

**UNIVERSIDADE DE CAXIAS DO SUL
ÁREA DO CONHECIMENTO DE CIÊNCIAS EXATAS E
ENGENHARIAS**

EFRAIN SCHMITT

UM ESTUDO SOBRE LOCALIZAÇÃO DE FONTE SONORA

**CAXIAS DO SUL
2023**

EFRAIN SCHMITT

UM ESTUDO SOBRE LOCALIZAÇÃO DE FONTE SONORA

Trabalho de conclusão de curso, como parte dos requisitos para à obtenção do grau de bacharel em Engenharia de Controle e Automação, Universidade de Caxias do Sul, área das ciências exatas e engenharia.

Orientador: Prof.: D.r Guilherme Holsbach Costa

CAXIAS DO SUL

2023

EFRAIN SCHMITT

UM ESTUDO SOBRE LOCALIZAÇÃO DE FONTE SONORA

Trabalho de conclusão de curso, como parte dos requisitos para à obtenção do grau de bacharel em Engenharia de Controle e Automação, Universidade de Caxias do Sul, área das ciências exatas e engenharia.

Aprovado(a) em 01/12/2023

Banca Examinadora

Prof.: D.r Guilherme Holsbach Costa
Universidade de Caxias do Sul

Prof.: M.e Ricardo Leal Costi
Universidade de Caxias do Sul

Prof.: D.r Marilda Machado Spindola
Universidade de Caxias do Sul

Dedico este trabalho aos meus pais, amigos e minha namorada que me incentivaram durante toda esta trajetória, assim como ao Professor Doutor Guilherme Holsbach Costa por seus ensinamentos e confiança ao longo da construção deste projeto.

AGRADECIMENTOS

A Deus por ter dado saúde e forças para superar as dificuldades deste projeto.

À universidade, seu corpo docente, direção e administração que oportunizaram a janela de conhecimento necessário.

Ao Professor Doutor Guilherme Holsbach Costa pela dedicação em suas orientações prestadas na elaboração deste trabalho, incentivando e colaborando no desenvolvimento de novas ideias.

E a todos que de alguma forma agregaram para a execução deste projeto.

RESUMO

Os avanços na área de localização de fonte sonora desempenham um papel crucial em diversos campos, tendo potencial de aplicação para melhorar a segurança e qualidade de vida. A localização de fonte sonora é de grande importância para as áreas de pesquisa de veículos autônomos, de sistemas de monitoramento, de robôs industriais, de análise de cena auditiva para aparelhos auditivos, entre outros. Diferentes abordagens podem ser utilizadas para a localização de uma fonte sonora. Dentre as principais estão o uso de Redes Neurais Convolucionais, Diferença de Tempo de Chegada, Correlação Cruzada Generalizada e Classificação de Sinais Múltiplos. Dentre essas possibilidades, as Redes Convolucionais surgem como tendência, porém apresentam alto custo computacional. A Classificação de Sinais Múltiplos é a abordagem mais indicada para localização de múltiplas fontes sonoras. Neste trabalho, é proposto um estudo sobre os fundamentos de localização de fonte sonora. Assim sendo, a localização de uma fonte sonora é explorada, com base na Correlação Cruzada Generalizada, ainda que o método seja inicialmente avaliado sem ponderação, recaindo no método de Correlação Cruzada. O erro de estimação de posição e a sensibilidade do método foram avaliados por simulação. Um sistema de aquisição com três microfones dispostos em geometria triangular foi construído e o algoritmo de localização de fontes foi implementado em duas plataformas distintas. Em testes práticos, chegou-se a um erro médio inferior a 14° , validando o conceito estudado.

Palavras-chave: Localização de Fonte Sonora, Método GCC, Correlação Cruzada.

LISTA DE FIGURAS

Figura 1 – Sistema de localização de fonte sonora (SSL)	11
Figura 2 – Estimativa de posição para dois microfones	19
Figura 3 – Estimativa de posição para três microfones	20
Figura 4 – Topologia da solução proposta	23
Figura 5 – Fontes sonoras posicionadas sinteticamente no plano cartesiano	26
Figura 6 – Sensibilidade aplicando defasagem de amostras no ponto FS_T1	29
Figura 7 – <i>Hardware</i> de aquisição	30
Figura 8 – Disposição das fontes sonoras em condições reais	31
Figura 9 – Sinal RS_T1 no ensaio real com <i>buzzer</i>	33

LISTA DE TABELAS

Tabela 1 – Métodos de generalização	16
Tabela 2 – Informação das fontes sonoras posicionadas sinteticamente	27
Tabela 3 – Resultados do método	28
Tabela 4 – Sensibilidade do método	30
Tabela 5 – Erro do método sob condições reais	32
Tabela 6 – Erro do método sob condições reais (Aquisitado com <i>notebook</i>)	34

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ALSA	<i>Advanced Linux Sound Architecture</i>
CNN	<i>Convolutional Neural Network</i>
CTB	Código de Trânsito Brasileiro
ESC	<i>Environmental Sound Classification</i>
ESR	<i>Environmental Sound Recognition</i>
GCC	<i>Generalized Cross-Correlation</i>
GCC-PHAT	<i>Generalized Cross-Correlation Phase Transform</i>
GMM	<i>Gaussian Mixture Model</i>
HMM	<i>Hidden Markov Model</i>
ILD	<i>Intensity Level Difference</i>
MFCC	<i>Mel Frequency Cepstral Coefficients</i>
MUSIC	<i>Multiple Signal Classification</i>
PBM	<i>Parts Based Models</i>
SNR	<i>Signal-to-Noise Ratio</i>
SRP-PHAT	<i>Steered-Response Power Phase Transform</i>
SSL	<i>Sound Source Localization</i>
SVM	<i>Support Vector Machine</i>
TDE	<i>Time Delay Estimation</i>
TDOA	<i>Time Difference of Arrival</i>

SUMÁRIO

1	INTRODUÇÃO	10
1.1	OBJETIVOS	13
1.1.1	Objetivo geral	13
1.1.2	Objetivos específicos	13
2	REFERENCIAL TEÓRICO	14
2.1	CORRELAÇÃO CRUZADA	14
2.2	ESPECTRO DE POTÊNCIA CRUZADA	15
2.3	CORRELAÇÃO CRUZADA GENERALIZADA	15
2.4	ESTIMAÇÃO DE POSIÇÃO DE UMA FONTE SONORA	18
2.4.1	Microfones Alinhados	19
2.4.2	Microfones em triângulo	21
3	MATERIAL E MÉTODOS	23
3.1	AVALIAÇÃO DO ERRO DO MÉTODO	24
3.2	AVALIAÇÃO DA SENSIBILIDADE DO MÉTODO	25
3.3	AVALIAÇÃO EM CONDIÇÕES REAIS	25
4	RESULTADOS	26
4.1	ERRO DO MÉTODO	26
4.2	SENSIBILIDADE DO MÉTODO	28
4.3	ERRO DO MÉTODO SOB CONDIÇÕES REAIS	30
5	CONSIDERAÇÕES FINAIS	35
	REFERÊNCIAS	36

1 INTRODUÇÃO

A compreensão e aprimoramento da localização de fontes sonoras desempenha um papel crucial em diversos campos, quando aplicada com o intuito de melhorar a segurança e a qualidade de vida. Na automação industrial, robôs autônomos podem se beneficiar da localização de fonte sonora para aprimorar suas tomadas de decisão (GENG; JUNG, 2008). Por outro lado, na área de veículos autônomos e semiautônomos, a capacidade de localizar e interpretar sinais acústicos, que carregam informações cruciais em cenários de trânsito, resulta em redundância ao sistema. O estudo de tecnologias e soluções desenvolvidas para a localização de fontes sonoras para fins automotivos tem sua importância evidenciada, por meio de patentes (DELP, 2016), e pela academia, por meio de artigos científicos (LIAQUAT *et al.*, 2021). A localização de fontes sonoras também pode ser aplicada para melhorar significativamente a funcionalidade dos sistemas de monitoramento de vigilância. As informações de direção da fonte sonora podem ser utilizadas para um controle automático da rotação de uma câmera, garantindo assim a captura na área de interesse (KOTUS, 2010). Da mesma maneira, outra área de pesquisa e desenvolvimento que tem grande avanço usando a localização de sinais acústicos, é a de aparelhos auditivos. A localização da fonte sonora em aparelhos auditivos permite a análise da cena auditiva, restaurando assim ao usuário uma melhor interação com o ambiente (FARMANI *et al.*, 2017).

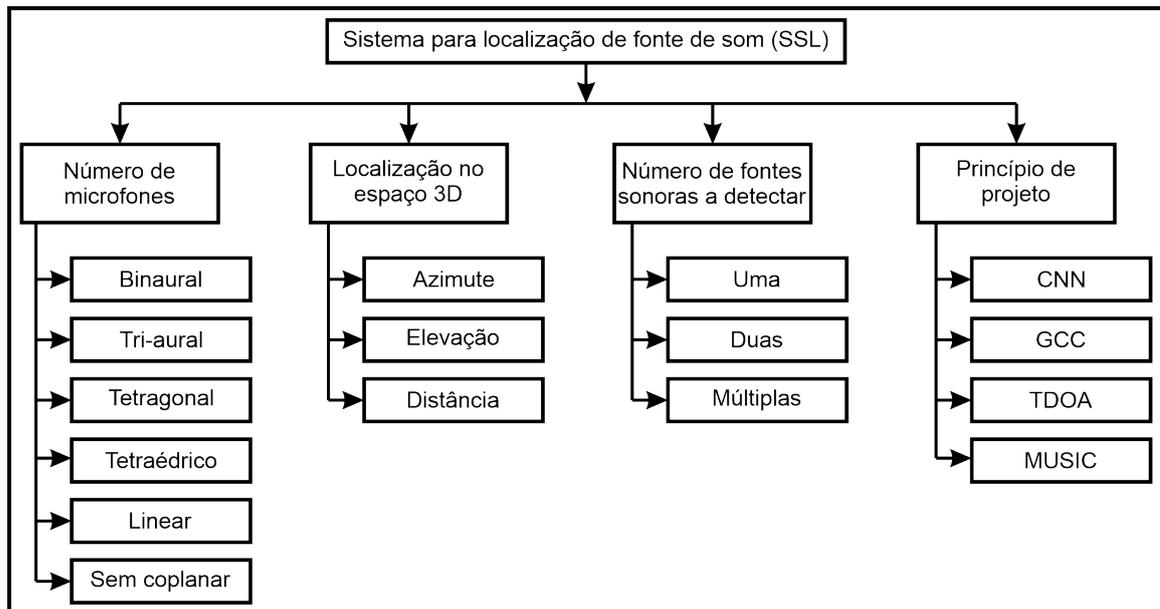
Métodos usados em um sistema de localização da fonte de som (SSL, *Sound Source Localization*) são apresentados na Figura 1, sendo classificados com base em quatro principais parâmetros: o número de microfones, localização no espaço 3D, número de fontes sonoras a serem detectadas e técnica principal de projeto.

Conforme (DESAI; MEHENDELE, 2022), as técnicas que se destacam no estado-da-arte são quatro: rede neural convolucional (CNN, *Convolutional Neural Network*), diferença de tempo de chegada (TDOA, *Time Difference of Arrival*), correlação cruzada generalizada (GCC, *Generalized Cross-correlation*) e classificação de sinais múltiplos (MUSIC, *Multiple Signal Classification*).

Em (MARCHEGIANI; NEWMAN, 2022) é proposto o uso de uma CNN motivado por técnicas de processamento de imagens. Usando espectrogramas dos sinais recebidos, um método de redução de ruído é aplicado. Esse método é baseado na segmentação semântica de gammatonegrams¹ dos sinais. A direção do sinal é obtida a partir de uma

¹ Conhecidos como Gammatone-like spectrograms, os gammatonegrams são uma uma imagem do tipo espectrograma. Os filtros gammatone são uma aproximação linear popular para a filtragem realizada

Figura 1 – Sistema de localização de fonte sonora (SSL)



Fonte: adaptado de (DESAI; MEHENDEALE, 2022)

CNN treinada com os gammatonegrams cruzados dos sinais sem ruído (segmentados). Como resultado, os autores obtêm robustez na localização de veículos de emergência em ambientes com razão sinal-ruído (SNR, *Signal-to-Noise Ratio*) extremamente baixa ($-40 \text{ dB} \leq \text{SNR} \leq 10 \text{ dB}$).

No trabalho de Schröder *et al.* (2013) são empregadas CNNs junto de modelos baseados em peças (PBM, *Parts Based Models*) e de modelo oculto de Markov (HMM, *Hidden Markov Model*) na identificação de fontes sonoras. Os autores concluem que PBMs aprendidos por máquina conseguem um melhor resultado para a maioria das condições de SNR. Entretanto, diferente do escopo deste trabalho, a proposta dos autores não está focada na localização da fonte de sinal, mas na identificação de sirenes de veículos de emergência.

Tanto em (MARCHEGANI; NEWMAN, 2022) quanto em (SCHRÖDER *et al.*, 2013), o aprendizado da máquina apresenta ótimos resultados para diferentes condições de SNR. Porém, as CNNs possuem um grande custo computacional tanto na aprendizagem do modelo quanto na operação.

No trabalho de Miao *et al.* (2014) o método de TDOA é empregado para a localização da fonte sonora. Utilizando uma matriz de quatro microfones os autores obtêm o TDOA usando o método GCC combinado à eliminação do efeito Doppler e à varredura do plano de origem. Simulações demonstram que este método pode localizar a

pelo ouvido (ELLIS, 2009).

fonte sonora desde que o SNR esteja acima de 5 dB, a distância de medição esteja abaixo de 10 m e a duração do sinal seja superior a 4 ms.

Outra abordagem, trabalhada em (WAN; WU, 2012), em que são usadas técnicas de processamento de sinais baseadas na discriminação de funções de correlação cruzada, calculadas pelo método de transformação de fase de correlação cruzada generalizada (GCC-PHAT, *Generalized Cross-Correlation Phase Transform*). Os resultados demonstraram que os algoritmos de localização com base na discriminação de funções de correlação cruzada fornecem maior precisão de localização em ambiente barulhento. Contudo, os resultados foram comparados apenas com os do método de transformação de fase de potência de resposta direcionada (SRP-PHAT, *Steered-Response Power Phase Transform*) que é bem conhecido por seu desempenho robusto em ambientes acústicos adversos.

Tanto em (MIAO *et al.*, 2014) quanto em (WAN; WU, 2012) a localização de fontes sonoras baseadas no método GCC apresentam bons resultados e como empregam técnicas de processamento de sinais requerem gastos computacionais moderados. Esse método foi inicialmente proposto por (KNAPP; CARTER, 1976), em que são aplicados pré-filtros com o intuito de acentuar o sinal passado para o correlacionador nas frequências para as quais SNR é maior e, simultaneamente, suprimir a potência do ruído.

No estudo de Gao *et al.* (2018) é proposto o uso do algorítmico MUSIC para localização de fontes sonoras. Os autores propõem uma modificação ao método MUSIC tradicional. Em vez de usar o autovalor máximo como o peso de cada banda de frequência, usam a relação sinal-ruído (SNR) como o coeficiente de peso de cada banda de frequência. Os resultados das simulação mostram que o método proposto tem o mesmo desempenho do algoritmo MUSIC tradicional na condição de uma única fonte. Quando o número de fontes é maior que um, esse método obtém um resultado mais preciso.

No trabalho de Zhang *et al.* (2020) um algoritmo conjunto de localização de fontes sonoras baseado na soma e no método MUSIC é proposto. Utilizando o algoritmo de soma os autores reduziram o espectro do espaço de busca do algoritmo MUSIC. Na simulação e nos resultados experimentais de laboratório mostraram que esse método pode não apenas localizar a fonte com precisão, mas também reduzir bastante o tempo de computação do algoritmo MUSIC.

Tanto em (GAO *et al.*, 2018) quanto em (ZHANG *et al.*, 2020) o método MUSIC apresenta excelente desempenho para a localização de múltiplas fontes sonoras, porém o método requer um grande gasto computacional, ainda que Zhang *et al.* (2020) tenham conseguido avanços significativos neste sentido.

Dentre as possibilidades para a localização de fonte sonoras mencionadas, o método CNN surge como tendência e já apresenta resultados de destaque. Por outro lado ainda é muito novo e demanda altos gastos computacionais durante a etapa de treinamento. O método MUSIC também apresenta gastos computacionais elevados, ainda que possam ser reduzidos por meio de diferentes técnicas. Entretanto, é mais indicado para situações em que se tem interesse em múltiplas fonte sonoras. Dentro deste contexto, este trabalho se propõe a estudar a localização de uma fonte sonora por meio de sinais sonoros obtidos de três sensores estrategicamente posicionados usando o método GCC. Nesse estudo, o método GCC será abordado na sua origem (KNAPP; CARTER, 1976).

1.1 OBJETIVOS

1.1.1 Objetivo geral

O objetivo do presente trabalho é desenvolver uma solução dedicada à localização automática de uma fonte sonora.

1.1.2 Objetivos específicos

O objetivo apresentado contempla os seguintes objetivos específicos:

- a) selecionar, a partir de uma revisão teórica, um método de processamento de sinais adequado à localização de uma fonte sonora;
- b) realizar a implementação do método por meio de uma simulação;
- c) implementar o hardware necessário;
- d) obter experimentos práticos que permitam avaliar o funcionamento e o desempenho do método em estudo.

2 REFERENCIAL TEÓRICO

Neste capítulo são apresentados os fundamentos teóricos e ferramentas matemáticas necessárias para o desenvolvimento do sistema em estudo.

2.1 CORRELAÇÃO CRUZADA

A correlação cruzada é uma medida da dependência entre dois sinais aleatórios diferentes. Sejam $x_1[\xi, n]$ e $x_2[\xi, n]$ dois processos estocásticos discretos, em que ξ indexa a realização do processo e n é a variável de tempo-discreto. A correlação cruzada entre eles pode ser definida por (OPPENHEIM; SCHAFER, 2013):

$$r_{x_1x_2}[n_1, n_2] = E\{x_1[n_1]x_2^*[n_2]\}, \quad (1)$$

em que $E\{\cdot\}$ denota a esperança matemática, o sobrescrito “*” representa o complexo conjugado, n_1 e n_2 denotam instantes fixos de tempo e o índice das realizações (ξ) é suprimido por simplicidade de notação¹.

Entretanto, é bastante razoável supor que um processo estocástico é estacionário (NALON, 2009). Em um processo estacionário, as estatísticas das variáveis aleatórias correspondentes a cada amostra são independentes do instante de amostragem. Considerando a estacionariedade, (1) se reduz a

$$r_{x_1x_2}[\tau] = E\{x_1[n + \tau]x_2^*[n]\}. \quad (2)$$

A notação de conjunto de sinais é um conceito matemático conveniente, porém em situações práticas tem-se um número finito de sequências (realizações do processo estocástico). Mais que isso, em muitos casos, têm-se apenas uma sequência disponível. Nesses casos, o que resta é estimar a média estatística (no conjunto das realizações) pela média amostral (média temporal). Isso apenas faz sentido se o processo aleatório puder ser modelado como um processo *ergódico*. Um processo ergódico é um processo aleatório cujas estatísticas no sentido do *ensemble* são as mesmas estatísticas temporais (para que isso se verifique, o processo precisa ser também estacionário). Assim sendo, (2) pode ser

¹ Deste ponto em diante, processos estocásticos serão representados suprimindo-se o índice das realizações (ξ), como comumente praticado na área (PAPOULIS; PILLAI, 2002). Fica ao encargo do leitor interpretar de que forma se aplica a variável $x[n]$, por exemplo, no contexto em que for apresentada.

determinada por

$$r_{x_1x_2}[\tau] = E\{x_1[n + \tau]x_2^*[n]\} = \langle x_1[n + \tau]x_2^*[n] \rangle = \lim_{L \rightarrow \infty} \frac{1}{2L + 1} \sum_{n=-L}^L x_1[n + \tau]x_2^*[n], \quad (3)$$

em que $\langle \cdot \rangle$ denota a média temporal.

Em (3) as médias temporais são funções de um conjunto infinito de variáveis aleatórias. Em condições práticas, entretanto, apenas um número finito de amostras, L , está disponível. Nesse caso, a correlação cruzada em (3) pode ser aproximada por

$$\hat{r}_{x_1x_2}[\tau] = \frac{1}{L} \sum_{n=0}^{L-1} x_1[n + \tau]x_2^*[n]. \quad (4)$$

A estimativa do atraso de tempo (TDE, *Time Delay Estimation*) é obtida a partir do valor da abscissa no instante que o resultado da correlação cruzada é máximo, ou seja, na posição de máxima semelhança entre os sinais (WAN; WU, 2012):

$$\hat{\tau} = \arg \max_{\tau} (\hat{r}_{x_1x_2}[\tau]). \quad (5)$$

2.2 ESPECTRO DE POTÊNCIA CRUZADA

O Espectro de Potência Cruzada (*Cross-Power Spectrum*) ou Densidade Espectral de Potência Cruzada (*Cross-Power Spectral Density*), $R_{x_1x_2}[\omega]$, é a transformada de Fourier da Correlação cruzada. Assim sendo,

$$\hat{R}_{x_1x_2}[\omega] = \mathcal{F}\{\hat{r}_{x_1x_2}[\tau]\}. \quad (6)$$

O espectro de potência cruzada pode ser obtido pela transformada de Fourier de tempo discreto (AGUIRRE, 2004), seguindo direto de (6) que:

$$\hat{R}_{x_1x_2}[\omega] = \sum_{k=-\infty}^{\infty} \hat{r}_{x_1x_2}[k]e^{-j\omega k}. \quad (7)$$

2.3 CORRELAÇÃO CRUZADA GENERALIZADA

O método de Correlação Cruzada Generalizada (GCC), proposto inicialmente por Knapp e Carter (1976), é calculado utilizando a transformada de Fourier inversa da função

Espectro de Potência Cruzada multiplicada por uma função de ponderação, conforme

$$\hat{r}_{x_1x_2}^{(GCC)}[\tau] = \mathcal{F}^{-1}\{\psi_g[\omega]\hat{R}_{x_1x_2}[\omega]\}, \quad (8)$$

em que $\psi_g[\omega]$ é uma função de ponderação definida como

$$\psi_g[\omega] = H_1[\omega]H_2^*[\omega], \quad (9)$$

em que $H_1[\omega]$ e $H_2[\omega]$ são filtros que, quando bem selecionados, facilitam a estimativa do atraso de tempo (TDE).

Conforme Knapp e Carter (1976), um sinal $s[n]$ proveniente de uma fonte sonora e monitorado, na presença de ruído, por dois sensores separados no espaço podem ser modelados matematicamente como

$$x_1[n] = s[n] + \eta_1[n] \quad (10)$$

$$x_2[n] = \alpha s[n + \tau] + \eta_2[n], \quad (11)$$

em que α , $s[n]$, $\eta_1[n]$ e $\eta_2[n]$ são reais e provenientes de um processo estocástico estacionário, $\eta_1[n]$ e $\eta_2[n]$ modelam ruídos aditivos nos sensores 1 e 2, respectivamente, e sendo $s[n]$ não correlacionado com os ruídos $\eta_1[n]$ e $\eta_2[n]$. Definidos os sinais em (10) e (11), algumas funções de ponderação, $\psi_g[\omega]$, definidas em (KNAPP; CARTER, 1976), são apresentadas na Tabela 1. Cada uma dessas funções será discutida na sequência.

Tabela 1 – Métodos de generalização

Nome do método	Ponderação $\psi_g[\omega] = H_1[\omega]H_2^*[\omega]$
Correlação Cruzada	1
Resposta impulsiva de Roth	$1/R_{x_1x_1}[\omega]$
SCOT	$1/\sqrt{R_{x_1x_1}[\omega]R_{x_2x_2}[\omega]}$
PHAT	$1/ R_{x_1x_2}[\omega] $
Filtro Eckart	$\alpha R_{ss}[\omega]/(R_{\eta_1\eta_1}[\omega]R_{\eta_2\eta_2}[\omega])$
ML ou HT	$\frac{ \gamma_{12}[\omega] ^2}{ R_{x_1x_2}[\omega] (1 - \gamma_{12}[\omega] ^2)}$

Fonte: adaptado de (KNAPP; CARTER, 1976)

Correlação Cruzada (*Cross-Correlation*) — Quando os filtros $H_1[\omega] = H_2[\omega] = 1 \forall \omega$, a TDE é simplesmente o valor da abcissa no qual acontece o pico da função de correlação cruzada, como já discutido na Seção 2.1.

Resposta impulsiva de Roth (*Roth Impulse Response*) — A ponderação por resposta impulsiva de Roth, inicialmente proposta em (ROTH, 1971), atenua as regiões de frequência em que o Espectro de Potência Cruzada da autocorrelação do ruído ($R_{\eta_1\eta_1}[\omega]$) é elevada, ou seja, em que é mais provável que a estimativa $\hat{R}_{x_1x_2}[\omega]$ esteja errada. Assim, $\psi_g[\omega]$ é definida em (KNAPP; CARTER, 1976) como,

$$\psi_g[\omega] = \frac{1}{R_{x_1x_1}[\omega]}. \quad (12)$$

Transformação de Coerência Suavizada (*Smoothed Coherence Transform – SCOT*)

— O método de ponderação nomeado como Transformação de Coerência Suavizada, inicialmente proposto em (CARTER; NUTTALL; CABLE, 1973), utiliza uma interpretação aproximada a do método Roth. Erros em $\hat{R}_{x_1x_2}[\omega]$ são provenientes das bandas de frequência onde $R_{\eta_1\eta_1}[\omega]$ tem um valor elevado, mas também podem ser provenientes das bandas de frequência onde $R_{\eta_2\eta_2}[\omega]$ tem um valor elevado. Assim, o método SCOT é definido como (KNAPP; CARTER, 1976):

$$\psi_g[\omega] = \frac{1}{\sqrt{R_{x_1x_1}[\omega]R_{x_2x_2}[\omega]}}. \quad (13)$$

Quando $G_{x_1x_1}[\omega] = G_{x_2x_2}[\omega]$, o método SCOT é equivalente ao Roth.

Transformação de Fase (*Phase Transform – PHAT*) — Diferente dos outros métodos abordados, a ponderação por meio da transformação de fase não utiliza atribuição de peso de acordo com características de SNR. Inicialmente proposta em (CARTER; NUTTALL; CABLE, 1972), foi desenvolvida puramente como uma técnica *ad hoc*, tendo como vantagem não sofrer espelhamento por ser descorrelacionado do ruído. Desse modo, $\psi_g[\omega]$ é definido por (KNAPP; CARTER, 1976) como,

$$\psi_g[\omega] = \frac{1}{|R_{x_1x_2}[\omega]|}. \quad (14)$$

Na prática, um defeito aparente do método PHAT é que ele pondera $R_{x_1x_2}[\omega]$ com o inverso de $R_{ss}[\omega]$, ocasionando erros acentuados onde a potência do sinal é menor.

Filtro Eckart (*Eckart Filter*) — O filtro de Eckart deriva seu nome dos trabalhos feitos pelo autor nesta área, a exemplo de (ECKART, 1952). Assim, $\psi_g[\omega]$ é definido em (KNAPP; CARTER, 1976) como

$$\psi_g[\omega] = \frac{\alpha R_{ss}[\omega]}{R_{\eta_1\eta_1}[\omega]R_{\eta_2\eta_2}[\omega]}. \quad (15)$$

O filtro de Eckart atua para suprimir bandas de frequência de alto ruído, assim como o SCOT. Também, ao contrário do método PHAT, o filtro de Eckart atribui peso zero a bandas onde $R_{ss}[\omega] = 0$. Na prática, o filtro de Eckart requer conhecimento ou estimativa dos espectros do sinal e do ruído. Conforme (KNAPP; CARTER, 1976), quando $\alpha = 1$ isso pode ser obtido com

$$\psi_g[\omega] = |\hat{R}_{x_1x_2}[\omega]| \{ |\hat{R}_{x_1x_1}[\omega] - |\hat{R}_{x_1x_2}[\omega]| | \hat{R}_{x_2x_2}[\omega] - |\hat{R}_{x_1x_2}[\omega]| | \}. \quad (16)$$

Estimador ML (*ML Estimator*) — O estimador ML proposto em (KNAPP; CARTER, 1976) mostra-se idêntico ao proposto em (HANNAN; THOMSON, 1973), e por isso também pode ser chamado de estimador HT. O estimador ML é o método PHAT inversamente ponderado de acordo com a variabilidade das estimativas de fase. Assim, o estimador HT é definido por Knapp e Carter (1976) como

$$\psi_g[\omega] = \frac{|\gamma_{12}[\omega]|^2}{|R_{x_1x_2}[\omega]|(1 - |\gamma_{12}[\omega]|^2)}, \quad (17)$$

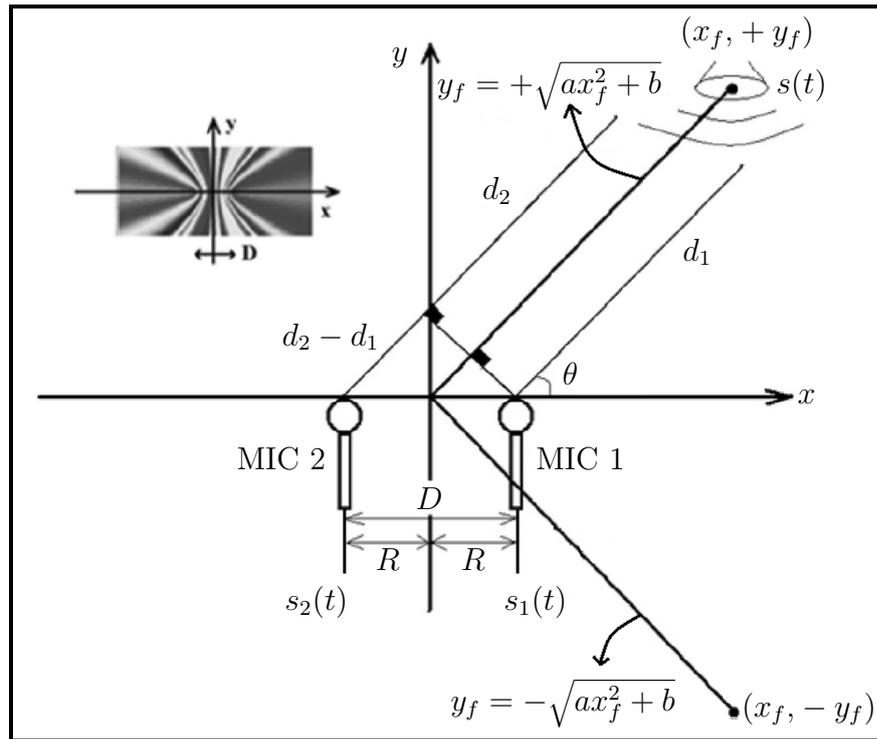
Conforme (KNAPP; CARTER, 1976), se $R_{\eta_1\eta_1}[\omega] = R_{\eta_2\eta_2}[\omega] = R_{\eta\eta}[\omega]$ é igual a uma constante no tempo $R_{ss}[\omega]$, então os últimos cinco métodos da Tabela 1 são os mesmos, exceto por uma constante. Porém, o método de correlação cruzada ($\psi_g[\omega] = 1 \forall \omega$) é uma função delta mesclada pela transformada de Fourier do Espectro de Potência do sinal(ruído).

2.4 ESTIMAÇÃO DE POSIÇÃO DE UMA FONTE SONORA

A estimação da posição em que se encontra uma fonte sonora pode ser realizada a partir dos atrasos temporais obtidos dos métodos de correlação cruzada. Conforme (PARSAYAN; AHADI, 2013), para a localização de uma fonte sonora em um plano horizontal (Azimute), são necessários dois microfones para uma varredura de 180°, que pode ser estendido para 360° com um microfone adicional. Existem diferentes configurações

de *arrays* de microfones que podem ser usadas para capturar o som de forma eficaz em diferentes situações. Para estimativas de posição em um plano horizontal são normalmente usados microfones alinhados ou distribuídos em um triângulo equilátero, conforme a Figura 2 e a Figura 3 respectivamente.

Figura 2 – Estimativa de posição para dois microfones



Fonte: adaptado de (PARSAYAN; AHADI, 2013)

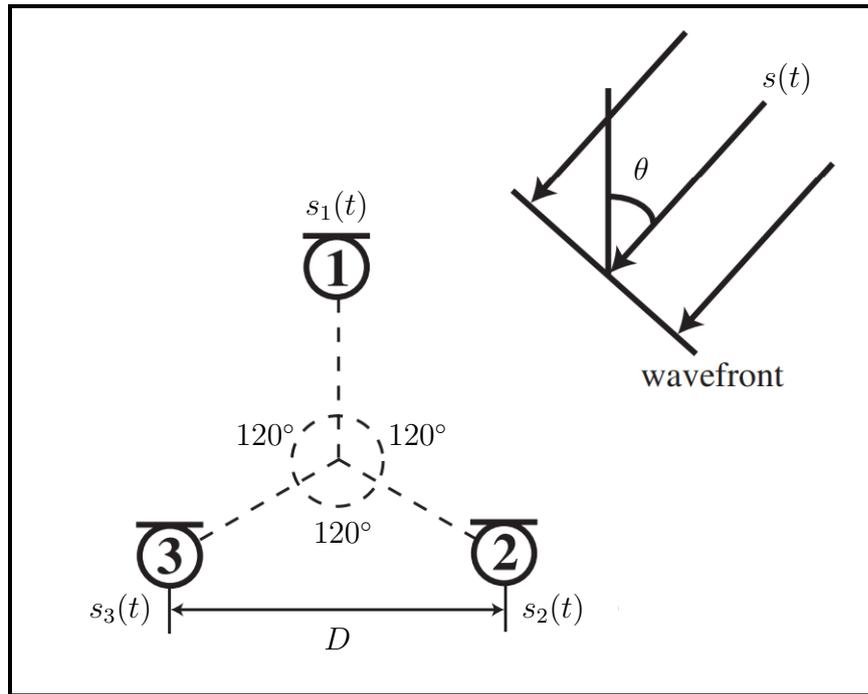
2.4.1 Microfones Alinhados

Para estimações de posição usando dois microfones alinhados, conforme ilustrado na Figura 2, o ângulo em que se encontra a fonte sonora em relação aos microfones pode ser calculado conforme (PARSAYAN; AHADI, 2013)

$$\theta = \arccos \frac{d_2 - d_1}{D}, \quad (18)$$

em que D é a distância entre os microfones e d_1 e d_2 são as distâncias da fonte sonora em relação a cada microfone. A diferença $d_2 - d_1$ pode ser reescrita em termos da velocidade de propagação do som no ar, v_{som} , como $d_2 - d_1 = (t_2 - t_1)v_{som}$. A estimativa de $t_2 - t_1$

Figura 3 – Estimativa de posição para três microfones



Fonte: adaptado de (HIOKA; MATSUO; HAMADA, 2009)

pode ser feita por meio dos métodos de correlação cruzada, sendo

$$d_2 - d_1 = \tau_{21} v_{som}, \quad (19)$$

em que τ_{21} é a estimativa de tempo encontrada a partir da divisão do resultado da correlação cruzada pela frequência de amostragem. Assim, (18) pode ser escrita como

$$\theta = \arccos \frac{\tau_{21} v_{som}}{D}. \quad (20)$$

A partir de (20) é possível encontrar apenas o ângulo em que se encontra a fonte sonora. Contudo, a localização da fonte sonora sobre uma curva hiperbólica em relação a x pode ser obtida por meio de (19), sendo que d_1 e d_2 podem ser reescritas usando Pitágoras. Assim, tem-se

$$\tau_{21} = \frac{\sqrt{(x_f - x_2)^2 + (y_f - y_2)^2} - \sqrt{(x_f - x_1)^2 + (y_f - y_1)^2}}{v_{som}}, \quad (21)$$

em que (x_1, y_1) , (x_2, y_2) e (x_f, y_f) são as coordenadas dos microfones e da fonte sonora, respectivamente, sendo o último par de coordenadas desconhecido. Assumindo-se que

as distâncias de ambos os microfones à origem seja R ($D = 2R$) e que ambos estejam localizados sobre o eixo x , (21) pode ser simplificada como

$$\tau_{21} = \frac{\sqrt{(x_f + R)^2 + y_f^2} - \sqrt{(x_f - R)^2 + y_f^2}}{v_{som}}. \quad (22)$$

Conforme (PARSAYAN; AHADI, 2013), (22) pode ser simplificada como

$$y_f^2 = ax_f^2 + b \quad (23)$$

$$a = \frac{4R^2}{v_{som}^2 \tau_{21}^2} - 1 \quad (24)$$

$$b = \frac{v_{som}^2 \tau_{21}^2}{4} - R^2 \quad (25)$$

em que y_f é a localização da fonte sonora sobre uma curva hiperbólica em relação a x_f , conforme mostrado na Figura 2.

Quando é considerado um terceiro microfone alinhado com os demais, pode-se derivar de (21) uma equação para a estimativa do atraso de tempo entre os microfones 1 e 3, levando a

$$\tau_{31} = \frac{\sqrt{(x_f - x_3)^2 + (y_f - y_3)^2} - \sqrt{(x_f - x_1)^2 + (y_f - y_1)^2}}{v_{som}}. \quad (26)$$

Desta forma, quando as curvas hiperbólicas obtidas a partir das diferenças de tempo dos microfones 1 e 2, e 1 e 3 se cruzam é possível estimar a localização (x_f, y_f) da fonte sonora. Este método não é linear, o que aumenta os tempos de processamento da localização e em alguns casos pode não convergir. Além disto, como os microfones ficam alinhados ainda existem ambiguidades, não diferenciando se a fonte sonora está na frente ou atrás.

2.4.2 Microfones em triângulo

Para estimações do ângulo em 360° utilizando três microfones em vértices de um triângulo equilátero, conforme ilustrado na Figura 3. O ângulo da fonte sonora em relação ao centro dos microfones pode ser calculado conforme (HIOKA; MATSUO; HAMADA, 2009) como

$$\theta = \arctan\left(\frac{x}{y}\right), \quad (27)$$

em que θ é obtido em radianos e as coordenadas $[x, y, z]$ são derivados do vetor coluna $\boldsymbol{\tau}'$, que conforme (HIOKA; MATSUO; HAMADA, 2009) é oriundo da multiplicação matricial

$$\boldsymbol{\tau}' = \mathbf{T}^T \boldsymbol{\tau}(\theta), \quad (28)$$

em que o sobrescrito “T” denota a transposição da matriz. A matriz \mathbf{T} é formada por três vetores coluna,

$$\mathbf{T} = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3] \quad (29)$$

$$\mathbf{e}_1 = \left[-\sqrt{\frac{1}{6}}, \sqrt{\frac{2}{3}}, -\sqrt{\frac{1}{6}} \right]^T \quad (30)$$

$$\mathbf{e}_2 = \left[\sqrt{\frac{1}{2}}, 0, -\sqrt{\frac{1}{2}} \right]^T \quad (31)$$

$$\mathbf{e}_3 = \left[\sqrt{\frac{1}{3}}, \sqrt{\frac{1}{3}}, \sqrt{\frac{1}{3}} \right]^T \quad (32)$$

e o vetor $\boldsymbol{\tau}(\theta)$ é formado pelas estimativas de tempo encontradas a partir da divisão do resultado da correlação cruzada pelo período de amostragem, sendo definido por (HIOKA; MATSUO; HAMADA, 2009)

$$\boldsymbol{\tau}(\theta) = [\tau_{12}(\theta), \tau_{23}(\theta), \tau_{31}(\theta)]^T \quad (33)$$

Os valores de x e y encontrados neste método não representam a real localização da fonte sonora, sendo apenas referências para calcular o ângulo onde se encontra a fonte sonora.

3 MATERIAL E MÉTODOS

Com a finalidade de testar as ferramentas matemáticas discutidas no capítulo anterior, este capítulo descreve os recursos necessários (*hardware* e *software*) e os métodos aplicados neste trabalho.

Inicialmente, para dar uma visão macro das etapas a serem seguidas, uma topologia do sistema é apresentada na Figura 4. Cada uma das etapas ilustradas na figura possuem importantes fundamentos que serão abordados na sequência. A primeira etapa

Figura 4 – Topologia da solução proposta



Fonte: Autor (2023)

apresentada na Figura 4 contempla a distribuição dos dispositivos para captura dos sinais de áudio. Quando se trata de uma aquisição em 360° , os trabalhos revisados utilizam a matriz de microfones composta por três microfones distribuídos em um triângulo equilátero (CATUR; SAPUTRA, 2019; HAMADA; KASPRZAK; PRZYBYSZ, 2012). Considerando o anteriormente exposto, são utilizados três microfones dispostos na forma de um triângulo equilátero, equidistantes em um metro.

A segunda etapa refere-se à aquisição dos sinais de áudio, a aquisição deve garantir o sincronismo entre os três canais de áudio. É proposto o uso de um recurso do sistema operacional Linux, em que os três sinais são virtualizados e aquisitados por meio da biblioteca de arquitetura de som Linux avançada (ALSA, *Advanced Linux Sound Architecture*) a uma frequência de amostragem de 48000 Hz . A frequência de amostragem escolhida é disponibilizada pela maioria dos microfones USB de mercado.

Na terceira etapa, é realizado o processamento dos sinais. Esse processamento resume-se a:

- Cálculo da correlação cruzada entre os sinais. Para tanto, é utilizado o método

- tradicional da correlação, ou seja, considerando em (9) $H_1[\omega] = H_2[\omega] = 1$, o que recai a implementação da equação (4);
- Estimação do deslocamento (em amostras) que maximiza a correlação cruzada, através da implementação da equação (5);
 - Conversão do deslocamento em amostras para um atraso temporal, com base na taxa de amostragem. Essa conversão é realizada pela divisão do atraso, estimado em amostras, pela frequência de amostragem;
 - Estimação do ângulo da fonte sonora. Esta estimativa é realizada por meio dos atrasos temporais, realizando a implementação da equação (27).

A avaliação do método é feita por meio da análise do erro, seguida pela avaliação de sua sensibilidade e, por último, através da aplicação em situações reais. A seguir, cada uma dessas avaliações será explicada em maior profundidade.

3.1 AVALIAÇÃO DO ERRO DO MÉTODO

O erro do método é avaliado através da análise em condições ideais, por meio de sinais sintéticos. Assume-se o conjunto de microfones posicionado na origem de um sistema bidimensional de coordenadas cartesianas. Define-se diferentes coordenadas nesse plano para posicionamento da fonte sonora. Para cada j -ésima posição, as distâncias entre a fonte e o i -ésimo microfone ($i = \{1, 2, 3\}$), d_{ij} , são calculadas. A partir delas, da frequência de amostragem (f_s) e da velocidade de propagação do som no ar (v), é possível calcular o atraso no tempo, em número de amostras (n_{ij}), decorrido entre a emissão do sinal de áudio na j -ésima posição e a chegada dele no i -ésimo microfone, conforme:

$$n_{ij} = \frac{d_{ij}}{v} f_s. \quad (34)$$

Uma determinada posição de fonte sonora pode então ser simulada gerando-se três versões de um sinal de áudio devidamente defasadas entre si.

No caso em estudo, a frequência de amostragem para a simulação é $f_s = 48000 \text{ Hz}$. Um sinal de áudio de cinco segundos de duração é utilizado para sofrer as devidas defasagens. Esse sinal foi gerado por meio da transição entre duas frequências. Iniciando em 600 Hz a frequência é incrementada linearmente até 800 Hz , com um tempo de transição de 3 segundos. Após chegar nos 800 Hz a frequência começa a ser decrementada linearmente. Escolheu-se este sinal, por estar disponível no aplicativo gerador de frequência utilizado. Qualquer sinal de áudio pode ser utilizado, desde que seja correlacionado.

Por meio dos sinais sintéticos obtidos é possível aplicar o método de identificação da posição da fonte sonora selecionado e verificar o desempenho da estimação sob condições ideais, isolando-se outras fontes de erro como: problemas de sincronismo entre a aquisição dos microfones, diferenças entre os sinais dada pela diferença na resposta em frequência dos microfones, interferências de outras fontes sonoras, reflexões do som no ambiente, entre outros.

3.2 AVALIAÇÃO DA SENSIBILIDADE DO MÉTODO

Nesta seção é proposta uma avaliação da sensibilidade do método a atrasos entre amostras, oriundos de eventuais problemas de sincronismo entre os canais de aquisição. Para tanto, são utilizados os mesmos sinais sintéticos descritos na Seção 3.1. Para a avaliação proposta, cada um dos sinais é atrasado, individualmente, em um número crescente de amostras, simulando um possível atraso em um dos canais de aquisição. Frente a esse cenário, avalia-se a variação do erro de estimação gerado pelo método.

3.3 AVALIAÇÃO EM CONDIÇÕES REAIS

Nesta avaliação utiliza-se como fonte sonora uma caixa som fabricada pela *House of Marley*, modelo *No Bounds XL*, com potência de 20W RMS ($2 \times 10W$). O reproduzido se trata do sinal caracterizado na Seção 3.1. A fonte sonora é, então, posicionada em diferentes posições, medidas em relação ao centro do triângulo de microfones, equidistantes a um metro. Os microfones são omnidirecionais USB da marca *Docooler*, modelo UM02. A implementação do *script* para aquisição se faz possível por meio de uma *Raspberry Pi 4* modelo B+ com sistema operacional Linux. Para cada posição uma aquisição de quatro segundos de áudio é realizada e posteriormente aplicada ao método para se estimar o ângulo. A avaliação se dá por meio da comparação dos ângulos reais, calculados a partir da posição da caixa de som, com os ângulos estimados pelo método.

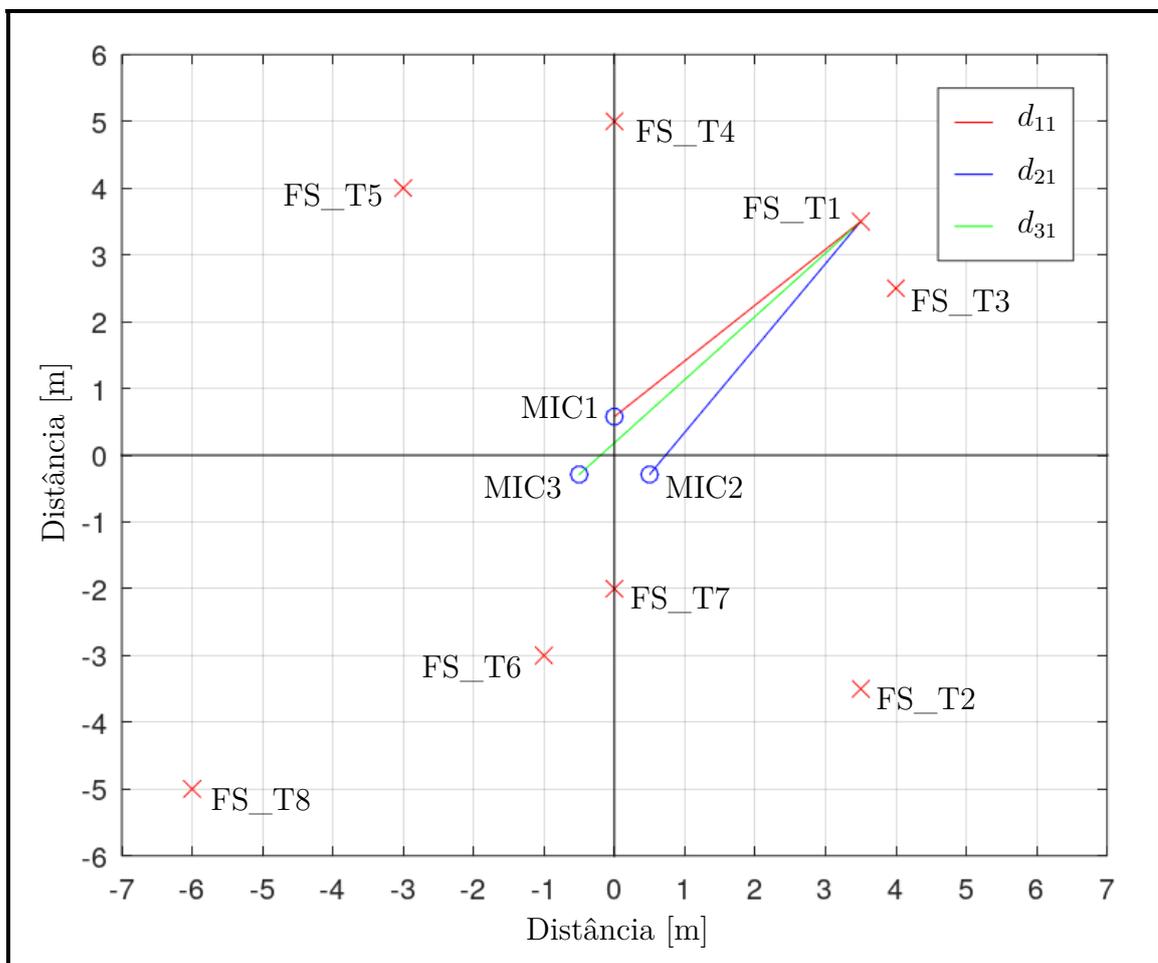
4 RESULTADOS

No decorrer deste capítulo são apresentados os resultados obtidos das avaliações propostas no capítulo anterior.

4.1 ERRO DO MÉTODO

Para realização da avaliação proposta na Seção 3.1, modelou-se um plano cartesiano com uma fonte sonora posicionada (sinteticamente) em oito coordenadas distintas, correspondendo aos pontos FS_T1 a FS_T8 ilustrados na Figura 5. As distâncias de cada um desses pontos até cada um dos microfones, d_{ij} , em que i é o índice do microfone, j é o índice da (posição da) fonte sonora e $\{1 \leq i \leq 3; 1 \leq j \leq 8\}$, foi determinada analiticamente, como *ground-truth*. Na Figura 5 são ilustradas as distâncias d_{11} , d_{21} e d_{31} entre o ponto FS_T1 e cada um dos três microfones.

Figura 5 – Fontes sonoras posicionadas sinteticamente no plano cartesiano



Fonte: o autor (2023)

De acordo com as posições das fontes sonoras e dos microfones, as distâncias foram convertidas em número de amostras, através de (34), assumindo-se $v = 340 \text{ m/s}$ e $f_s = 48.000 \text{ Hz}$. Na Tabela 2 são apresentados os resultados da conversão, assim como as distâncias da fonte em relação aos três microfones e o ângulo de cada fonte sonora. Os valores calculados do ângulo da fonte sonora em relação à origem do plano de referência, θ , são referentes ao eixo das abcissas, iniciando no primeiro quadrante, em sentido anti-horário. Por meio dos números de amostras n_{1j} , n_{2j} e n_{3j} , três sinais de

Tabela 2 – Informação das fontes sonoras posicionadas sinteticamente

	$\theta[^\circ]$	$d_{1j}[m]$	$d_{2j}[m]$	$d_{3j}[m]$	$n_{1j}[\text{amostras}]$	$n_{2j}[\text{amostras}]$	$n_{3j}[\text{amostras}]$
FS_T1	45°	4,56	4,83	5,51	644	682	778
FS_T2	315°	5,37	4,39	5,13	759	620	724
FS_T3	32°	4,44	4,48	5,29	627	632	747
FS_T4	90°	4,42	5,31	5,31	624	750	750
FS_T5	127°	4,55	5,54	4,96	643	782	701
FS_T6	251°	3,71	3,10	2,76	524	437	389
FS_T7	270°	2,58	1,78	1,78	364	252	252
FS_T8	220°	8,19	8,03	7,24	1156	1133	1022

Fonte: o autor (2023)

áudio, com cinco segundos de duração cada um, são criados a partir de um sinal cuja frequência varia de 600 Hz a 800 Hz , conforme discutido na Seção 3.1. Deste modo, são criados 24 sinais de áudio, ou seja, 3 sinais (referentes a MIC1, MIC2 e MIC3) para cada posição da fonte sonora no plano cartesiano. O *script* usado para plotar os pontos no plano cartesiano, calcular as distâncias, convertê-las em defasagens amostrais e gerar os sinais de áudio está disponível no repositório GitHub “TCC_Efrain Schmitt” com o nome “Gerador_Sinais_Sintéticos.m”. Para executar o *script* é necessário que na mesma pasta se encontre o arquivo “Sinal_Base.wav”, disponível para *download* no repositório.

Com os sinais sintéticos gerados para cada fonte sonora foi possível implementar o método e obter os ângulos calculados. A implementação do *script* foi feita na seguinte sequência:

- Implementação da equação (4) considerando $H_1[\omega] = H_2[\omega] = 1$;
- Implementação da equação (5);
- Conversão do deslocamento em amostras para um atraso temporal, usando a taxa

de amostragem;

- Implementação da equação (27).

O arquivo da implementação se encontra no mesmo repositório do GitHub, tendo como nome “Avaliação_Erro_Método.m”. O resultado dos ângulos estimados é apresentado na Tabela 3, assim como os ângulos verdadeiros (de referência) calculados a partir do plano cartesiano. Para obter o erro percentual, o valor absoluto da diferença entre os ângulos é calculada, sendo também apresentada na tabela. Para fins de comparação com os ângulos verdadeiros, θ , apresentados na Tabela 2, os ângulos estimados, $\hat{\theta}$, são convertidos usando como referência a posição em relação ao eixo y , ou seja,

$$\begin{cases} y > 0 & \Rightarrow \hat{\theta} \leftarrow 90 - \theta \\ y < 0 & \Rightarrow \hat{\theta} \leftarrow 270 - \theta \end{cases} \quad (35)$$

Tabela 3 – Resultados do método

Nº teste	θ [°]	$\hat{\theta}$ [°]	Erro absoluto [°]	Erro percentual [%]
FS_T1	45	45,97	0,97	0,27
FS_T2	315	315,99	0,99	0,28
FS_T3	32	32,11	0,11	0,03
FS_T4	90	90	0	0
FS_T5	127	125,46	1,54	0,43
FS_T6	251	249,46	1,54	0,43
FS_T7	270	270	0	0
FS_T8	220	219,24	0,76	0,21

Fonte: o autor (2023)

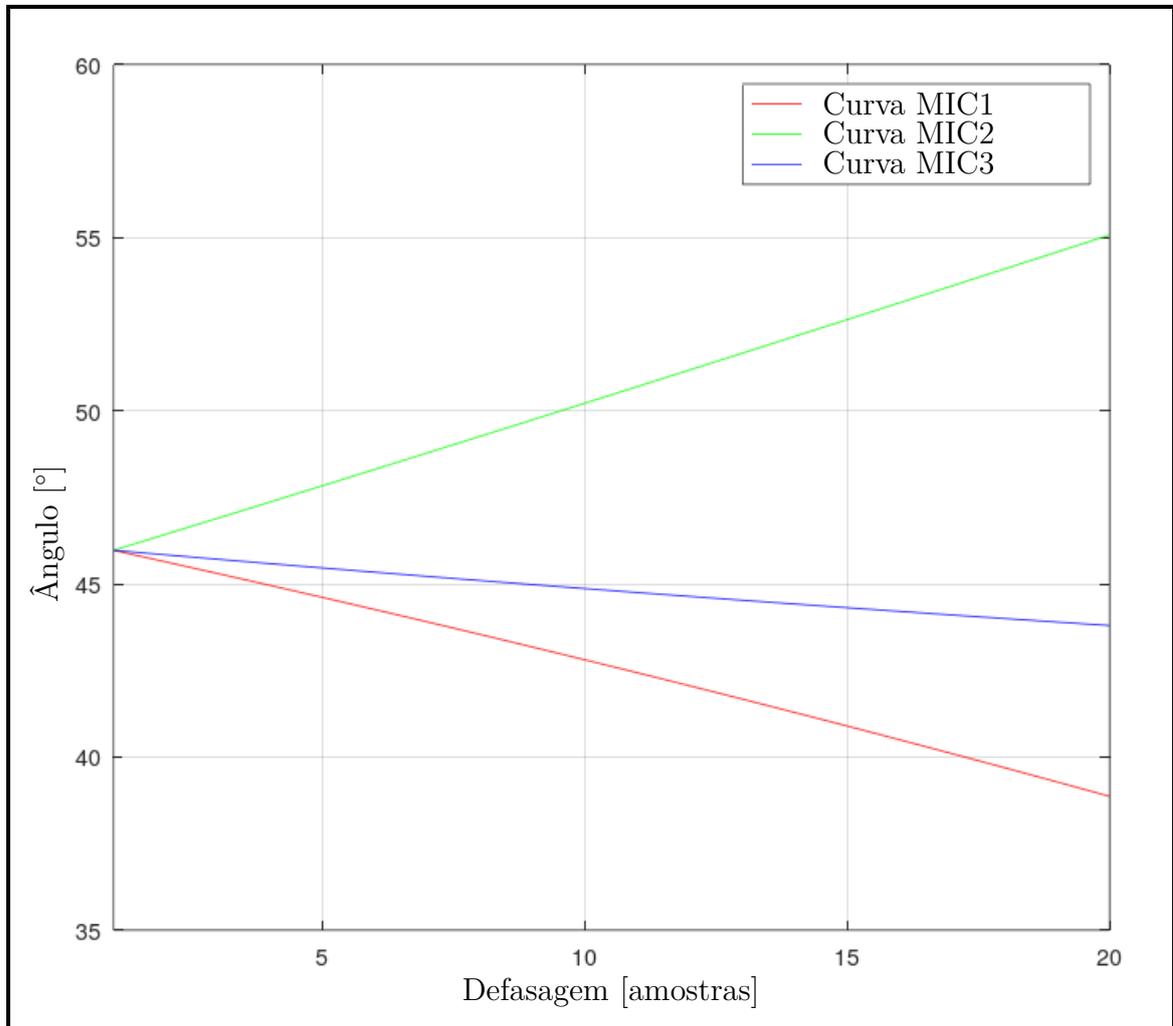
Portanto, é possível verificar que os erros numéricos do método em estudo apresentam-se inferiores a 0,5%, para um intervalo de identificação de 360°.

4.2 SENSIBILIDADE DO MÉTODO

Esta seção visa identificar a sensibilidade do método conforme proposto em 3.2. A análise de sensibilidade se dá sobre os sinais sintéticos gerados no item anterior, por meio da defasagem, forçosamente errônea, dos sinais de cada um dos três microfones. Na Figura 6 é mostrado o resultado da estimação do ângulo, assumindo-se a fonte so-

nora na posição FS_T1, frente a uma defasagem de até 20 amostras, ou seja, a consequência de uma eventual estimação errônea dos valores de n_{11} , n_{21} , n_{31} . O mesmo

Figura 6 – Sensibilidade aplicando defasagem de amostras no ponto FS_T1



Fonte: o autor (2023)

procedimento foi realizado com as demais pontos de teste, do FS_T2 ao FS_T8. O *script* usado para traçar os gráficos é disponibilizado no repositório com o nome “Avaliação_Sensibilidade_Método.m”. Na Tabela 4 são apresentados os erros em graus para uma defasagem de 20 amostras. Em que θ_{MIC1} , θ_{MIC2} , θ_{MIC3} , é o resultado do ângulo para uma defasagem errônea de 20 amostras no sinal do *MIC1*, *MIC2*, *MIC3*, respectivamente. Neste caso, tem-se erros abaixo de 3%, para uma condição de defasagem de 20 amostras.

A defasagem em amostras em aplicações reais é ocasionada pela aquisição dessincronizada dos canais de áudio (microfones), portanto a análise de sensibilidade em

Tabela 4 – Sensibilidade do método

Nº teste	θ [°]	θ_{MIC1} [°]	θ_{MIC2} [°]	θ_{MIC3} [°]	Máximo erro absoluto [°]	Erro percentual [%]
FS_T1	45	38,86	55,07	43,80	10,07	2,80
FS_T2	315	310,35	313,53	323,96	8,96	2,49
FS_T3	32	23,59	40,89	31,82	8,89	2,47
FS_T4	90	90,00	96,92	83,08	6,92	1,92
FS_T5	127	131,42	128,73	116,43	10,57	2,94
FS_T6	251	252,27	240,25	256,10	10,76	2,99
FS_T7	270	270,00	260,88	279,12	9,12	2,53
FS_T8	220	225,41	211,50	220,89	8,50	2,36

Fonte: o autor (2023)

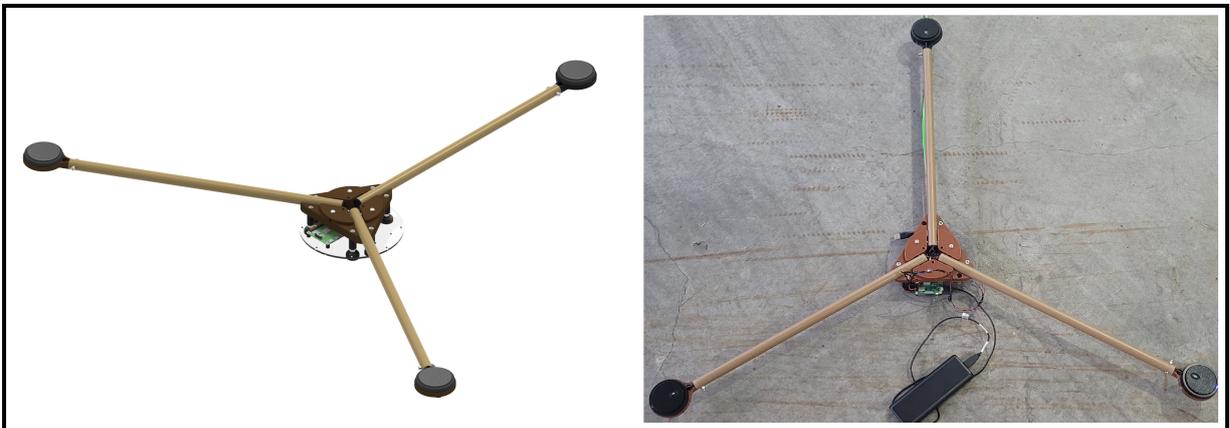
questão é de grande importância para a construção de projetos de hardware, servindo como balizador.

4.3 ERRO DO MÉTODO SOB CONDIÇÕES REAIS

Esta seção visa a avaliação do desempenho do método considerando-se sinais e condições reais de aplicação. Para isso três microfones foram afixados em uma disposição de triângulo, espaçados em um metro entre si, conforme descrito no Capítulo 3.

A Figura 7 mostra o modelo da estrutura de fixação, desenvolvido em software de CAD 3D, à esquerda, e o hardware montado, à direita. A construção das partes estruturais é constituída de manufatura aditiva, tubos de PVC e chapa de acrílico. Para

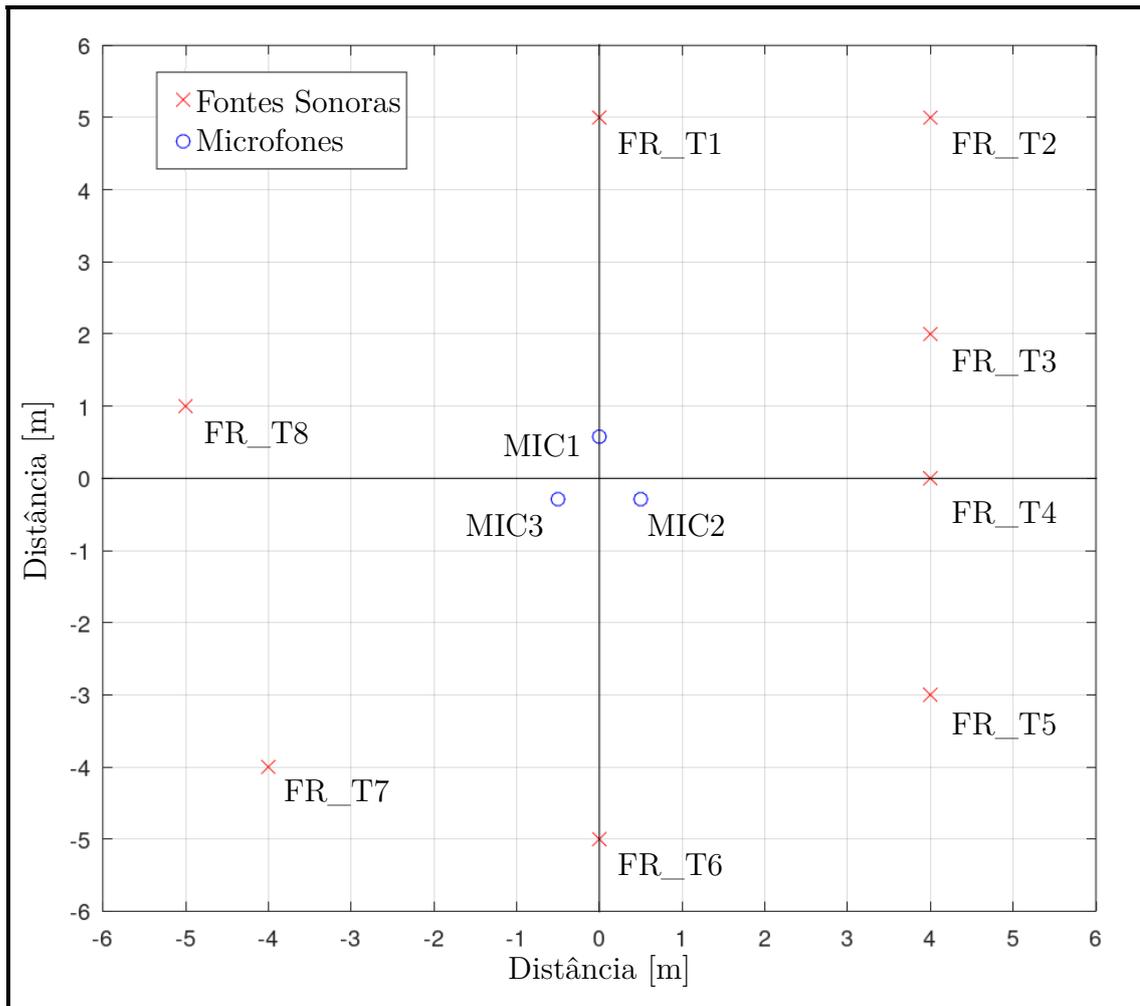
Figura 7 – *Hardware* de aquisição



Fonte: o autor (2023)

realizar a aquisição implementou-se um *script* que utiliza a biblioteca ALSA, nativa para aquisição de áudio no sistema operacional Linux. As aquisições foram realizadas em um ambiente aberto (terreno gramado) com baixas perturbações acústicas, utilizando o sinal sonoro descrito na Seção 3.3. A Figura 8 mostra as posições da fonte e dos microfones consideradas no ensaio.

Figura 8 – Disposição das fontes sonoras em condições reais



Fonte: o autor (2023)

Optou-se usar um tempo de aquisição do sinal considerado para a estimação de cada posição de quatro segundos, a uma frequência de amostragem de 48000 Hz , gerando assim sinais de áudio com 192.000 amostras. A aquisição usou o recurso da biblioteca ALSA para virtualização de canais, sendo necessário criar o arquivo “.asoundrc” no usuário do sistema operacional Linux. Esse arquivo realiza a configuração da virtualização de canais, dependendo do microfone utilizado, deve ser alterado. O seguinte comando no

console do sistema operacional deve ser executado para adquirir o áudio e utilizar da virtualização:

```
1 arecord -D array3 -d 4 -c 3 -r 48000 -f S16_LE ./name.wav
```

Nesse comando, “array3” é o nome atribuído a configuração de virtualização criada no arquivo “.asoundrc” , a informação após “-d” é a duração da aquisição, após “-c” é o número de canais adquiridos, após “-r” é a frequência de amostragem, após “-f” é o formato da amostra e “./name.wav” é o nome e a extensão do arquivo de saída.

Para analisar os sinais foi utilizado um *script* similar ao usado na avaliação do erro feita na Seção 4.1, diferenciando-se apenas pela modificação da leitura dos sinais, que, agora, a cada arquivo de saída possui os três sinais de áudio. Desta forma, após a leitura de cada arquivo de saída, os canais foram separados em MIC1, MIC2, MIC3 para serem analisados. Os sinais de áudio e o *script* estão disponibilizados no repositório, na pasta “Avaliação Condições Reais”. Os resultados sob condições reais de aplicação são mostrados na Tabela 5. Os valores dos ângulos, θ e $\hat{\theta}$, são referentes ao eixo das abcissas, iniciando no primeiro quadrante em sentido anti-horário.

Tabela 5 – Erro do método sob condições reais

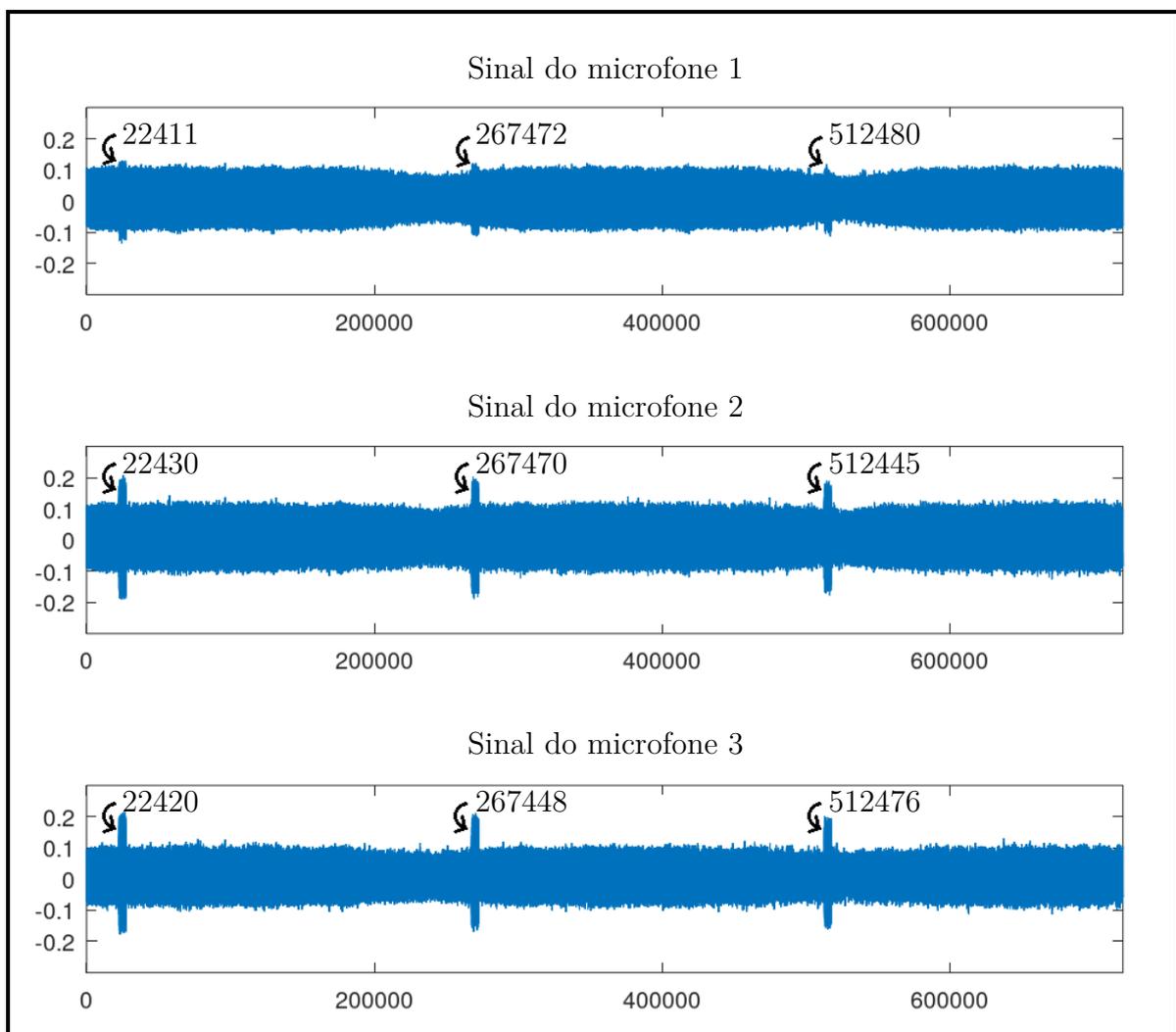
Nº teste	θ [°]	$\hat{\theta}$ [°]	Máximo erro absoluto [°]	Erro percentual [%]
FR_T1	90	22,3	67,7	18,8
FR_T2	51,3	176,9	125,6	34,9
FR_T3	26,5	91,8	65,3	18,1
FR_T4	0	62,1	62,1	17,3
FR_T5	323,1	260,1	63	17,5
FR_T6	270	305,4	35,4	9,8
FR_T7	225	238,6	13,6	3,8
FR_T8	168,7	57	111,7	31

Fonte: o autor (2023)

Em condições de ensaio que envolvem a virtualização de canais sem garantia de sincronismo, o método produz resultados adequados à uma prova de conceito. Isso constitui um sólido ponto de partida, mas há espaço para investigar formas de aprimoramento adicionais. Omitindo-se os valores de FR_T2 a FR_T8, que podem ter sido afetados por influências do ambiente, a maior porcentagem de erro observada é de 18,8%, equivalente a um ângulo de 67,7°. De forma a identificar se os erros obtidos se devem a problemas de

sincronismo na aquisição dos microfones, um *buzzer* foi adicionado ao centro da estrutura triangular de fixação mostrada na Figura 7. Uma lógica, em linguagem Python, foi implementada para acionar o *buzzer* por 50 *ms* a cada 5 *s*, durante o processo de aquisição. Um novo ensaio foi realizado no mesmo local e nas mesmas posições mostradas na Figura 8, porém agora com 15 *s* de aquisição. Para exemplificar o sinal do *buzzer*, na Figura 9 são traçados os sinais referentes ao ensaio da posição FR_T1 e indicada a primeira amostra de incidência referente a cada ocorrência do *buzzer*.

Figura 9 – Sinal RS_T1 no ensaio real com *buzzer*



Fonte: o autor (2023)

Através deste teste, observou-se que não há um padrão consistente. Ou seja, a virtualização não segue um erro de defasagem constante. O período entre cada *beep* emitido pelo *buzzer* dentro dos sinais avaliados varia. Isso impossibilita a correção da defasagem das amostras resultantes da virtualização, tornando o sincronismo dos sinais

difícil de ser realizado.

Com o objetivo de validar o método e solucionar os problemas de sincronismo identificados, foi utilizado um *notebook* com sistema operacional *Windows* como hardware de aquisição, e uma aquisição em múltiplos núcleos foi implementada através da abertura de *threads* em um *script* desenvolvido em linguagem Python.

O script desenvolvido está disponível no repositório como “Aquisição_Notebook”. Esse *script* abre três rotinas de aquisição distintas usando a biblioteca *threading*. Após cada aquisição, ele calcula a diferença de tempo entre o início e o fim da aquisição. Quando os valores das três *threads* são iguais, o sinal de áudio pode ser analisado, pois as aquisições dos três microfones foram realizadas simultaneamente.

Os sinais foram analisados usando o mesmo *script* de análise que foi utilizado anteriormente para avaliar o método em condições reais. A única diferença é que agora os sinais dos microfones são gravados em arquivos separados, em vez de estarem todos em um único arquivo. Assim a leitura dos sinais de áudio no *script* teve que ser alterada. Todos os arquivos se encontram no repositório na pasta “Aquisição Notebook”. Os resultados obtidos são apresentados na Tabela 6.

Tabela 6 – Erro do método sob condições reais (Aquisitado com *notebook*)

Nº teste	θ [°]	$\hat{\theta}$ [°]	Máximo erro absoluto [°]	Erro percentual [%]
FR_T1	90	90,3	0,1	0,1
FR_T2	51,3	70,6	19,3	5,4
FR_T3	26,5	4,2	22,3	6,2
FR_T4	0	31,9	31,9	8,9
FR_T5	323,1	315,1	8	2,2
FR_T6	270	285,8	15,8	4,4
FR_T7	225	217,4	7,6	2,1
FR_T8	168,7	172,4	3,7	1

Fonte: o autor (2023)

Agora o método apresenta resultados mais satisfatórios, com erros abaixo de 10%. Comparando os resultados da Tabela 6 com os da Tabela 5, percebe-se uma melhora significativa, com uma redução do máximo erro percentual de 34,9% para 8,9%.

5 CONSIDERAÇÕES FINAIS

Este trabalho propôs um estudo sobre localização de fonte sonora, aplicação que proporciona avanços significativos na redundância de sistemas, na segurança e na qualidade de vida das pessoas. Neste estudo o método GCC é abordado na sua origem.

Para se atingir o desenvolvimento de uma solução dedicada à localização automática de uma fonte sonora, definiu-se três objetivos específicos. O primeiro seleciona, a partir de uma revisão teórica, um método de processamento de sinais adequado à localização de uma fonte sonora. Verificou-se que diversas funções de ponderação já foram desenvolvidos para GCC, porém o trabalho aborda uma implementação sem o uso da ponderação, o que recai a implementação da correlação cruzada. Depois, para o segundo objetivo se faz necessária a implementação do hardware, com o intuito de adquirir amostras de áudio. Nesta etapa, verificou-se que a escolha de microfones USB não foi a mais adequada, pois necessitou uma aquisição em sistema operacional, que acaba operando com multi tarefas, sendo difícil realizar aquisições com sincronismo. Por fim, o último objetivo visa obter experimentos práticos para avaliar o método em estudo. A análise das aquisições permitiu concluir que o método apresenta resultados consistentes em condições reais, sendo um dos principais desafios a aquisição sincronizada dos sinais de áudio. Quando o sincronismo é alcançado, em condições ambientes com baixas perturbações acústicas, o erro percentual médio é de 3,8%, sendo esse um erro absoluto de $13,68^\circ$, para um intervalo de identificação de 360° . De toda maneira, os objetivos foram alcançados, tendo o método apresentado resultados concordantes com a literatura estudada.

Os *scripts* disponibilizados no repositório GitHub são aplicáveis a outras distâncias entre microfones, desde que os microfones fiquem posicionados em forma de triângulo, e equidistantes.

Em pesquisas futuras, pode-se avaliar a construção de um hardware dedicado a aquisição de multi-canais. Sendo esta configuração de hardware, com sincronismo entre canais de áudio, uma solução com poucos fornecedores no mercado, e normalmente com um alto valor agregado.

REFERÊNCIAS

- AGUIRRE, L. A. *Introdução à identificação de sistemas*. 2. ed. Belo Horizonte: Editora UFMG, 2004.
- CARTER, G.; NUTTALL, A.; CABLE, P. The smoothed coherence transform (scot). *Naval Underwater Systems Center, New London Lab., New London, CT, Tech. Memo TC-159-72*, 1972.
- CARTER, G.; NUTTALL, A.; CABLE, P. The smoothed coherence transform. *Proceedings of the IEEE*, v. 61, n. 10, p. 1497–1498, 1973.
- CATUR, H. A. H.; SAPUTRA, H. M. Azimuth estimation based on generalized cross correlation phase transform (gcc-phat) using equilateral triangle microphone array. *2019 International Conference on Radar Antenna Microwave Electronics and Telecommunications (ICRAMET)*, p. 89–93, 2019.
- Michael J. Delp. *AUTONOMOUS VEHICLE DETECTION OF AND RESPONSE TO EMERGENCY VEHICLES*. 2016. 9278689. Disponível em: <<https://patents.google.com/patent/US9278689B1/en>>.
- DESAI, D.; MEHENDALE, N. A review on sound source localization systems. *Archives of Computational Methods in Engineering*, p. 4631–4642, May. 2022.
- ECKART, C. Optimal rectifier systems for the detection of steady signals. 1952.
- ELLIS, D. Gammatone-like spectrograms. *Columbia University*, 2009. Disponível em: <<https://www.ee.columbia.edu/~dpwe/resources/matlab/gammatonegram/#5>>. Acesso em: 11 mar. 2023.
- FARMANI, M. *et al.* Informed sound source localization using relative transfer functions for hearing aid applications. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 25, n. 3, p. 611–623, 2017.
- GAO, S. *et al.* A modified frequency weighted music algorithm for multiple sound sources localization. *IEEE 23rd International Conference on Digital Signal Processing (DSP)*, p. 1–4, 2018.
- GENG, Y.; JUNG, J. Sound-source localization system for robotics and industrial automatic control systems based on neural network. In: *2008 International Conference on Smart Manufacturing Application*. [S.l.: s.n.], 2008. p. 311–315.
- HAMADA, N.; KASPRZAK, W.; PRZYBYSZ, P. Auditory scene analysis by time-delay analysis with three microphones. p. 1–8, 2012.
- HANNAN, E. J.; THOMSON, P. J. Estimating group delay. *Biometrika*, v. 60, n. 2, p. 241–253, 08 1973.

- HIOKA, Y.; MATSUO, M.; HAMADA, N. Multiple-speech-source localization using advanced histogram mapping method. *Acoustical Science and Technology*, v. 30, n. 2, p. 143–146, 2009.
- KNAPP, C.; CARTER, G. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 24, n. 4, p. 320–327, 1976.
- KOTUS, J. Application of passive acoustic radar to automatic localization, tracking and classification of sound sources. In: *2010 2nd International Conference on Information Technology, (2010 ICIT)*. [S.l.: s.n.], 2010. p. 67–70.
- LIAQUAT, M. U. *et al.* Localization of sound sources: A systematic review. *MDPI*, v. 14, n. 3910, Jun. 2021.
- MARCHEGANI, L.; NEWMAN, P. Listening for sirens: Locating and classifying acoustic alarms in city scenes. *IEEE Transactions on Intelligent Transportation Systems*, v. 23, n. 10, p. 17087–17096, Oct. 2022.
- MIAO, F. *et al.* A moving sound source localization method based on tdoa. *INTERNOISE 2014 - 43rd International Congress on Noise Control Engineering: Improving the World Through Noise Control*, 01 2014.
- NALON, J. A. *Introdução ao processamento digital de sinais*. 1. ed. Rio de Janeiro: LTC, 2009.
- OPPENHEIM, A. V.; SCHAFER, R. W. *PROCESSAMENTO EM TEMPO DISCRETO DE SINAIS*. 3. ed. São Paulo: Pearson Education do Brasil, 2013.
- PAPOULIS, A.; PILLAI, S. U. *Probability, Random Variables, and Stochastic Processes*. 4. ed. Boston: McGraw Hill, 2002.
- PARSAYAN, A.; AHADI, S. M. N-dimensional n-microphone sound source localization. *EURASIP Journal on Audio, Speech, and Music Processing*, v. 2013, n. 27, Dec. 2013.
- ROTH, P. R. Effective measurements using digital signal analysis. *IEEE Spectrum*, v. 8, n. 4, p. 62–70, 1971.
- SCHRÖDER, J. *et al.* Automatic acoustic siren detection in traffic noise by part-based models. *IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 493–497, Oct. 2013.
- WAN, X.; WU, Z. Sound source localization based on discrimination of cross-correlation functions. *Applied Acoustics*, v. 74, p. 28–37, Jan. 2012.
- ZHANG, Y. *et al.* Joint acoustic source localization algorithm based on summation and music algorithm for power equipment in substations. *2020 5th Asia Conference on Power and Electrical Engineering (ACPEE)*, p. 26–31, 2020.