

1º Congresso Brasileiro de Redes Neurais

Escola Federal de Engenharia de Itajubá
Itajubá, 24 a 27 de outubro de 1994

Uso da Quantização Vetorial e Mapas de Kohonen para Traçado de Trajetórias do Sinal da Fala

*Ana Cristina S. Antunes
Gilberto A. Carrijo*

Departamento de Engenharia Elétrica
Universidade Federal de Uberlândia
38.400-902 - Uberlândia - MG

Resumo. Desenvolveu-se um trabalho com o objetivo de analisar os melhores métodos para visualização dos sinais da fala, distribuindo-os sobre mapas fonotópicos. Utilizou-se ora um algoritmo desenvolvido por Teuvo Kohonen, o qual baseia-se na organização dos sistemas nervosos biológicos, ora a junção deste com um Quantizador Vetorial. Os resultados obtidos comprovam a eficácia dos meios empregados, e sugerem a sua utilização num projeto de auxílio a deficientes auditivos.

I - INTRODUÇÃO

Acredita-se, pelo conhecimento que se tem até hoje sobre sistemas nervosos biológicos, que as organizações de baixo nível de tais sistemas sejam determinadas geneticamente. Com o aprendizado, durante a vida, as organizações de níveis mais altos vão se definindo, através de um processo de **auto-organização**. Este princípio de auto-organização implica na colocação ordenada dos neurônios, de forma que possam refletir algumas características físicas do estímulo externo.

Um algoritmo, desenvolvido por Teuvo Kohonen [6], é capaz de produzir

mapas auto-organizáveis com princípios similares aos das células nervosas, cujas aplicações concentram-se principalmente nas tarefas de reconhecimento do sinal da fala. Em sua versão mais simplificada, representa um classificador não supervisionado, bastante semelhante ao algoritmo de **Quantização Vetorial** [5]. É capaz de extrair as características mais importantes dos padrões de entrada, situados num espaço multidimensional, e representá-las num sistema de poucas coordenadas.

Com estas características, esta rede revela-se bastante útil na visualização de sinais da fala num mapa bidimensional, representando-os numa forma mais genérica que os mapas formantes, pois leva em conta o espectro completo de cada quadro do sinal [4].

II - MAPAS DE KOHONEN

O processamento do sinal da fala, notadamente no caso do reconhecimento, representa um problema bastante complexo.

Trabalha-se com padrões de entradas contínuos, sem uma definição precisa do início e fim deste dentro do sinal, e a saída normalmente é definida por um símbolo.

Os mapas de Kohonen são capazes de representar, em poucas dimensões, padrões de entrada que pertencem a um espaço multidimensional. Kohonen compôs uma rede, com elementos neuronais, capaz de trabalhar com um número indefinido de padrões de entrada. Esta rede, após treinamento, pode ser utilizada para identificação de novos padrões.

Os mapas auto-organizáveis de Kohonen são constituídos basicamente de nós de entrada, uma matriz de pesos que conecta a entrada à saída, e pelos nós de saída. Tais nós de saída são arranjados numa grade, normalmente bidimensional, na forma de uma janela retangular ou hexagonal. Um exemplo de topologia desta rede pode ser visto na Figura 1.

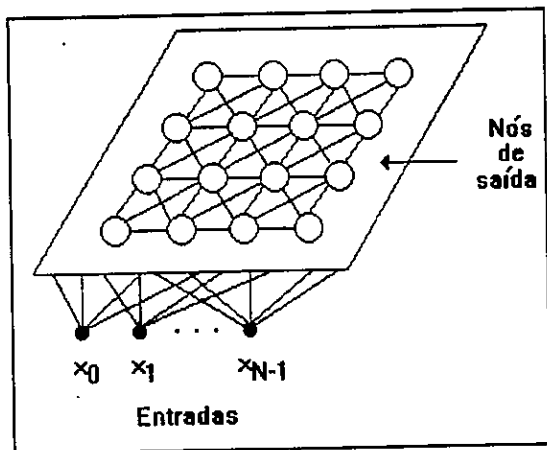


Figura 1: Mapas auto-organizáveis de Kohonen.

As entradas para a rede são contínuas, seqüenciais no tempo, e sem especificação da saída desejada, sendo, portanto, classificada como uma **rede de treinamento não supervisionado** [3]. Os nós de saída mantêm-se altamente interconectados, e serão ordenados naturalmente na fase de aprendizado. A adaptação dos pesos se fará de forma que, após um número suficiente de entradas, as células da saída topologicamente próximas sejam sensíveis às entradas fisicamente

semelhantes. Sendo assim, tais pesos definirão grupos que amostram o espaço de entrada. Esse processo de adaptação é também conhecido com **Treinamento Competitivo** [2]. Seu principio básico provém de estudos de problemas estatísticos, que envolvem análise de agrupamentos.

Para que seja possível o ordenamento natural dos neurônios, o algoritmo de Kohonen requer que uma vizinhança seja definida ao redor de cada nó. O raio de tal vizinhança vai diminuindo com o tempo, conforme pode ser visto na Figura 2.

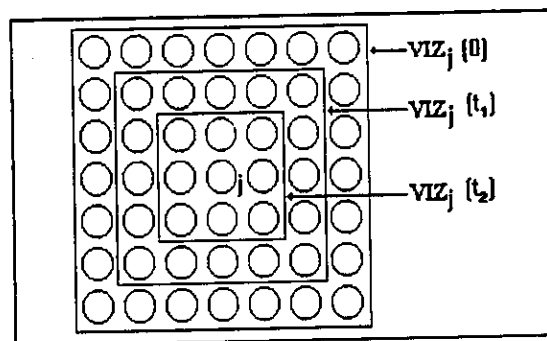


Figura 2: Exemplo de vizinhanças formadas no Mapa de Kohonen.

Também é necessário uma forma de se calcular distâncias entre nós; o que se utiliza normalmente é a **Distância Euclidiana**:

$$d_j = \sum_{i=0}^{N-1} (x_i(t) - w_{ij}(t))^2 \quad (1)$$

onde

- $x_i(t)$ é a entrada do nó i no tempo t
- $w_{ij}(t)$ é o peso do nó de entrada i ao nó de saída j no tempo t .

Quando um nó de saída é escolhido por ter a menor distância ao nó de entrada, os pesos deste e de sua vizinhança serão modificados, usando:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t) (x_i(t) - w_{ij}(t)) \quad (2)$$

$$\forall j \in VIZ_j(t) \quad 0 \leq i \leq N-1$$

A função $\alpha(t)$ representa o termo de ganho ($0 < \alpha(t) < 1$), o qual decresce com o tempo t .

Esta alteração nos pesos dos nós da vizinhança é feita com a finalidade de que tais nós tornem-se mais sensíveis aquele tipo de entrada. O processo é repetido para todas as entradas. Os pesos normalmente convergirão e serão fixados após o número de iterações desejado.

Nas tarefas que envolvem processamento do sinal da fala, as entradas geralmente correspondem aos componentes espectrais da Transformada Rápida de Fourier (FFT) [7, Cap. 6], ou ainda, a coeficientes preditos linearmente (LPC) [7, Cap. 8].

III - IMPLEMENTAÇÃO

Neste trabalho, o mapa de Kohonen foi implementado para atuar como uma ferramenta que possibilitasse a visualização do sinal da fala [1]. Os sinais da fala são inicialmente preprocessados, a fim de que se possa extrair suas características. A seguir, os parâmetros obtidos são aplicados ao algoritmo de Kohonen, efetuando assim o treinamento da rede. Como último passo, para cada palavra é definida uma trajetória dentro deste mapa, a qual é composta pelos nós aos quais cada quadro mais se aproxima. Isto é feito com o intuito de verificar que trajetórias fisicamente próximas podem representar as mesmas palavras, pronunciadas em momentos diferentes.

Paralelamente ao algoritmo de Kohonen, implementou-se também o algoritmo da Quantização Vetorial [5]. Formulou-se uma maneira de trabalhar com os dois métodos em conjunto. Esta junção procede da seguinte maneira. Inicialmente,

são calculados centróides, a partir dos sinais amostradas das palavras, pelo algoritmo de Quantização Vetorial. Os quadros de tais palavras passam, então, a ser representados apenas pelos centróides das células nas quais se localizam. A seguir, os centróides gerados são submetidos, como dados de entrada, ao treinamento do mapa de Kohonen. Finalmente, são calculadas e traçadas as trajetórias das palavras, cujos quadros são representados apenas pelos centróides.

Para efetuar a implementação dos algoritmos citados, e das rotinas gráficas necessárias, utilizou-se um computador da IBM, a RISC 6000/520H.

Até agora apenas foi falado sobre o trabalho com palavras isoladas. No entanto, dentro do módulo de Kohonen, ainda existe a possibilidade de se trabalhar com vogais e com frases.

IV - AQUISIÇÃO DOS DADOS

Os dados utilizados nos testes foram, em sua maioria, obtidos de arquivos contendo amostras de sinais da fala.

Para gravação dos arquivos que fizeram parte dos testes, utilizou-se um sistema de aquisição de dados, implantado num IBM PC 286. Tal sistema é composto basicamente por um filtro passa-baixa de 5 kHz, e um conversor A/D de 12 bits, com uma frequência de amostragem de 10 kHz. Na Figura 3 encontra-se um diagrama representando o processo ao qual o sinal original é submetido.

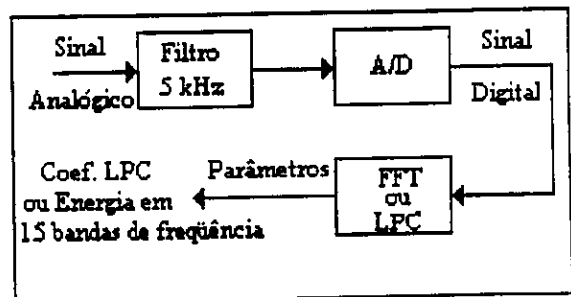


Figura 3: Diagrama do preprocessamento do sinal da fala.

O banco de dados utilizado nos testes pode ser visto na Tabela 1.

Elemento da fala	Num. amostras	Tempo gravação [s]
vogal	1.800	0,18
palavras isoladas	4.000	0,40
frases	80.000	8,00

Tabela 1: Tipos de elementos da fala utilizados nos testes.

As palavras utilizadas correspondem a 5 pronúncias isoladas de cada um dos seguintes termos:

- bola (*bola*);
- casa (*casa*);
- casa (*ccasa*);
- colo (*ccolo*);
- humano (*chumano*);
- sonhos (*csonhos*).

Entre parênteses encontram-se os nomes atribuídos aos arquivos das respectivas palavras.

V - RESULTADOS OBTIDOS

Os resultados aqui apresentados correspondem, numa primeira parte, a aplicação do algoritmo de Kohonen, e, numa segunda parte, a junção dos métodos apresentados.

Mapas de Kohonen

Os sinais aqui utilizados foram divididos em quadros de 256 amostras cada, com sobreposição de 128 amostras entre os quadros adjacentes. Para a Transformada de Fourier, usou-se 16 filtros, com 15 bandas de frequência. Conseqüentemente, o número de nós de entrada foi mantido em 16. Ajustou-se um mapa retangular com dimensão 10x10, contendo, portanto, 100 nós de saída.

Para as vogais, a seqüência de treinamento foi composta de 50 quadros de sinal, ou seja, 5 vogais de 10 quadros cada. As trajetórias de um representante de cada

vogal podem ser vistas na Figura 4. Pode-se verificar que as trajetórias geradas mantiveram-se bem distribuídas sobre o mapa, sem haver, sequer, ocorrência de sobreposição entre as mesmas.

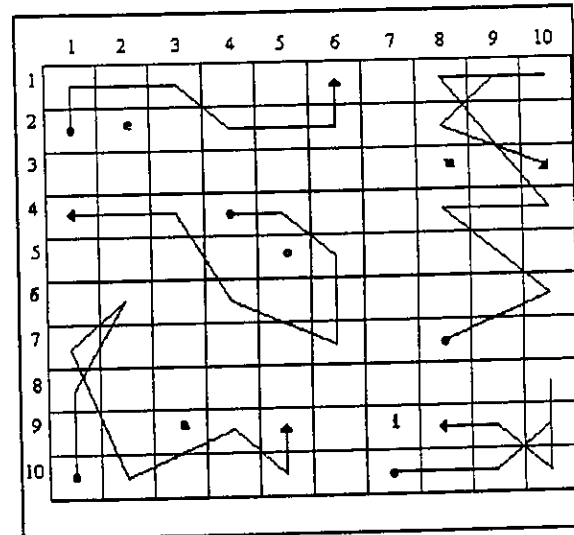


Figura 4: Trajetórias das vogais sobre o Mapa de Kohonen

Com relação as palavras isoladas, foram utilizadas as 30 palavras já citadas, sendo cada palavra composta de 25 quadros de amostras. Efetuou-se um total de 500 iterações no treinamento. As Figuras 5 e 6 mostram as trajetórias de algumas das palavras que participaram do aprendizado. Pode-se constatar novamente a propriedade dos mapas auto-organizáveis em espalhar os dados durante seu treinamento, ressaltando que, tais dados, representam os fonemas que compõem as palavras treinadas.

Como prosseguimento, foi feito o treinamento do mapa com 3 frases, sendo cada uma composta de 500 quadros. Foi efetuado um total de 100 iterações. As respectivas trajetórias correspondem a palavras contidas nestas frases, conforme pode ser visto nas Figuras 7 e 8. Não nota-se, com esse novo conjunto de treinamento, grande diferença no comportamento das trajetórias.

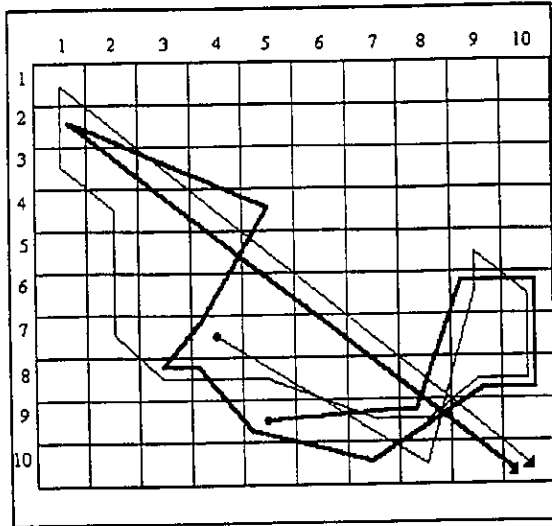


Figura 5: Trajetórias de 2 exemplares da palavra *bola* sobre os mapas auto-organizáveis treinados com 30 palavras.

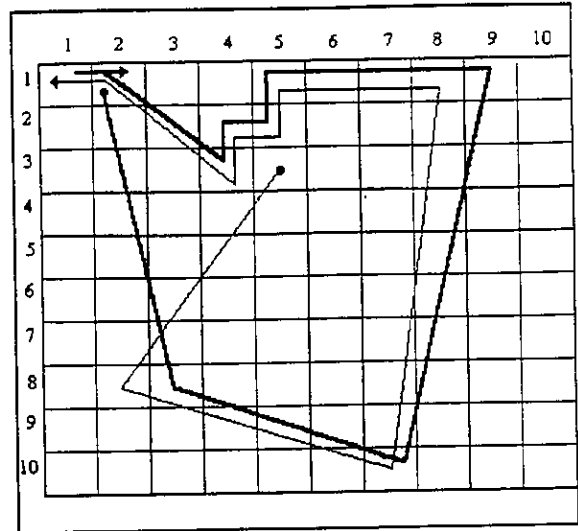


Figura 7: Trajetórias de 2 exemplares da palavra *casa* sobre os mapas auto-organizáveis treinados com 3 frases.

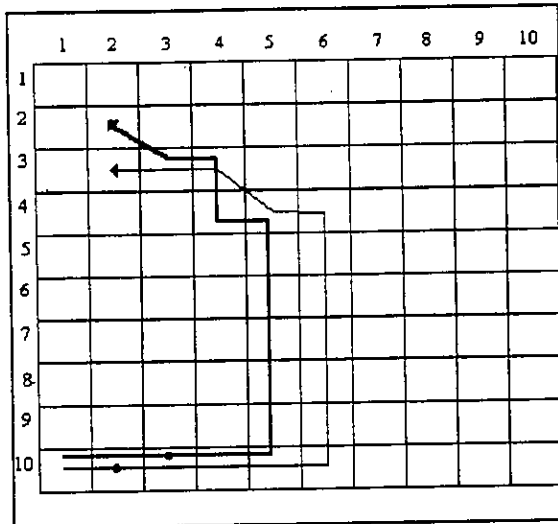


Figura 6: Trajetórias de 2 exemplares da palavra *chumano* sobre os mapas auto-organizáveis treinados com 30 palavras.

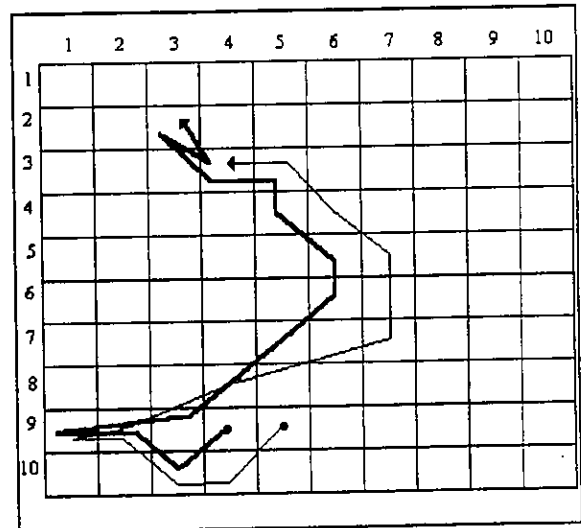


Figura 8: Trajetórias de 2 exemplares da palavra *chumano* sobre os mapas auto-organizáveis treinados com 3 frases.

Quantização Vetorial e Mapas de Kohonen

Neste módulo, as 30 palavras citadas participaram dos testes. A partir de tais palavras, foram gerados 64 centróides, com o algoritmo de Quantização Vetorial. Estes centróides formaram, então, o conjunto de treinamento do mapa de Kohonen. A dimensão do mapa continuou em 10x10 e o número de iterações foi aumentado para 1.000. Para o cálculo das trajetórias foram utilizadas as mesmas palavras, porém com seus quadros sendo

representados agora apenas pelos centróides gerados. As Figuras 9 e 10 mostram alguns destes resultados. Nenhuma melhora substancial foi obtida, embora, no geral, as trajetórias tenham mantido-se bastante próximas para palavras de mesma natureza.

VI - CONCLUSÃO

Pode-se verificar, com base nos resultados obtidos, a forte propriedade do algoritmo de Kohonen em agrupar os dados de entrada, de acordo com suas

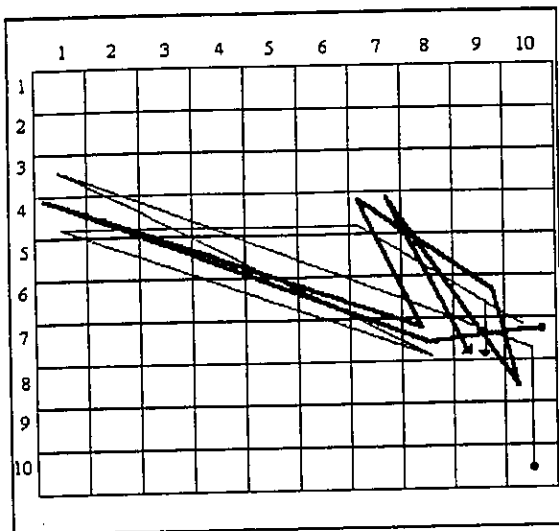


Figura 9: Trajetórias de 2 exemplares da palavra *bola* sobre os mapas auto-organizáveis treinados com 64 centróides.

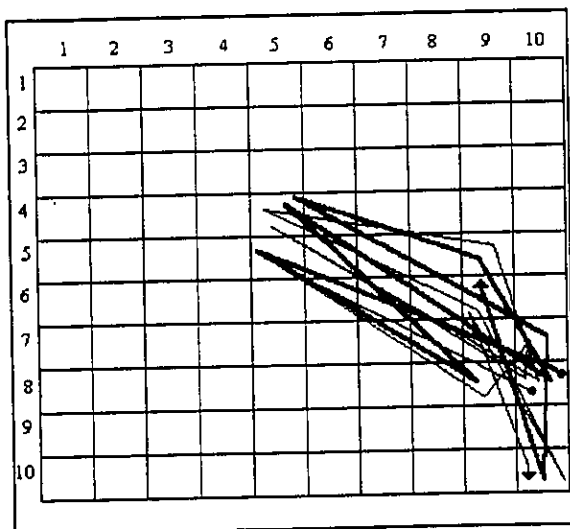


Figura 10: Trajetórias de 2 exemplares da palavra *chumano* sobre os mapas auto-organizáveis treinados com 64 centróides.

características físicas, e de espalhar estes grupos por sobre um mapa, denominado Mapa de Kohonen. Pôde-se notar, quando aplicou-se a Quantização Vetorial em conjunto com Kohonen, que essa propriedade de distribuição não se manteve de forma tão homogênea. A grande vantagem que se obteve neste caso foi em termos de velocidade, visto que o tamanho do conjunto de treinamento diminuiu consideravelmente.

Conclue-se que é totalmente viável aplicar os métodos vistos em tarefas que envolvam o reconhecimento de elementos

da fala, notadamente no caso de palavras isoladas, simplesmente pela análise das trajetórias geradas. Isto é possível devido a observação de que, trajetórias fisicamente próximas, normalmente representam o mesmo elemento falado.

O algoritmo de Kohonen possibilita o desenvolvimento de um bom sistema para visualização do sinal da fala. Tal sistema poderia ser usado, por exemplo, como auxílio a deficientes auditivos, no treinamento da reprodução de palavras isoladas.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ANTUNES, Ana C. S. *O Uso do Mapa de Kohonen para Traçado de Trajetórias do Sinal da Fala*. Uberlândia: DEENE da UFU, 1993. 135 p. (Dissertação de Mestrado).
- [2] KANGAS, Jarj A.; KOHONEN, Teuvo K.; LAAKSONEN, Jorma T. Variants of self-organizing maps. *IEEE Transactions on Neural Networks*, v.1, n. 1, p. 93-99, 1990.
- [3] KEPUSKA, Veton Z. & GOWDY, John N. Investigation of phonemic context in speech using self-organizing feature maps. *IEEE*, p. 504-507, 1989.
- [4] KOHONEN, Teuvo; MÄKISARA, Kai; SARAMÄKI, Tapio. Phonotopic maps - insightful representation of phonological features for speech recognition. In: IEEE - INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 7., 1984, Montreal, Canada. *Proceedings...* [S.l.:S.n.], 1984. p. 182-185.
- [5] LINDE, Yoseph; BUZO, Andrés; GRAY, Robert M. An algorithm for vector quantizer design. *IEEE Transaction on Communications*, v. COM-28, n. 1, p. 84-95, Jan. 1980.
- [6] LIPPMANN, Richard P. An introduction to computing with neural nets. *IEEE ASSP MAGAZINE*. p. 4-23, Apr. 1987.
- [7] RABINER, Lawrence R. & SCHAFER, Ronal W. *Digital processing of speech signals*. New Jersey: Prentice-Hall, Inc., 1978.