music. Although these persons could not see each other, they could hear when the other had required a change – this led to a situation where the both started to „play" with each other, resulting in a wonderful and energetic performance.

References

Essl, K. (**1989**). Zufall und Notwendigkeit. Anmerkungen zu Gottfried Michael Koenigs Streichquartett 1959 vor dem Hintergrund seiner kompositionstheoretischen Überlegungen. *Musik-Konzepte*, 66 „Gottfried Michael Koenig". Munich: edition text + kritik.

Essl, K. (**1992**). Kompositorische Konsequenzen des Radikalen Konstruktivismus. *Positionen. Beiträge zur neuen Musik*, 11.

Henck, H. (**1980**). *Karlheinz Stockhausens Klavierstück X. Ein Beitrag zum Verständnis der seriellen Kompositionstechnik.* Cologne: Neuland Musikverlag, 19–23.

Koenig, G.M. (**1965**). Serielle und aleatorische Verfahren in der elektronischen Musik. *Die Sonde*, 5 (1).

Nelson, T. (**1970**). Literary Machines. Swarthmore/PA: self-published

Okopenko, A. (**1970**). *Lexikon-Roman einer sentimentalen Reise zum Exporteurtreffen in Druden.* Salzburg: Residenz Verlag

Koenig, G.M. (**1970**). Project 1. *Electronic Music Reports*, 2. Utrecht: Instituut voor Sonology.

Stockhausen, K. (**1957**). ...wie die Zeit vergeht.... *die Reihe. Informationen über serielle Musik*, 3. Vienna: Universal Edition

Acknowledges

I want to thank Bob Willey (University of California San Diego), Dennis Patrick (University of Toronto, Faculty of Music), and Gerhard Eckel (IRCAM, Paris) for their discussions which helped me to formulate this article.

Appendix

The software described in the paper runs on an Apple Macintosh computer and requires Max 2.5 (© by IRCAM /Opcode) or later. It is in the public domain and available via ftp or WWW.

Real Time Composition Library for Max 2.5

Currently a version 2.0 of the„*Real Time Composition Library*" for MAX 2.5 is available from the following ftp-sites:

(**a**) ftp.ircam.fr, /pub/IRCAM/programs/max/patches/composition/RTClib2.0.sea.hqx

(**b**) kahless.isca.uiowa.edu, /ftp/pub/max/RTC-lib_2.0.sea.hqx

(**c**) ftp.mars.let.uva.nl, /pub/software/RTC-lib_2.0.sea.hqx

Lexikon-Sonate is available as:

(**1**) MAX program: A diminished version specially designed for the „Yamaha DisKlavier" can be obtained from the „Disklavier Archive" which is maintained by Bob Willey (http://crca-www.ucsd.edu/bobw/disklavier.html). It can be retrieved via anonymous ftp from: wendy.ucsd.edu, /pub/midi/disklavier/essl/LexiconSonate.sit.Hqx

(**2**) MIDI file: 5 different MIDI-files generated by *Lexikon-Sonate* can be obtained via anonymous ftp from: kahless.isca.uiowa.edu, /pub/max/lexicon/

(**3**) Disklavier disk: A recording of *Lexikon-Sonate* as a Disklavier disk can be found at: http://crca-www.ucsd.edu/bobw/disk3.html

(**4**) Audio on CD: An excerpt of the premiere of *Lexikon-Sonate* (featuring the „Bösendorfer SE Grand Piano") was released on the CD „Karlheinz Essl: Rudiments" (1995). It can be ordered from my publisher: TONOS Musikverlags GmbH, Ahastr. 9, D-64285 Darmstadt / Germany / Europe, Tel: +49–6151–31 23 47,Fax: +49–6151–31 32 78.

Reconhecimento de timbres musicais através da rede neural auto-organizável de Kohonen

MARCELO JARA PÉREZ[1], JOSE EDUARDO FORNARI[1], FURIO DAMIANI[1,2]

[1]Departamento de Semicondutores, Instrumentos e Fotônica, Faculdade de Engenharia Elétrica-UNICAMP
[2]Núcleo Interdisciplinar de Comunicação Sonora-UNICAMP

DSIF/FEE/Unicamp, 13081-970, Campinas, SP, Brasil.
e-mail: Jara@dsif.fee.unicamp.br

Resumo

Foi realizada a simulação de uma rede neural para a discriminação das diferenças timbrísticas de tons musicais. O método consiste em treinar uma rede neural auto-organizável de Kohonen com uma sequência de 17 amostras de instrumentos orquestrais. No final da fase de treinamento formam-se mapas auto-organizados onde ocorrem agrupamentos das amostras por família instrumental. Verificou-se a capacidade de reconhecimento utilizando-se todas as amostras. O sucesso do reconhecimento e a classificação do timbre dos instrumentos está diretamente relacionada à geração de mapas cuja qualidade é fortemente dependente das propriedades de convergência e da estabilidade do modelo. Esta rede neural é adequada para o reconhecimento de padrões timbrísticos com pequena taxa de erro.

Introdução

Diversos trabalhos nas áreas de psico-linguística, acústica fisiológica e psico-acústica tem abordado a discriminação de timbres (Plomp,1976; Grey & Moorer, 1977; Singh,1987). A sua percepção pelo sistema auditivo humano é um fenómeno complexo, que envolve grande capacidade de processamento para ser analisado e classificado no cérebro, de acordo com regras não sempre bem compreendidas. O reconhecimento do timbre musical depende de uma série de condições, tais como o contexto em que o sinal é percebido (Grey, 1978), sua complexidade, a amplitude e a forma como os harmônicos estão distribuídos no espectro de frequência. A forma do ataque do sinal e a variação do espectro de energia nos instantes iniciais são fundamentais na percepção (Gordon, 1987). Neste trabalho avaliamos a capacidade da rede neural de Kohonen (1982) de reconhecer e classificar timbres sonoros de instrumentos musicais, tocados isoladamente. O modelo de Kohonen foi utilizado com sucesso no reconhecimento de fonemas na língua finlandesa e na geração automática destes fonemas num computador, em tempo real (Kohonen, 1987).

A rede neural recorrente simples (SRNN) foi já utilizada para o reconhecimento de tons dos fonemas da língua Mandarin (Wang & Chen, 1994), reconhecendo variantes de tons de uma mesma estrutura fonética. Existem poucas pesquisas disponíveis na literatura sobre o reconhecimento de características timbrais usando redes neurais. A características de auto-organização e classificação de sinais sensoriais dos Mapas de Kohonen (1990), determinaram a escolha deste modelo, pois possibilitam o treinamento da rede sem supervisão. A seqüência de padrões de treinamento é aleatoriamente apresentada à rede. As respostas aos padrões são automaticamente mapeadas pelos neurônios. A fase de reconhecimento é feita após a elaboração auto-organizada dos mapas.

Na próxima secção descrevem-se suscintamente estudos sobre a discriminação e percepção do timbre. Na secção 2 especifica-se a arquitetura da rede neural e a forma como foi empregada no reconhecimento das amostras. Na secção 3 descreve-se a metodologia empregada nas simulações, o pré-processamento dos sinais e a forma como

eles foram utilizados no treinamento da rede. Na secção 4 apresentam-se os resultados. Descreve-se também o processo utilizado no reconhecimento das amostras. As conclusões e sugestões para futura pesquisa na área são apresentadas na ultima secção.

1. Noções sobre o timbre sonoro e o seu reconhecimento natural

Nos estudos sobre a percepção e o reconhecimento do timbre sonoro tem sido abordados principalmente os aspectos da **altura** (*pitch*) e **sonoridade** (*loudness*). Plomp (1978) publicou o resultado dos seus estudos, baseados em diversas experiências e simulações computacionais que lhe permitiram gerar tons complexos com espectro de fase variável e estudar o seu efeito sobre a percepção do timbre. Propôs uma técnica chamada de *escalamento multidimensional*, que permite descrever de forma mais eficiente a natureza do timbre. Um sinal sonoro é descrito por :

$$p(t) = \Sigma^{m}_{n=1} \{\alpha_n \sin(2\pi nft + \phi_n)\} \qquad \text{(e.1.1)}$$

O timbre é, então, determinado pelo espectro das amplitudes $\alpha_1\alpha_2,...\alpha_m$ e o espectro de fase dado pelos componentes ϕ_1, ϕ_2, ..., ϕ_n dos harmônicos sucessivos. O propósito da técnica de escalamento multidimensional é obter informação de alguma estrutura ou padrão de dados e poder representa-la de forma que possa ser facilmente visualizada. No caso do timbre, esta informação é obtida a partir da derivação de uma matriz de dados, cujas entradas representam a dissemelhança em timbre de um par de estímulos sonoros. Utiliza-se um espaço, em que as distancias entre pontos representam índices de dissemelhança. Plomp chegou à conclusão que, embora os componentes da fase de um tom influenciam de algum modo no reconhecimento do timbre, a fase não exerce quantitativamente tanta influência quanto à distribuição dos harmônicos no espectro e as suas amplitudes. Nas experiências, ouvintes comparavam as dissemelhanças entre uma série de estímulos sonoros de mesma sonoridade e altura. Não se constatou uma correlação significativa entre as diversas formas de onda de fase aleatória e a sua dissemelhança no timbre, concluindo que o efeito da fase na percepção natural de diferenças timbrísticas é relativamente pequeno.

Plomp procurou uma relação entre o timbre e o espectro de frequência. Foram utilizados 17 instrumentos diferentes que tocaram a mesma nota, com frequência de 349 Hz, para um auditorio treinado musicalmente. O espectro foi analizado por um conjunto de filtros de larguras de faixa de 1/3 de oitava, obtendo-se assim uma representação 15-dimensional. A dissemelhança entre dois timbres foi relacionada à diferença de magnitude entre seus harmônicos.

Outros aspectos, como a relação entre o timbre e a altura, efeitos espúrios como a reverberação e a interação entre harmônicos vizinhos influem no reconhecimento do timbre. Plomp concluiu que, na percepção de dissemelhanças de timbres sonoros, a fase tem alguma influência nas baixas frequências. O efeito da fase no timbre depende da relação entre as componentes seno e coseno, diminuindo à medida em que aumenta a frequência fundamental; julga-se que as diferenças na envoltória das formas de onda sejam a principal causa da sensibilidade à fase. O timbre depende fortemente da envoltória do espectro, mais do que da frequência fundamental.

Outros trabalhos têm pesquisado a influência do contexto musical na discriminação do timbre. Grey (Grey, 1978) demonstrou que esses efeitos variam dependendo dos instrumentos utilizados. Fez estudos comparativos utilizando ouvintes que julgavam se o timbre variava ou permanecia inalterado durante uma sessão. Comparou-se a discriminação do timbre de tons isolados em relação a diversos contextos musicais, utilizando até 12 contextos diferentes: 4 tons isolados, 4 padrões musicais monofônicos e 4 padrões musicais polifônicos. Grey fez experiências com três instrumentos distintos: clarinete, trompete e fagote. Os seus resultados mostraram que os padrões musicais parecem destacar as diferenças espectrais existentes entre as versões de um timbre, enquanto os tons isolados parecem permitir uma melhor comparação dos detalhes temporais. A discriminação é prejudicada quando são utilizados padrões polifônicos, Grey especula que a causa disto pode ser devida às complicadas formas de interação na distribuição do espectro da energia acústica e/ou a efeitos de mascaramento em algumas vozes por outras mais fortes. Os resultados do seu estudo variam, de acordo com o instrumentos musical utilizado.

Um aspecto relevante na percepção do timbre é a forma como a energia acústica varia no tempo e a variação associada do espectro. Grey e Gordon(1987) mostraram que modificações espectrais afetam a percepção do timbre. Em gravações digitais de 16 instrumentos musicais, fizeram diversas transformações, alterando as características da envoltória do espectro original. Geraram novos timbres usando síntese aditiva digital, adicionando à gravação original um conjunto de harmônicos modulados no tempo amplitude e frequência. Das16 amostras, 8 foram modificados em 4 pares, de forma que a envoltória do espectro de uma, amostra modificada fosse correspondente

à de seu par associado. Essas modificações da envoltória demonstraram ter forte influência na percepção do timbre sonoro. Outro trabalho (Gordon, 1987) mediu e modelou o ataque de instrumentos musicais, definindo o PAT (*Perceptual Attack Time*) para medir as características e a precisão rítmicas de uma execução musical.

2. O modelo de rede neural auto-organizável de Kohonen

O modelo e o algoritmo propostos originalmente por Kohonen (1982), conhecido como Mapa Auto-Organizado (*Self-Organizing Feature Map*), tem a propriedade de gerar representações internas das características dos sinais que são apresentados à sua entrada durante a fase de aprendizado. Os mapas têm sido utilizados como filtros adaptativos no reconhecimento de voz e imagens, no: controle de processos, robótica e representação de estruturas complexas de dados. Uma de suas características é preservar internamente a ordem topológica dos sinais de entrada. O aprendizado da rede é não-supervisionado, mas Kohonen (1990) apresentou variantes de sua proposta

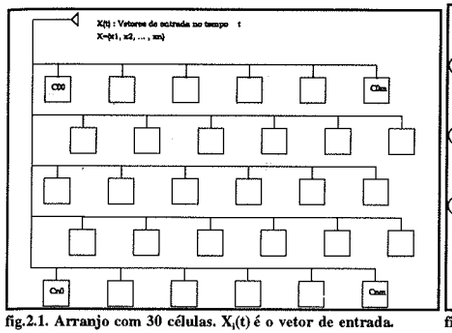

fig.2.1. Arranjo com 30 células. $X_i(t)$ é o vetor de entrada.

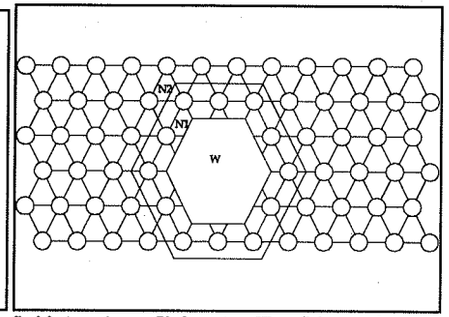

fig.2.2 Arranjo com 72 elementos.; W - célula vencedora.

original, inserindo um processo supervisionado de ajuste fino do mapa, o algoritmo LVQ (*Learning Vector Quantization*).

Na fase de aprendizado, as células da rede são automatica e gradualmente *sintonizadas* aos sinais de entrada apresentados seqüencialmente, formando internamente aglomerados de padrões de classes semelhantes. Esta sintonia é quantificada pela mudança de valor dos pesos de cada célula. Na fig.2.1, vemos que cada célula C_{ij}, possui um peso, representado por um vetor n-dimensional $\mathbf{W}=\{w_1,w_2,...,w_n\}$, cujos componentes variam à medida que uma sequência de sinais, representada por vetores $\mathbf{X}(t)=\{x_1,x_2,...,x_n\}$, é apresentada na sua entrada. Cada vetor de entrada é apresentado simultaneamente à toda a rede. A cada iteração (apresentação de um vetor), são modificados os pesos da célula vencedora e os das células pertencentes a sua vizinhança, de acordo com o algoritmo descrito a seguir. Na fig. 2.2, vemos a forma como as células componentes da rede são interconectadas, formando uma vizinhança de tipo hexagonal, onde cada uma das células (exceto as das bordas) possui seis vizinhos. As seis células em torno da vencedora formam a vizinhança de raio 1, N_1, as outras 12 pertencem a vizinhança de raio 2, as 18 células formam a vizinhanca total N_2; W, a vencedora, é a célula central da vizinhança hexagonal. Uma vizinhança pode ter forma retangular.

2.1 Algoritmo de operação.

Um vetor de entrada $\mathbf{X}=\{x_1, x_2, ..., x_n\}^T \hat{I}\, \mathbf{R}^n$, é fornecido no instante *t* a todas as células da rede. O vetor de pesos de uma célula é $\mathbf{W}=\{w_1,w_2,..., w^n\}^T \hat{I}\, \mathbf{R}^n$. O algoritmo de aprendizado muda os pesos de conexão de cada uma das células, em função dos vetores de entrada. Inicialmente os valores dos pesos w_i de cada célula são inicializados com valores aleatórios. A geração do mapa auto-organizado é feita utilizando-se um número suficiente de vetores de entrada distribuídos estatisticamente no espaço de entrada, pelo algoritmo seguinte:

Passo 1: Achar a unidade (célula) **c**, cujo vetor de pesos $\mathbf{W}(t)$ tenha a mínima distância da entrada $\mathbf{X}(t)$:

$$d_c(\mathbf{X}) = \| \mathbf{X}(t) - \mathbf{W}_c(t) \| = \min_i \{\| \mathbf{X}(t) - \mathbf{W}_i(t)\|\} \qquad \text{(e.2.1)}$$

Neste caso, diz-se que a unidade **c** responde a **X**(t). No caso mais simples, a distância $d_c(\mathbf{X})$, entre **X** e **W**, pode ser a *Distância Euclidiana.*

Passo 2: Modificar os vetores de peso da unidade ganhadora **c** e dos seus vizinhos topológicos. A vizinhança topológica N_c, tal como a ilustrada na fig.2.2, pode assumir formas variadas (por exemplo: hexagonal ou retangular). No processo de aprendizado, N_c diminui continuamente ao longo do tempo.

O raio de N_c, é inicializado com um valor amplo (p/ ex., 75% do raio total da rede). O raio vai diminuindo, de acordo com uma função, que pode ser linear ou não-linear. Os pesos são atualizados de acordo com as expressões seguintes:

$$m_i(t+1) = m_i(t) + \alpha(t)[x(t) - m_i(t)] \; \forall \; i \in N_c \qquad \text{(e.2.2)}$$

$$m_i(t) = m_i(t) \; \forall \; i \notin N_c \qquad \text{(e.2.3)}$$

Na expressão (e.2.2), a função a(t), é real e positiva, variando entre os valores $(0 < \alpha(t) < 1)$. A função $\alpha(t)$ é conhecida como o *ganho* e pode ir decrescendo de forma linear ou não-linear até chegar a valores próximos a zero.

2.2. Mapas auto-organizados de características topológicas.

O algoritmo descrito acima é a base da operação da rede para utilizá-la como um sistema de reconhecimento de padrões. O componente x_j do vetor de entrada pode ser considerado como a atividade de um neurônio *j*, numa camada sensorial de entrada, considerando-se o modelo de Kohonen com 2 camadas: uma sensorial (as entradas) e uma de mapeamento bi-dimensional (as células de processamento) . Uma célula **c** possui uma resposta forte a um sinal **X**, quando $d_c(\mathbf{X})$ é pequena. Em consequência, o vetor $\mathbf{W}_c$ aponta diretamente para àquela posição no espaço de entrada *n*-dimensional, na qual a célula **c** está melhor *sintonizada*. Alguns autores chamam o $\mathbf{W}_c$ a *posição virtual* da célula **c** no espaço de características de dimensão *n*. O processo de aprendizado dos vetores $\mathbf{W}_c$, representa a evolução do mapa em intervalos de tempo discreto $t = 0, 1, ..., t_{max}$. Como resultado deste processo, os pesos evoluem até abrangerem um espaço característico virtual, de dimensão *n*, constituindo um **mapa auto-organizado**.

3. Metodologia e procedimento experimental

Neste trabalho utilizou-se a arquitetura descrita nas fig.2.1 e 2.2 num arranjo de 12 x 8 células, onde se aplicou o algoritmo apresentado na secção 2.1. A rede foi estimulada por uma sequência de vetores de dimensão 15 e 16, representando o sinal sonoro. Os estímulos foram gerados a partir de amostras de instrumentos acústicos gerados por um sintetizador ROLAND D-20. Os 17 arquivos que representam os timbres escolhidos abrangeram uma ampla gama de instrumentos musicais, pertencentes às famílias ou classes das cordas, madeiras, metais e percussão. Os timbres foram rotulados de acordo com a seguinte convenção: **A**=Apito, **B**=Bumbo, **C**=Contrabaixo, **D**=Corne Inglês, **E**=Violoncello, **F**=Clarinete, **G**=Prato Orquestral, **H**=Caixa, **I**=Flauta, **J**=Trompa, **K**=Oboe, **L**=Pizzicato de violino, **M**=Queixada, N=Trompete, **O**=Trombone, **P**=Tuba, **Q**=Violino.

3.1 Estímulos

Os sons foram amostrados numa *workstation*, no padrão U-LAW (8 bits; fs=8 Khz), com duração de cerca de 500 mseg. Utilizando o *software* MATLAB 4.2 para UNIX, foi desenvolvido um programa que recebe a amostra com a duração indicada e a divide em 3 trechos, correspondentes ao ataque (**A**), sustentação (**S**) e decaimento (**D**) do sinal. Para cada trecho, calculou-se o espectro discreto, através da FFT. A função *spectrum* do MATLAB implementa o método Welch de estimativa de Espectro de Potência (Oppenheim & Schafer, 1975). Utilizando esta função foram gerados vetores de 16 componentes, representando o espectro de potência de cada um dos trechos A, S e D. Nas figs.3.1-fig.3.4, a imagem superior representa a forma de onda do sinal no tempo. **A**, **S** e **D** indicam o trecho em que foi calculado o espectro, para cada amostra instrumental. A figura 3-d inferior representa a evolução do espectro de energia do sinal, ao longo do tempo, durante o intervalo de 500 ms.

3.2. Formato dos padrões de entrada

Os vetores correspondentes aos 3 trechos amostrados foram concatenados num só vetor de 48 componentes (3 x 16), deslocando cada sub-vetor à esquerda (ver fig. 3.5). Um método similar, aplicado a redes do tipo TDNN

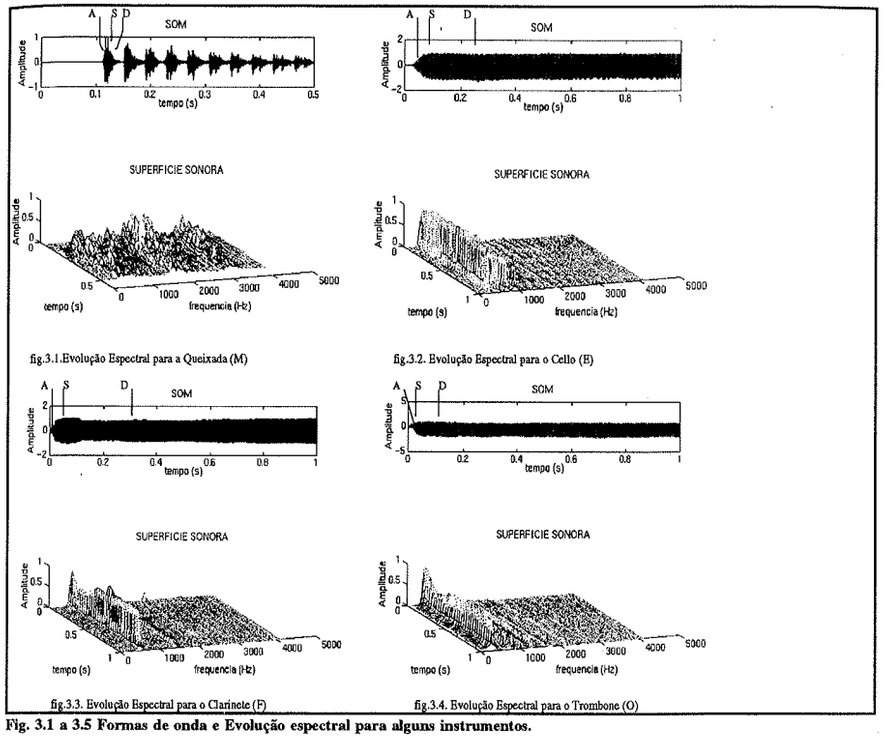

fig.3.1.Evolução Espectral para a Queixada (M)

fig.3.2. Evolução Espectral para o Cello (E)

fig.3.3. Evolução Espectral para o Clarinete (F)

fig.3.4. Evolução Espectral para o Trombone (O)

Fig. 3.1 a 3.5 Formas de onda e Evolução espectral para alguns instrumentos.

(*Time-Delay Neural Network*) foi utilizado por J. Kangas (Kangas, 1994) para acrescentar informação de contexto em problemas de reconhecimento de fonemas. A estrutura dos vetores de entrada, **X**(t), que serão os padrões definitivos de ensino da rede, é apresentada na fig.3.5.

A rede é treinada por uma sequência de vetores de padrões de timbre com a estrutura mostrada na fig.3.5.

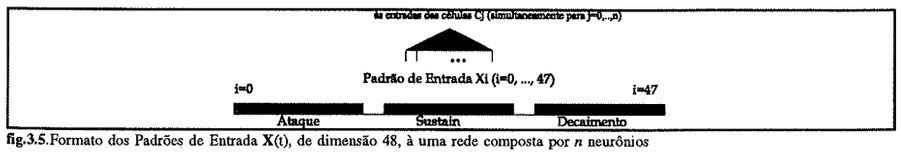

fig.3.5.Formato dos Padrões de Entrada **X**(t), de dimensão 48, à uma rede composta por *n* neurônios

O número destes padrões deve ser grande, para ter uma boa representação do espaço amostral. Nem sempre é possível alcançar esse número de vetores de entrada, quando os padrões são obtidos a partir de dados experimentais. Neste caso, o conjunto desses vetores deve ser reapresentado obter mapas com as características desejadas.

4. Mapas auto-organizados resultantes

A classificação das famílias de instrumentos, utilizada neste trabalho obedece à seguinte convenção :

Classe A - Cordas : violino, violoncello, contrabaixo, pizzicato de violino.
Classe B - Madeiras : oboe, flauta, clarinete, corne inglês
Classe C - Metais : tuba, trombone, trompete, trompa.
Classe D - Percussão : queixada, bumbo, caixa, prato orquestral, apito.

O mapa da fig.4.1 foi obtido após 81600 passos de treinamento. Vê-se a formação de regiões, correspondentes às diferentes classes ou famílias instrumentais. Cada unidade do mapa representa uma célula de uma rede de 12 x 8 neurônios. Cada célula, ou neurônio, ficou sintonizado num determinado padrão.

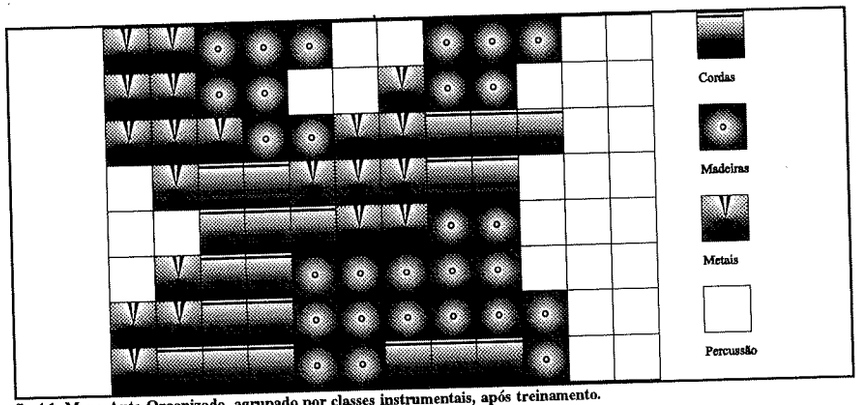

fig.4.1. Mapa Auto-Organizado, agrupado por classes instrumentais, após treinamento.

Numa segunda fase, após a formação do mapa, apresentou-se à rede uma série de 68 vetores de entrada, representando uma sequência de vetores *desconhecidos*, obtidos de forma similar ao procedimento descrito no item 3.1. O modelo conseguiu reconhecer esses padrões *desconhecidos*, com uma taxa de erro de aproximadamente 7%, utilizando-se os mesmos parâmetros do mapa da fig.4.1. Os padrões de entrada desconhecidos representam pequenas variantes dos vetores de treinamento que foram inicialmente apresentados à rede. Uma série de mapas similares ao apresentado na fig. 4.1, foram obtidos em cada simulação. A evolução no tempo e a forma final do mapa (após o treinamento) é fortemente dependente dos parâmetros utilizados no algoritmo descrito na secção 2.1. A forma como a função $\alpha(t)$ evolui no tempo exerce una enorme influência na formação final do mapa. A rapidez com que a vinzinhança N_c diminui, a medida que o tempo evolui na fase de treinamento, também demonstrou ser fundamental. Os mapas com características topológicas pobres possuem muitas regiões inconexas, ao contrário daquele mostrado na fig.4.1.

5.Conclusões

Nas redes similares às estudadas neste trabalho, o sucesso do reconhecimento de um padrão, previamente ensinado, está associado à geração de mapas auto-organizados com características topológicas bem definidas. Nas nossas simulações obtiveram-se resultados muito variáveis; desde taxas de erro razoavelmente boas (da ordem do 4%) até resultados bastante pobres (taxas de erro superiores a 70%). Estes resultados são dependentes de muitos fatores. Primeiramente, do número de passos de treinamento (ou aprendizado). O processo é melhor comportado enquanto se tiver um número de passos de treinamento superiores a 1000. Kohonen utilizou, em seus trabalhos, um número de passos da ordem de 10000, para conseguir bons resultados. Outro fator são os parâmetros que controlam o comportamento da função $\alpha(t)$ e a sua evolução no tempo. Ao contrario do sugerido em algumas publicações, a escolha de uma função (linear ou não-linear) exerce grande influencia nos resultados, como foi comprovado nas nossas simulações. Do mesmo modo, a forma como a vizinhança da célula vencedora evolui no tempo de treinamento também é muito importante. Para o efeito do reconhecimento de timbres sonoros, é fundamental um adequado pre-processamento dos sinais. Apesar de se tentar introduzir informação relativa às características transientes do sinal acústico, utilizando vetores concatenados, o uso só de informação do espectro de energia parece ser insuficiente. A informação da fase como informação adicional seria um fator interessante a estudar, embora, como foi mencionado, para o caso da percepção humana de diferenças timbrísticas não seja tão importante. No relativo ao estudo de novas técnicas de reconhecimento de sinais acústicos, utilizando sistemas neurais artificiais, existem muitas possibilidades. Um aspecto particularmente interessante é o estudo da influência do contexto musical e da utilização de tons polifônicos, em lugar de tons isolados.

No final deste trabalho, verificou-se a capacidade de reconhecimento de padrões timbrísticos pelo modelo de Kohonen, e identificaram-se alguns importantes aspectos que permitem diminuir a taxa de erro no reconhecimento e melhorar a formação de mapas auto-organizados de características topológicas. Outra característica da arquitetura do modelo é ser especialmente adequada para implementação em sistemas VLSI (*Very Large Scale Integration*), o que torna o modelo particularmente interessante para aplicações em tempo real.

Bibliografia

Erwin ,E., Obermayer ,K. & Schulten, K. (1992). Self-Organizing Maps: stationary states, metastability and convergence rate. *Biol. Cybernetics*, 67, 35-45.

Gordon, J.(1987).The perceptual attack time of musical tones. *Journal of the Acoustical Society of America* (82)1, 88-105.

Grey, J.(1978). Timbre discrimination in musical patterns. *Journal of the Acoustical Society of America*, 64(2), 467-472.

Grey, J. & Gordon, J.(1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America*, 63(5), 1493-1500.

Grey & Moorer (1977). Perceptual evaluations of synthesized musical instrument tones. *Journal of the Acoustical Society of America*, 62(2), 454-462.

Kangas,J. (1994). *On the Analysis of Pattern Sequences by Self-Organizing Maps*, Phd dissertation, Helsinki. Univ.

Kohonen, T. (1982).Self-organizing formation of topologically correct feature maps.*Biol. Cybernetics* 43,59-69.

Kohonen (1988). The Neural Phonetic Typewriter. *IEEE Computer magazine*, 21(3), 11-22.

Ritter , H. & Kohonen, T.(1989). Self-Organizing Maps, *Biological Cybernetics*, 61, 241-254.

Kohonen, T.(1990).The Self-Organizing Map, *Proceedings of the IEEE*, 78(9),1464-1480.

Lo, Z. & Bavarian, B. (1991). On the rate of convergence in topology preserving neural networks. *Biological Cybernetics*, 65, 55-63.

Plomp, R. (1976). *Aspects of Tone Sensation - A Psychophysical Study.* Academic Press, London.

Ritter & Schulten, 1986.

Singh, P. (1987). Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre? *Journal of the Acoustical Society of America*, 82(3), 886-899.

Oppenhein, A. & Schafer, R.(1975). *Digital Signal Processing*, Prentice-Hall.

Wang,Y. & Chen,S. (1994).Tone recognition of continuous Mandarin speech assisted with prosodic information. *Journal of the Acoustical Society of America*, 96(5), 2637-2645.