

# CLASSIFICAÇÃO DE SINAIS ULTRASSONOROS DE OBJETOS TRIDIMENSIONAIS UTILIZANDO REDES NEURAIAS ARTIFICIAIS

J. R. G. Manzan\* \*\*, S. Nomura\*, E. L. Flores\* \*\*\*, J. B. D. Filho\* e A. C. P. Veiga\*

\*Universidade Federal de Uberlândia, Uberlândia, Brasil

\*\*Instituto Federal do Triângulo Mineiro, Uberaba, Brasil

\*\*\*Universidade Federal de Goiás, Goiânia, Brasil

e-mail: josercardo@iftm.edu.br

**Resumo:** Este trabalho propõe um classificador de sinais sonoros altamente ruidosos baseado em redes neurais artificiais (RNAs). Estes sinais são refletidos de objetos-alvo tridimensionais e capturados por um dispositivo denominado K-Sonar. O K-Sonar emite e recebe sinais ultrassonoros e ainda converte-os em sinais de frequência audível pelos seres humanos. A coleta dos sinais é feita com um aparato experimental constituído do K-Sonar em uma extremidade e do objeto-alvo a ser classificado na outra extremidade. Os formatos dos objetos-alvo tridimensionais são triangular côncavo, triangular convexo, "U" côncavo e "U" convexo. Dois tipos de vetores de características são extraídos dos sinais coletados para servirem como dados de entrada para o classificador. O classificador é baseado em RNAs do tipo multilayer perceptron (MLP) treinadas com vetores-alvo bipolares ortogonais (diferentes dos convencionais). Os resultados preliminares mostraram que o classificador pode atingir taxas significativas de acerto da ordem de 91,7% utilizando sinais sonoros altamente ruidosos. A proposta mostrou-se vantajosa por não exigir um pré-processamento sofisticado dos sinais e assim, reduzir o esforço computacional. E foi incrementada a expectativa de desenvolver interfaces em tempo real para o auxílio a deficientes visuais baseadas na ecolocalização humana.

**Palavras-chave:** Classificação de sinais sonoros, dispositivo de ultrassom, interface para ecolocalização humana, processamento de sinais sonoros, redes neurais artificiais.

**Abstract:** *This paper proposes a classifier of highly noisy sound signals based on artificial neural networks (ANNs). Those signals are reflected from three-dimensional target objects and they are captured by a device known as K-Sonar. The K-Sonar sends and receives ultrasound signals and it also converts them into the audible frequency signals for human beings. The capture of the signals is done by an experimental apparatus constituted by the K-Sonar at one end and the target object to be classified at another end. The 3D*

*target object shapes are concave triangular, convex triangular, concave "U" and convex "U". Two types of feature vectors have been extracted from the captured signal to be used as input data to the classifier. The classifier is based on ANN multilayer perceptron (MLP) type trained with orthogonal bipolar target vectors (different from the conventional ones). The preliminary results showed that the classifier can achieve significant success rates of approximately 91.7% through highly noisy sound signals. The proposal has been advantageous because of avoiding sophisticated signal preprocessing and reducing the computational effort. And the expectation of developing real time interfaces for supporting visually impaired persons based on human echolocation was increased.*

**Keywords:** *Signal sound classification, ultrasound device, interface for human echolocation, sound signal processing, artificial neural networks.*

## Introdução

O sentido humano mais utilizado para a locomoção é a visão. Desta forma o mundo real configura-se em um ambiente repleto de obstáculos para pessoas com deficiência visual. A cada dia cresce o apelo social por mecanismos, procedimentos e equipamentos que permitam integrar essas pessoas ao ambiente que vivemos. São várias áreas de estudos dedicadas ao assunto. Uma das áreas de estudos é denominada ecolocalização [1].

Na natureza, espécies de animais como golfinhos e morcegos [2][3], possuem a habilidade extremamente aguçada de se guiar por meio da emissão e da recepção dos sons emitidos e refletidos sobre diversos objetos. Desta maneira, seres humanos também podem utilizar o sentido da audição como técnica para se locomoverem em seu dia-a-dia. Existe um caso de ecolocalização humana conhecido mundialmente em Los Angeles, do jovem Daniel Kish [4] que perdeu a visão aos 2 anos de idade. Ele estala a sua língua para emitir sons e por meio da audição consegue reconhecer

eficientemente o ambiente que o cerca. Isso permite que ele se locomova como qualquer outra pessoa.

Vale destacar alguns sistemas que têm sido desenvolvidos como suporte para a localização espacial utilizando sinais sonoros e ambientes acústicos virtuais [5][6]. O sistema proposto por Jones [5] utiliza uma câmera para capturar imagens do ambiente e convertê-las em sinais sonoros. Isso torna o sistema bastante complexo em termos de processamento de sinais. Por outro lado, a abordagem das interfaces baseadas em sinais de áudio não verbais [6] fundamenta a proposta deste trabalho.

Diferentemente dos sistemas existentes, o objetivo deste trabalho é obter um sistema que utiliza sinais sonoros que foram capturados sem nenhuma isolação a ruídos e classificá-los usando um classificador de baixo custo e tolerante a ruídos. A expectativa é que sirva de suporte à criação de interfaces em tempo real baseadas nas habilidades de ecolocalização humana que possam auxiliar na interação de deficientes visuais com o ambiente.

**Materiais e métodos**

**Aparato Experimental**– Ao utilizar um aparelho conhecido comercialmente como K-Sonar, os sons podem ser emitidos por ele e os sons refletidos dos objetos podem ser captados por esse mesmo aparelho.

Nos experimentos realizados neste trabalho, conforme o esquema (vista superior) do aparato mostrado na Figura 1, em uma extremidade de uma prancha horizontal fixou-se o dispositivo K-Sonar que pode ser movimentado horizontalmente e a uma distância média de 70 cm do objeto-alvo tridimensional colocado na outra extremidade.

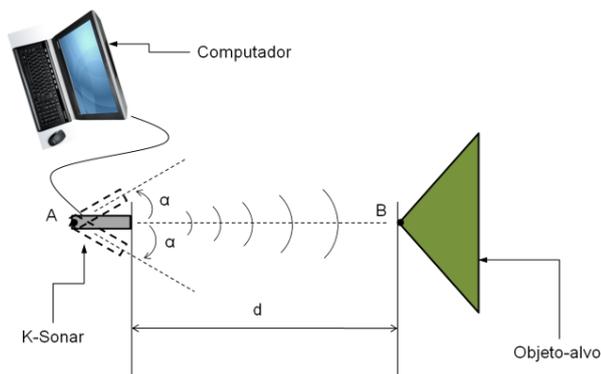


Figura 1: Esquema do aparato experimental.



Figura 2: Aparato experimental montado.

A Figura 2 mostra o aparato experimental montado e preparado para a coleta de dados. Os objetos tridimensionais são denominados da seguinte forma: triangular côncavo (objeto 1), triangular convexo (objeto 2), “U” côncavo (objeto 3) e “U” convexo (objeto 4). A Figura 3 ilustra esses objetos.

**Coleta de sinais**– Para que o sistema a ser desenvolvido seja o mais robusto possível, a coleta dos sinais é realizada sem nenhuma isolação ou proteção a ruídos.



Figura 3: Imagem dos objetos tridimensionais.

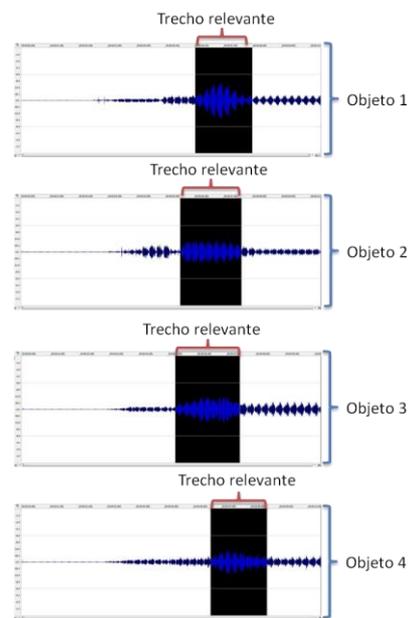


Figura 4: Amostras de sinais referentes à varredura 1.

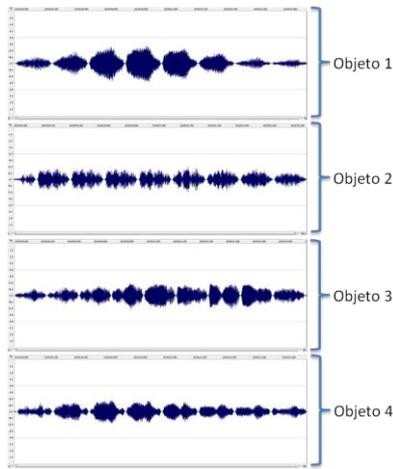


Figura 5: Amostras dos trechos relevantes ampliados para sinais referentes à varredura 1.

A coleta dos sinais correspondentes aos objetos-alvo é realizada mediante a varredura do dispositivo apontado a cada objeto. As varreduras 1 e 2 de meio ciclo são lentas (com duração média de 2 s) da esquerda para a direita e da direita para a esquerda, respectivamente. As rápidas (com duração média de 1 s) da esquerda para a direita e da direita para a esquerda, respectivamente são as varreduras 3 e 4 também de meio ciclo. E as varreduras 5 e 6 são as de ciclo completo (com duração média de 2 s) que começam do lado esquerdo e terminam deste lado e que começam do lado direito e terminam deste lado, respectivamente.

Cada tipo de varredura é repetida 6 vezes. Assim, tem-se para cada objeto, 36 amostras de sinais. O tempo total de coleta de cada amostra é da ordem de 10 s. A Figura 4 apresenta as amostras de sinais coletados pela varredura do tipo 1 sobre os objetos a serem classificados.

**Processamento de sinais**– O primeiro passo é recortar o trecho relevante da amostra que corresponde ao período em que o objeto é varrido pelo K-Sonar. A Figura 5 apresenta os trechos relevantes ampliados referentes às amostras apresentadas na Figura 4. Esse período é dado em função do tipo de varredura feito para cada objeto. A detecção dos tempos inicial e final de cada trecho é realizada visualmente. O passo seguinte é calcular as seguintes medidas em cada trecho relevante recortado: média aritmética, mediana e desvio-padrão dos sinais sonoros; Transformada Discreta de Fourier (DFT), médias, medianas e desvio-padrões dos coeficientes reais e imaginários da DFT dos sinais sonoros; médias do módulo e da fase da DFT e potência dos sinais sonoros.

A média aritmética, a mediana e o desvio-padrão são calculados pelas Equações (1), (2) e (3), respectivamente. Nessas equações  $\bar{x}$ ,  $x_i$  e  $F_i$  representam a média amostral, o  $i$ -ésimo dado amostral e a frequência absoluta para  $i$ -ésimo dado amostral, respectivamente.  $n$ ,  $\tilde{x}$  e  $l_{Md}$  representam o tamanho da amostra, a mediana amostral e o limite inferior da classe mediana, respectivamente.  $\sum f$ ,  $h$ ,  $F_{Md}$  e  $s$

representam o somatório das frequências das classes anteriores à classe mediana, a amplitude da classe mediana, a frequência absoluta da classe mediana e o desvio-padrão amostral, respectivamente.

$$\bar{x} = \frac{\sum x_i \cdot F_i}{n} \quad (1)$$

$$\tilde{x} = l_{Md} + \frac{\left(\frac{n}{2} - \sum f\right) \cdot h}{F_{Md}} \quad (2)$$

$$s = \sqrt{\frac{\sum x_i^2 \cdot F_i - \frac{(\sum x_i \cdot F_i)^2}{n}}{n-1}} \quad (3)$$

As Equações (4), (5), (6), (7) e (8) mostram os cálculos dos coeficientes da transformada discreta de Fourier (DFT) do sinal sonoro, do módulo, da fase e da potência desses coeficientes, respectivamente.

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\left(\frac{2\pi}{N}\right)nk} \quad (4)$$

$$X[k] = X_r[k] + jX_i[k] \quad (5)$$

$$|X[k]| = \sqrt{X_r^2[k] + X_i^2[k]} \quad (6)$$

$$\theta[k] = \arctan\left(\frac{X_i[k]}{X_r[k]}\right) \quad (7)$$

$$P[k] = |X[k]|^2 = X_r^2[k] + X_i^2[k] \quad (8)$$

Nas equações (4) a (8),  $N$ ,  $n$ ,  $X[k]$ ,  $X_r[k]$  e  $X_i[k]$  são o número de amostras, a variável tempo, o coeficiente da DFT, a parte real do coeficiente da DFT e a parte imaginária do coeficiente da DFT, respectivamente.  $|X[k]|$ ,  $\theta[k]$  e  $P[k]$  são o módulo do coeficiente da DFT, a fase da DFT e a potência, respectivamente.

**Reconhecimento de padrões**– No sistema proposto a classificação dos sons dos objetos é realizada utilizando uma rede perceptron de múltiplas camadas (MLP). São realizados dois testes com a MLP. E para a realização destes testes, dois tipos de vetores de características dos sinais sonoros são extraídos para servirem de entrada para o classificador heurístico.

No primeiro teste, utiliza-se um tipo de vetor de características cujos parâmetros referem-se às medidas calculadas pela aplicação das Equações (1) a (5), conforme descritas anteriormente. Este tipo de vetor é identificado como tipo I no teste.

No segundo teste, utiliza-se o vetor de características identificado como tipo II composto pelas medidas calculadas com as Equações (1) a (8). Ressalta-se que a principal diferença deste tipo de vetor em relação ao tipo I é a consideração dos espectros de potência dos sinais.

Os vetores-alvo usados no treinamento são bipolares e ortogonais (OBVs) conforme [7]. Estes vetores sempre têm dimensão equivalente a potência de 2. Como são 4 tipos de padrões a serem classificados,

OBVs de dimensão 4 são suficientes para a classificação deste problema. Contudo, foram testados também o desempenho de classificação para OBVs de dimensões 8, 16 e 32. A escolha de vetores-alvo com dimensões maiores foi feita com o intuito de verificar um possível aumento de desempenho de classificação. Considerando que cada tamanho de vetor-alvo corresponde a um tipo de topologia de rede neural, significa que foram construídas 4 arquiteturas de rede diferentes para fins de análise de desempenho do classificador heurístico.

A formação do conjunto de treinamento contém os vetores de características referentes aos 6 tipos de varredura de cada objeto. Para cada combinação (tipo de varredura x tipo de objeto), foram selecionadas as 4 primeiras repetições (amostras) das 6 disponíveis. Como cada vetor de características possui 9 valores, a matriz de treinamento é composta de 96 linhas e 9 colunas.

A formação do conjunto de teste segue o mesmo princípio da formação do conjunto de treinamento. Contudo, das 6 repetições (amostras) disponíveis, foram selecionadas as duas últimas repetições. Fazendo isso para as 6 varreduras têm-se as 24 amostras para treinamento e 12 para teste como foi mencionado acima.

A rede neural MLP foi treinada utilizando a toolbox `traindx` do software Matlab R2008. Esta toolbox usa os critérios de momentum e taxa de aprendizagem adaptativa. As redes foram treinadas com a taxa de aprendizagem inicial de 0,01. Após diversos testes preliminares, optou-se pelo critério de parada do treinamento por alcance do erro quadrático médio de 0,05. As simulações foram realizadas em um computador com processador i7 3537U e com 8GB de memória RAM.

## Resultados

Considerando-se a geração de pesos sinápticos aleatórios iniciais, são realizados 10 treinamentos e testes para cada uma das 4 arquiteturas de rede: (9 x 100 x 4), (9 x 100 x 8), (9 x 100 x 16) e (9 x 100 x 32). Os números entre parênteses correspondem ao número de neurônios nas camadas de entrada, intermediária e de saída, respectivamente. Quanto à definição do tamanho da camada intermediária, foram feitos vários testes e análises para se concluir que as melhores taxas de acerto ocorrem com 100 neurônios nesta camada.

A Figura 6 mostra para os dois tipos de vetores de características e para cada tamanho do vetor-alvo as médias (dos 10 treinamentos e testes) de desempenho de classificação. Para cada tamanho do vetor-alvo há uma arquitetura correspondente da rede MLP.

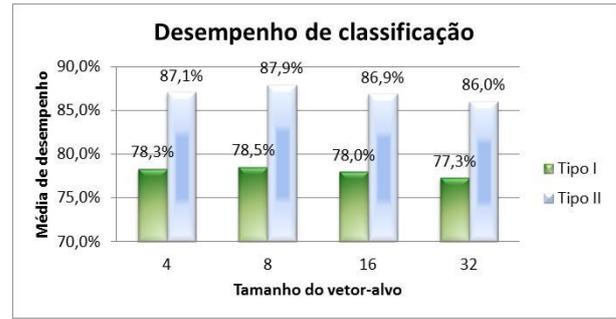


Figura 6: Desempenhos de classificação alcançados por MLPs para dois tipos de vetores de características.

## Discussão

Para os experimentos com o primeiro tipo de vetor de características (tipo I), a maior média das taxas de classificação obtidas com os sinais dos objetos foi de 78,5% conforme a Figura 6. O erro padrão da proporção, calculado pela equação (9), é de 11,6% ao nível de 95% de confiança, o que resulta em um intervalo de confiança da proporção de acerto entre 66,9% e 90,1%.

$$erro = Z \frac{\alpha}{2} \sqrt{\frac{p(1-p)}{n}} \quad (9)$$

Na fórmula da equação (9),  $Z \frac{\alpha}{2}$  é o coeficiente de probabilidade da tabela normal padronizada,  $p$  é a proporção obtida com o conjunto amostral e  $n$  é o número de amostras.

Para os experimentos com o segundo tipo de vetor de características (tipo II), a taxa de classificação dos sinais relacionados aos objetos foi de 87,9% de acordo com a Figura 6. O erro padrão da proporção ao nível de 95% de confiança foi de 9,2%, resultando em um intervalo de confiança para a proporção de acerto entre 78,7% e 97,1%.

Pelos resultados obtidos, as taxas de classificação dos objetos foram melhores com a utilização do vetor de características do tipo II. As médias gerais de desempenho no uso dos vetores de características dos tipos I e II foram de 78,0% e 86,9%, respectivamente. Observa-se que a consideração dos espectros relacionados à transformada discreta de Fourier, proporcionou um aumento médio no desempenho de classificação da ordem de 8,9%. Isso evidencia a sua importância na descrição das características dos sinais dos objetos neste caso.

## Conclusão

Este trabalho propôs um classificador heurístico de sinais ultrassonoros correspondentes a objetos tridimensionais. Experimentos realizados com o referido classificador levaram a desempenhos significativos.

Os resultados mostraram que o sistema de classificação é bastante robusto e pouco suscetível à ação dos ruídos. Conclui-se isso da consideração de que não foi utilizada nenhuma proteção contra ruídos na etapa de coleta dos sinais. Também da consideração de que não foi realizado nenhum pré-processamento de sinais para a remoção dos ruídos.

Além disso, com um vetor-alvo de tamanho reduzido de apenas 4 elementos se obteve um desempenho próximo ao do vetor com o dobro do tamanho de 8 elementos. Isso é uma grande vantagem em termos de redução de custo computacional para implementação de sistemas heurísticos de tempo real.

Com este trabalho, aumentou-se a expectativa de desenvolver interfaces baseadas nas técnicas de ecolocalização humana que possam fornecer suporte (acessibilidade) aos deficientes visuais para que eles possam interagir com o ambiente em que vivem.

### **Agradecimentos**

Agradecimento à FAPEMIG pelo apoio financeiro através do convênio TEC – APQ01344-10.

### **Referências**

- [1] Griffin D. Echoes of bats and men. New York: Anchor Books Doubleday & Co; 1959.
- [2] Stebbins W. The acoustic sense of animals. Harvard University Press; 1983.
- [3] Schnitzler HU, Henson OW. Performance of airborne animal sonar systems: I. Microchiroptera. Em: R.G. Busnel, J.F. Fish, editors. Animal Sonar Systems. New York: Plenum Press; 1980. p. 109-181.
- [4] Kish D. Evaluation of an echo-mobility training program for young blind people [thesis]. San Bernardino: California State University; 1995.
- [5] Jones WD. Sight for sore ears. IEEE Spectrum. 2004; 41(2):13-14.
- [6] Nomura S, Shiose T, Kawakami H, Katai O. Nonspeech audio-based interfaces. Encyclopedia of Information Science and Technology. 2009; 6(2): 2840-2849.
- [7] Manzan JRG, Nomura S, Yamanaka K, Carneiro MBP, Veiga ACP. Improving iris recognition through new target vectors in MLP artificial neural networks. In: Proceedings of Lecture Notes in Computer Science – Artificial Neural Networks in Pattern Recognition; 2012 Sep; Trento, Italy. 2012. p. 115-126.