Sistema Especialista para Detecção do Timbre da Clarineta

Leonardo Carneiro de Araújo

[1]CEFALA – Centro de Estudos da Fala, Acústica, Linguagem e Música
UFMG – Universidade Federal de Minas Gerais

leoca@cefala.org

Abstract. *It is presented bellow a method for automatic identification of the clarinet. The schema described bellow is based on the analysis of the decomposed acoustic signal using wavelets functions. For each sample of an instrument it is automatically identified the attack, sustain and release regions. In each of the regions a decomposition process is carried out and finally it is got as a result the mean and variance of the wavelets coefficients. Those are the input data of the fuzzy or neural network used in the identification of the clarinet. The experiments show it is possible to acquire a low level of error rate.*

Resumo. *É apresentado um método para identificação automática da clarineta. O esquema proposto é baseado na analise da decomposição do sinal acústico através de funções wavelets. Para cada amostra de um instrumento são identificados automaticamente a região de 'attack', 'sustain' e 'release'. Para cada uma dessas regiões é feita a decomposição e é obtida a média e a variâncias dos coeficientes de wavelet. Esses serão os dados de entrada da rede fuzzy ou neural para identificação da clarineta. Os experimentos mostram que é possível obter uma boa taxa de acertos.*

1. Introdução

Timbre é uma característica sonora multidimensional e, por conseguinte, de difícil definição, em contraposição à altura e a intensidade percebida que podem ser facilmente definidas por serem características unidimensionais. O timbre é comumente definido como tudo o que caracteriza um som que não é a sua intensidade e sua altura. Esta inclusive é a própria definição adotada pela ANSI (American National Standards Institute) "... that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different". Sabemos também que o timbre é uma característica que não pode ser tomada isoladamente das demais. A altura, a duração e a intensidade influenciam o timbre sonoro percebido.

A tarefa de extrair características timbrísticas do sinal acústico não é trivial. Muitos estudos de pscicoacústica mostram que a parte estacionária do sinal acústico não é, muitas vezes, suficiente para prover toda a informação necessária ao reconhecimento de um instrumento. Existe então características importantes nos transientes do sinal (ataque e decaimento). Como não sabemos, precisamente, definir o que é timbre, é de se esperar que a tarefa de classificar ou segregar sons com relação as suas informações timbrísticas seja, por conseguinte, imprecisa. No entanto, isso não nos impede de formular estratégias e extrair características, que, embora possam não se assemelhar ao processo realizado pela mente humana para fazer esse tipo de segregação, podem ser tão eficientes quanto.

Sabemos que existem alguns fatores que são determinantes na constituição da informação timbrística. O conteúdo espectral e, também, a forma como este varia no

tempo é extremamente importante na determinação do timbre. Algumas possíveis abordagens seriam: determinar, através de uma a transformada de Fourrier com janela, a forma como o espectro do sinal evolui. Uma segunda possibilidade seria obter para uma janela deslizante sobre o sinal os coeficientes mel cepstrais que representam o sinal sob a janela. A forma de evolução desses coeficientes poderia, de alguma forma, conter a informação timbrística e assim ser utilizada na separação e identificação de padrões.

Utilizaremos aqui uma terceira opção a transformada wavelets. Veremos adianta algumas características dessa transformada (sessão 2.1). A forma como os coeficientes da decomposição em wavelet variam no tempo pode nos dar um bom parâmetro que capte a informação timbrística de um som musical.

Inicialmente utilizaremos uma forma bem simples de averiguar como variam os coeficientes da transformada wavelet do sinal. Tomaremos apenas as médias e as variâncias dos coeficientes wavelets nas primeiras escalas de decomposição. Escolhidos os parâmetros que utilizaremos, resta-nos decidir de que maneira faremos a classificação com base nesses parâmetros. As redes neurais artificiais podem conseguir bons resultados na classificação dos vetores de características timbrísticas, uma vez que são capazes de aprender tudo o que são capazes de representar.

1.1. Dados utilizados na pesquisa

O conjunto de dados utilizados nesta pesquisa consiste em amostras de notas dos seguintes instrumentos: canto, clarineta, flauta transversal, trombone, trompete e viola. Todas as amostras foram gravadas no estúdio da Escola de Música da UFMG. Como dispomos de uma quantidade muito superior de amostras de clarineta do que dos demais instrumentos, optamos por fazer um sistema especialista que distingue a clarineta dos demais instrumentos.

2. Abordagem do Problema

Uma das vantagens em se utilizar wavelets é que sua teoria já está bem consolidada, principalmente com relação a construção de bases e algoritmos eficientes para transformação, graças a trabalhos de Mallat, Daubechies, dentre outros. Uma outra característica peculiar das wavelets é a existência de diversas bases, sendo possível escolhe a base mais adequado para o trabalho em questão, embora tal escolha seja feita com base apenas experimental. Além dessas características citadas, a complexidade computacional das wavelets é linear com o número (N) de coeficientes computados $(O(N))$ enquanto outras transformações obtêm complexidade da ordem de $N \times log_2(N)$.

2.1. Decomposição wavelet do sinal de áudio

As funções wavelets formam uma base para um subespaço das funções quadrado integráveis e possuem ainda uma característica peculiar que será aqui importante. As wavelets fornecem uma análise de múltiplas resoluções através de um seccionamento do subespaço inicial em diversos subespaços encaixantes. Esses subespaços contém as diferenças de informações entre duas versões de subespaços em resoluções diferentes. Se os subespaços de baixa resolução são obtidos a partir de filtragens passa-baixas, os subespaços gerados pelas funções wavelets em cada nível são obtidos a partir de filtragens passa-banda. Além disso as funções wavelets possuem suporte compacto, dessa

forma cada função da base terá uma contribuição local para o sinal a ser decomposto. Ou seja, a contribuição das funções wavelets é local tanto em tempo, quanto em freqüência.

Existe então um algoritmo rápido que utiliza os filtros espelhados em quadratura para realizar a mudança de escala. Tipicamente um filtro passa-baixas e um filtro passa-banda (H e G respectivamente) são utilizados. A convolução com o filtro passa-baixas resulta em uma representação de baixa resolução do sinal A e a convolução com o filtro passa-banda resulta na versão de detalhes do sinal D.

A transformação de escalas é feita através do seguinte algoritmo rápido: $a_{j,n} = \sum_k \overline{h_k} a_{j-1,2n+k}$ e $d_{j,n} = \sum_k \overline{g_k} a_{j-1,2n+k}$ onde $a_{j,n}$ e $d_{j,n}$ são os coeficientes correspondentes a análise de multiresolução, representando o sinal na escala j, obtidos da filtragem da versão do sinal na escala $j-1$.

Optamos por fazer uma análise do sinal, efetuando apenas duas etapas de decomposição. Teremos assim, para cada sinal três vetores que o representam: o vetor da representação de baixa resolução, o vetor dos detalhes de nível 2 e o vetor dos detalhes de nível 1.

Cada amostra de áudio foi segmentada automaticamente para separar o sinal em regiões transientes (ataque e decaimento) e de estado estacionário (”sustain”). Cada uma dessas regiões foi tratada como uma sinal independente para o qual foi efetuada a análise de multiresolução, obtendo assim os três vetores acima citados.

Teremos assim um total de 9 vetores por amostra. Mas cada vetor possuí muitos coeficientes o que tornaria impraticável utilizá-los como entrada de uma rede fuzzy. Optamos então por apresentar para a rede apenas as médias e as variâncias de nossos coeficientes. Teremos então 18 entradas na rede, sendo 9 médias e 9 variâncias, 3 dessas para cada um dos segmentos da nota.

3. Resultados experimentais

As amostras amostras originais foram normalizadas para que todas tivessem a mesma amplitude eficaz. Após serem normalizadas elas foram automaticamente segmentadas. Após obtidos os segmentos das amostras de áudio, para cada um deles foi obtido a média e a variância dos coeficientes de wavelets na primeira e segunda escalas.

O conjunto de dados foi separada em conjunto de treinamento, conjunto de validação e conjunto de teste. Este conjunto foi montado de forma a ter o mesmo número de padrões para a clarineta e os demais instrumentos.

Com a rede fuzzy utilizada (que possuí 18 entradas e 18 funções de pertinência), obtivemos alguns resultados muitos bons que são ilustrados nas figuras 3. Nesses resultados a taxa de acerto chegou a atingir 100%. Como limiar de separação adotamos o valor 0.5, de forma que qualquer resultado obtido abaixo de 0.5 é considerado zero (acerto, para o caso ilustrado na figura) e qualquer resultado acima de 0.5 é considerado um (erro, para o exemplo abaixo).

Uma rede MLP com 3 camadas intermediárias foi treinada utilizando a técnica ‘early stoping’. Para a melhor rede obtida, cujo resultado de treinamento consta na figura 4, obtivemos 100% de acerto nos testes utilizando o conjunto de teste. Mas a grande

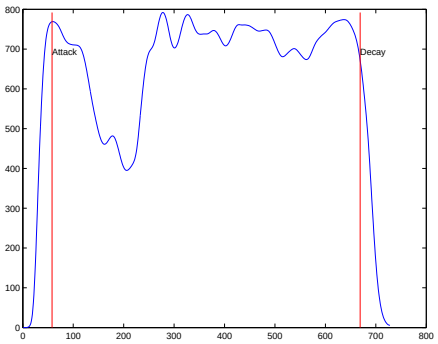

Figura 1

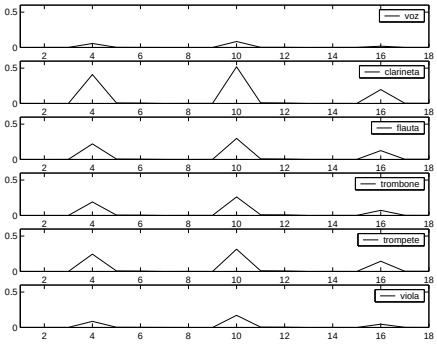

Figura 2

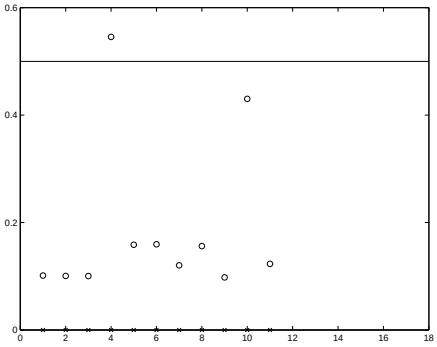

Figura 3

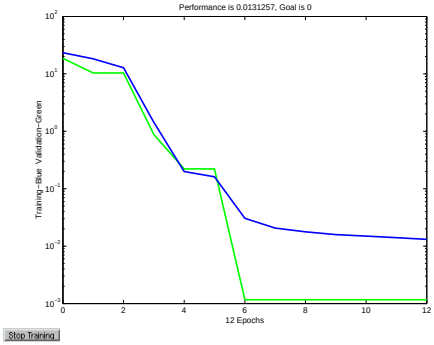

Figura 4

maioria das outras redes obtidas possuíam percentagem de acerta de aproximadamente 90%.

4. Conclusão

Os parâmetros extraídos das amostras de som (notas) aparentemente mostraram-se eficientes para representar a característica timbrística de instrumentos, uma vez que com base nesses parâmetros apenas foram obtidas altas taxas de acertos utilizando quer uma rede neural MLP ou uma rede fuzzy. Um próximo trabalho será coletar mais amostras dos demais instrumentos e tentar fazer um classificador capaz de reconhecer os diferentes instrumentos.

Referências

Garcia, C., Zikos, G., and Tziritas, G. (2000). Wavelet packet analysis for face recognition. *Image Vision Computing*, 18:289–297.

Haykin, S. (1999). *Neural Networks*. Prentice Hall, 2nd edition.

Jang, J.-S. R., Sun, C.-T., and Mizutani, E. (1996). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Pearson Education, 1st edition.

Mallat, S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press, 2nd edition.