhor rejeitar um locutor erroneamente do que identificá-lo de forma incorreta.

Outro aspecto importante a ser observado na avaliação de medidas de desempenho é a dispersão dos valores individuais em torno da média, o qual pode ser medido por meio do coeficiente de variação (C.V.), definido como a razão percentual entre a média e o desvio padrão de uma distribuição [20]. A Tabela 3 mostra que, para os coeficientes Cepstrais, as taxas médias de identificação obtidas com o algoritmo QV-SSC são

Tabela 3: Coeficiente de Variação (C.V.) das taxas médias de identificação do sistema utilizando os três algoritmos: QV-LBG, QV-KMVVT e QV-SSC para os coeficientes Cepstrais (CEP) e Delta Cepstrais (DCEP).

ALGORITMO	CEP	DCEP
QV-LBG	14,4%	19,7%
QV-KMVVT	7,0%	20,8%
QV-SSC	4,5%	21,2%

as mais representativas (menor C.V.), seguidas das obtidas com o algoritmo QV-KMVVT e por fim pelo algoritmo QV-LBG. No que se refere aos coeficientes Delta Cepstrais, observa-se que a representatividade dos valores médios obtidos é bem menor quando comparada aos coeficientes Cepstrais e além disso que existe pouca variabilidade em relação aos três algoritmos.

4. Conclusões

Neste trabalho foi apresentado um algoritmo para identificação de locutor utilizando a quantização vetorial paramétrica. Os parâmetros utilizados para compor o vetor de características de cada locutor são os coeficientes obtidos a partir da análise por predição linear. Foram extraídos dois tipos de coeficientes (Cepstrais e Delta Cepstrais), visando determinar qual destes melhor representa as características vocais de cada locutor.

Os algoritmos LBG, KMVVT e SSC foram utilizados para o projeto de dicionários de padrões acústicos (obtidos a partir do vetor de características) destinados à tarefa de reconhecimento da identidade vocal de locutores.

Os resultados obtidos evidenciam a superioridade dos algoritmos que utilizam redes neurais sobre o tradicional algoritmo LBG no projeto de dicionários de padrões acústicos para identificação de locutor. Verificou-se, também, que dentre os algoritmos KMVVT e SSC, o segundo apresenta desempenho superior, além de ser mais simples a sua implementação.

Dentre os tipos de coeficientes utilizados, observou-se que os coeficientes Cepstrais são os que apresentam maior eficiência no que se refere à identificação de pessoas a partir de suas vozes, pois proporcionam as maiores taxas médias de identificação e confiabilidade, bem como as menores taxas médias de falsa rejeição e de falsa aceitação. Os coeficientes Cepstrais estabelecem, portanto, um bom compromisso entre as taxas de falsa rejeição (locutor cadastrado não é identificado) e as taxas de falsa aceitação (locutor é identificado de forma incorreta).

A identificação de locutor não se justifica apenas por possibilitar uma interação mais confortável entre o homem e a máquina, mas sobretudo pela segurança que pode proporcionar. Os resultados apresentados neste trabalho satisfizeram as exigências dos sistemas de reconhecimento de locutor, tendo sido obtidos, para a maioria dos locutores, pequenos índices de falsa aceitação e de falsa rejeição.

Referências

- [1] Rabiner, L. R. and Schafer, R. W. *Digital Processing of Speech Signals*. Prentice-Hall, Upper Saddle River, New Jersey, 1978.
- [2] Campbell, J. P. Speaker Recognition: A Tutorial. *Proceedings of the IEEE*, Vol. 85, No. 9, pages 1437–1462, September 1997.
- [3] Rabiner, L. R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, Vol. 77, No. 2, pages 257–286, February 1989.

- [4] Oglesby, J. and Mason, J. S. Optimisation of Neural Models for Speaker Identification. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'90)*, pages 261–264, 1990.
- [5] Bennani, Y., Fogelman-Soulié, F., and Gallinari, P. Text-Dependent Speaker Identification Using Learning Vector Quantization. *Proceedings of the International Neural Network Conference*, Vol. 2, Dordrecht, Netherlands, pages 1087–1090, 1990.
- [6] Linde, Y., Buzo, A., and Gray, R. M. An Algorithm for Vector Quantizer Design. *IEEE Transactions on Communications*, Vol. COM - 28, No. 1, pages 84–95, January 1980.
- [7] Madeiro, F., Vilar, R. M., Fechine, J. M., and Aguiar Neto, B. G. A Self-Organizing Algorithm for Vector Quantizer Design Applied to Signal Processing. *International Journal of Neural Systems*, Vol. 9, No. 3, Special Issue on Neural Networks in Brazil: V Brazilian Symposium on Neural Networks, pages 219–226, June 1999.
- [8] Madeiro, F., Fechine, J. M., and Aguiar Neto, B. G. Algoritmo Modificado de Kohonen Aplicado ao Projeto de Dicionários de Padrões Acústicos para Reconhecimento de Locutor. *Anais do V Simpósio Brasileiro de Redes Neurais (SBRN'98)*, Belo Horizonte – MG, Brasil, pages 22–26, Dezembro, 1998.
- [9] Vilar França, R. M. and Aguiar Neto, B. G. Voice Waveform Vector Quantization Using a Competitive Algorithm. “Records of the IEEE GLOBECOM'94”, pages 872–875, November 1994.
- [10] Deller Jr., J. R., Proakis, J. G., and Hansen, J. H. L. *Discrete-time Processing of Speech Signals*. Macmillan Publishing Co., 1993.
- [11] Farrell, K. R., Mammone, R. J., and Assaleh, K. T. Speaker Recognition Using Neural Networks and Conventional Classifiers. *IEEE Transactions on Speech and Audio Processing (Special Issue on Neural Networks for Speech)*, Vol. 2, pages 194–205, January 1994.
- [12] Mammone, R. J., Zhang, X., and Ramachandran, R. P. Robust Speaker Recognition – A Feature-Based Approach. *IEEE Signal Processing Magazine*, Vol. 13, No. 5, pages 58–71, September 1996.
- [13] Reynolds, D. A. and Rose, R. C. Robust Text-independent Speaker Identification Using Gaussian Mixture Speaker Models. *IEEE Transactions on Speech and Audio Processing*, Vol. 3, No. 1, pages 72–83, January 1995.
- [14] Tolba, H. and O'Shaughnessy, D. Voiced-Unvoiced Classification Using the First Mel Frequency Cepstral Coefficient. *Proceedings of the International Conference on Speech Processing*, Vol. 1, Seoul - Korea, pages 137–142, August 1997.
- [15] Gray, R. M. Vector Quantization. *IEEE ASSP Magazine*, pages 4–29, April 1984.
- [16] He, J., Liu, L., and Palm, G. A Discriminative Training Algorithm for VQ-based Speaker Identification. *IEEE Transactions on Speech and Audio Processing*, Vol. 7, No. 3, pages 353–356, May 1999.
- [17] Lee, D., Baek, S., and Sung, K. Modified K-means Algorithm for Vector Quantizer Design. *IEEE Signal Processing Letters*, Vol. 4, No. 1, pages 2–4, January 1997.
- [18] Yair, E., Zeger, K., and Gersho, A. Competitive Learning and Soft Competition for Vector Quantizer Design. *IEEE Transactions on Signal Processing*, Vol. 40, No. 2, pages 294–309, February 1992.
- [19] Kohonen, T. *Self-Organization and Associative Memory (3rd ed)*. Springer-Verlag, Berlin, 1989.
- [20] Costa Neto, P. L. de O. *Estatística*. Editora Edgard Blucher LTDA., São Paulo, SP, 1977.