

## LOCALIZAÇÃO DE FONTES SONORAS COM TOLERÂNCIA A AMBIENTES REVERBERANTES

José Vieira, José Fonte, Daniel Gonçalves, Luís Almeida

vieira@det.ua.pt, etjfonte@ua.pt, a14632@alunos.det.ua.pt, lda@det.ua.pt

DET/IEETA - Universidade de Aveiro

Portugal

Resumo: A construção de um sistema artificial capaz de realizar a tarefa de localização de fontes sonoras assume no contexto da robótica móvel uma importância que uma análise sumária não deixa antever. Na interacção com seres humanos, em sistemas de televigilância, em sistemas de reconhecimento de orador e de reconhecimento de fala, a correcta localização do orador constitui uma mais valia importante, por exemplo, para orientar correctamente um sistema de visão.

Neste artigo descreve-se um sistema de localização de fontes sonoras utilizando apenas dois sensores acústicos. O sistema proposto implementa uma versão do algoritmo proposto em (Huang *etal*, 1999a) que inspirando-se no efeito de precedência do sistema auditivo Humano consegue realizar de forma robusta a localização em ambientes reverberantes.

Palavras Chave: Localização de fontes sonoras, processamento de sinal, robótica, navegação.

### 1. INTRODUÇÃO

As câmaras de vídeo são, hoje em dia, sensores comumente utilizados em robótica para identificação do ambiente circundante de um robot, para navegação, para determinação de movimento, etc. No entanto, as câmaras de vídeo têm tipicamente um campo visual limitado, o que pode introduzir dificuldades por não permitir detectar eventos relevantes fora desse campo. Contudo, na maior parte dos casos, os objectos de interesse num determinado cenário estão associados a som ou ruído próprio que, se for correctamente localizado, permite orientar a câmara para a zona de interesse.

Neste artigo descreve-se um sistema de localização de fontes sonoras que utiliza apenas dois microfones e que se baseia na medição da chamada Diferença de Tempo Interaural ou DTI. O sistema é ainda capaz de realizar a tarefa de localização em ambientes reverberantes utilizando um processo inspirado no efeito de precedência (Huang *etal*, 1999a). O algoritmo de detecção referido foi utilizado num pequeno robot que serve de demonstrador do funcionamento da solução proposta. Mais informações sobre este trabalho, realizado no âmbito de um projecto de fim de curso no DET, podem ser encontradas em:

<http://www.ieeta.pt/~vieira/proj5ano/2001-02/goodears/index.html>

A próxima secção faz uma breve revisão sobre o problema da localização de fontes sonoras com especial incidência nos sistemas auditivos de animais e humanos. Na secção 3 aborda-se de forma breve a localização de fontes sonoras utilizando a DTI. Na secção 4 descreve-se o algoritmo utilizado para implementar o efeito de precedência de modo a resolver o problema de localização em ambientes reverberantes. A secção 5 descreve muito sucintamente o robot construído para servir de demonstrador, sendo alguns resultados experimentais apresentados na secção 6. A secção 7 refere as conclusões deste artigo.

### 2. ENQUADRAMENTO

A capacidade de localizar fontes sonoras reveste-se de grande importância para a maior parte dos mamíferos. Estudos sobre a evolução animal revelaram que a gama de frequências audível está relacionada com a necessidade de localização de fontes sonoras e que a acuidade da detecção sonora está relacionada com a dimensão do campo de visão central que possui a maior acuidade (Heffner and Heffner, 1992). Além disso, a rápida e correcta detecção da direcção e distância de uma fonte sonora constitui uma vantagem competitiva importante para sobrevivência.

O sistema auditivo humano é capaz de detectar o azimute e a elevação de uma fonte sonora, mesmo que

esteja colocada atrás do ouvinte e não perde grande acuidade em ambientes reverberantes (Hartmann, 1999; Blauert, 1996). Esta notável capacidade, e apesar de possuímos apenas dois sensores acústicos, é conseguida com recurso a um elaborado processamento por parte do nosso sistema nervoso, dos sinais acústicos captados. Para determinação do azimute de uma fonte sonora o sistema auditivo humano utiliza duas técnicas distintas (Blauert, 1996): Diferença de Nível Interaural (DNI) e a Diferença de Tempo Interaural (DTI). O primeiro método é utilizado para frequências com um comprimento de onda inferior ao diâmetro da cabeça (cerca de 2kHz) e tira partido do efeito de “sombra” provocado por esta sobre o ouvido na posição oposta à fonte sonora. O segundo método baseia-se na medição do atraso relativo com que um dado sinal chega a cada um dos ouvidos e é utilizado pelo sistema auditivo para sinais com um comprimento de onda superior ao diâmetro da cabeça. Estes dois métodos, por si só, não permitem explicar a robustez do sistema auditivo humano na detecção da direcção de uma fonte sonora em ambientes reverberantes. Esta capacidade é explicada recorrendo ao efeito de precedência conhecido também pela “lei da primeira frente de onda” e que foi verificado experimentalmente pela primeira vez pelo Físico de Princeton Joseph Henry (1797-1878) (Schroeder 1999). Henry verificou que quando dois sons similares, ligeiramente desfasados e provenientes de duas direcções diferentes chegam a um ouvinte, estes são percebidos como um único som proveniente da direcção do primeiro a chegar.

### 3. LOCALIZAÇÃO DE FONTES SONORAS UTILIZANDO A DTI

Considere-se um sistema de medição da direcção de uma fonte sonora utilizando apenas dois microfones tal como se ilustra na figura 1. Se a fonte sonora se encontrar a uma distância  $l$  muito superior à distância  $d$  entre os dois micros pode-se considerar que as ondas acústicas junto dos micros são planas. Se a fonte sonora estiver localizada a uma distância  $d_E$  e  $d_D$  do microfone esquerdo e direito respectivamente, então, a partir da diferença temporal  $\Delta t$  na chegada da frente sonora podemos determinar o ângulo  $\theta$  a partir da expressão (1) onde  $v_c$  é a velocidade do som e  $d$  a distância entre os microfones.

$$\sin(\theta) = \frac{d_E - d_D}{d} = \frac{v_c \Delta t}{d}, \quad (1)$$

Uma das formas de determinar o atraso relativo  $\Delta t$  entre os sinais emitidos por uma fonte sonora, consiste em realizar a correlação cruzada (2) entre os sinais do canal esquerdo e direito com  $K$  amostras.

$$R(n) = \sum_{k=0}^{K-1} x_E(k)x_D(k+n) \quad n \in [-N \dots N] \quad (2)$$

Se considerarmos que os sinais captados pelos microfones são amostrados com uma frequência de amostragem  $f_s$ , o número máximo de amostras que é

necessário calcular na correlação será dado por  $N$  de acordo com a equação (3).

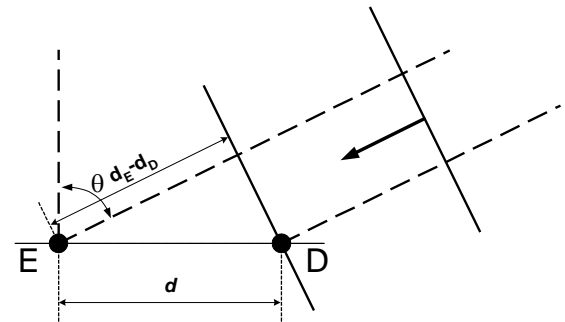


Fig. 1. Geometria do problema de determinação da direcção do som. Os microfones estão colocados nas posições E e D a uma distância  $d$  um do outro.

$$N = f_s d / v_c \quad (3)$$

Na figura 2 podemos observar a correlação para uma fonte sonora localizada a  $33^\circ$  do sistema, utilizando uma frequência de amostragem de 8 KHz e com uma distância entre micros de 33cm. O pico de correlação ocorre para a amostra  $-4$  da correlação o que pela equação (1) dá um ângulo de

$$\theta = \frac{v_c n}{f_s d} = \frac{340 \times 4}{8000 \times 0.33} \approx 31^\circ$$

Uma forma alternativa de determinar a DTI consiste em dividir o sinal dos dois canais em blocos de  $M$  amostras, calcular a DFT de cada um dos canais e medir a diferença de fase das componentes espectrais de mais baixa frequência para evitar as ambiguidades da DFT (Huang *et al*, 1999a). A utilização deste método sai fora do âmbito deste artigo e será objecto de trabalho futuro.

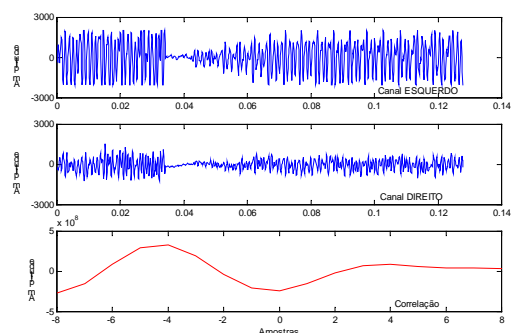


Fig. 2. Exemplo da correlação (em baixo) entre o canal esquerdo e direito.

### 4. LOCALIZAÇÃO EM AMBIENTES REVERBERANTES

Qualquer um dos métodos descritos na secção anterior para realizar a localização de fontes sonoras, falha quando é utilizado em ambientes reverberantes. Esta limitação é particularmente relevante dado que a maior parte dos ambientes de operação reais geram

reverberação. No entanto, conforme referido na Secção 2, é possível detectar fontes sonoras na presença de reverberação utilizando o efeito de precedência. Apesar deste efeito ser um tema bastante estudado (Jeffress, 1948; Gaik, 1993; Lindemann, 1986; Bodden, 1995; Schauer *et al*, 2000; Hofman and Opstal, 1997, 1998), não existem muitos trabalhos sobre modelos artificiais para o simular. Os trabalhos de Huang *et al* (1999, 1999a, 2001) propõe um modelo computacional baseado no estudo psicoacústico do efeito de precedência. Neste efeito ocorre uma inibição da capacidade de localizar fontes sonoras imediatamente a seguir à chegada de uma frente de onda. Esta característica permite evitar que o sistema de localização funcione com sinais acústicos onde a reverberação predomina. Verifica-se igualmente que não existe um nível absoluto para que este efeito ocorra. Além disso, quando uma frente de onda de pequena amplitude é seguida de uma outra de maior amplitude, a inibição é cancelada e o sistema de localização volta a funcionar seguindo-se a inibição (Huang *et al*, 1999a).

O modelo utilizado baseia-se numa modelização dinâmica do nível de reverberação existente numa sala devido aos sinais ocorridos anteriormente. Considera-se neste modelo que o decaimento da reverberação é exponencial. A equação (4) modeliza a resposta a impulso de uma sala. Entre a chegada da frente de onda, e o início da reverberação, decorre um tempo  $T$ . Esta porção de sinal será a utilizada para efeitos de correlação, uma vez que se encontra livre de reverberação. Após este período, considera-se no modelo, que a reverberação terá um decaimento exponencial com uma constante de tempo  $\tau$ .

$$h_p(t) = \begin{cases} 0 & 0 \leq t \leq T \\ g_i e^{(t-T)/\tau} & t > T \end{cases} \quad (4)$$

Na figura 3 podemos observar a resposta impulsional deste modelo para uma sequência de impulsos. Pode-se verificar que os impulsos (traços verticais) dão origem ao aparecimento de reverberação.

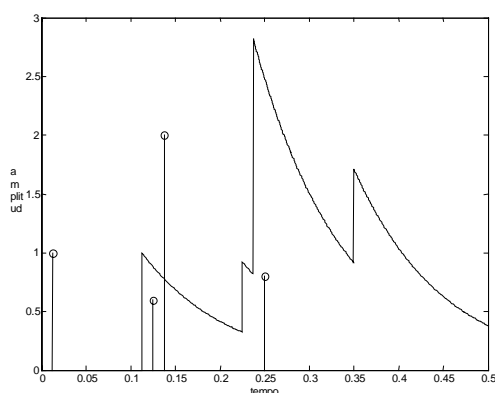


Fig. 3. Simulação da resposta a impulso do modelo dado pela equação (4). São aplicados 4 impulsos assinalados pelos traços verticais.

No robot descrito na secção 5 este modelo é utilizado para estimar a reverberação presente no

ambiente. Quando a relação entre o sinal de entrada e a estimativa da amplitude da reverberação ultrapassa um dado limiar, considera-se que se está na presença de uma nova frente de onda, designado por *onset* (Huang *et al*, 1999), e que o sinal em torno desse ponto se encontra “livre” de reverberação.

A figura 4 apresenta um diagrama de blocos do sistema de localização com estimação e inibição da reverberação. O funcionamento deste sistema é ilustrado mais adiante, na secção 6 – figura 7, onde podemos observar a detecção de *onsets* a funcionar e as porções de sinal utilizadas no cálculo da correlação.

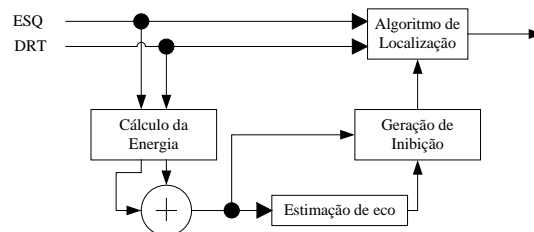


Fig. 4. Diagrama de blocos que descreve o sistema da geração da inibição.

## 5. ROBOT LOCALIZADOR DE FONTES SONORAS

Para demonstrar a capacidade de localização de fontes sonoras do sistema, construiu-se um pequeno robot autónomo e móvel, utilizando como unidade de processamento dos sinais sonoros e controlo da navegação o DSPC31 da Texas Instruments. Na figura 5 podemos observar um diagrama com os principais blocos constituintes do sistema.

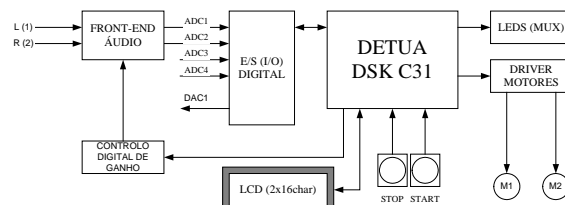


Fig. 5. Diagrama de blocos do sistema completo do robot.

Todo o hardware foi montado numa pequena base circular de madeira, equipada com duas rodas motrizes independentes, que permitem ao robot perseguir uma fonte sonora. Na figura 6 podemos observar o protótipo desenvolvido.

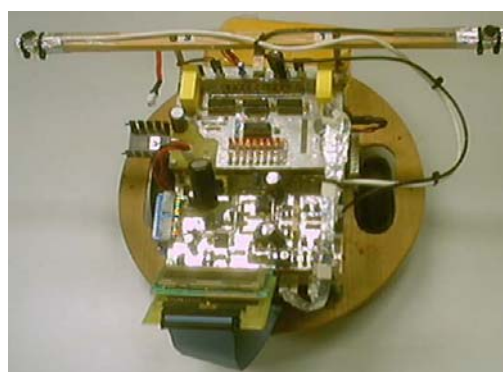


Fig. 6. Aspecto do robot onde se podem ver os dois microfones (cantos superiores da fotografia).

O sistema de aquisição foi desenvolvido com 4 canais para permitir futuros testes de algoritmos de localização com mais de dois microfones.

O algoritmo de navegação utilizado é muito simples e consiste apenas na seguinte sequência: escuta parado, roda para alinhar com a fonte e avança durante um certo tempo. Esta estratégia foi adoptada para evitar a interferência do ruído dos motores.

## 6. RESULTADOS EXPERIMENTAIS

Para validar o algoritmo de localização foram adquiridos sinais em estéreo em vários tipos de salas com dimensões e tempos de reverberação diferentes. A sequência de gravação foi: um bater de palmas, a vogal “a” sustentada, um “pssst” e a frase “experiência de som”, a uma distância de um metro dos microfones. Na figura 7 pode ver-se o resultado obtido num dos vários locais (escadas com um tempo de reverberação elevado). Podemos observar a detecção dos *onsets*, o sinal utilizado na correlação e o ângulo medido.

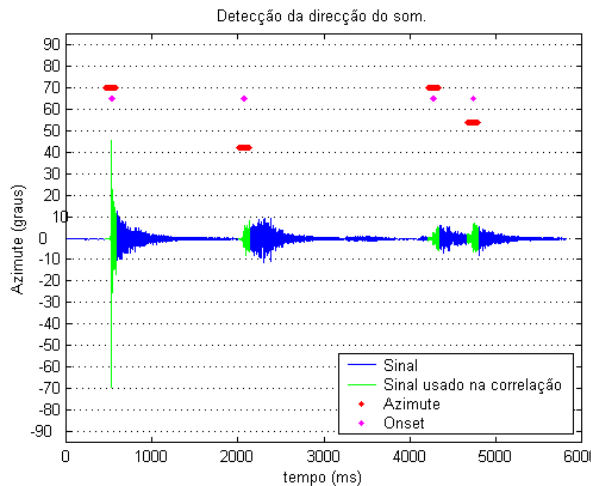


Fig. 7. Determinação dos *onsets* num local com bastante reverberação e com a fonte sonora colocada 60° graus à esquerda.

Nestes ensaios a fonte sonora foi deslocada em torno do sistema de medição e foram realizadas várias medidas.

Tabela 1 Resultados dos ensaios de medição da direção da fonte sonora

$\theta_{\text{emissão}}$	$\theta_{\text{medido}}$	Erro( $\theta$ )
-90° R	-90° R	0°
-60° R	-70° R	10°
-45° R	-54° R	9°
-30° R	-33° R	3°
-15° R	-16° R	1°
0°	0°	0°
15° L	16° L	1°
30° L	33° L	3°
45° L	42° L	3°
60° L	70° L	10°
90° L	90° L	0°

Na tabela 1 apresentam-se os resultados obtidos, onde se apresenta o ângulo da fonte, o ângulo medido e o respectivo erro. Para a frequência de amostragem utilizada de 8kHz os ângulos possíveis de detectar são os seguintes: 0°, ±8°, ±16°, ±24°, ±33°, ±42°, ±54°, ±70°, ±90°.

Apesar dos resultados serem satisfatórios e o sistema ser capaz de localizar as fontes sonoras mesmo em ambientes reverberantes, verificou-se que a performance pode variar bastante com as condições acústicas do ambiente onde é realizada a detecção. Para um melhor desempenho é importante efectuar uma correcta estimativa dos parâmetros da acústica da sala onde é realizada a detecção.

Contudo, note-se que a zona de menor erro é precisamente na direcção frontal. Em sistemas de orientação para a fonte sonora, como o robot construído, os erros verificados são pouco significativos. De facto, se por erro, o sistema não se orienta exactamente para a fonte sonora com um só movimento acabará por fazê-lo no movimento seguinte.

## 7. CONCLUSÕES

Este artigo discute o interesse da localização de fontes sonoras bem como a sua aplicação na área da robótica. São também brevemente analisados os mecanismos de localização de fontes sonoras dos mamíferos, em particular no que diz respeito à tolerância a ambientes reverberantes. Esta tolerância é conseguida com o chamado efeito de precedência que permite discriminar as frentes de onda originais dos respectivos ecos.

O cerne do artigo é a construção de um sistema artificial de localização de fontes sonoras no plano azimutal, com tolerância a ambientes reverberantes. Este sistema, baseado na medição da Diferença de Tempo Interaural (DTI) conjugada com o efeito de precedência, é utilizado para guiamento de um robot autónomo e móvel.

Os resultados experimentais obtidos são bastante encorajadores com detecção aproximadamente correcta da direcção do som para sons diversos e em ambientes fortemente reverberados. Embora se tenha registado um erro máximo de cerca de 12% nos azimutes em torno de +/- 70°, esse erro não afecta sensivelmente a capacidade de orientação para a fonte sonora e deve-se em parte à resolução na detecção do ângulo.

Ainda assim, para melhorar o desempenho do sistema em ambientes dinâmicos e diversificados, estão em estudo alguns métodos para realizar a estimação dos respectivos parâmetros acústicos sem recurso a sinais de estímulo, utilizando apenas os sons captados.

## REFERÊNCIAS

- Blauert, Jens, (1996), *Spatial Hearing: The Psychophysics of Human Sound Localization*, Mit Press.
- Bodden, Markus, (1995), *Binaural Modeling and Auditory Scene Analysis*, IEEE 1995 Workshop on

- Applications of Signal Processing to Audio and Acoustics, IEEE, Mohonk Mountain House, New Paltz, New York.
- Gaik, Werner, (1993), *Combined Evaluation of Interaural Time and Intensity Differences: Psychoacoustic Results and Computer Modeling*, Journal of Acoustical Society of America, **94**, N.1, pp.98-110, Jul.
- Hartmann, William M., (1999), *How We Localize Sound*, Physics Today, pp.24-29, Nov.
- Heffner, R. S., and Heffner, Henry E. (1992), *The Evolutionary Biology of Hearing*, Springer-Verlag, Fay, R. R., Webster, D. B., and Popper, A. N.
- Hofman, Paul, and Opstal, John van, (1997), *Identification of Spectral Features as Sound Localization Cues in the External Ear Acoustics*, Proceedings IWANN'97, Heidelberg: Springer, Mira-Mira, J.J., Moreno-Diaz, R., and Cabestany, J., pp.1126-1135.
- Hofman, Paul M., and Opstal, John Van, (1998), *Spectro-temporal Factors in Two-dimensional Human Sound Localization*, Journal of Acoustical Society of America, **103**, N.5, pp.2634-2648, May.
- Huang, J., Supaongprapa, et al.. (1999), *A Model Based Sound Localization Systems and Its Application to Robot Navigation*, Robotics and Autonomous Systems (Elsevier Science), **27**, N.4, pp.199-209.
- Huang, J., Ohnishi, N., Guo, X., and, Sugie, N. (1999a), *Echo Avoidance in a Computational Model of the Precedence Effect*, Speech Communication (Elsevier Science), **27**, N.3-4, pp.223-233.
- Huang, Yiteng, and Benesty, Jacob, and Elko, Gary W., and Mersereau, Russel M. (2001), *Real-Time Passive Source Localization: A Practical Linear-Correction Least-Squares Approach*, IEEE Transactions on Speech and Audio Processing, **9**, N.8, pp.943-955, Nov.
- Jeffress, Lloyd A., (1948), *A place Theory of Sound Localization*, Journal of Comparative and Physiological Psychology, **61**, pp.468-486.
- Lindemann, W., (1986), *Extension of a Binaural Cross-Correlation Model by means of Contralateral Inhibition, II: The Law of the first Wave Front*, Journal of Acoustical Society of America, **80**, N.6, pp.1623-1630, Dec.
- Schauer, C., Zahn, T., Paschke, P., and Gross, H.-M., (2000), *Binaural Sound Localization in a Artificial Neural Network*, ICASSP 2000, pp.865-868, IEEE, Instambul, Turkey, Jun.
- Schroeder, Manfred R., (1999), *Computer Speech - Recognition, Compression, Synthesis*, Springer, Information Sciences, Berlin.