Sociedade de Engenharia e Áudio

Artigo de Congresso

Apresentado no 6º Congresso da AES Brasil
12ª Convenção Nacional da AES Brasil
5 a 7 de Maio de 2008, São Paulo, SP

Classificação Automática de Sons de Instrumentos Musicais usando Discriminantes Lineares

Jorge Costa Pires Filho[1], Paulo Antonio Andrade Esquef [2], Luiz Wagner Pereira Biscainho [2]
[1]Instituto de Pesquisas da Marinha
Rio de Janeiro, RJ, Brasil
[2]PEE/COPPE & DEL/Poli, UFRJ
Rio de Janeiro, RJ, Brasil
jcpfilho@gmail.com, pesquef@yahoo.com, wagner@lps.ufrj.br

RESUMO
Este artigo apresenta um estudo da utilização de discriminantes lineares e transformações não-lineares do espaço de entrada num sistema para reconhecimento da assinatura sonora de instrumentos musicais. Utilizaram-se no trabalho amostras da base de dados MIS, da Universidade de Iowa. O sistema proposto é avaliado para uma seleção de escolhas de técnicas de pré-processamento de dados; formação do vetor de atributos; e classificação. Com o uso combinado de *Line Spectral Frequencies* (LSF) e Discriminantes Lineares, os desempenhos obtidos foram em torno de 92% e 86%, respectivamente, para a taxa de acerto da família do instrumento e do instrumento individualmente. Esses resultados são compatíveis com a faixa de taxas de acertos reportada na literatura.

0 INTRODUÇÃO

No atual contexto de reconhecimento de instrumentos musicais, ainda não há consenso quanto à melhor abordagem em sinais polifônicos (os quais apresentam simultaneamente sons de diversos instrumentos musicais). Atualmente, a maior parte dos estudos desta área contempla o caso monofônico, seja em notas isoladas, seja em trechos de música solo. Um breve levantamento em [6] destaca os seguintes trabalhos e resultados: Marques e Moreno [1] reportam taxas de acerto de 70% com um sistema que analisa segmentos de 0,2 s, utiliza *Linear Prediction Coding* (LPC), *Line Spectral Frequencies* (LSF), FFT e *Mel-Frequency Cepstrum Coefficients* (MFCC) para formar o vetor de características, e emprega modelos de misturas gaussianas (GMM) e *Support Vector Machines* (SVM) para classificar os trechos analisados de 9 instrumentos. Martin [2] usa um conjunto de características perceptuais derivadas de um correlograma lag-log para classificar notas isoladas de 27 instrumentos e reporta taxas de acerto de cerca de 86% para a família do instrumento e cerca de 71% para o instrumento individualmente. Eronen e Klapuri [3] também usam um conjunto de características perceptuais para classificar notas isoladas de 30 instrumentos e reportam taxas de acerto de cerca de 94% e 85%, respectivamente, para família e instrumento individualmente. Agostini *et al.* [4] empregam somente características espectrais para classificar 27 instrumentos e reportam taxas de acerto de cerca de 96% e 92% para família e instrumento, respectivamente. Kitahara *et al.* [5] apresentam um classificador para tons de 19 instrumentos que utiliza uma distribuição normal de diversos parâmetros, dependente da freqüência fundamental, obtendo taxas de acerto de cerca de 90% e 80% para família e instrumento, respectivamente. Finalmente, Krishna e Sreenivas [6] descrevem um sistema que usa LSF como características representativas de segmentos obtidos a partir de notas isoladas e modelos de misturas gaussianas para a classificação. Relatam ter obtido taxas de acerto de 95% e 90% para família e instrumento, respectivamente. Exceto para Brown *et al*. [7] e Marques e Moreno [1], todos os outros resultados

reportados se referem a sistemas classificadores que utilizam notas isoladas.

Uma das motivações dessa área de estudo são aplicações comerciais que visam a catalogar discotecas através de um processo automático, etiquetando cada música com a presença dos instrumentos musicais que a compõem, facilitando assim uma busca seletiva. Outras aplicações de interesse são a transcrição automática de música [8] e a codificação de áudio em alto nível, usando modelagem de fonte sonora [9].

A escolha de se abordar a classificação de instrumentos musicais a partir de notas isoladas nesse estudo pode ser justificada por diversos motivos. Primeiramente, ela pode ser adaptada tanto para classificar trechos de música monofônica quanto para outros sinais de áudio oriundos de uma única fonte. No mais, a identificação de instrumentos a partir de notas isoladas, apesar de não ser a mais apropriada para resolver o problema na sua concepção mais geral (sinais de música contendo superposição no tempo e na freqüência de vários instrumentos musicais), não é restritiva caso se queira identificar sinais que já tenham passado por um processo de separação de fontes. Pode-se assumir que o escopo do presente trabalho é identificar qual é o instrumento associado a um sinal que, tendo sido previamente separado, pertence a uma única fonte.

O fato de se usar nesse trabalho discriminante linear para a função de classificação tem como principal meta avaliar o desempenho de um algoritmo mais simples, mais robusto e mais rápido na fase de treinamento do que os classificadores mais elaborados propostos na literatura, tais como o SVM e o GMM.

1 METODOLOGIA

Uma das preocupações deste trabalho foi obter resultados que permitissem a avaliação comparativa da eficiência do uso da técnica de discriminação linear na identificação de instrumentos frente a soluções concorrentes. Assim, para se traçar uma avaliação de desempenho utilizaram-se como paradigmas os resultados apresentados por diversos autores, sumarizados em [6]. Isso permite avaliar quão bom é o desempenho que se obtém com o uso do discriminante linear combinado com a forma de obtenção do vetor de características proposta neste artigo.

Para atacar o problema de reconhecimento da assinatura sonora de notas isoladas de instrumentos musicais, subdividiu-se o problema em módulos funcionais independentes. O sistema que os integra para formar o classificador é ilustrado na Figura 1.

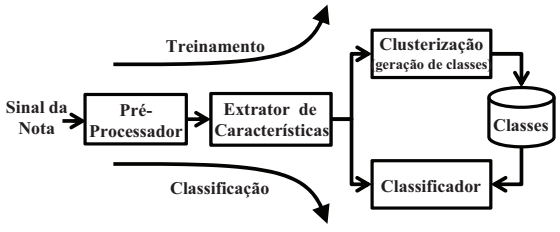

Figura 1: Sistema de Identificação de Instrumentos Musicais.

Como pode ser visto, o procedimento se divide em 4 blocos, de forma bem próxima ao modelo em [9]:

1. Pré-processador;
2. Extrator de Características;
3. Gerador de Classes;
4. Classificador.

O Pré-processador é responsável pelo escalonamento dinâmico dos sinais e sua segmentação em trechos de interesse. O Extrator de Características é responsável pela obtenção do vetor de características que representa o sinal entregue pelo Pré-processador. O Gerador de Classes é responsável pela definição das classes em que serão agrupados os sinais da base de dados, tanto no estágio de treinamento como no de testes do classificador. E o Classificador é o módulo responsável por decidir sobre qual a classe a que pertence um dado sinal de teste, representado por seu vetor de características.

Durante a etapa de classificação, dois conjuntos distintos de sinais (amostras) foram utilizados: um conjunto de treinamento para o classificador e outro conjunto de teste para a avaliação da taxa de acerto.

2 BANCO DE DADOS

As amostras de sons de instrumentos musicais (*Musical Instrument Samples* - MIS) da Universidade de Iowa foram gravadas em uma câmara anecóica no ***"Wendell Johnson Speech and Hearing Center"*** na Universidade de Iowa [10] com os seguintes equipamentos:

1. Microfone Neumann KM 84;
2. Mixer Mackie 1402-VLZ;
3. Gravador DAT Panasonic SV-3800.

As únicas exceções foram os sons de piano, cuja gravação não-anecóica ocorreu em um pequeno estúdio.

Os sinais são monofônicos (com exceção do piano, gravado em estéreo), amostrados a 44,1 kHz e representados em 16 bits, e foram armazenados em formato AIFF.

Para cada instrumento, os sinais gravados englobam a tessitura usual do instrumento em escalas cromáticas tocadas em três níveis dinâmicos não normalizados: *pp*, *mf* e *ff*, ou seja, *pianissimo*, *mezzo forte* e *fortissimo*. Quando cabível, foram gravados estilos diferentes de execução, por exemplo, com e sem *vibrato*, com arco e *pizzicato*. Cada nota tem aproximadamente 2 segundos de duração, sendo imediatamente precedida e seguida de silêncio.

Dos sinais disponíveis na base de dados, este trabalho utilizou somente os listados na Tabela 1, apresentada na Seção 6.

3 PRÉ-PROCESSAMENTO DO SINAL

O módulo Pré-processador recebe o sinal de áudio correspondente a uma nota isolada e devolve um conjunto de trechos do sinal já pré-processados.

O objetivo principal do estágio de pré-processamento é segmentar o sinal em três trechos que correspondem ao ataque, sustentação e decaimento da nota. Para isso foram utilizados 2 limiares sobre um sinal de detecção, escolhido como a potência instantânea do sinal da nota gravada: o primeiro localizado a 10% e o segundo a 90% da média do sinal de detecção. Portanto, o ataque corresponderá ao trecho compreendido entre o instante de tempo em que o sinal de detecção ultrapassa pela primeira vez o primeiro limiar (10%) e o instante em que o sinal de detecção ultrapassa pela primeira vez o segundo limiar (90%). Já para obtenção do trecho de decaimento faz-se um

procedimento análogo, porém começando do final do sinal de detecção para seu inicio. O trecho de quase-estacionariedade, ou seja, de sustentação, é considerado como o intervalo de tempo entre o final do ataque e o início do decaimento. Na Figura 2, observa-se a envoltória do sinal de detecção correspondente a uma nota, com os dois limiares usados para a segmentação.

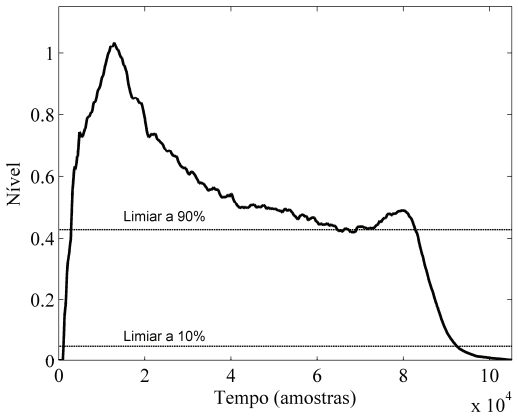

Figura 2: Envoltória do sinal de detecção (potência instantânea do sinal) contra os limiares de segmentação. Para melhor legibilidade, optou-se por substituir o sinal de detecção por sua envoltória.

4 EXTRAÇÃO E PROCESSAMENTO DE CARACTERISTICAS

O módulo Extrator de Características é responsável pela obtenção de um conjunto de características representativas do trecho de sustentação do sinal segmentado pelo módulo anterior. Seguindo a estratégia utilizada em [11] e [6], no presente estudo avaliaram-se características estatísticas (os momentos de segunda e terceira ordem) e parâmetros relacionados à análise espectral paramétrica do sinal (os coeficientes de predição linear ou as LSFs, para ordens predefinidas). O número de elementos do vetor de características será variado numa faixa arbitrariamente predefinida.

Antes da extração do vetor de características, o trecho de sustentação sofre um escalamento [11] de forma a assumir média zero e desvio-padrão unitário. Para esse procedimento, o momento de segunda ordem já teve de ser calculado. Calcula-se depois o momento de terceira ordem.

Em seguida, são estimados os coeficientes de predição linear (LPC) e aqueles associados aos polinômios das LSFs. Para um LPC de ordem *N*, o procedimento consiste em encontrar um conjunto de coeficientes a_k que minimiza o erro quadrático médio do seguinte preditor *forward*, aplicado em uma seqüência s_n:

$$\hat{s}_n = \sum_{k=1}^{N} a_k s_{n-k} + e_n. \quad (2)$$

O preditor da Equação (2) pode ser visto como a saída de um filtro gerador só-pólos $H(z) = 1/A(z)$, com

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_N z^{-N}, \quad (3)$$

excitado por e_n. Dois polinômios, simétrico e assimétrico, associados às LSFs podem ser definidos a partir de $A(z)$, respectivamente, por.

$$P(z) = A(z) + z^{-(N+1)} A(z^{-1}) \quad (4)$$

$$Q(z) = A(z) - z^{-(N+1)} A(z^{-1}). \quad (5)$$

As raízes de $P(z)$ e $Q(z)$ se localizam na circunferência unitária, e suas fases definem os valores das LSFs.

Por fim, o vetor de características é formado pelo conjunto de coeficientes LSF ou LPC mais os momentos de ordens 2 e 3 do trecho de sustentação.

Vale ressaltar que redundâncias presentes no vetor de características definido acima poderiam ser eliminadas com a aplicação de técnicas de redução de dimensionalidade. Entretanto, isso fica além do escopo deste trabalho.

5 CLASSIFICADOR

Foram avaliados três tipos distintos de classificadores, a saber:

1. Vizinho mais próximo (VMP)
2. Máquina de Vetor Suporte (SVM)
3. Discriminante Linear Generalizado (DLG)

Como os classificadores SVM e DLG somente conseguem separar duas classes, optou-se por avaliar apenas uma forma específica de se obter o resultado no caso multiclasse.

A generalização para discriminação multiclasse adotada neste trabalho utilizou o procedimento um-contra-um (*one-against-one*) [2,6]: calculam-se *P* discriminantes, onde *P* representa o número de duplas possíveis no total de classes que estão sendo avaliadas. Uma dada amostra é testada segundo todos os *P* discriminantes, e o número de atribuições da amostra a cada classe é contabilizado. A amostra é classificada como pertencente à classe que recebeu mais votos.

5.1 Vizinho mais Próximo (VMP)

Este método estima a classe mais provável de uma dada amostra a ser classificada pela sua distância (segundo alguma métrica) a um conjunto de treinamento formado por amostras cujas classes são previamente conhecidas.. Percorre-se o conjunto de treinamento, calculando a distância de cada uma de suas amostras à amostra a classificar, em busca da que apresenta a menor distância. A classe atribuída à amostra sob teste será a desse 'vizinho mais próximo'.

Neste trabalho arbitrou-se como métrica de distância a distância Euclidiana de ordem 2 entre a amostra **X** e a amostra $\mathbf{M}^j$ do conjunto de treinamento:

$$D_x^j = \sqrt{\sum_{i=1}^{n} \left(x_i - M_i^j\right)^2}, \quad (6)$$

onde:

x_i = elemento *i* do vetor de características da amostra **X**.

M_i^j = elemento *i* do vetor de características da amostra $\mathbf{M}^j$ do conjunto de treinamento $\mathcal{M}$.

5.2 Máquina de Vetor Suporte

Concisamente, uma SVM implementa discriminantes lineares (hiperplanos) no espaço obtido por uma

transformação não-linear do espaço de entrada. A Figura 3 ilustra a separação por discriminantes lineares.

Na sua forma tradicional, uma SVM diferencia uma classe, a positiva, de outra, a negativa, em um esquema binário de classificação. Para isso a SVM constrói um hiperplano que maximiza a margem de separação entre os exemplares positivos e os negativos. Esse objetivo é atingido através de uma abordagem baseada na Teoria Estatística de Aprendizagem [12,13], implementando aproximadamente o método de minimização do risco estrutural [12].

Apesar de utilizar discriminantes lineares, uma SVM não necessita, para efeitos de generalização, de classes linearmente separáveis. Tal propriedade se deve ao fato de a discriminação ser empregada num espaço de características já submetido a uma transformação não-linear. Esta operação pode ser justificada invocando-se o célebre Teorema de Cover [12], que afirma que padrões não-linearmente separáveis pertencentes a um dado espaço de características são, com alta probabilidade, linearmente separáveis num espaço de características transformado, desde que: (a) a transformação seja não-linear; (b) a dimensão do espaço transformado seja alta o suficiente.

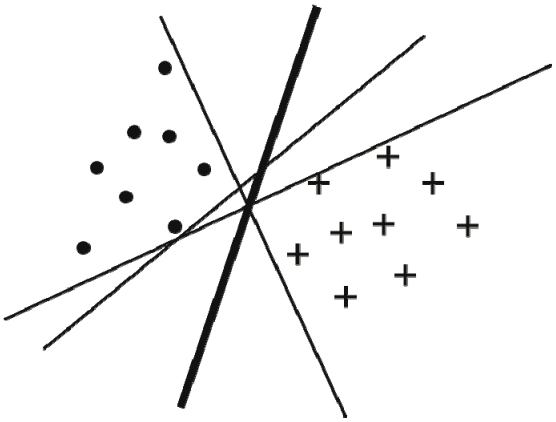

Figura 3: Hiperplano separador ótimo (linha grossa), de duas classes indicadas pelos marcadores • e +.

Para se obter a transformação não-linear do espaço de características requerida pelo SVM, utiliza-se o conceito de *kernel*. A idéia dá função *kernel* é aplicar operações no espaço de características ao invés de no espaço transformado, cuja dimensão é potencialmente maior. Assim, torna-se desnecessário calcular explicitamente o produto interno no espaço transformado para realizar as projeções, na tentativa de contornar o problema da dimensionalidade.

A SVM resolve um problema de programação não-linear que maximiza a margem entre os vetores transformados e o hiperplano separador, de modo a posicioná-lo de forma eqüidistante em relação aos chamados vetores-suporte.

Neste trabalho utilizou-se um *toolbox* SVM para o MATLAB (em dll) denominado SVM-KM e disponibilizado na Internet via licença geral de uso público da GNU. Em particular, foram utilizadas as funções svmclasslib e svmvallib. O *kernel* utilizado nesse trabalho para o SVM foi o gaussiano.

5.3 Discriminante Linear Generalizado (DLG)

Analogamente a uma SVM, o DLG tenta encontrar um hiperplano que separe duas classes. O objetivo é achar a partir de um conjunto de treinamento o vetor $\vec{w}$ que define um hiperplano separador, pela minimização do quadrado do erro de classificação

$$\varepsilon = t_{\vec{x}} - \widetilde{y}(\vec{x}), \tag{10}$$

onde $t_{\vec{x}}$ (que pode assumir os valores $\{-1,1\}$) é a classe da amostra $\vec{x}$, e $\widetilde{y}$ é uma função estimadora da classe. Assim, espera-se que se $\vec{w}^{\mathrm{T}}\vec{x} > 0$, sendo T o operador transposição, a amostra $\vec{x}$ pertença à classe 1; caso contrário, pertencerá à classe –1. Portanto, a classe da amostra $\vec{x}$ é determinada por

$$y(\vec{x}) = \operatorname{sign}\left(\vec{w}^{\mathrm{T}}\vec{x}\right). \tag{11}$$

Para viabilizar a minimização por métodos que utilizam a direção do gradiente, substituiu-se a função sinal em (11) pela função tangente hiperbólica. A mudança se justifica, uma vez que esta, assim como a função sinal, possui sua imagem limitada pelos valores $\{-1,1\}$, sendo contudo totalmente diferenciável em seu domínio. Redefine-se, então, a classe da amostra $\vec{x}$ como

$$\widetilde{y}(\vec{x}) = \operatorname{tgh}\left(\vec{w}^{\mathrm{T}}\vec{x}\right). \tag{12}$$

5.4 Transformação do Espaço de Características

Também foi investigado o efeito de uma extensão do espaço de características, consistindo na incorporação das potências, até um inteiro *k*, de cada parâmetro do vetor de características. Desta forma, se *n* é a dimensão do vetor de características associado a uma amostra, após a extensão *kn* será a nova dimensão tanto deste vetor de características transformado, agora definido por

$$\vec{x}_{\mathrm{p}} = \begin{cases} \left[\vec{x}\right], & \text{se} \quad k=1 \\ \left[\vec{x}^{\mathrm{T}} \;\; \vec{x}^{2^{\mathrm{T}}}\right]^{\mathrm{T}}, & \text{se} \quad k=2 \\ \vdots & \vdots \quad \vdots \\ \left[\vec{x}^{\mathrm{T}} \;\; \vec{x}^{2^{\mathrm{T}}} \;\; \cdots \;\; \vec{x}^{\Psi^{\mathrm{T}}}\right]^{\mathrm{T}}, & \text{se} \quad k=\Psi \end{cases}, \tag{13}$$

quanto do hiperplano separador, agora dado por

$$\vec{w} = \left[\vec{w}_1^{\mathrm{T}} \quad \vec{w}_2^{\mathrm{T}} \quad \cdots \quad \vec{w}_k^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{14}$$

Tomou-se na Eq. (13) a liberdade de notar por $\vec{v}^{\,i}$ a versão de $\vec{v}$ com todos os seus elementos elevados à potência i.

Nesse caso, a nova função estimadora da classe passa a ser

$$\widetilde{y}'(\vec{x}) = \operatorname{tgh}\left(\vec{w}^{\mathrm{T}}\vec{x}_{\mathrm{p}}\right). \tag{15}$$

Esta transformação não-linear foi usada em particular com o classificador DLG. Como se verá mais adiante, ela provocou um aumento na taxa de acerto das classes.

6 CLUSTERIZAÇÃO E FORMAÇÃO DAS CLASSES

Os tipos mais tradicionais de instrumentos musicais podem ser classificados de diversas formas, sendo uma das mais comuns a que se baseia no processo de produção de

som. O estudo dos instrumentos musicais designa-se por organologia, que foge ao escopo deste artigo.

Nesse trabalho, seguindo o procedimento de [6], quatro famílias de instrumentos foram definidas: flautas, palhetas, metais e cordas (FRBS – *Flutes, Reeds, Brass, and Strings*) agregando os instrumentos da Tabela 1 conforme a Tabela 2.

Tabela 1: Classe *Default* (Instrumento).

Instrumento	**# Notas**
Clarinete em Mi Bemol	119
Clarinete em Si Bemol	139
Fagote	122
Flauta	227
Flauta Baixo	102
Flauta Contralto	99
Oboé	104
Saxofone Contralto	192
Saxofone Soprano	192
Trombone Baixo	131
Trombone Tenor	99
Trompa	96
Violino	601
Violoncelo	668

Tabela 2: Classe FRBS (Família).

Família	**Instrumentos**
Flautas	Flauta, Flauta Baixo, Flauta Contralto
Palhetas	Clarinete em Mi Bemol, Clarinete em Si Bemol, Fagote, Oboé
Metais	Saxofone Contralto, Saxofone Soprano, Trombone tenor e Trompa
Cordas	Violino, Violoncelo

7 EXPERIMENTOS E RESULTADOS

Para realizar as simulações com o sistema classificador que utiliza o DLG, variou-se a quantidade dos coeficientes da parametrização LPC e LSF. Mais especificamente, foram predefinidas parametrizações com 8, 16 e 24 coeficientes.

Para cada uma das seis combinações possíveis, modificou-se também o grau de potenciação (conforme definido na Seção 5.4) para a transformação do espaço de entrada entre $k = 1$ e $k = 4$. Tais configurações de processamento foram adotadas para ambas as classes (*Default* e FRBS), perfazendo assim, um total de 48 simulações usando o DLG.

As demais simulações usando VMP e SVM não fizeram uso de transformação direta do espaço de características, conforme a Eq. (13). No entanto, para o classificador SVM foi usado um *kernel* gaussiano.

Todas as simulações, independentemente do classificador empregado, foram feitas usando o mesmo conjunto de treinamento e teste. Mais especificamente, aproximadamente 90% do total das amostras foram usadas para treinamento dos classificadores. O conjunto complementar restante foi usado para os testes de aferição de desempenho.

A seguir são apresentados os resultados das simulações realizadas, conforme as configurações experimentais supracitadas. Cada tabela contém uma descrição sumária destes resultados, tanto para a classe FRBS quanto para a classe *Default*.

A Tabela 3 e a Tabela 4 apresentam os resultados associados ao desempenho do classificador DLG, para discriminar os instrumentos musicais nas classes FRBS e *Default*, respectivamente. São mostradas as taxas médias de acerto conforme a combinação do valor da potenciação *k,* usado na transformação no espaço de características (ver Eq. (13)), com o tipo e o número de coeficientes da representação paramétrica (LPC ou LSF) do sinal no trecho de sustentação.

Tabela 3: Classe FRBS - DLG. As maiores taxas de acerto em cada coluna são indicadas em negrito. A maior taxa de acerto é indicada na célula com fundo cinza.

DLG	*k*= 1	*k*= 2	*k*= 3	*k*= 4
LSF-8	57,84%	73,87%	78,40%	**86,06%**
LPC-8	63,41%	70,38%	71,78%	78,05%
LSF-16	64,11%	**89,20%**	88,15%	75,26%
LPC-16	66,90%	68,64%	75,96%	80,49%
LSF-24	71,78%	88,85%	**91,99%**	74,22%
LPC-24	**74,91%**	76,66%	81,53%	80,49%

Tabela 4: Classe *Default* – DLG. As maiores taxas de acerto em cada coluna são indicadas em negrito. A maior taxa de acerto é indicada na célula com fundo cinza.

DLG	*k*= 1	*k*= 2	*k*= 3	*k*= 4
LSF-8	73,40%	80,85%	81,91%	82,98%
LPC-8	67,02%	72,70%	78,01%	77,30%
LSF-16	78,01%	85,11%	82,62%	77,30%
LPC-16	72,70%	80,50%	79,08%	80,14%
LSF-24	**80,50%**	**85,46%**	**85,82%**	**85,11%**
LPC-24	77,30%	83,33%	80,14%	83,33%

Os melhores resultados alcançados pelos classificadores DLG, SVM e VMP para discriminação na classe FRBS, em um sistema que usou a parametrização LSF com 8, 16 e 24 coeficientes são mostrados, respectivamente, na Tabela 5, na Tabela 6 e na Tabela 7. Note que $k = 3$ foi adotado para o classificador DLG.

Tabela 5: Classe FRBS – LSF com 8 coeficientes.

LSF-8	Flautas	Palhetas	Metais	Cordas	Taxa
DLG	78,57%	64,58%	92,96%	92,86%	86,06%
SVM	88,10%	85,42%	94,37%	92,86%	**91,29%**
VMP	85,71%	81,25%	88,73%	88,89%	87,11%

Tabela 6: Classe FRBS – LSF com 16 coeficientes.

LSF-16	Flautas	Palhetas	Metais	Cordas	Taxa
DLG	75,87%	79,17%	90,14%	96,03%	89,20%
SVM	78,57%	89,58%	85,92%	95,24%	**89,55%**
VMP	88,10%	77,08%	95,77%	88,89%	88,50%

Tabela 7: Classe FRBS – LSF com 24 coeficientes.

LSF-24	Flautas	Palhetas	Metais	Cordas	Taxa
DLG	88,10%	68,75%	98,59%	98,41%	**91,99%**
SVM	78,57%	79,17%	87,32%	96,83%	88,85%
VMP	85,71%	75,00%	98,59%	81,75%	85,37%

A Tabela 8 apresenta as taxas médias de acerto alcançadas pelos classificadores avaliados, em um cenário de teste no qual a parametrização LSF com 24 coeficientes foi adotada para discriminar os instrumentos presentes na classe *Default.*

Tabela 8: Classe *Default* – LSF com 24 coeficientes.

LSF-24	DLG	SVM	VMP
Flauta Contralto	100,00%	100,00%	77,78%
Flauta Baixo	90,00%	100,00%	70,00%
Flauta	100,00%	86,36%	81,82%
Clarinete em Si Bemol	46,15%	53,85%	30,77%
Clarinete em Mi Bemol	36,36%	45,45%	45,45%
Oboé	40,00%	70,00%	20,00%
Saxofone Contralto	94,74%	100,00%	100,00%
Trombone Baixo	69,23%	53,85%	69,23%
Trompa	77,78%	77,78%	44,44%
Saxofone Soprano	63,16%	84,21%	89,47%
Trombone Tenor	100,00%	100,00%	88,89%
Violoncelo	98,48%	96,97%	98,48%
Violino	95,00%	95,00%	83,33%
Fagote	91,67%	91,67%	75,00%
Taxa Global	85,82%	**87,59%**	79,43%

8 DISCUSSÃO

Como se pode observar pela Tabela 3 e pela Tabela 4, a utilização de parametrização por LSF para formar o vetor de características no sistema de classificação com DLG apresentou um desempenho superior melhor ao observado quando do uso de LPC. Isto corrobora os resultados obtidos por Krishna e Sreenivas [6].

Também pela Tabela 3 e pela Tabela 4 constata-se que uso da transformação não-linear sobre o espaço de características ($k > 1$) tende a favorecer o aumento das taxas de acerto. Os melhores resultados para ambas as classes ocorreram para a combinação $k = 3$ e LSF-24. Isso sugere ser desnecessário usar potenciação de grau maior que 3 em associação com 24 LSFs. Contudo, esse ponto merece melhor investigação.

Ainda pela Tabela 3 e pela Tabela 4 verifica-se que, tanto com LPC quanto com LSF, as taxas de acerto tendem a crescer com o número de coeficientes. Portanto, pode-se especular que taxas ainda maiores de acerto possam ser alcançadas com a utilização de parametrização de mais alta ordem.

Com base nos dados mostrados nas Tabelas 5, 6 e 7, a análise comparativa dos desempenhos dos classificadores para discriminar famílias de instrumentos (classe FRBS) revela que o uso de LSF-16 implica nas mais altas taxas médias de acerto para o SVM e o VMP. Já o melhor desempenho do DLG, que representa o melhor desempenho global, é alcançado quando se utiliza parametrização LSF-24. Nota-se que a diferença de desempenho entre o primeiro e o segundo (SVM com LSF-8) colocados é menor que 1%.

Ao se repetir a análise anterior para a discriminação de instrumentos (classe *Default*) observou-se que o uso de LSF-24 implica a obtenção dos melhores desempenhos de classificação para o DLG e o SVM. Como visto na Tabela 8, o SVM alcança a maior taxa média de acerto, sendo seguido pelo DLG com uma diferença de menos de 2%.

Ainda com referência à Tabela 8, verifica-se que o as taxas de acerto obtidas pelo DLG (com $k = 3$) mostraram-se iguais ou superiores às do VMP e do SMV, respectivamente, para 78% e 57% dos instrumentos. Logo, pode-se alegar que, para a base de dados e às condições de teste descritas, o desempenho do DLG com $k = 3$ é similar ao do SVM com *kernel* gaussiano.

Na tarefa de discriminação de família de instrumentos, a mais alta taxa de acerto alcançada pelo DLG foi de aproximadamente 92%. Entretanto, o desempenho cai para aproximadamente 86% quando o objetivo é identificar os instrumentos. Esses resultados ficaram bem próximos daqueles obtidos por Krishna e Sreenivas [6], em que as taxas de acerto são de 95% e 90%, respectivamente, para a identificação das mesmas família e classes de instrumentos individuais.

Uma vantagem do emprego de DLG em relação ao SVM é a menor complexidade computacional daquele, fator que se manifesta principalmente na maior rapidez para a obtenção da convergência no estágio de treinamento do classificador.

9 CONCLUSÕES

Este artigo abordou o problema de identificação automática de sons de instrumentos musicais a partir de sinais acústicos correspondentes a gravações anecóicas de notas isoladas. Em particular, o foco de interesse foi avaliar o desempenho de um sistema classificador que utiliza transformações não-lineares do espaço de entrada que alimenta um classificador do tipo Discriminante Linear Generalizado (DLG). Para fins comparativos, os resultados obtidos são confrontados com os alcançados por sistemas classificadores do tipo SVM e Vizinho mais Próximo.

Na tarefa de discriminação da família de instrumentos definida no trabalho, a mais alta taxa de acerto alcançada pelo DLG foi de aproximadamente 92%. Entretanto, o desempenho cai para aproximadamente 86% quando o objetivo é identificar os instrumentos individualmente.

Os resultados acima ficaram bem próximos daqueles obtidos por Krishna e Sreenivas [6], que obtêm com um classificador GMM taxas de acerto de 95% e 90%, respectivamente, para a identificação das mesmas classes de família e instrumentos individuais. Assim pode-se concluir que o classificador DLG apresentou um desempenho satisfatório, visto que, dentro das configurações adotadas neste estudo, seus resultados ficaram próximos ou melhores que os classificadores concorrentes avaliados.

Por fim, uma vantagem do emprego do DLG em relação ao SVM é a menor complexidade computacional daquele, fator que se manifesta principalmente na maior rapidez para a obtenção da convergência no estágio de treinamento do classificador.

Aspectos associados às outras etapas do sistema completo de classificação, como a operação ao longo do tempo, a formação das classes etc., estão sendo investigados em conjunto com a classificação propriamente dita.

10 AGRADECIMENTOS

Os trabalhos de pesquisa de Paulo Esquef e Luiz Biscainho são financiados pelo CNPq, respectivamente, através de bolsas de Pós-Doutorado Júnior (Processo 152042/2007-5) e de Produtividade em Pesquisa.

11 REFERÊNCIAS

[1] J. Marques and P. Moreno, *A Study of Musical Instrument Classification Using Gaussian Mixture Models and Support Vector Machines*, Cambridge Research Labs Technical Report Series CRL/4, 1999

[2] K. D. Martin, *Sound Source Recognition: A Theory and Computational Model*, PhD Thesis, Massachusetts Institute of Technology, Cambridge, MA, 1999.

[3] A. Eronen and A. Klapuri, "Music instrument recognition using cepstral coefficients and temporal features," in Proc. of ICASSP, pp. 753-756, 2000.

[4] G. Agostini, M. Longari and E. Pollastri, "Music instrument timbres classification with spectral features," *in* Proc. of ICME, pp. 97-102, 2001.

[5] T. Kitahara, M. Goto and H. G. Okuno, "Music instrument identification based on F0-dependent multivariate normal distribution," in Proc. of ICASSP, pp. 421-424, 2003.

[6] A. G. Krishna and T. V. Sreenivas, "Music instrument recognition: from isolated notes to solo phrases," *in* Proc. of ICASSP, pp. 265-268, 2004.

[7] J. C. Brown, O. Houix and S. McAdams, "Feature dependence in the automatic identification of musical woodwind instruments," J. Acoust. Soc. Am., Vol. 109, No. 3, pp. 1064-1072, 2001.

[8] A. Klapuri and M. Davy, *Signal Processing Methods for Music Transcription,* Springer, pp. 3-17, 2006

[9] Kim, H.-G., *Introduction to MPEG-7 Audio,* John Wiley & Sons, Inc., New York, 2005.

[10] *MIS – Musical Instruments Samples of IOWA University*.

[11] J. C. P. Filho, D. B. Haddad, L. P. Calôba, "Classificação de padrões de varredura de radares," in Anais do VIII Congresso de Redes Neurais, 2007.

[12] S. Haykin, *Neural Networks: a Comprehensive Foundation,* Prentice Hall, 2nd. Ed., 1999.

[13] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.

[14] Steve R. Gunn, *Support Vector Machines for Classification and Regression*, Technical Report - Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, Southampton University, 1998.