

# On the Semblance Based TDOA Algorithm for Sound Source Localization: a parametric study

Aldeia, G. S. I., Ferreira, H., Nose-Filho, K.

Universidade Federal do ABC (UFABC)  
Laboratório de Sinais e Sistemas (LSS)

Florianópolis/SC  
22-25 de novembro de 2020

# Índice

- ➊ Introdução
- ➋ Algoritmo SB-SSL
- ➌ Metodologia
- ➍ Resultados e discussão
- ➎ Conclusões

# Hiper-parâmetros

Muitos algoritmos possuem parâmetros que podem ser previamente definidos.

## Por quê ajustá-los?

- ✓ Possibilitam um ajuste fino do algoritmo
- ✓ Podem maximizar o desempenho em tarefas específicas
- ✗ Custo adicional
- ✗ Pode não ser tarefa trivial

# Como ajustar hiper-parâmetros

A ideia por trás do ajuste prévio é encontrar um conjunto de valores para os parâmetros que torna o algoritmo mais eficiente, sendo uma maximização de performance.

## Maneira mais simples

Utilizar valores *ad-hoc*, que podem ser baseados em outros valores vistos na literatura ou em heurísticas.

## Buscas exploratórias de melhor configuração

Processo de ajuste de parâmetros, como o *gridsearch* ou *manual search*

# Algoritmo do estudo

Recentemente, estudamos um algoritmo para localização de fontes sonoras (SSL) utilizando uma função de correlação cruzada comum no processamento de sinais sísmicos, denominado *Semblance Based TDOA Algorithm for Sound Source Localization* (SB-SSL), com três parâmetros ajustáveis pelo utilizador.

# Objetivos

O objetivo deste trabalho é analisar os parâmetros do algoritmo SB-SSL e:

- i) Verificar se existe um subconjunto de parâmetros que pode ter um valor fixado sem impactar na performance;
- ii) Avaliar o impacto de cada parâmetro no desempenho final;
- iii) Chamar atenção à importância e ganhos que uma análise de sensibilidade pode gerar para o uso de um método computacional.

# Justificativa

Vemos essa análise como um passo natural na criação de um algoritmo - aumentar o entendimento do seu funcionamento e simplificar o processo de uso.

# Índice

- ① Introdução
- ② Algoritmo SB-SSL
- ③ Metodologia
- ④ Resultados e discussão
- ⑤ Conclusões



# Algoritmo SB-SSL

Dado um arranjo de  $k$  microfones ( $k \geq 2$ ), a determinação da direção  $\mathbf{k}_d \in \mathcal{R}^3$  parametrizada por azimuth  $\Theta_d \in [-\pi, \pi]$  e elevação  $\Phi_d \in [-\frac{\pi}{2}, \frac{\pi}{2}]$  é feita criando-se uma grade uniformemente espaçada (com espaçamento dado por  $\Delta$ ) de possíveis direções.

O atraso do microfone localizado em  $m_i$ , utilizando como ponto de referência  $\mathbf{0} = [\mathbf{0}, \mathbf{0}, \mathbf{0}]^T$ , pode ser calculado por:

$$\tau_{d,k} = -\frac{\mathbf{k}_d \cdot \mathbf{m}_k}{v}, \quad (1)$$

onde  $v$  é a velocidade do som, e  $\cdot$  denota a operação de produto interno.

# Cálculo da correlação cruzada

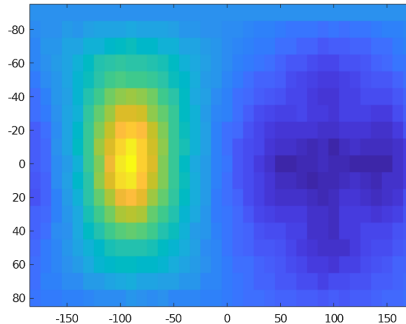
Em seguida é calculada a função de Coerência Semblance  $Z_d$  para cada par  $(\Theta_d, \Phi_d)$  de direções:

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

onde  $k$  denota os microfones,  $n$  denota as amostras temporais,  $N_r$  é o número total de sensores,  $\hat{s}_k(n) = s_k(n - \tau_k)$  é o sinal no instante de tempo  $n$  do  $k$ -ésimo microfone após a correção de tempo  $\tau_k$ .

# Painel Semblance

Ao final, temos uma matriz de valores de coerência para cada direção testada, que pode ser interpretada como uma imagem.



# Combinando vários painéis

Foi observado que o áudio em janelas (com tamanho dado por  $w$  e sobreposição dada por  $\delta$ ) e combinar cada painel com o *max pooling* apresenta melhores resultados.

Dessa forma, temos os seguintes parâmetros:

- $\Delta$ : Espaçamento da grade;
- $w$ : Tamanho da janela;
- $\delta$ : Sobreposição entre janelas.

# Índice

- ① Introdução
- ② Algoritmo SB-SSL
- ③ Metodologia**
- ④ Resultados e discussão
- ⑤ Conclusões

# Bases de dados

Uso de duas bases de dados:

- **Validação (150 áudios sintéticos)** para obter resultados de cada possível configuração permitida durante o processo de *gridsearch*;
- **Testes (150 áudios reais do DREGON dataset)** para obter resultados em dados não artificiais.

# Dados reais

- 150 áudios selecionados aleatoriamente de gravações com drone estático do *DREGON dataset*;
- Baixa variedade de direções da fonte sonora, podendo enviesar o modelo e favorecer algumas configurações específicas;
- Uso apenas para obtenção de resultados que correspondem ao desempenho em um cenário real.

# Dados sintéticos

- Necessidade de dados com direções variadas;
- Uso de 3 áudios do *DREGON dataset* sem ruído, com voz ativa e posição conhecida;
- Combinação dos áudios anteriores com gravações de ruído puro dos rotores do drone



# Criação dos dados sintéticos

Os sinais originais disponibilizados já possuem uma direção, sendo necessário centralizá-lo antes de mudar a direção.

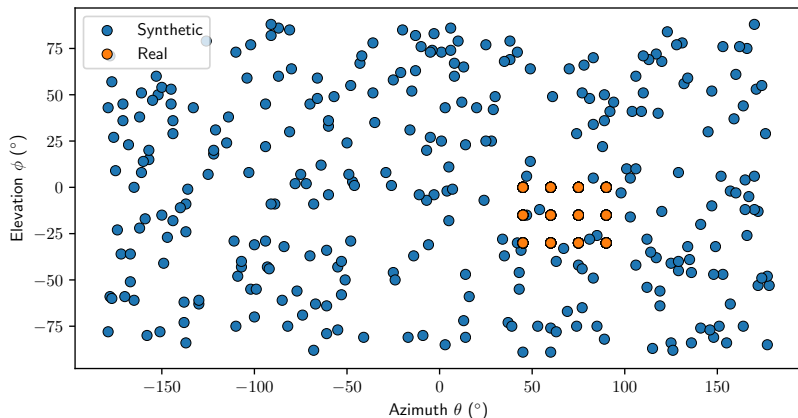
- Cada canal passa por um *upsample*;
- Os atrasos para cada canal são aplicados para a direção original conhecida (centralização);
- Uma direção aleatória é sorteada;
- O atraso para cada canal é aplicado para corresponder à nova direção;
- *Downsample* é feito em cada canal.

# Criação dos dados sintéticos

É preciso adicionar ruídos para simular um cenário real.

- Velocidades aleatórias para cada rotor são sorteadas;
- Áudios de cada rotor individual na velocidade sorteada são combinados;
- Sinal e ruído são normalizados e combinados com uma SNR aleatória entre  $[-25, 12]$ .

# Direções reais dos dados originais e sintéticos



# Medida de desempenho

Como medida de desempenho, foi utilizada a Equação *Great Circle Distance* (GCD), que retorna o menor ângulo entre dois pontos na superfície de uma esfera.

# Obtenção da performance dos parâmetros

Determinação de valores que cada parâmetro pode assumir, cobrindo o intervalo com menores valores com maior resolução (onde acredita-se que se concentrem as melhores configurações):

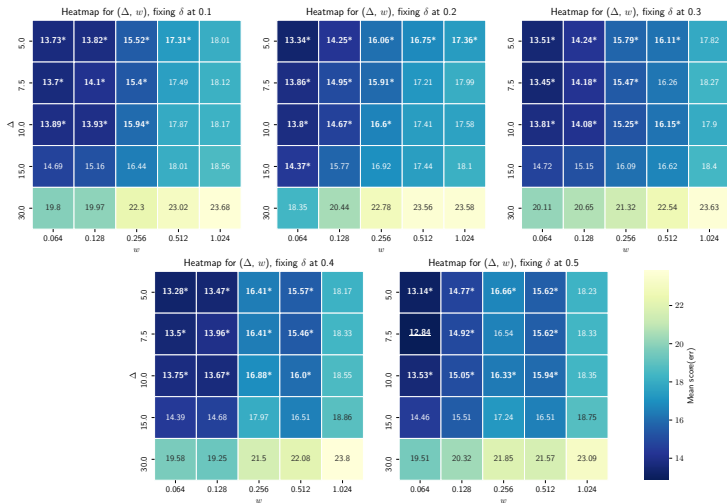
Parâmetro	Valores permitidos
$\Delta$	$\{5, 7.5, 15, 30\}$ ( $^{\circ}$ )
$w$	$\{0.064, 0.128, 0.256, 0.512, 1.024\}$ ( $s$ )
$\delta$	$\{10, 20, 30, 40, 50\}$ (%)

Ao todo, existem 125 configurações

# Índice

- ① Introdução
- ② Algoritmo SB-SSL
- ③ Metodologia
- ④ Resultados e discussão
- ⑤ Conclusões

# Mapas de calor



# Melhor configuração

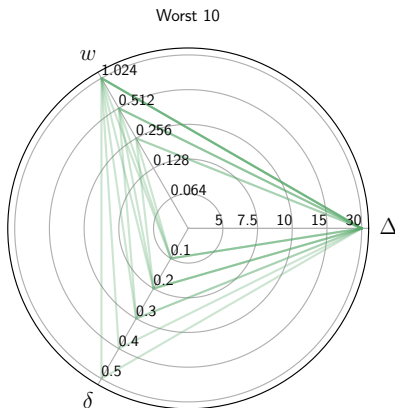
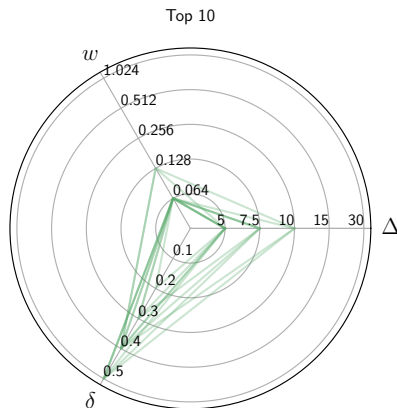
Na melhor configuração, os parâmetros assumem os valores  $\Delta = 7.5$ ,  $w = 0.064$ , e  $\delta = 0.5$ , com um erro médio de  $12.84^\circ$ . Executando, para os dados de teste, o algoritmo com essa configuração, o erro médio obtido é de  $15.20^\circ$ .



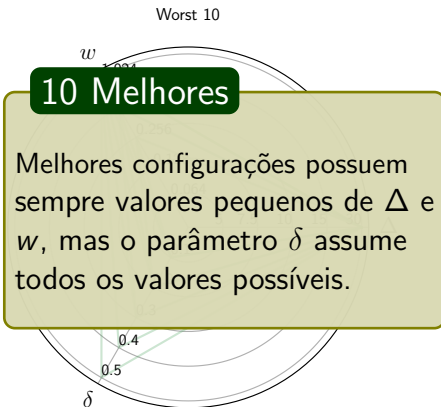
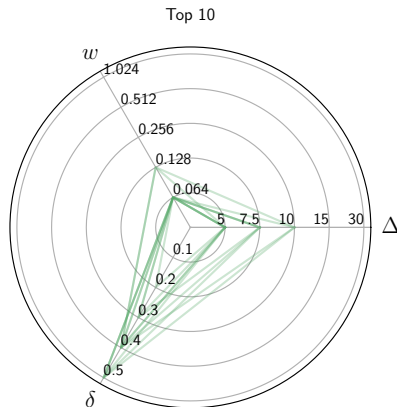
# Heatmaps

- Para  $\Delta$ , é justificado um valor pequeno ter um bom desempenho, pois aumenta a resolução espacial e permite estimar a direção com um menor erro;
- para  $w$ , considerando-se que são criadas várias janelas, por conta do sinal de fala não estar ativo durante todo o tempo, podendo ser mascarado pelo ruído em janelas grandes, é justificável a dominância de valores baixos;
- Por outro lado, o  $\delta$  acaba funcionando incluindo em janelas sinais que já foram processados, funcionalidade que aparentemente não justifica ser utilizada neste algoritmo.

# Radar plot das melhores e piores configurações



# Radar plot das melhores e piores configurações



# Radar plot das melhores e piores configurações

Top 10

 $w$ 

## 10 Piores

Apresentam sempre valores altos para os dois primeiros parâmetros, e todos os valores permitidos para  $\delta$ .

0.4

0.5

 $\delta$ 

Worst 10

 $w$ 

1.024

0.512

0.256

0.128

0.064

0.1

0.2

0.3

0.4

0.5

 $\delta$  $\Delta$

# Índice

- ① Introdução
- ② Algoritmo SB-SSL
- ③ Metodologia
- ④ Resultados e discussão
- ⑤ Conclusões

# Conclusões

- Resultados com um conjunto maior de dados que no artigo que propõe o SB-SSL, embora os achados se preservem;
- Parâmetro  $\delta$  apresenta fortes indícios de ser desnecessário, simplificando o uso do algoritmo, além de simplificar o processo de ajuste fino, devido à menor quantidade de parâmetros para ajustar.

# Contato

**Obrigado!**

# Informações adicionais

Guilherme Seidyo Imai Aldeia

✉ [guilherme.aldeia@ufabc.edu.br](mailto:guilherme.aldeia@ufabc.edu.br)

Henrique Ferreira

✉ [hfsantos@ufabc.edu.br](mailto:hfsantos@ufabc.edu.br)

Kenji Nose-Filho

✉ [kenji.nose@ufabc.edu.br](mailto:kenji.nose@ufabc.edu.br)

Link da apresentação

🔗 [galdeia.github.io/presentations/SBrT\\_2020.pdf](https://galdeia.github.io/presentations/SBrT_2020.pdf)

Link do repositório com códigos e resultados

🔗 <https://github.com/gAldeia/sensitivity-analysis-SEMBLANCE>