

A semblance based TDOA algorithm for sound source localization

Aldeia, G. S. I., Crispim, A. E., Barreto, G., Alves, K., Ferreira, H., Nose-Filho, K.

Guilherme Seidyo Imai Aldeia

Universidade Federal do ABC

Petrópolis/RJ
2019

Contextualização

O uso de drones em cenários de busca e resgate está ganhando interesse na pesquisa. Podemos fazer o uso de microfones para localizar pessoas pedindo por ajuda, porém:

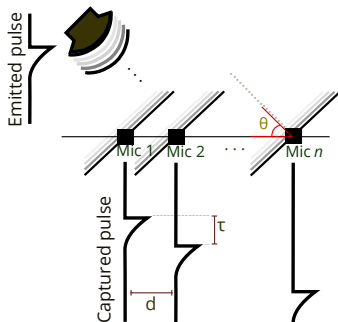
- ✗ A fonte sonora precisa estar ativa;
- ✗ O *ego-noise* (ruído produzido pelo próprio drone) tem grande interferência.

Objetivos

Propor uma abordagem para a tarefa de localização de fonte sonora e compará-la com um algoritmo do estado-da-arte.

Determinando a direção da fonte sonora

Sound Source



Temos:

- Um *array* de microfones com posições conhecidas (podemos medir d);
- Conhecimento da velocidade de propagação do som (v).

Assumimos que:

- O som se propaga com uma frente de onda plana;
- O som chega nos diferentes microfones em diferentes instantes de tempo.

Determinando a direção da fonte sonora

Seja θ a direção da origem do som em relação à um ponto de referência (e.g. o primeiro microfone), e τ a correção dos tempos de atraso (que se baseia em uma direção para encontrar o atraso relativo em cada canal para a correção):

- Conhecendo d, τ, v podemos achar θ ;
- Conhecendo d, θ, v podemos achar τ .

Mas não conhecemos nem θ nem τ .

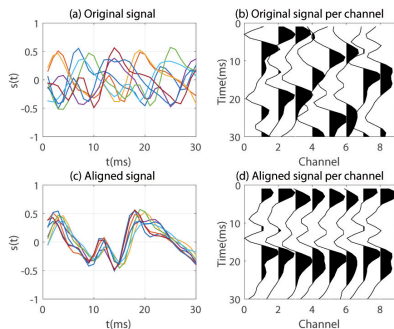
Corrigindo os atrasos dos sinais no tempo

O atraso τ para o microfone de índice i e localização m_i , assumindo que a fonte do som esteja em k_d , é dado por:

$$\tau(m_i) = -\frac{k_d \cdot m_i}{v}, \quad (1)$$

onde k_d é um vetor que aponta para uma direção parametrizada por azimuth ($\Theta_d \in [-\pi, \pi]$) e elevação ($\Phi_d \in [-\frac{\pi}{2}, \frac{\pi}{2}]$).

Corrigindo os atrasos dos sinais no tempo



A correção que gera o melhor alinhamento dos sinais possivelmente está associada à direção da fonte sonora.

Semblance

A função de correlação cruzada *Semblance* (utilizada em física de reflexão) mede o nível de similaridade entre os sinais de diferentes sensores.

$$Z_d = \frac{\sum_n |\sum_k \hat{s}_k(n)|^2}{N_r \sum_n \sum_k |\hat{s}_k(n)|^2}, \quad (2)$$

onde k é o índice do microfone, n as amostragens no tempo, N_r o número de sensores, e $\hat{s}_k(n) = s_k(n - \tau_k)$ o sinal na amostragem no tempo n do k -ésimo microfone após ter a correção τ_k (para um dado Θ e Φ) aplicada.

Global Semblance

Nossa proposta inicial foi o algoritmo **Global Semblance**.

Algoritmo proposto - Global Semblance

- ① Criar uma grade de “chutes” uniformemente espaçados;
- ② Estimar o atraso para cada possível direção;
- ③ Medir a correlação nos sinais corrigidos para cada direção;
- ④ Escolher direção com maior correlação.

Global Semblance

Algoritmo 1: Find semblance global (find_global)

input : Δ : interval between angles to be tested
 SoS : speed of sound on the medium
 Fs : sampling rate
 s : matrix containing the audio of the 8-channel microphones
 $micPos$: array with coordinates $[x, y, z]$ of the microphones positions

output: z : matrix mapping correlation with angles
 Θ : tested values for elevation
 Φ : tested values for azimuth

```

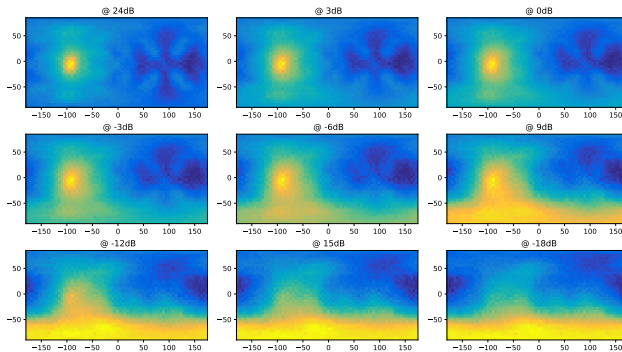
step =  $\Delta * \pi / 180$ ;
 $\Theta = [\theta \mid \theta \leftarrow [-\pi, -\pi + step, \dots, \pi]]$ ;
 $\Phi = [\phi \mid \phi \leftarrow [-\pi/2, -\pi/2 + step, \dots, \pi/2]]$ ;
 $\tau = []$ ;

for ( $i, \theta$ ) in (range( $\Theta$ ),  $\Theta$ ) do
    for ( $j, \phi$ ) in (range( $\Phi$ ),  $\Phi$ ) do
         $kd = [\cos(\theta) * \cos(\phi), \sin(\theta) * \cos(\phi), \sin(\phi)]$ ;
        for ( $k, mic$ ) in (range( $micPos$ ),  $micPos$ ) do
             $\tau[i, j, k] = \text{round}(((kd * mic') / SoS) * Fs)$ ;
for  $i$  in range( $\Theta$ ) do
    for  $j$  in range( $\Phi$ ) do
        for  $k$  in range(numMic) do
             $\hat{s}_k(n) = s_k(n - \tau[i, j, k])$ 
             $z[j, i] = \text{semblance}(\hat{s})$ 
return  $z, \Theta, \Phi$ ;

```

Painel de Semblance

Um heatmap obtido calculando a correlação Semblance para todas as combinações de (θ, ϕ) — chamado de painel de Semblance — permite visualizar o resultado da grade de chutes:



Local Semblance

Uma variação que foi estudada em testes preliminares, e chamada de **Local Semblance**, se mostrou mais eficiente.

Variação - Local Semblance

- ① Dividir o áudio em vários *frames* (podendo ter sobreposição ou não);
 - ② Aplicar o Global Semblance para cada *frame*;
 - ③ Combinar os painéis em um único utilizando os maiores valores (*pooling max*);
 - ④ Escolher direção de maior correlação.
- ✓ Capaz de realçar a relação SNR nos *frames* onde a fonte sonora está ativa, obtendo melhores resultados.

Local Semblance

Algoritmo 2: Find semblance local (find_local)

input : *frameSize*: size of the frames
 overlap: overlap between frames
 Δ : interval between angles to be tested
 SoS: speed of sound on the medium
 Fs: sampling rate
 s: the data of the 8-channel
 micPos: array with coordinates [x, y, z] of the microphones positions
output: *z*: matrix mapping correlation with angles
 Θ : tested values for elevation
 Φ : tested values for azimuth

 sTotal = length(*x*); // total samples
 sSize = round(*frameSize* * *Fs*); // sample size
 sOverlap = round(*overlap* * *sSize*); // sample overlap
 nFrames = ceil((*sTotal* - *sSize*) / (*sSize* - *sOverlap*)) + 1;
 painels = [];
 for *i* in range(*nFrames*) **do**
 begFrame = *i* * (*sSize* - *sOverlap*);
 endFrame = *begFrame* + *sSize*;
 sFrame = *s*[*begFrame* : *endFrame*, :];
 painels[*i*] = find_global(Δ , *SoS*, *Fs*, *sFrame*);
 return pooling(*painels*), Θ , Φ ;

Dados utilizados

Temos:

- 3 gravações de um drone com 8 microfones acoplados, com apenas sinal de fala (com direção conhecida);
- 1 gravação de puro ruído dos rotores do drone, obtida com os mesmos microfones.

Foram feitas diversas combinações com diferentes relações SNR: $[24, 21, 18, \dots, 3, 0, -3, -4, -5, \dots, -19, -20, -21]$

Isso permite encontrar um limiar de SNR para dizer onde o algoritmo começa a apresentar problemas.

Métrica de erro

O erro foi calculado pela equação *Great circle distance*:

$$\Delta\sigma = \arctan \frac{\sqrt{(\cos \phi_2 \sin(\Delta\theta))^2 + (\cos \phi_1 \sin \phi_2 - \sin \phi_1 \cos \phi_2 \cos(\Delta\theta))^2}}{\sin \phi_1 \sin \phi_2 + \cos \phi_1 \cos \phi_2 \cos(\Delta\theta)}. \quad (3)$$

Essa equação dá o ângulo entre 2 pontos na superfície de uma esfera. Foi considerado um erro aceitável até 10° .

Ajuste de hiper-parâmetros

Os algoritmos possuem um grupo de hiper-parâmetros que influenciam na qualidade do resultado:

- Em comum aos algoritmos Local Semblance e Global Semblance:
 - Δ - tamanho do espaçamento da grade de chutes;
- No caso do Local Semblance:
 - *Overlap* - sobreposição entre os frames;
 - *FrameSize* - tamanho de cada frame.

Para encontrar a melhor configuração, foi utilizado o *gridsearch*.

Gridsearch

O *gridsearch* é uma técnica comum no campo de aprendizado de máquina para ajuste de hiper-parâmetros. Faz uma busca exaustiva entre todas as possíveis combinações, obtendo aquela que minimiza o erro.

① $\text{overlap} = [0, 0.1, 0.2, 0.3, 0.4, 0.5]$

② $\Delta = [17.5, 15, 12.5, 10, 7.5, 5]$

③ $\text{frameSize} = [0.064, 0.128, 0.256, 0.512, 1.024]$

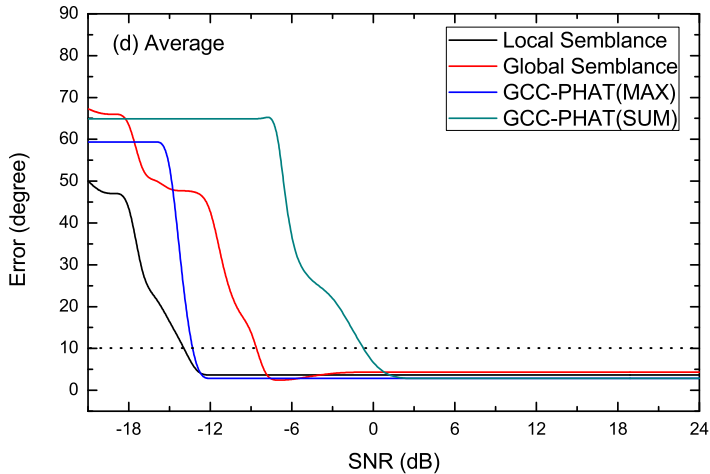
Comparação com o estado-da-arte

Os resultados foram comparados com o GCC-PHAT (*Generalized Cross Correlation PHAse-Transform method*), utilizando dois *poolings* diferentes - *max* e *sum*.

Resultados do Gridsearch

- Melhor resultado: $\Delta = 10$, $FrameSize = 0.064s$, 20% *Overlap* entre os *frames*;
- No geral, valores de Δ menores que 10 não tem diferença significativa no resultado;
- Para o *FrameSize*, quanto menor o valor, melhor o resultado;
- Um pequeno grau de *Overlap* (20%) tem melhores resultados.

Desempenho médio dos métodos



Discussão

À respeito dos métodos propostos:

- Apresentaram um erro de $\approx 2^\circ$ em vários casos;
- Local Semblance leva cerca de 6.8 segundos para executar em cada áudio, com duração de 5s;
- Abordagem Global usa 1 *core*, a Local é paralelizada em todos os *cores* de um i7@1.3GHz;
- O Local Semblance é capaz de acertar a direção até em casos de SNR de -16dB.

Discussão

Comparado com o GCC-PHAT:

- GCC-PHAT foi comparado nas mesmas condições: $\Delta=10$ e *frames* de 0.064s;
- o Local Semblance superou o GCC-PHAT;
- O Global Semblance tem um resultado intermediário entre o GCC-PHAT (max pooling) e o GCC-PHAT (sum pooling);

Conclusões

- Nesse artigo propomos uma nova técnica baseada na função de coerência Semblance para o problema de localização de fonte sonora;
- O algoritmo foi testado com 3 áudios em diferentes configurações de SNR;
- Algoritmo apresenta boa performance.

Continuação da jornada

Um próximo artigo sobre o método está para ser publicado!

*4th Workshop on Communication Networks and Power Systems
(WCNPS 2019)*

**On the application of SEGAN for the attenuation of the
ego-noise in the speech sound source localization
problem**

Conferência dias 3 e 4 de outubro 2019

<https://ieee-wcnps.org/>

Contato

Guilherme Aldeia (discente e apresentador do trabalho)

✉ `guilherme.aldeia@aluno.ufabc.edu.br`

Kenji Nose-Filho (professor docente orientador):

✉ `kenji.nose@ufabc.edu.br`

Link da apresentação:

🔗 `galdeia.github.io/presentations/SBrT.pdf`