Modelagem da variação do timbre musical utilizando modelos auditivos e mapas de Kohonen

Maurício A. Loureiro[1], Hugo B. de Paula[2], Tairone N. Magalhães[1]
[1]CEFALA - Centro de Estudos da Fala, Acústica, Linguagem e Música, UFMG
[2]PUC Minas – Pontifícia Universidade Católica de Minas gerais
mauriciloureiro@ufmg.br, hugo@pucminas.br, tairone@cefala.org

1. Introdução

Timbre Musical

O timbre é percebido a partir da interação de propriedades estáticas e dinâmicas do som, agregando não apenas um conjunto complexo de atributos auditivos, mas também uma gama de aspectos psicológicos e musicais. Em 1964, o timbre foi definido pela ASA (*American Standard Association*) como "aquele atributo do sentido auditivo em termos do qual o ouvinte pode julgar que dois sons similarmente apresentados e tendo a mesma intensidade e altura, são dissimilares". Metodologias de testes subjetivos de similaridade emprestados da psicologia experimental permitiram uma redução da dimensionalidade deste atributo em representações de menor complexidade, com a introdução da noção de "taxa de similaridade" entre estímulos auditivos, o que possibilitou mapear geometricamente o conceito de similaridade acústica, fornecendo uma quantificação psicológica de uma estrutura relativamente complexa a partir de dados bem simples. Estudos mais recentes estabeleceram correlações entre fatores perceptivos relacionados ao timbre e grandezas físicas mensuráveis diretamente do som, possibilitando outras abordagens na investigação de timbre baseadas em análise acústica (Misdariis, Smith *et al.*, 1998).

Os instrumentos acústicos tradicionais permitem a produção e controle de uma vasta gama de timbres, que dependem da altura da nota, da maneira como é tocada e da habilidade do músico em seu controle dinâmico. Entretanto, não há uma escala que o permita ser especificado quantitativamente pelo sistema tradicional de notação musical assim como

volume e altura, descritos em escalas fraco-forte e de gamas alturas. Investigações mais extensivas sobre a variação intencional do timbre e sobre a capacidade do ouvinte em perceber e compreender esta intenção tornam-se cada vez mais necessárias para a compreensão da contribuição de parâmetros acústicos , entre eles a diferenciação de timbre para a condução e percepção da expressividade de uma performance musical.

Modelos Auditivos

Pesquisas em neuropsicologia e neurofisiologia da audição humana têm possibilitado a construção de modelos que buscam representar o comportamento dos componentes responsáveis pela audição, tais como ouvido externo e médio, membrana basilar, células ciliares do ouvido interno e fibras nervosas do oitavo nervo craniano do sistema nervoso. Modelos auditivos ou cocleares vêm oferecendo uma alternativa promissora para pesquisa em timbre musical. Estes modelos transformam o sinal em padrões de disparos neuronais, conhecidos como neurogramas, resultando numa representação que não traduz necessariamente a distribuição de energia do sinal ao longo de seu espectro de freqüência. Estudos comprovaram uma maior eficiência destes modelos frente aos tradicionais, em casos em que o sinal se apresenta altamente corrompido por ruído. Esta propriedade os torna adequados para a pesquisa em timbre, por implementarem uma característica complexa da percepção auditiva, que é a capacidade de discriminação de timbres mesmo com a presença de altos níveis de ruído. Estudos mais recentes comprovam a robustez destes modelos na investigação dos mecanismos perceptivos do timbre musical (Cosi, De Poli *et al.*, 1994; Toiviainen, Kaipainen *et al.*, 1995; Toiviainen, 1996; De Poli e Prandoni, 1997).

Mapas Auto-Organizativos de Kohonen

Mapas Auto-Organizativos de Kohonen, conhecidos como SOM (Self-Organizing Maps), são algoritmos de redes neurais não supervisionados, capazes de mapear dados de entrada de grandes dimensões em espaços de baixa dimensão, preservando as relações

topológicas essenciais dos dados originais (Kohonen, 1995; Braga, Ludermir *et al.*, 2000). Por não se fundamentar em suposições a priori sobre as características dos dados analisados, SOM se mostra como uma ferramenta poderosa para experimentos que envolvem análise de dados de grande complexidade e com altos índices de não linearidade, tais como sons musicais.

Leman propôs uma comparação entre os resultados obtidos por mapeamentos timbrísticos resultantes de redes SOM com aqueles espaços construídos por MDS a partir de medições psicológicas, com o objetivo de estabelecer uma plataforma que incorporasse fenômenos auditivos e neuronais para a pesquisa musicológica (Leman, 1991; Leman, 1994). Toiviainen comparou a eficiência das representações do timbre musical em espaços construídos por distâncias topológicas computacionalmente calculadas por SOM e por medições de estimação subjetiva de similaridade (Toiviainen, Kaipainen *et al.*, 1995). Os resultados deste trabalho comprovaram um alto grau de correlação entre os dois domínios, sugerindo uma adequação do modelo de Kohonen para representar complexos perceptivos multidimensionais A equipe de pesquisa do CSC da Universidade de Pádua desenvolveu estudos de classificação de timbres musicais utilizando SOM (De Poli, Prandoni *et al.*, 1993; De Poli e Tonella, 1993).

2. Objetivos

Este trabalho buscou explorar as possibilidades de utilização de modelos auditivos computacionais para investigar o significado dos parâmetros físicos determinantes nas variações de timbre que ocorrem na execução de uma nota musical. A rica variação timbrística da clarineta foi utilizada para a construção de mapas timbrísticos, nos quais similaridades entre sonoridades pudessem ser representadas por distâncias geométricas em espaços de baixa dimensionalidade.

3. Métodos

Dados de Análise

Embora o timbre possa variar independentemente da intensidade ou da duração, o alto grau de correlação entre timbre e intensidade facilita a amostragem de "valores" distintos de timbre para uma mesma nota a partir da especificação da intensidade. Assim, um músico foi instruído a executar cada nota em quatro níveis de intensidade distintos: *pianissimo*, *mezzo-piano*, *mezzo-forte* e *fortíssimo*, com o mínimo de variação possível. As amostras foram obtidas a partir de gravações de todas as notas dos dois registros mais graves da Clarineta em Si bemol, variando de Ré 3 (147 Hz) a Lá 5 (880 Hz), executadas nos quatro níveis de intensidade acima, com uma duração média de 3 segundos. Utilizamos também neste estudo amostras de grande conteúdo expressivo e larga variação de dinâmica e timbre, em diferentes registros do instrumento, extraídas da Abertura do 1º movimento do *Quintetto op. 115* em Si menor para clarineta e quarteto de cordas de Brahms (compassos 5 a 17). As gravações foram reamostradas a 22,05 kHz, e suas amplitudes foram normalizadas, para que fossem classificados em função de sua variação dinâmica.

Modelos Auditivo

Estas amostras foram então representadas em cocleagramas, extraídos com a utilização do modelo auditivo desenvolvido por Richard Lyon (1988) implementado em Matlab por Malcolm Slaney (1998). Neste modelo, um filtro linear simples simula a filtragem que ocorre no canal auditivo. Na modelagem da cóclea, são combinados vários canais de bancos de filtros, que modelam a propagação das ondas de pressão, com ressonadores, as convertem em movimento da membrana basilar. O movimento da membrana basilar é detectado pelas células ciliares internas, que capturam apenas a fase positiva do sinal. Estas são simuladas através de Retificadores de Meia Onda (Half Wave Rectifiers ou HWRs), que fazem a representação neural do sinal. Controles de Ganho Automático modelam efeitos

como mascaramento, diferenças de tempo de adaptação no ouvido interno e não-linearidades da cóclea (Lyon e Mead, 1988). A saída gerada por esse modelo é uma matriz com a probabilidade de ocorrência de disparos ao longo do nervo auditivo, denominada *cocleagrama*. A fim de reduzir a quantidade de dados, os dados de entrada foram decimados a cada 500 amostras.

SOM Toolbox

Os cocleagramas foram então classificados por um mapa auto-organizável de Kohonen bidimensional com topologia hexagonal. Os arranjos destes mapas neurais podem tomar várias formas, definindo as relações de vizinhança entre os neurônios. Este estudo utilizou uma implementação do SOM em Matlab desenvolvida por Juha Vesanto e equipe, da Helsinki Unversity of Technology, o SOM Toolbox (Versanto, Himberg *et al.*, 2000).

4. Resultados alcançados e discussão

Inicialmente, um mapa SOM hexagonal de tamanho 16x10 foi usado para mapear as notas mais graves do instrumento. O mapa da esquerda da Figura 1 mostra os sons *ff* e *pp* destas 7 notas. Observa-se que o SOM foi capaz de discriminar os pontos pertencentes a cada som, agrupando-os em poucas células contíguas. Percebe-se também que tanto as alturas quanto as intensidades foram adequadamente discriminadas pelo SOM e que notas *pp* foram agrupadas mais compactamente que notas *ff*. Notas *pp* tiveram a maioria de seus pontos concentrados em uma célula. Isto pode ser melhor verificado no mapa do centro que indica em cada célula a porcentagem da extensão total de cada som nela mapeada, proporcional ao tamanho do ponto. Esta diferenciação de espalhamento no mapeamento das intensidades também pode ser verificada pela distribuição de distância métrica do mapa, mostrada no gráfico da direita da Figura 1, no qual distâncias entre hexágonos representam distâncias entre neurônios do mapa. Este resultado é consistente com estudos anteriores, nos quais espaços

espectrais construídos a partir de Analise de Componente Principais (PCA) da distribuição espectral foram utilizados para representar o timbre (Loureiro, de Paula *et al.*, 2004a; Loureiro, de Paula *et al.*, 2004b).

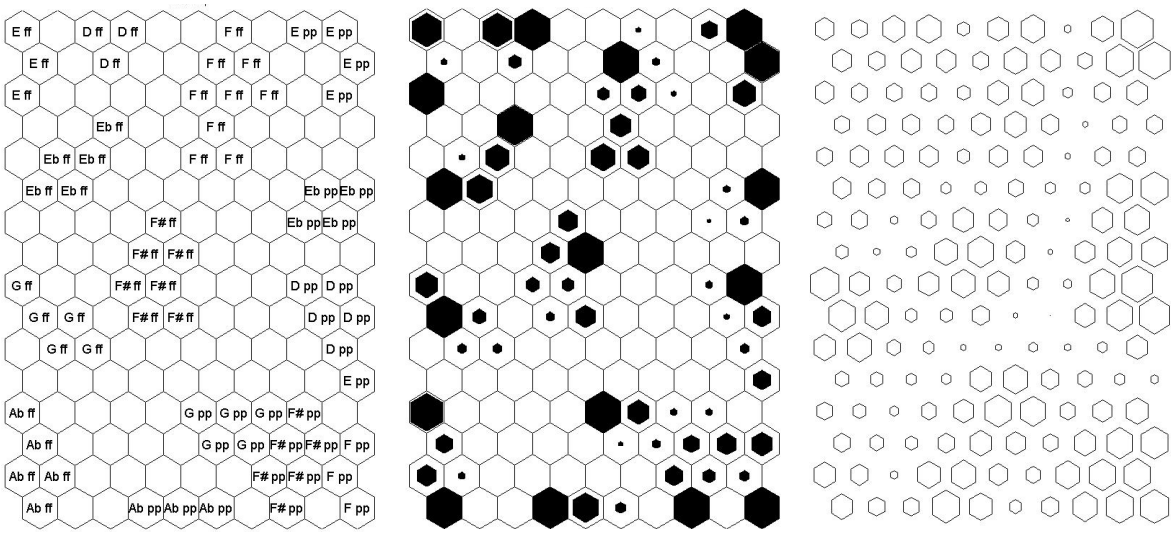

Figura 1: Mapeamento das amostras em neurônios do SOM (à esquerda); porcentagem de amostras presentes em cada neurônio (ao meio); matriz de distâncias entre os neurônios, onde quanto maior a célula, menor a distância entre suas vizinhas (à direita).

No entanto, a influencia da altura da nota neste tipo de representação a torna inadequada para as finalidades deste estudo, que buscou caracterizar as variações sutis do timbre, tais como aquelas que ocorrem ao longo de uma mesma nota durante uma performance musical. Vários sons de uma mesma altura foram então classificados em mapas SOM individuais para cada altura. O mapa da Figura 2 mostra 4 sons da nota Fá# 3, amostradas fora do contexto de uma performance musical, nas 4 intensidades, *pp*, *mp*, *mf*, *ff*. Nota-se que houve uma boa discriminação das intensidades, cada uma ocupando um dos 4 cantos do mapa.

Com a finalidade de caracterizar a variação de timbre durante uma performance musical, adicionou-se ao mapa anterior 3 notas de mesma altura, extraídas da Abertura do 1º movimento do *Quintetto op. 115* em Si menor para clarineta e quarteto de cordas de Brahms, dos compassos 15, 16 e 17. Uma comparação auditiva entre estas 3 notas revela que a nota do

compasso 15 é a que exibe maior variação timbrística, enquanto que a nota do compasso 17 corresponde a um final de frase em pianíssimo (*ppp*), com pouca variação de timbre. Podemos observar a mesma tendência de mapear os níveis de intensidade ao longo do eixo vertical e de distribuir sons de maior intensidade ao longo de regiões mais extensas do mapa. Além disso, é interessante notar que os sons musicais foram mapeados no lado esquerdo do mapa, enquanto que os sons não musicais foram deslocados para o lado direito. Isto pode estar relacionado com o fato de que os sons musicais foram tocados por um músico distinto em um instrumento também distinto, mesmo que gravados em condições idênticas.

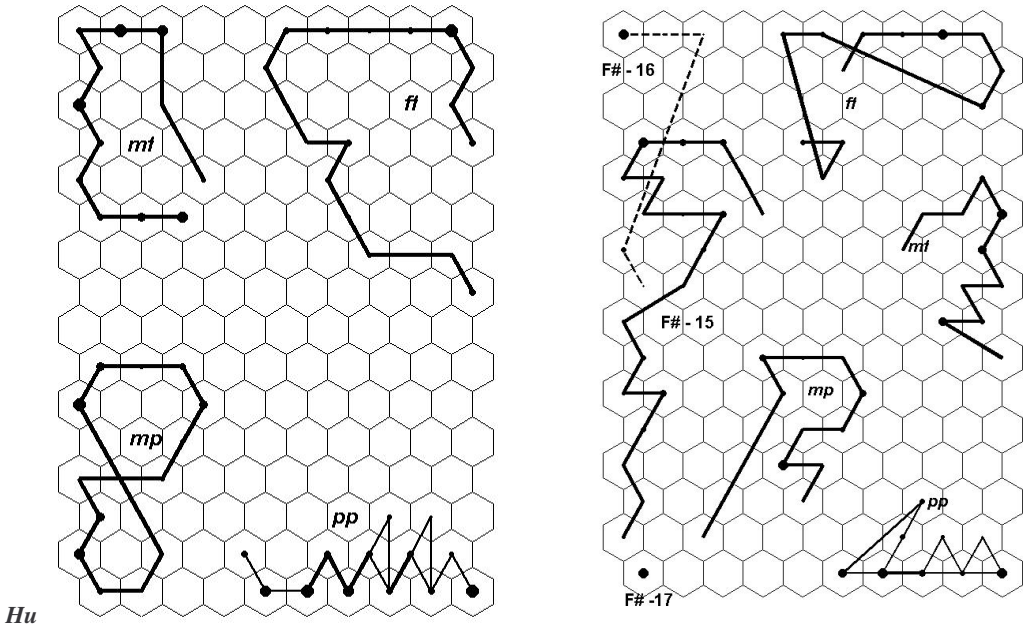

Figura 2. Trajetórias de amostras do F# na oitava 3 em 4 níveis de intensidade (à esquerda); sons não musicais acrescidos de três amostras extraídas do Quintetto de Brahms (à direita; tocadas por instrumentista e clarineta distintos.

Conclusão

Esta representação gráfica do timbre e de sua evolução temporal oferece caminhos para a compreensão do controle dinâmico e instantâneo que o instrumentista detém sobre o som dos instrumentos acústicos e de como este controle é percebido pelo ouvinte. Entretanto, a interação entre a não linearidade dos modelos auditivos e a topologia não uniforme dos mapas auto-organizáveis de Kohonen ainda precisam ser melhor entendida para que esta representação reflita mecanismos de produção intencional e percepção do timbre.

6. Referências

Braga, A. P., T. B. Ludermir e A. C. P. L. F. Carvalho. (2000). *Redes Neurais Artificiais: Teoria e Aplicações*. Rio de Janeiro, RJ: Editora LTC.

Cosi, P., G. De Poli e G. Lauzzana. (1994). Auditory Modeling and Self-Organizing Neural Networks for Timbre Classification. *Journal of New Music Research*, v.23, p.71-98.

De Poli, G. e P. Prandoni. (1997). Sonological Models for Timbre Characterization. *Journal of New Music Research*, v.26, p.170-197.

De Poli, G., P. Prandoni e P. Tonella. (1993). Timbre Clustering by Seff-Organizing Neural Networks. *Proceedings of the X Colloquium on Musical Informatics*, Milan: University of Milan.

De Poli, G. e P. Tonella. (1993). Self-Organizing Neural Networks and Grey's Timbre Space. *Proceedings of the International Computer Music Conference*, Tokyo: International Computer Music Association. p.441-444.

Kohonen, T. (1995). *Self-Organizing Maps - Springer Series in Information Sciences*. Berlim: Springer Verlag. v.30.

Leman, M. (1991). Emergent Properties of Tonality Functions by Self-Organization. *Interface - Journal of New Music Research*, v.19, n.2-3, p.85-106.

_____. (1994). Schema-Based Tone Center Recognition of Musical Signals. *Journal of New Music Research*, v.23, n.2, p.169-204.

Loureiro, M. A., H. B. de Paula e H. C. Yehia. (2004a). Representation and Classification of the Timbre Space of a Single Musical Instrument. *International Speech Communication Association - Tutorial and Research Workshop on Statistical and Perceptual Audio Processing (SAPA 2004)*, Jeju, Korea, 2004: International Conference Center. p.546-549.

_____. (2004b). Timbre Classification of a Single Music Instrument. *5th International Conference on Music Information Retrieval*, Barcelona: University Pompeu Fabra. p.546-549.

Lyon, R. F. e C. Mead. (1988). An analog electronic cochlea. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v.36, n.7, p.1119–1134.

Misdariis, N. R., B. K. Smith, D. Pressnitzer, P. Susini e S. McAdams. (1998). Validation of a Multidimensional Distance Model for Perceptual Dissimilarities Among Musical Timbres. *Proceedings of the 16th International Congress on Acoustics*, Woodbury, New York: ASA - The Acoustical Society of America.

Slaney, M. (1998). *Toolbox for Auditory Modeling Work*. Apple Computer, Inc, 2nd edition. http://rvl4.ecn.purdue.edu/%7Emalcolm/interval/1998-010/.

Toiviainen, P. (1996). Optimizing Auditory Images and Distance Metrics for Self-Organizing Timbre Maps. *Journal of New Music Research*, v.25, p.1-30.

Toiviainen, P., M. Kaipainen e J. Louhivuori. (1995). Musical Timbre: Similarity Ratings Correlate with Computational Feature Space Distances. *Journal of New Music Research*, v.24, p.282-298.

Versanto, J., J. Himberg, E. Alhoniemi e J. Parhankangas. (2000). *Self-Organizing Map in Matlab: the SOM Toolbox*. Helsinki University of Technology. http://www.cis.hut.fi/projects/somtoolbox/.