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 veículos aéreos não tripulados em cenários de busca e resgate está ganhando interesse na pesquisa.

- Pode ser feito uso de câmeras de alta resolução, mas teríamos problemas com:
  - ✗ Falta de iluminação;
  - ✗ Neblina;
  - ✗ Obstáculos.

Alternativa

Uso de microfones para captar sinais sonoros de pessoas pedindo por ajuda.

Contextualização

Uso de sinais sonoros evita os problemas anteriores, mas possui os seus próprios:

- ✗ Requer que a fonte sonora esteja ativa;
- ✗ Tem interferência do *ego-noise*, ruído não estacionário produzido pelo próprio veículo aéreo.

Na literatura

Uso de algoritmos de estado-da-arte que se baseiam no **domínio da frequência** para a tarefa de estimar a direção utilizando microfones.

Contextualização

Podemos citar os artigos:

- *Robust acoustic source localization of emergency signals from micro air vehicles*;
- *DREGON: Dataset and Methods for UAV-Embedded Sound Source Localization*;

Que fizeram o estudo e avaliação de desempenho de várias técnicas do estado-da-arte para a tarefa de localização de fonte sonora.

Porém...

Não foram comparadas técnicas baseadas no **domínio do tempo**.

Nosso trabalho

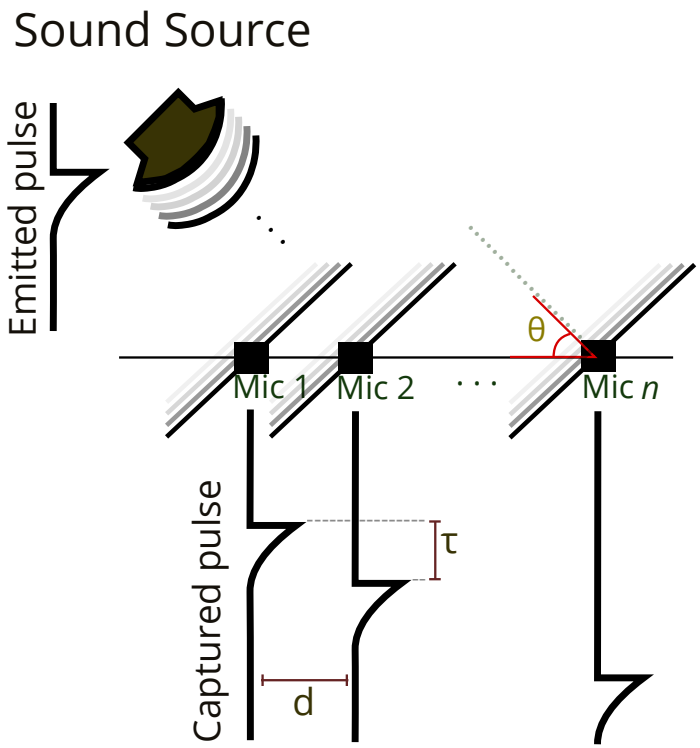
Objetivos

Estudar uma abordagem baseada no **domínio do tempo** para a tarefa de localização de fonte sonora e compará-la com um algoritmo do estado-da-arte.

Proposta

Propor o uso de uma técnica baseada na diferença de tempo de chegada do som em diferentes sensores, com a função de coerência *Semblance*, amplamente utilizada em processamento de sinais sísmicos, para o problema de localização de fonte sonora.

Semblance



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.

Semblance

Seja θ a direção da origem do som, 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;

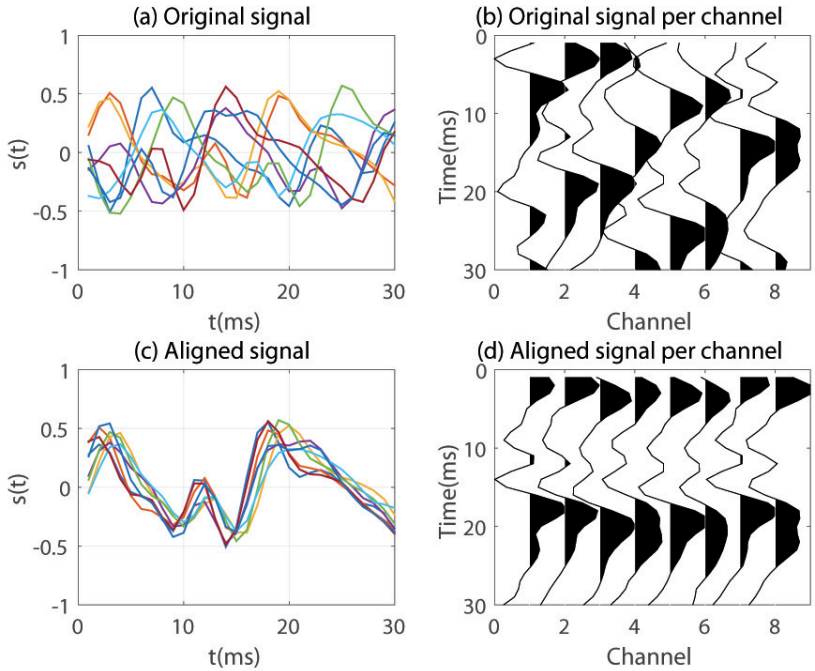
Semblance

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:

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

onde k_d é um vetor que aponta para uma direção parametrizada por azimuth ($\Theta_d \in [-\pi, \pi]$) e elevação ($\Phi_d \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$) em relação à um ponto de referência (i.g. o primeiro microfone, o centro de massa dos microfones).

Semblance



A correção τ é associada a um par (θ, ϕ) que aponta para uma direção. A correção de maior correlação provavelmente é a direção da fonte sonora.

Semblance

A função de correlação cruzada *Semblance* mede o nível de similaridade entre os sinais de diferentes sensores.

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

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

Em física de reflexão (sísmica) temos um caso parecido com o de localização de fonte sonora, que faz o uso do Semblance.

Algoritmo proposto - Global Semblance

Criar uma grade de "chutes" uniformemente espaçados, estimar o atraso para cada possível direção e medir a correlação nos sinais corrigidos, para cada direção. Aquela com maior correlação é escolhida como direção da fonte sonora.

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

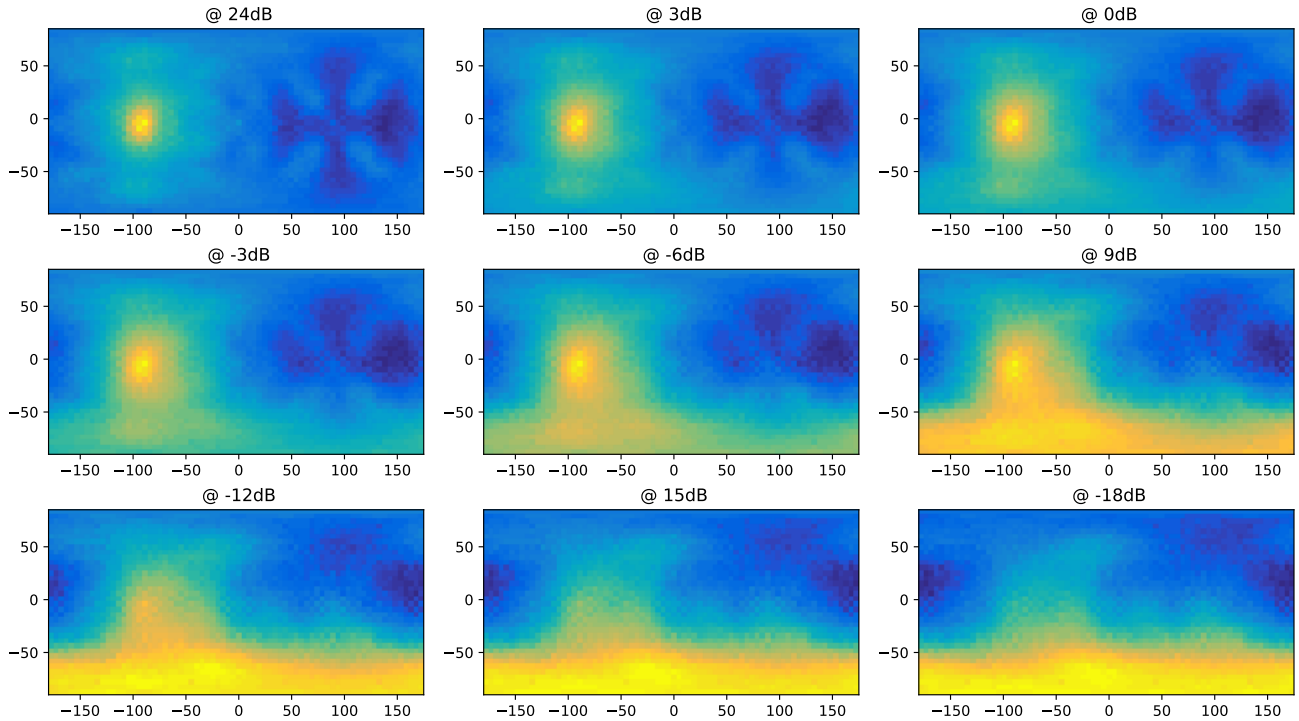
output : z: matrix mapping correlation with angles
Θ: tested values for elevation
Φ: tested values for azimuth

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

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

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.

- Capaz de realçar a relação SNR nos *frames* onde a fonte sonora está ativa, obtendo melhores resultados.

Local Semblance

Divide o áudio em vários *frames*, obtém o painel de Semblance para cada um deles e então combina os painéis em um único (técnica chamada de *pooling*), selecionando para cada par (θ, ϕ) a maior correlação entre os painéis.

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**
    $begFrame = i * (sSize - sOverlap);$
    $endFrame = begFrame + sSize;$
    $sFrame = s[begFrame : endFrame, :];$
    $painels[i] = find\_global(\Delta, SoS, Fs, sFrame);$

return $pooling(painels), \Theta, \Phi;$

Dados utilizados

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

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

Resultando em 26 configurações para cada um dos áudios de fala, 78 medidas de erro.

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

Dados utilizados

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. Foram utilizados como possíveis valores:

1. overlap $=$ [0, 0.1, 0.2, 0.3, 0.4, 0.5]
2. $\Delta =$ [17.5, 15, 12.5, 10, 7.5, 5]
3. frameSize $=$ [0.064, 0.128, 0.256, 0.512, 1.024]

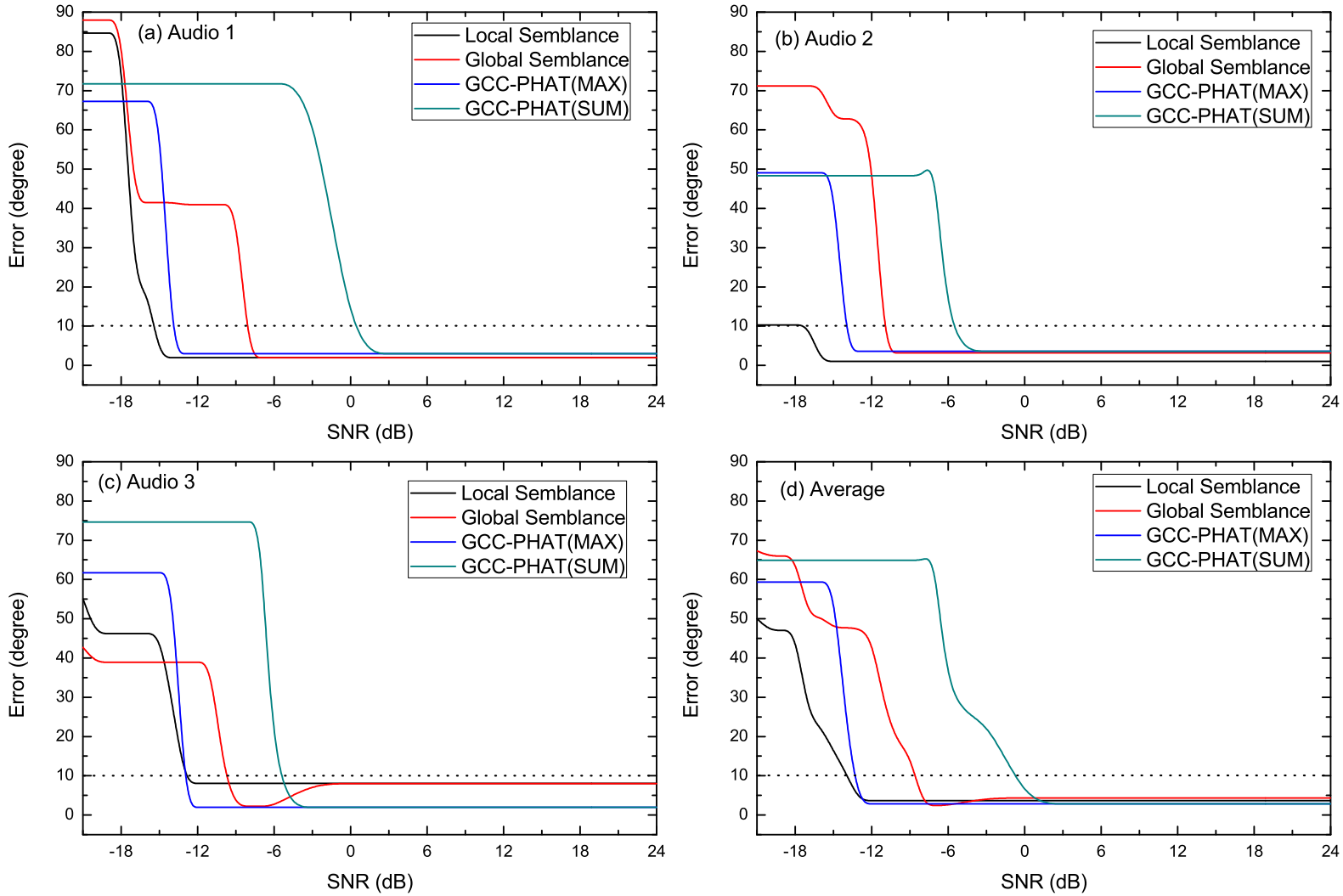
Comparação com o estado-da-arte

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

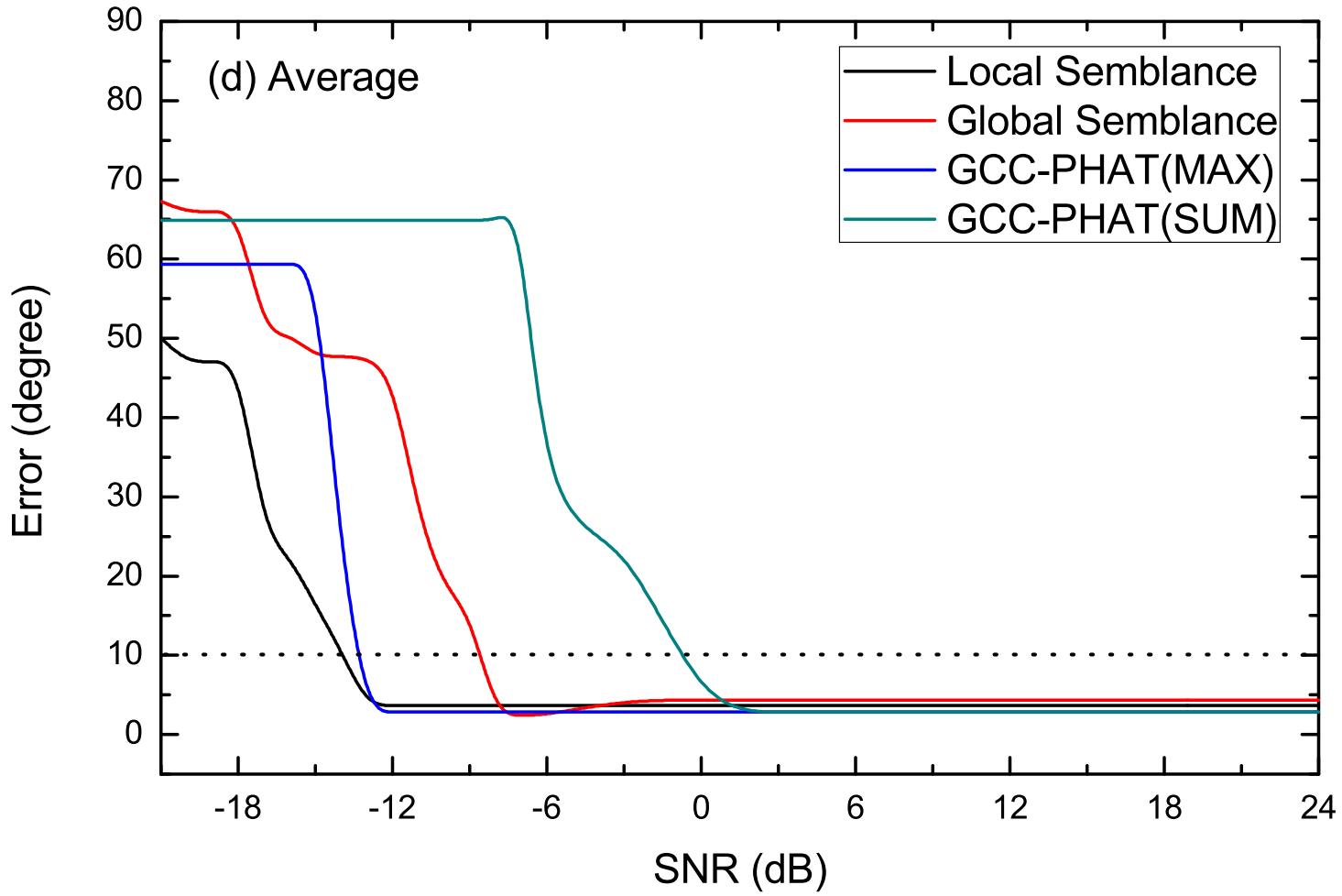
Resultados do Gridsearch

- 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.

Resultados com o estado-da-arte



Resultados com o estado-da-arte



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: Δ=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.

Perspectivas futuras

- Análise de complexidade
- Testar técnicas com filtro
- Otimizar a busca, implementando heurísticas baseadas em busca em árvore para ser aplicável em tempo real.

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