

# LOCALIZADOR DE FONTE SONORA PARA VIDEOCONFERÊNCIA

<sup>1</sup>Ricardo A. Andriani, <sup>2</sup>Miguel A. Sovierzoski

Centro Federal de Educação Tecnológica do Paraná – CEFET-PR

<sup>2</sup>Departamento Acadêmico de Eletrônica

<sup>1</sup>Curso de Engenharia Industrial Elétrica – ênfase Eletrônica e Telecomunicações

Av. Sete de Setembro, 3165

CEP 80230-901 Curitiba, PR – Brasil

<http://www.cefetpr.br>

[miguelp@cefetpr.br](mailto:miguelp@cefetpr.br)

**Resumo** – Neste trabalho é discutida e apresentada a implementação de um sistema de automatização do posicionamento de câmera para ambientes de videoconferência. O sistema tem seu funcionamento baseado na localização do participante ativo através da sua voz, que é captada por um par de microfones e analisada. É então estimada uma localização e esta informação é utilizada para movimentar a câmera na direção do participante em questão. A implementação apresentada visa baixo custo e menor complexidade que outros sistemas já existentes, facilitando sua futura comercialização.

**Abstract** - In this work the implementation of a system of automatization of the position of camera for teleconferencing environments is discussed and presented. The system has its functioning based on the location of the active participant through its voice, that is received by a pair of microphones and analyzed. Then is estimated a location and this information is used to put into motion the camera in the route of the participant in question. The implementation discussed here aims at to have a low cost and minor complexity that other existing systems, facilitating its future commercialization.

## I. INTRODUÇÃO

Um ponto fundamental para o sucesso de sistemas automatizados é a sua interação com os humanos. Um caso em que destaca-se esta necessidade é o ambiente de videoconferência, onde câmeras e microfones devem ser direcionados à pessoa que está discursando no momento. Este direcionamento pode ser executado de forma manual, entretanto, isto representa um inconveniente aos participantes que necessitam movimentar-se constantemente para posicionar a câmera ou permanecer em uma determinada posição a fim de não sair do campo visual da câmera, limitando a sua capacidade de expressão. Um sistema de direcionamento automático de câmera baseado no posicionamento do locutor dentro do ambiente torna-se

então bastante conveniente nesta situação. Este sistema deve cobrir um campo de 180 graus e ter sensibilidade suficiente para um alcance de no mínimo 5 metros. Diversas técnicas foram usadas para localizar a origem do som, algumas baseadas na quantidade de energia dos sinais, outras em análise espectral [1], entretanto a implementação do localizador de fonte sonora abordada neste trabalho baseia-se na técnica denominada *Time Delay of Arrival* (TDOA) [1], que consiste em determinar o atraso na chegada do som em um par de microfones. O localizador de fonte sonora aqui abordado visa automatizar a movimentação horizontal de uma câmera, utilizando uma implementação simples, de menor custo e complexidade, com dois microfones adequadamente posicionados.

## II. PROPOSTA DO SISTEMA

O sistema localizador de fonte sonora visa atender às necessidades de um ambiente de videoconferência, sendo necessária fundamentalmente a movimentação horizontal de uma câmera acompanhando a movimentação dos participantes na sala. Visou-se implementar este sistema com um baixo custo e complexidade moderada, utilizando-se ferramentas e componentes disponíveis no mercado. O campo de movimentação da câmera pretendido é de 180°, que foi considerado suficiente para cobrir todas as áreas de uma sala com o sistema corretamente posicionado.

Além da movimentação automática, existe a possibilidade da movimentação manual controlada através de uma interface em um computador, proporcionando maior flexibilidade aos participantes em momentos onde é necessário um ajuste manual da posição da câmera mantendo-a fixa nesta posição.

## III. MÉTODO MATEMÁTICO

O sistema localizador de fonte sonora baseia-se nas características de propagação do som para determinar a origem do mesmo. A posição da câmera é determinada pela posição estimada da origem do som.

Para obter-se a posição estimada de origem do som, utiliza-se dois microfones estrategicamente posicionados com distância determinada entre eles e calcula-se a diferença de tempo entre os sinais que chegam aos microfones. Além da diferença de tempo sabe-se também em qual microfone o sinal chega primeiro, definindo assim se a movimentação da câmera deve ser para a direita ou para a esquerda. Considerando-se a velocidade de propagação do som de 331 m/s pode-se estimar a distância entre a fonte sonora e cada um dos microfones, conseguindo-se assim a informação necessária para o correto posicionamento da câmera.

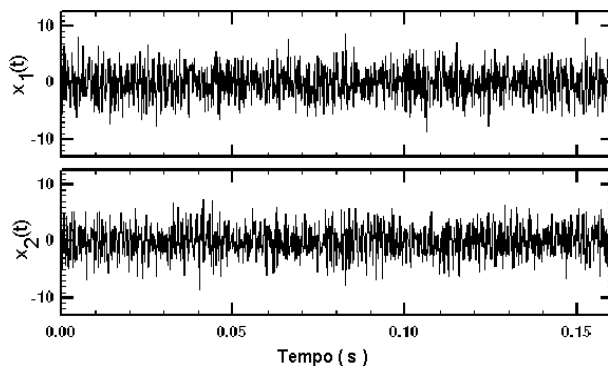
Para determinar a diferença de tempo de chegada dos sinais nos dois microfones necessita-se de um sistema de processamento de alto desempenho e também de cálculos matemáticos que permitam expressar essa diferença em unidades conhecidas. Sendo assim, foi utilizada a técnica de correlação cruzada [3] para determinar a posição, processada por um sistema baseado em DSP (processador digital de sinais).

A fim de ilustrar esta técnica, considere-se dois sinais  $p_1(t)$  e  $p_2(t)$  ambos formados a partir de uma fonte sonora pontual  $s(t)$ . Os sinais estão atrasados diferentemente em relação à fonte, sendo que  $p_1(t)$  está atrasado 25ms em relação a  $s(t)$  e  $p_2(t)$  está atrasado 14ms em relação a  $s(t)$ . Para representar a dispersão e outros efeitos, dois sinais não relacionados também são considerados, cada um com aproximadamente 20% da energia do sinal original, representados por  $n_1(t)$  e  $n_2(t)$ . Tem-se então os dois sinais que chegam aos microfones, dados pelas equações (1) e (2).

$$x_1(t) = p_1(t) + n_1(t) \quad (1)$$

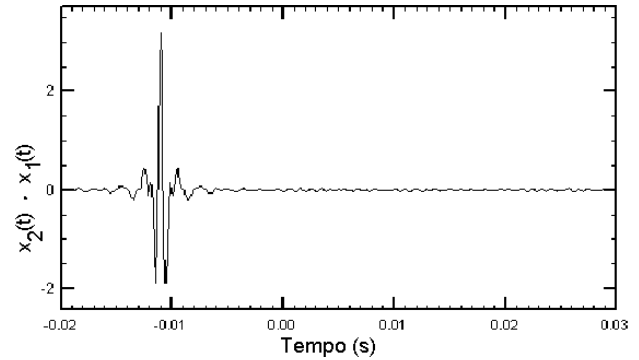
$$x_2(t) = p_2(t) + n_2(t) \quad (2)$$

A partir dos sinais individuais captados em cada um dos microfones, representados na *Figura 1*, será utilizado o cálculo de correlação para obter-se a informação temporal de chegada dos sinais.



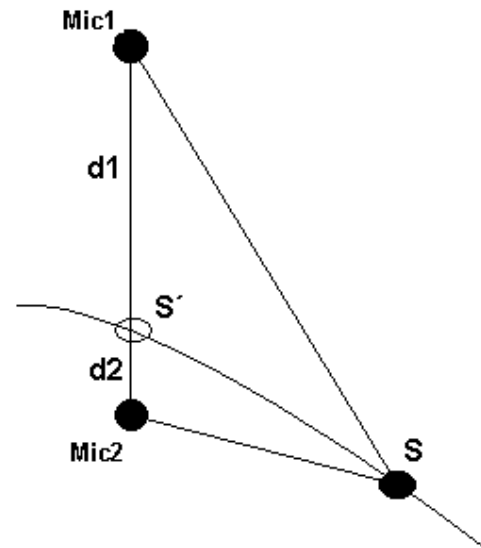
**Figura 1 – Sinais Captados nos Microfones**

Correlacionando o sinal  $x_2(t)$  tendo  $x_1(t)$  como referência obtêm-se um alto nível de correlação em  $-11\text{ms}$  como pode ser visto na *Figura 2*, indicando que o sinal  $x_2(t)$  está adiantado 11ms em relação ao sinal  $x_1(t)$ .



**Figura 2 – Correlação de  $x_2(t)$  em relação a  $x_1(t)$**

Considerando-se que os microfones encontram-se 11,89 m distantes um do outro e que um atraso de 11ms representa 3,64 m, quando considerar a velocidade do som de 331 m/s, pode-se estimar a distância da fonte ao microfone 1, distância  $d_1$ , e a distância da fonte ao microfone 2, distância  $d_2$ , através das *Equações* (3) e (4). É conveniente destacar que o sinal na informação temporal obtida da correlação indica qual sinal está atrasado em relação ao outro, fornecendo o quadrante de origem do som. Esta informação também foi considerada na formação da *Equação* (4), além da geometria representada pela *Figura 3*.



**Figura 3 – Diagrama de Propagação do Som da Fonte aos Microfones, onde S é a fonte sonora, Mic1 e Mic2 são os microfones e a curva S-S' representa as possíveis posições da fonte sonora S.**

A informação necessária para a movimentação horizontal da câmera é a direção de origem do som. Esta informação está representada na *Figura 3* pela curva que passa por S e S', que são duas possíveis origens do som. O ponto desta curva que cruza a linha reta que passa pelos dois microfones fornece a informação de direção aproximada, suficiente para o correto posicionamento da câmera. Supondo que a origem do som fosse exatamente o ponto S' tem-se que:

$$d_1 + d_2 = 11,89 \text{ m} \quad (3)$$

e que

$$d_2 - d_1 = -3,64 \text{ m} \quad (4)$$

Resolvendo o sistema de equações tem-se que  $d_1 = 7,765 \text{ m}$  e  $d_2 = 4,125 \text{ m}$ .

A partir das distâncias dos microfones pode-se obter a informação angular para o direcionamento da câmera. Se  $d_1$  for zero pode-se dizer que a fonte está alinhada com o eixo que passa pelos microfones e que está no quadrante onde está posicionado o microfone 1. Já se  $d_2$  for zero então a fonte também está alinhada ao eixo mas está no quadrante onde está posicionado o microfone 2. Se  $d_1$  for igual a  $d_2$  então a fonte está posicionada perpendicularmente ao eixo dos microfones, indicando que a câmera deverá movimentar-se para sua posição central. Em posições intermediárias a proporção entre  $d_1$  e  $d_2$  será convertida em um ângulo de zero a noventa graus que será utilizado para definir a posição de câmera.

Considerando o mesmo exemplo, correlacionar  $x_1(t)$  tendo  $x_2(t)$  como referência obtêm-se um pico na correlação cruzada em +11ms como pode ser visto na *Figura 4*.

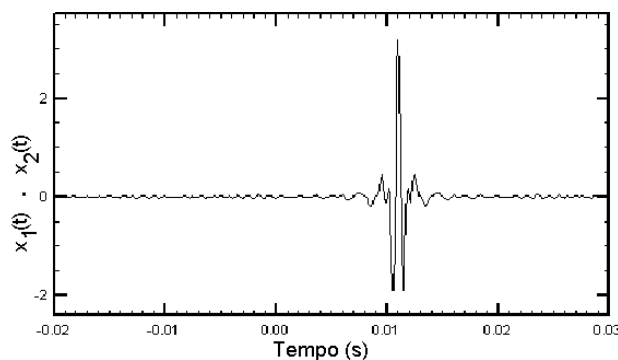


Figura 4 – Correlação de  $x_1$  em relação a  $x_2$

Neste caso chega-se à equação (5), que leva às mesmas equações e aos mesmos resultados obtidos da correlação de  $x_2$  em relação a  $x_1$ . Verifica-se então que

tomando-se qualquer um dos dois sinais como referência pode-se chegar aos resultados corretos.

$$d_1 - d_2 = 3,64 \text{ m} \quad (5)$$

#### IV. IMPLEMENTAÇÃO

O sistema localizador de fonte sonora implementado consiste em uma câmera de vídeo com suporte giratório controlado por um motor de passo, um módulo de controle que utiliza um DSP (processador digital de sinais) e um par de microfones para captar os sinais, conforme representado na *Figura 5*. Opcionalmente um computador pode também fazer parte do sistema para permitir o controle manual à distância do posicionamento da câmera de vídeo.

Os sinais recebidos pelos microfones são amostrados por conversores analógico-digitais controlados por um kit de desenvolvimento do DSP TMS320C5402 da Texas Instruments. Estes sinais amostrados são analisados obtendo-se o ponto estimado de origem do som e então sinais de controle são enviados ao suporte giratório da câmera posicionando-a corretamente. O módulo de controle também recebe comandos do computador via interface serial, interrompendo o cálculo de posições se recebido um comando de posicionamento manual e reiniciando o cálculo se recebido o comando de posicionamento automático.

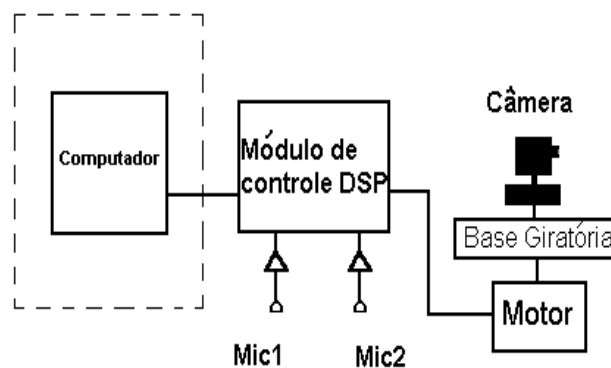


Figura 5 – Principais Componentes do Sistema

Nesta implementação foi escolhida uma distância entre os microfones de 30 cm. Este valor permite uma precisão suficiente de posições de câmera e não torna o sistema mecanicamente desajeitado, com um tamanho que dificulte sua manipulação e instalação. Os microfones foram dispostos nos extremos de uma régua, como apresentado pela *Figura 6*.

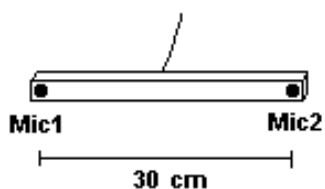


Figura 6 – Disposição dos Microfones

O módulo de controle captura os sinais dos microfones a uma taxa de amostragem de 16 kHz e com uma resolução de 15 bits, características estas que devem-se à escolha da utilização de conversores A/D compatíveis com o kit de desenvolvimento DSP, destinados à aquisição de áudio via microfones de eletreto. Blocos de 256 amostras dos sinais de cada um dos microfones são analisados em conjunto e correlacionados, a fim de se obter a posição rapidamente. A seguir a posição encontrada é relacionada com o respectivo posicionamento de câmera necessário para cobrir aquela posição. Considerando a posição atual da câmera sabe-se então qual movimentação deve ser executada na câmera.

Considerando a distância de 30 cm entre os microfones, a taxa de amostragem de 16 kHz e a velocidade do som de 331 m/s, temos uma precisão de 14,5 amostras para a direita da câmera e 14,5 amostras para a esquerda da câmera, como mostra a Equação (6). Existem então 28 posições diferentes de câmera, tendo em vista que podemos ter o sinal do microfone 1 atrasado em relação ao microfone 2 ou ter o sinal do microfone 2 atrasado em relação ao microfone 1.

$$d_{mic}/v_{som} * f_s = 0,3/331 * 16000 = 14,5 \quad (6)$$

O controle via computador tem a capacidade de definir o modo de operação da câmera, que pode ser manual ou automático e, caso o modo esteja em automático, de movimentar a câmera para a direita e para a esquerda, apertando-se um botão na interface de controle.

## V. CONCLUSÕES

O sistema localizador de fonte sonora apresentado é uma das possíveis soluções para o ambiente de videoconferência e para outras aplicações que necessitem a localização da origem de um determinado som. A implementação apresentada tem como vantagens seu baixo custo e menor complexidade de implementação do que outros sistemas, que são baseados em oito ou mais microfones. O direcionamento apenas horizontal, restrição imposta pela utilização de apenas dois microfones, mostra-se

suficiente na maioria dos ambientes, pois normalmente os participantes não se movimentam verticalmente e o campo de visão da câmera consegue suprir pequenas movimentações verticais. O sistema foi testado em um ambiente de tamanho médio, obtendo uma pequena porcentagem de erro de posicionamento, causada pela reverberação, reflexão e pela existência de múltiplas fontes sonoras no ambiente. Em ambientes muito grandes como um auditório a reverberação torna-se muito grande, depreciando o desempenho do sistema. A simplicidade desta implementação permite que o sistema seja instalado rapidamente, apenas posicionando o equipamento em um local apropriado da sala, diferentemente de outros sistemas que exigem a instalação de microfones por toda a sala. Futuramente pretende-se melhorar o desempenho do sistema através da substituição dos conversores analógico-digitais e também, se possível, do aumento da precisão de controle do motor responsável pela movimentação da câmera. A precisão de movimentação está também relacionada às características de amostragem e distância entre os microfones, o que impõe um limite de 28 posições diferentes de câmera para a configuração atual, mas que pode ser ampliado se necessário. Pretende-se também implementar um algoritmo de rejeição de aproximações de atraso incorretas, de modo a proporcionar um sistema mais estável em diversos ambientes, e que proporcione aos participantes de uma videoconferência toda a liberdade de expressão dada pela sua movimentação.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] D.V. Rabinkin, R.J. Renomeron, A. Dahl, J. French, J. Flanagan and M. Bianchi, "A DSP Implementation of Source Location Using Microphone Arrays", *J. Acous. Soc. Am.*, Vol 99, No. 4 Pt. 2, p. 2503, Abril de 1996 (*131st Meeting of the Acoustical Society of America*, Maio de 1996).
- [2] D.V. Rabinkin, R.J. Renomeron and J. Flanagan, "Microphone Array Sensor Placement Optimization in Reverberant Environments", *J. Acous. Soc. Am.*, Vol. 102, No. 5 Pt. 2, p. 3207, Novembro de 1997. (*134th Meeting of the Acoustical Society of America*, Dezembro de 1997).
- [3] C. Knapp and G. Carter, "The generalized Correlation Method for Estimation of Time Delay", *IEEE Transactions on Acoustics Speech and Signal Processing*. Bol ASSP-24. n. 4, Agosto de 1976.