CRIPTOANÁLISE DE SINAIS DE VOZ CIFRADOS POR TSP USANDO REDES NEURAIS E *MEAN-FIELD ANNEALING*

J. A. Apolinário Jr.
UFRJ/COPPE - Rio de Janeiro/RJ
apolin@coe.ufrj.br

P. R. S. Mendonça
UFRJ/COPPE - Rio de Janeiro/RJ
mendonca@coe.ufrj.br

L. P. Calôba
UFRJ/COPPE-Rio de Janeiro/RJ
caloba@coe.ufrj.br

***Resumo* - Este trabalho apresenta um esquema de criptoanálise de sinais de voz cifrados por permutação de segmentos temporais (TSP). São estimadas distâncias espectrais entre extremidades de segmentos e uma distância total mínima é buscada com uma abordagem do tipo *Problema do Caixeiro Viajante* usando Rede Neural.**

***Abstract* - This work presents a scheme of cryptoanalysis of speech signals ciphered by Time Segment Permutation (TSP) . Spectral distances are estimated between borders of segments and a minimum total distance is searched with a *Traveling Salesman Problem* approach using Neural Network.**

1. Introdução

A CRIPTOANÁLISE de sinais de voz cifrados por permutação de segmentos temporais (TSP) de tamanho fixo e bloco a bloco (*jumping window*) foi apresentada em artigo anterior [Apolinário, 1993] onde bons resultados foram obtidos para o caso de 8 segmentos. A idéia básica era a escolha da permutação que apresenta a menor soma das distâncias espectrais das bordas de segmentos adjacentes. Trata-se, pois, de um problema de otimização combinatória muito parecido com o tradicional Problema do Caixeiro Viajante. A abordagem usada anteriormente foi a busca exaustiva das 8! permutações possíveis sendo que sua principal limitação é o número segmentos que pode ser bem maior em *scramblers* reais. Este trabalho busca uma solução baseada numa rede neural de modo que possa ser usada posteriormente no caso de termos 16 segmentos.

Será brevemente apresentada na Seção 2 o esquema de criptoanálise. A seguir, encontra-se na Seção 3 a formulação do problema com uma rede neural (Hopfield) de 8x8 neurônios onde cada um está associado a um segmento numa dada posição de maneira análoga a uma cidade no roteiro de um Caixeiro Viajante. A Seção 4 mostra como resultados válidos e com distâncias mínimas (não garantidamente globais mas bem razoáveis) foram obtidos. A Seção 5 mostra alguns resultados obtidos em simulações e a Seção 6 algumas conclusões.

2. Esquema de Criptoanálise

Observamos na fig. 1 os conceitos de bloco e segmentos. Os segmentos foram permutados dentro de um bloco; para que isto possa ser realizado, é necessário que todos os segmentos deste bloco sejam armazenados numa memória antes de serem transmitidos numa ordem diferente da original. Isto implica num retardo total de comunicação igual a duas vezes o tamanho do bloco (transmissão e recepção). Este retardo é uma das limitações do processo, bem como o número de segmentos em cada bloco: um número grande de segmentos diminuiria a inteligibilidade residual e aumentaria a resistência à criptoanálise, mas ao mesmo tempo causaria expansão de banda, necessidade de um sincronismo mais preciso e um efeito mais acentuado de superposição de segmentos quando o sinal passa por um canal.

Neste trabalho, é de interesse considerar-se a distância entre os espectros do final e do início de dois segmentos de sinal de voz para verificar-se o quanto os mesmos se assemelham. Isto é feito levantando-se uma medida de distância espectral, ou seja, um número não negativo que meça o quanto o espectro do fim de um segmento está próximo ou não do espectro do início de outro.

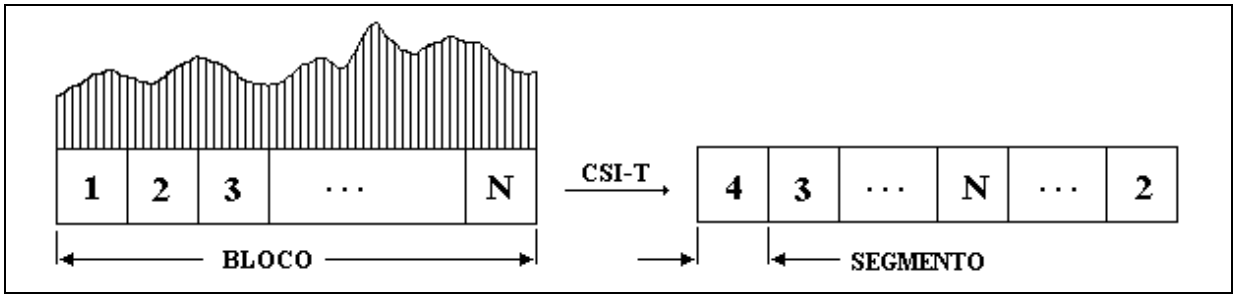

Fig. 1 - Sinal de voz dividido em blocos e segmentos.

Dada a necessidade de levantar-se os coeficientes da modelagem do sinal de voz numa análise a curto tempo para que seja possível comparar os espectros em dois instantes de tempo (antes e após a transição dos segmentos), optou-se pela utilização de uma estrutura reticulada (*lattice*) que, além de oferecer a possibilidade de obtenção dos coeficientes de modo adaptativo (amostra a amostra), apresenta encadeamento de seções idênticas, coeficientes com magnitude menor que 1, teste de estabilidade por inspeção e obtenção dos coeficientes direto das amostras de voz sem um cálculo intermediário da função de autocorrelação [Cowan e Grant, 1985]. Os coeficientes da estrutura reticulada, usualmente chamados coeficientes de reflexão $\{\mathbf{k_i}\}$, independem da ordem **p** do filtro e podem ser transformados em coeficientes $\{\mathbf{a_i}\}$ (LPC) ou $\{\mathbf{c_i}\}$ (cepstrais). O algoritmo utilizado para a estimação dos coeficientes de reflexão neste trabalho foi uma versão normalizada do RLSL (*Recursive Least-Squares Lattice*) conhecida como SQNLSL (*Square-Root-Normalized Least-Square Lattice*). Este algoritmo apresenta uma diminuição da complexidade das recursões e uma melhoria nas propriedades numéricas das variáveis [Cowan e Grant, 1985].

Encontra-se na literatura [Apolinário Jr., 1993] várias medidas de distância espectral. Todas elas são amplamente usadas em processamento de voz e particularmente em criptofonia [Sridharan, Dawson e Goldburg, 1991] para medir objetivamente inteligibilidade residual e qualidade de voz recuperada. Serão usados neste trabalho os coeficientes cepstrais e a partir deles define-se uma distância (cepstral) que está associada a uma média quadrática das diferenças de dois espectros em magnitude logarítmicos [Markel e Gray, 1976], dada por:

$$dc(D,E) = \left[c_D(0) - c_E(0)\right]^2 + 2.\sum_{i=1}^{p} \left[c_D(i) - c_E(i)\right]^2 \qquad (1)$$

onde os c's são os coeficientes cepstrais em D e E, e p é a ordem do modelo. Observa-se na eq. 1 que já encontra-se embutida a informação de energia (c's índice zero).

O esquema proposto é aplicável num sinal de voz que foi cifrado pela permutação de N segmentos temporais de mesmo tamanho dentro de blocos transmitidos sequencialmente. Será assumido inicialmente que N = 8 para fins de comparação com um esquema de busca exaustiva cujos resultados são disponíveis. Entretanto o mesmo esquema pode ser aplicado a sinais cifrados com 16 segmentos sendo esta a razão primordial do uso de redes neurais uma vez que a busca exaustiva é inviável (16!≈2.1E13) neste caso.

O algoritmo apresentado na fig. 2 mostra como será processada a reordenação dos segmentos. Para cada bloco (8 segmentos) lido são estimados os coeficientes $\{\mathbf{k_i}\}$ à direita e à esquerda de cada segmento. Para os coeficientes à esquerda, roda-se um filtro SQNLSL do final para o início do segmento. Os coeficientes à direita são obtidos com o mesmo filtro rodando do início para o final do segmento. Os coeficientes estimados são armazenados e as distâncias do final para o início de dois segmentos são calculadas e armazenadas numa matriz contendo todas as distâncias possíveis entre os segmentos do bloco.

```
/*  ALGORITMO CSITOVOZ (CRIPTOANÁLISE DE CSI-T)  */
INÍCIO  CSITOVOZ (ENTRA:SINAL.CSI, SAI:SINAL.VOZ)
   "DECLARAÇÃO DE ARQUIVOS E VARIÁVEIS";
   "ABERTURA DE ARQUIVOS";
   "LEITURA  E ESCRITA DE CABEÇALHOS";
   "INICIALIZAÇÃO DE VARIÁVEIS";
   ENQUANTO  NÃO (FIM DE ARQUIVO)
      "LER  UM BLOCO DO ARQUIVO DE ENTRADA":
      PARA  SEG DE  1  ATÉ  NR_SEGS PASSO 1
         "CALCULAR  Kd  (COEFICIENTES À DIREITA DO SEGMENTO )";
         "CALCULAR  Ke  (COEFICIENTES À ESQUERDA DO  SEGMENTO)";
      FIM-PARA;
      "CALCULAR AS DISTÂNCIAS d[i][j]  (DIR SEG i PARA ESQ SEG  j)";
      "ACHAR A PERMUTAÇÃO DE MENOR DISTÂNCIA (USANDO  REDE NEURAL)";
      "GUARDAR OS COEFICIENTES DO FINAL DO BLOCO";
      "REORDENAR OS  SEGMENTOS  E ESCREVER NO ARQUIVO DE SAÍDA";
   FIM-ENQUANTO;
   "FECHAR ARQUIVOS";
FIM  {CSITOVOZ}.
```

Figura 2 - Algoritmo de criptoanálise (BLOCO com NR_SEGS = 8 segmentos).

A seguir, busca-se uma solução (permutação) com distância mínima conforme será visto nas seções 3 e 4. A permutação escolhida é guardada num vetor que será a ***chave*** para a reordenação dos segmentos. Uma vez

efetuada a reordenação, é escrito o bloco no arquivo de saída e o processamento continua com a leitura do próximo bloco.

3. Formulando o Problema

O uso de Redes de Hopfield (fig. 3) na busca de soluções para problemas de otimização combinatória já é corrente. Pode-se provar [Hopfield, 1984] que uma Rede de Hopfield operada de forma apropriada é estável segundo o Critério de Lyapunov, ou seja, existe uma função de energia, ou função de Lyapunov, que decresce até um ponto de mínimo local durante a operação da rede. A função de Lyapunov usada neste trabalho é dada por

$$E = -\frac{1}{2}\sum_{\substack{ijkl \\ k\neq i \\ l\neq j}} w_{ijkl}\, y_{ij}\, y_{kl} - \frac{1}{2}\sum_{\substack{ijk \\ k\neq i}} w_{ijkj}\, y_{ij}\, y_{kj} - \frac{1}{2}\sum_{\substack{ijl \\ l\neq j}} w_{ijil}\, y_{ij}\, y_{il} - \sum_{\substack{ij \\ k\neq i \\ l\neq j}} w_{ijij}\, y_{ij}^2 - \sum_{ij} t_{ij00}\, y_{ij} + cte, \qquad (2)$$

onde y_{ij} corresponde à saída do neurônio ij, t_{ij00} corresponde ao *bias* aplicado ao neurônio ij e w_{ijkl} corresponde à sinapse conectando a saída do neurônio ij à entrada do neurônio kl.

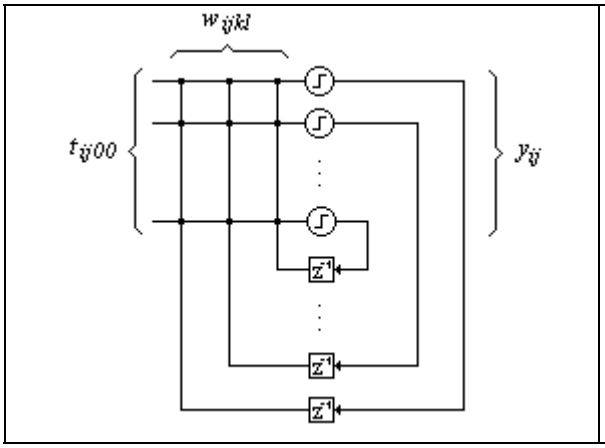

Fig. 3. Rede de Hopfield usada na busca de uma solução para o problema do Caixeiro Viajante, onde não há a presença de entradas externas

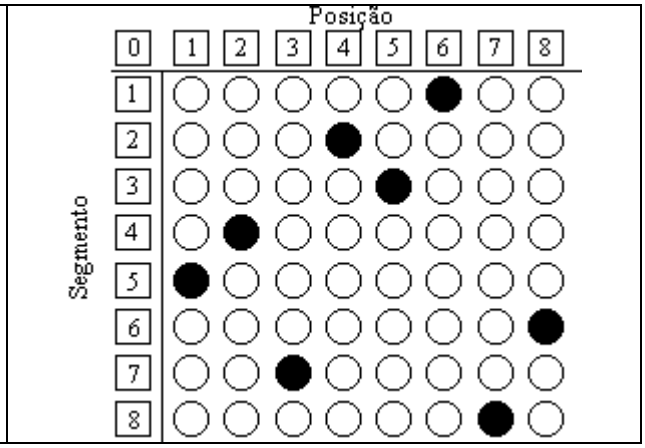

Fig. 4. Cada círculo corresponde a um neurônio, e os círculos hachurados indicam os neurônios ativados. A permutação (válida) resultante da configuração acima é 5 - 4 - 7 - 2 - 3 - 1 - 8 - 6

É necessário, então, formular-se o problema do Caixeiro Viajante de forma que sua solução seja equivalente à minimização de uma função objetivo com a forma da eq. 2. Uma proposta para tal função é $F = F_A + F_B + F_C + F_D + F_E$, onde, observando-se ainda a fig. 4,

$$F_A = \sum_{ij} A_{ij} y_{ij}\left(1 - y_{ij}\right), \qquad (3)$$

(mínimo quando $y_{ij} \in \{0, 1\}$; fornece um parâmetro a mais para o ajuste das sinapses)

$$F_B = \sum_{i} B_i \left(1 - \sum_{j} y_{ij}\right)^2 \qquad (4)$$

(mínimo quando há exatamente um neurônio ativo por coluna (fig. 4))

$$F_C = \sum_{j} C_j \left(1 - \sum_{i} y_{ij}\right)^2, \qquad (5)$$

(mínimo quando há exatamente um neurônio ativo por linha (fig. 4))

$$F_D = \sum_{\substack{ijk \\ i\neq k}} D_{ijk}\, y_{ij}\left(d_{ki}\, y_{k,j-1} + d_{ik}\, y_{k,j+1}\right) \text{ e} \qquad (6)$$

(minimizados os termos anteriores, este termo é mínimo quando a permutação encontrada para os segmentos produz a menor distância)

$$F_E = \frac{1}{2}\sum_{i} E_i d_{0i} y_{i1} \qquad (7)$$

(este termo é similar ao anterior exceto por referir-se a distância entre o último segmento do bloco anterior e o primeiro do bloco atual; o fator 1/2 deve-se a incerteza na permutação tomada para o bloco anterior)

Fazendo $A_{ij} = A$, $B_i = B$, $C_j = C$ e $D_{ijk} = E_i = D$ $\forall$ (i, j, k), desenvolvendo F e comparando-a termo a termo com E (função de Lyapunov), obtém-se os pesos das sinapses da Rede de Hopfield desejada, mostrados a seguir:

$$t_{ij00} = \begin{cases} -A + 2(B + C) \text{ se } j \neq 1; \\ -A + 2(B + C) - (1/2)Dd_{0i} \text{ se } j = 1; \end{cases} \quad w_{ijij} = A - B - C; \quad \begin{matrix} w_{ijil} = -2B; \\ w_{ijkj} = -2C; \end{matrix} \quad w_{ijkl} = \begin{cases} Dd_{ki} \text{ se } j = l + 1; \\ Dd_{ik} \text{ se } j = l - 1; \\ 0, \text{ caso contrá rio} \end{cases} \qquad (8)$$

A etapa seguinte foi o ajuste dos valores de *A*, *B*, *C* e *D*. É natural que *B* e *C* tenham o mesmo valor, pois são responsáveis por haver exatamente 1 neurônio por linha e 1 neurônio por coluna no diagrama da fig. 4, havendo o mesmo custo associado à violação de qualquer uma dessas duas condições. Após vários experimentos computacionais verificou-se que o melhor valor para *A* é aquele que anula os pesos das sinapses de auto-realimentação, dados por w_{ijij}. Assim encontra-se $A = B + C$ e fez-se $B = C = 1$ e $A = 2$.

O ajuste do coeficiente *D* mostrou-se uma tarefa difícil. Um valor muito pequeno resulta num custo baixo para uma permutação que corresponda a uma distância elevada, enquanto que um valor elevado implica um custo relativamente baixo para a violação da condição de haver somente um neurônio ativo por linha e por coluna levando a uma solução não válida. Assim, foi desenvolvido um esquema adaptativo para *D*, descrito a seguir.

Inicialmente é encontrada uma permutação chamada *permutação não-péssima*, que consiste em se tomar a partir do último segmento do bloco anterior, o segmentos imediatamente mais próximo, segundo a distância cepstral. A rede é operada inicialmente com *D* igual ao inverso do valor médio das distâncias cepstrais entre os segmentos. Quando a solução encontrada não é válida decrementa-se *D* através da expressão $D_{novo} = (1 - \alpha_{down})D_{anterior}$. Já quando a solução encontrada é válida mas não é menor que a distância correspondente à da *permutação não-péssima*, incrementa-se o valor de *D* segundo uma expressão do tipo $D_{novo} = (1 + \alpha_{up})D_{anterior}$. Os parâmetros α_{up} e α_{down} foram encontrados empiricamente, sendo iguais a 0.1 e 0.5, respectivamente.

O esquema descrito acima encontrou uma solução superior a não péssima em 70% das tentativas e a solução ótima (obtida, para comparação, através de busca exaustiva) em 25% das tentativas de criptanálise de um sinal cifrado por um *scrambler* de 8 segmentos.

4. Melhorando os Resultados

O grande problema na otimização combinatória e na otimização convexa de modo geral é a presença de mínimos locais, nos quais os algoritmos de minimização usuais congelam. Uma solução para esse problema é o uso de técnicas de *Simulated Annealing* e *Mean-Field Annealing* [Cichocki e Unbehauen, 1993]. Ambas as técnicas baseiam-se em fenômenos termodinâmicos. Um metal quando derretido e subseqüentemente resfriado terá suas moléculas reordenadas segundo uma configuração de mínima energia se o resfriamento for feito de forma suficientemente lenta. A idéia é então perturbar a rede através da injeção de ruído (aumento da temperatura), e lentamente diminuir a amplitude do ruído (resfriamento), de forma que a rede convirja para uma configuração correspondente a um mínimo global de sua função de Lyapunov.

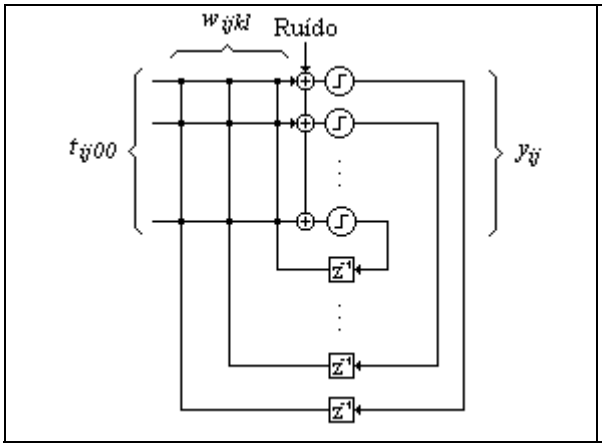

Fig. 5. Esquema para otimização através da técnica de *Simulated Annealing*.

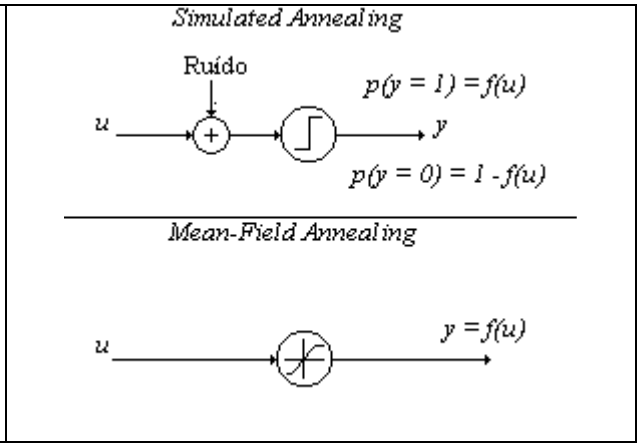

Fig. 6. Comparação entre os neurônios da técnica de *Simulated Annealing* e *Mean-Field Annealing*.

Na técnica de *Simulated Annealing* é adicionado ao sinal da entrada dos neurônios da Rede de Hopfield ruído com uma distribuição especificada e média nula (fig. 5), de forma que na saída do neurônio não mais se terá 1 ou 0 segundo uma função determinística do sinal de entrada, e sim 1 ou 0 com probabilidades determinadas pela entrada e pela distribuição usada. Caso se use a distribuição de Boltzmann tem-se então a

Máquina de Boltzmann. Na técnica de *Mean-Field Annealing* o neurônio discreto é substituído por um neurônio analógico que ao ser apresentado ao sinal de entrada produz na saída o valor esperado da saída de um neurônio discreto que recebesse além do próprio sinal de entrada um ruído com a distribuição dada (fig. 6).

Estudos comparativos [Haykin, 1994] mostram que as técnicas chegam a resultados similares, com uma maior velocidade de convergência para o método de *Mean-Field Annealing*, sendo, portanto, esta a técnica escolhida para o problema em questão.

Impondo-se que o ruído satisfaça uma distribuição de Boltzmann, obtém-se para a função $y = f(u)$ do neurônio analógico uma função do tipo sigmóide:

$$y = \frac{1}{1 + e^{-u/T}} . \tag{9}$$

O "resfriamento" da rede é feito de acordo com a expressão $T_{novo} = 0.93 T_{anterior}$ [Cichocki e Unbehauen, 1993]. É preciso ainda determinar T_0, a temperatura inicial do processo. Isso é feito verificando-se a máxima excursão dos parâmetros u, de forma que a excursão correspondente de y seja da ordem de 0.2 ou 0.1 em torno do centro do hipercubo, o que equivale a um agrupamento dos mínimos da função objetivo. O valor final de T foi escolhido como 0.2, para o qual a sigmóide já é praticamente igual a uma função degrau. Para cada valor de T a função é posta em operação, e assume-se que houve convergência se a energia acumulada na rede, normalizada, não variou mais que 10^{-4}, sendo a variação medida através da expressão abaixo:

$$\Delta E = \frac{\left\| E_{anterior} - E_{atual} \right\|}{E_{anterior}} . \tag{10}$$

Como a complexidade computacional do cálculo da variação da energia acumulada na rede é muito alta, foi desenvolvido um novo critério de parada, baseado na norma euclidiana da diferença do vetor de saída da rede após a atualização de seus neurônios. Essa alteração diminuiu em cerca de seis vezes o tempo de processamento para a Otimização das distâncias em cada bloco cifrado.

Com o uso combinado das técnicas de *Mean-Field Annealing* e ajuste adaptativo de D, encontrou-se uma solução superior à *solução não-péssima* em 100% das tentativas e a solução ótima em 40% das tentativas de criptanálise de um sinal cifrado por um *scrambler* de 8 segmentos.

5. Resultados Experimentais

Os resultados das simulações são a seguir apresentados por meio de medidas objetivas de desempenho do quanto o sinal criptoanalisado aproxima-se do sinal original, apesar de poder-se conjecturar que o objetivo da criptoanálise é a obtenção do conteúdo da mensagem e não a recuperação perfeita do sinal de voz que foi cifrado.

Os sinais de voz usados nos testes foram os mesmos do artigo anterior [Apolinário Jr., 1993] de modo a podermos comparar os resultados obtidos com a Rede Neural proposta e os anteriormente obtidos com busca exaustiva que, no caso do problema de otimização combinatória, são os ótimos.

Observamos, a seguir, a tab. 1 que apresenta uma medida de distorção espectral relativa (foi chamada de distorção para não confundir com as medidas de distância espectral usadas na criptoanálise e corresponde à distância Euclideana não ponderada dos coeficientes $\{\mathbf{a_i}\}$ estimados segmento a segmento dos sinais original e criptoanalisado em relação ao original) para os sinais cifrados *versus* cada técnica utilizada.

	SCHOOL	SINTO	TELEF	VEGA
SOLUÇÃO COM BUSCA EXAUSTIVA	23.3 %	24.0 %	21.3 %	5.3 %
SOLUÇÃO COM REDE NEURAL	41.0 %	39.1 %	34.7 %	40.7 %
SOLUÇÃO NÃO PÉSSIMA	41.2 %	92.1 %	39.6 %	68.4 %

Tab. 1 - Distorção espectral relativa.

Uma outra medida objetiva de desempenho, a ***taxa de acertos*** (número de segmentos recolocados em seus locais de maneira acertada pelo número total de segmentos do sinal), é mostrada na tab. 2.

	SCHOOL	SINTO	TELEF	VEGA
SOLUÇÃO COM BUSCA EXAUSTIVA	72.7 %	65.6 %	63.1 %	91.2 %
SOLUÇÃO COM REDE NEURAL	56.2 %	40.6 %	32.5 %	56.6 %
SOLUÇÃO NÃO PÉSSIMA	52.3 %	12.5 %	22.5 %	19.8 %

Tab. 2 - Taxa de acertos (# segmentos certos / # total de segmentos).

6. Conclusão

Dos resultados obtidos na Seção 5 e após ouvir-se os sinais criptoanalizados, podemos concluir que o método proposto consegue recuperar a inteligibilidade do sinal cifrado e apresenta um resultado bem melhor que o método da busca da sequência não-péssima, embora não tenha a mesma qualidade da busca exaustiva. A grande vantagem que podemos apontar é a possibilidade de usar a mesma idéia básica, com algumas modificações (uso de *Simulated Annealing* com distribuição de Cauchy tem apresentado expressivos resultados), no caso de termos um número maior de segmentos (16 é um número consideravelmente maior no contexto e encontrado na prática) onde a busca exaustiva não é factível em tempo computacional.

Referências

Apolinário Jr., José Antonio, Criptoanálise de Sinais de Voz Cifrados por Permutação de Segmentos Temporais, Dissertação de Mestrado, UnB, Brasília, Junho de 1993.

Apolinário Jr., José Antonio, Criptoanálise de Sinais de Voz Cifrados por Permutação de Segmentos Temporais Baseada Em Distâncias Espectrais, Anais do 11o Simpósio Brasileiro de Telecomunicações, Natal, 1993.

Cichoki, A. e **Unbehauen**, R., Neural Networks for Optimization and Signal Processing, John Wiley, 1993.

Cowan, C.F.N. e **Grant**, P.M., Adaptive Filters, Englewood Cliffs, Prentice-Hall, 1985.

Haykin, Simon, Neural Networks - A Comprehensive Foundation, Macmillan College Publishing Company, Inc., 1994.

Hopfield, J.J., Neural networks and physical systems with emergent collective computational abilities, in "Procedings of the National Academy of Sciences - USA", vol. 79, abril de 1982, pp. 2554-2558.

Hopfield, J.J., Neurons with graded response have collective computational properties like those of two-state neurons, in "Procedings of the National Academy of Sciences - USA", vol. 81, maio de 1984, pp. 3088-3992.

Markel, J.D. e **Gray** Jr., A.H., Linear Prediction of Speech, Berlin, Springer-Verlag, 1976.

Sridharan, S., **Dawson**, E. e **Goldburg**, B., Fast Fourier Transform Based Speech Encryption System, in "IEE Proceedings-I", vol. 138, n°. 3, junho 1991, pp 215-223.

Wasserman, Phillip D., Neural Computing - Theory and Pratice, Van Nostrand Reinhold, 1989.