



**UNIVERSITA' POLITECNICA DELLE MARCHE**  
**FACOLTA' DI INGEGNERIA**

---

Corso di Laurea magistrale in **Ingegneria Meccanica**

**Valutazione dell'accuratezza di algoritmi di processamento di segnali audio nel  
supportare l'interazione tra l'uomo e robot**

**Accuracy evaluation of audio signal processing to support interaction between a  
person and a mobile robot**

Relatore: Chiar.mo

**Prof. Ing. Gian Marco Revel**

Correlatrici:

**Dott. Ing. Sara Casaccia**

**Dott. Ing. Ilaria Ciuffreda**

Tesi di Laurea di:

**Agostino Piacente**

**A.A. 2020 / 2021**

*Alla mia famiglia, e a tutti coloro che mi hanno  
accompagnato in questo percorso*

## Sommario

Capitolo 1.....	8
1.1 Riassunto .....	8
1.2 Background letterario.....	9
1.2.1 Array signal processing.....	9
1.2.1.1 Steered Response Power (SRP) .....	12
1.2.1.2 Multiple Signal Classification (MUSIC).....	15
1.2.1.3 Coherent Signal Subspace Method (CSSM) e Weighted Average of Signal Subspaces (WAVES)....	16
1.2.1.4 Test of Orthogonality of Projected Subspace (TOPS).....	19
1.2.1.5 SRP, MUSIC, CSSM, WAVES, TOPS a confronto .....	21
1.2.1.6 Nostro contributo alla letteratura .....	21
1.2.2 Array signal processing applicato ai robot .....	22
Capitolo 2.....	25
2.1 Registrazione degli audio utilizzando il Robot Misty II.....	25
2.1.1 Materiali .....	25
2.1.1.1 Misty II.....	25
2.1.1.1.1 Array di microfoni.....	26
2.1.1.2 Casse altoparlanti .....	27
2.1.2 Testing room.....	28
2.1.3 Protocollo di misura .....	29
2.1.3.1 Protocollo di misura nell'ambiente reale .....	29
2.1.3.2 Protocollo di misura nell'ambiente controllato .....	31
2.2 Elaborazioni e valutazione dell'incertezza.....	33
2.2.1 Prima catena algoritmica.....	33
2.2.1.1 Prima catena algoritmica in ambiente reale .....	34
2.2.1.2 Prima catena algoritmica in ambiente controllato.....	36
2.2.2 Seconda catena algoritmica .....	37
2.2.3 Terza catena algoritmica .....	38
2.2.4 Quarta catena algoritmica .....	40
Capitolo 3.....	41
3.1 Risultati.....	41
3.1.1.1 Risultati ottenuti per la prima catena algoritmica nell'ambiente reale .....	42
3.1.1.2 Risultati ottenuti per la prima catena algoritmica nell'ambiente controllato .....	42
3.1.2 Risultati ottenuti dall'implementazione della seconda catena algoritmica.....	43
3.1.2.1 Risultati ottenuti per la seconda catena algoritmica nell'ambiente reale .....	43
3.1.3 Risultati ottenuti dall'implementazione della terza catena algoritmica .....	44

3.1.3.1 Risultati ottenuti per la terza catena algoritmica nell'ambiente reale .....	44
3.1.4 Risultati ottenuti dall'implementazione della quarta catena algoritmica.....	45
3.1.4.1 Risultati ottenuti per la quarta catena algoritmica nell'ambiente reale.....	45
3.1.5 Risultati relativi all'accuratezza dell'algoritmo Ina Speech .....	45
3.2 Analisi dei dati ottenuti .....	46
3.2.1 Analisi dei dati ottenuti: confronto degli algoritmi nei due ambienti.....	46
3.2.2 Analisi dei dati ottenuti: confronto catene algoritmiche .....	49
3.2.3 Analisi dei dati ottenuti: confronto distanze.....	50
3.2.3 Analisi dei dati ottenuti: confronto test .....	53
Capitolo 4.....	56
4.1 Conclusioni .....	56
4.1.1 Prestazioni ottimali al variare dell'algoritmo e della catena algoritmica utilizzati .....	56
4.1.2 Confronto prestazioni nei due diversi ambienti di prova.....	57
4.1.3 Prestazioni ottimali al variare delle condizioni di prova .....	58
4.1.4 Prestazioni ottimali al variare della distanza.....	58
4.2 Sviluppi futuri .....	59
5 Bibliografia.....	60
5 Bibliografia.....	61

## Indice delle figure

<b>Figura 1:</b> struttura generale di un beamformer .....	10
<b>Figura 2:</b> calcolo del ritardo temporale tra i microfoni l e q, che distano dalla sorgente $r^{(s)}_l$ e $r^{(s)}_q$ , per un'onda che si muove con velocità c. ....	13
<b>Figura 3:</b> errore quadratico medio per le stime non anomale in funzione di p per l'algoritmo proposto e quello tradizionale.....	14
<b>Figura 4:</b> percentuale di riduzione dell'onere computazionale in funzione di p per l'algoritmo proposto ....	14
<b>Figura 5:</b> errore di localizzazione del MUSIC in funzione del numero di onde d'onda incidenti .....	15
<b>Figura 6:</b> errore quadratico medio in funzione del rapporto segnale rumore per il Root-MUSIC (simulato e teorico) e l'ESPERIT.....	16
<b>Figura 7:</b> deviazione standard della stima in funzione del rapporto segnale rumore .....	17
<b>Figura 8:</b> probabilità di rilevazione delle due sorgenti distinte in funzione del rapporto segnale rumore ....	18
<b>Figura 9:</b> deviazione standard delle stime della sorgente a 33 gradi eseguite col WAVES (linea continua) e col BICCSM (linea tratteggiata).....	19
<b>Figura 10:</b> deviazione standard delle stime della sorgente a 37 gradi eseguite col WAVES (linea continua) e col BICCSM (linea tratteggiata).....	19
<b>Figura 11:</b> errore quadratico medio delle stime della sorgente a 33 gradi per gli algoritmi riportati in legenda .....	20
<b>Figura 12:</b> errore quadratico medio delle stime della sorgente a 33 gradi per gli algoritmi riportati in legenda .....	20
<b>Figura 13:</b> Misty II.....	25
<b>Figura 14:</b> geometria dell'array di microfoni presente sulla testa di Misty II .....	27

<b>Figura 15:</b> casse altoparlanti utilizzate per riprodurre la voce umana e il rumore domestico .....	27
<b>Figura 16:</b> ambiente di prova reale – stanza a quota 185 nell’Università Politecnica delle Marche .....	28
<b>Figura 17:</b> camera semi anecoica .....	29
<b>Figura 18:</b> configurazione delle registrazioni del test 1 nell’ambiente reale .....	30
<b>Figura 19:</b> configurazione delle registrazioni del test 2 nell’ambiente reale .....	30
<b>Figura 20:</b> configurazione delle registrazioni del test 3 nell’ambiente reale .....	31
<b>Figura 21:</b> configurazione delle registrazioni del test 1 condotte nell’ambiente controllato.....	32
<b>Figura 22:</b> configurazione delle registrazioni del test 2 nell’ambiente controllato .....	32
<b>Figura 23:</b> configurazione delle registrazioni del test 3 nell’ambiente controllato .....	33
<b>Figura 24:</b> schema funzionale della prima catena algoritmica.....	34
<b>Figura 25:</b> retta di taratura relativa alle stime ottenute dall’elaborazione delle registrazioni del test 3 nell’ambiente reale alla distanza di 0,5m tramite l’algoritmo MUSIC implementato secondo la prima catena algoritmica.....	35
<b>Figura 26:</b> retta di taratura relativa alle stime ottenute dall’elaborazione delle registrazioni del test 3 nell’ambiente controllato alla distanza di 1,5m tramite l’algoritmo SRP implementato secondo la prima catena algoritmica.....	36
<b>Figura 27:</b> schema funzionale della seconda catena algoritmica.....	37
<b>Figura 28:</b> input e output di Ina Speech .....	37
<b>Figura 29:</b> segmentazione dei segnali audio sulla base delle informazioni provenienti da Ina Speech .....	38
<b>Figura 30:</b> filtro butterworth del secondo ordine con frequenza di taglio pari a 400 e 2000 Hz.....	39
<b>Figura 31:</b> schema funzionale della quarta catena algoritmica .....	40
<b>Figura 32:</b> confidenza statistica con fattore di copertura due dei 5 algoritmi implementati secondo la prima catena algoritmica sulle registrazioni eseguite nell’ambiente controllato e reale alle distanze di 1,5m e 2,5m nelle condizioni di prova del test 1.....	47
<b>Figura 33:</b> confidenza statistica con fattore di copertura due dei 5 algoritmi implementati secondo la prima catena algoritmica sulle registrazioni eseguite nell’ambiente controllato e reale alle distanze di 1,5m e 2,5m nelle condizioni di prova del test 2.....	48
<b>Figura 34:</b> confidenza statistica con fattore di copertura due dei 5 algoritmi implementati secondo la prima catena algoritmica sulle registrazioni eseguite nell’ambiente controllato e reale alle distanze di 1,5m e 2,5m nelle condizioni di prova del test 3.....	49
<b>Figura 35:</b> valori della confidenza statistica con fattore di copertura due ottenuta dall’implementazione della prima catena algoritmica mediata sui test condotti.....	51
<b>Figura 36:</b> valori della confidenza statistica con fattore di copertura due ottenuta dall’implementazione della seconda catena algoritmica mediata sui test condotti.....	51
<b>Figura 37:</b> valori della confidenza statistica con fattore di copertura due ottenuta dall’implementazione della seconda catena algoritmica mediata sui test condotti.....	52
<b>Figura 38:</b> valori della confidenza statistica con fattore di copertura due ottenuta dall’implementazione della quarta catena algoritmica mediata sui test condotti .....	52
<b>Figura 39:</b> valori della confidenza statistica con fattore di copertura due mediati sulle distanze relativo all’implementazione della prima catena algoritmica .....	53
<b>Figura 40:</b> valori della confidenza statistica con fattore di copertura due mediati sulle distanze relativo all’implementazione della seconda catena algoritmica .....	54
<b>Figura 41:</b> valori della confidenza statistica con fattore di copertura due mediati sulle distanze relativo all’implementazione della terza catena algoritmica .....	54
<b>Figura 42:</b> valori della confidenza statistica con fattore di copertura due mediati sulle distanze relativo all’implementazione della quarta catena algoritmica.....	55

## Indice delle tabelle

<b>Tabella 1:</b> classificazione degli algoritmi usati nel seguente lavoro di tesi .....	12
<b>Tabella 2:</b> confronto caratteristiche algoritmi utilizzati nel seguente lavoro di tesi .....	21
<b>Tabella 3:</b> confidenza statistica con fattore di copertura 2 ottenuta implementando i 5 algoritmi secondo la prima catena algoritmica sulle registrazioni condotte nell'ambiente reale .....	42
<b>Tabella 4:</b> confidenza statistica con fattore di copertura 2 ottenuta implementando i 5 algoritmi secondo la prima catena algoritmica sulle registrazioni condotte nell'ambiente controllato.....	43
<b>Tabella 5:</b> confidenza statistica con fattore di copertura 2 ottenuta implementando i 5 algoritmi secondo la seconda catena algoritmica sulle registrazioni condotte nell'ambiente reale.....	43
<b>Tabella 6:</b> confidenza statistica con fattore di copertura 2 ottenuta implementando i 5 algoritmi secondo la terza catena algoritmica sulle registrazioni condotte nell'ambiente reale.....	44
<b>Tabella 7:</b> confidenza statistica con fattore di copertura 2 ottenuta implementando i 5 algoritmi secondo la prima catena algoritmica sulle registrazioni condotte nell'ambiente reale .....	45
<b>Tabella 8:</b> valori della confidenza statistica con fattore di copertura due mediati sui test condotti e sulle distanze.....	50



# Capitolo 1

## 1.1 Riassunto

Questo lavoro di tesi si basa sull'utilizzo di un robot sociale per l'identificazione di persone in un ambiente domestico tramite l'elaborazione di segnali audio registrati da microfoni collocati sulla testa del robot. A tal fine è stata calcolata l'incertezza di misura con la quale gli algoritmi di processamento dei segnali audio SRP, MUSIC, CSSM, TOPS, WAVES sono in grado di localizzare, stimando la direzione di arrivo del suono, la persona che sta parlando nella traccia audio.

In questo lavoro si è utilizzato il robot Misty II. Dotato di tre microfoni omnidirezionali a campo lontano Misty II ha consentito di eseguire le registrazioni di tracce audio che riproducono una voce umana.

Le registrazioni sono state condotte sia in ambiente reale, nella stanza Q185\_012 dell'Università Politecnica delle Marche, sia in un ambiente controllato, la camera semi anecoica presente nella stessa struttura dell'ambiente reale, e sono state eseguite in diverse condizioni di prova al fine di riprodurre condizioni riscontrabili in un reale ambiente domestico.

I segnali audio registrati nelle condizioni descritte nel capitolo 2 sono stati processati dagli algoritmi SRP, MUSIC, CSSM, TOPS, WAVES, già presenti in letteratura, ai quali è stato aggiunto un sistema implementato ad hoc nel quale si è effettuato un pre-processing dei segnali audio atto a rimuovere le componenti del segnale non contenenti la voce umana nel dominio del tempo e della frequenza.

Le stime sull'angolo di arrivo del suono ottenute dopo l'elaborazione del segnale audio con gli algoritmi e con il sistema di pre-processing sono state confrontate con i valori noti di queste ottenuti sviluppando un protocollo di misura, così da valutare l'incertezza dell'intero sistema testato nelle diverse condizioni di prova.

**A seguito dei risultati ottenuti si è concluso che l'algoritmo che ha esibito la minore incertezza è stato, con un valore medio della confidenza statistica con fattore di copertura due pari al 6,17%, l'SRP quando questo è stato applicato ad un segnale audio che ha subito in pre-processing una segmentazione temporale atta a rimuovere gli istanti temporali in cui non risulta presente nell'audio considerato la voce umana.**



## 1.2 Background letterario

### 1.2.1 Array signal processing

Il presente lavoro di tesi è collocato nell'ambito dell'array signal processing.

L'array signal processing è quel campo di ricerca che studia come i dati ricevuti da un insieme di microfoni, nel quale è possibile eseguire un campionamento sia temporale che spaziale del campo sonoro, possano essere analizzati per produrre stime dei parametri quali il numero di sorgenti sonore e il DOA (direzione di arrivo delle onde sonore) delle sorgenti sonore che hanno generato il campo sonoro ricevuto dall'array.

Il DOA è un parametro di fondamentale importanza in quanto rappresenta la direzione dalla quale le onde sonore arrivano all'array di microfoni. La conoscenza di questo parametro corrisponde dunque alla conoscenza della posizione angolare della sorgente rispetto all'array di microfoni e consente così di eseguire la localizzazione della sorgente.

A partire dal 1948 alcuni lavori scientifici [1] hanno iniziato a trattare di sistemi basati sul filtraggio spaziale e beamforming alla base dei sistemi radar e sonar.

L'introduzione di una nuova metodologia di processamento dei dati in uscita dall'array di sensori, chiamata stima basata sul sottospazio [2], ha fatto in modo che la stima delle posizioni angolari delle sorgenti sonore, a partire dai dati raccolti da un array di sensori, crescesse esponenzialmente d'interesse scientifico portando i ricercatori a proporre un numero molto elevato di tecniche di stima finalizzate ad operare nel miglior modo possibile in determinate condizioni di utilizzo, come diverse condizioni di riverbero e di rumore ambientale, diverso numero di sorgenti da localizzare, diversi array di sensori disponibili o diverso onere computazionale ammissibile.

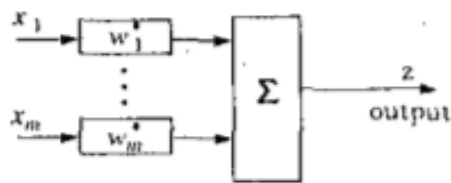
Il modo più naturale per classificare le varie tecniche di stima della localizzazione delle sorgenti sonore o, più in generale, dei parametri delle sorgenti sonore, è effettuare una classificazione basata sul modo operandi degli algoritmi che prendono come ingresso i dati in uscita dall'array di microfoni [3]. Considerato questo background, gli algoritmi utilizzati in tale lavoro di tesi possono essere suddivisi in due macrogruppi:

- 1) tecniche di beamforming
- 2) metodi basati sul sottospazio del segnale e del rumore.

Nelle sezioni seguenti i due macrogruppi sono spiegati più in dettaglio.

### **1) Tecniche di beamforming**

Come evidenziato dalla figura 2, nella quale viene riportata la struttura generale di un beamformer, il beamforming è definito come un processo di filtraggio e somma che, applicando un filtro temporale ai segnali dei microfoni prima di sommarli, produce in uscita un unico segnale focalizzato [7]. In figura 1 vengono indicati con  $x_1, \dots, x_m$  i segnali ricevuti dai microfoni e con  $w_1, \dots, w_m$  i filtri applicati a tali segnali.



*Figura 1: struttura generale di un beamformer*

Il beamforming ha una duplice utilità in quanto può essere applicato sia per catturare al meglio il segnale della sorgente sonora, sia per localizzarla. Nel primo caso la posizione della sorgente è nota e l'obiettivo del processo di filtraggio è utilizzare tale informazione per amplificare il segnale ricevuto da una determinata direzione. Nel secondo caso invece il beamformer può essere utilizzato, come vedremo, per generare un'opportuna funzione dalla cui massimizzazione si può eseguire la localizzazione della sorgente.

L'algoritmo utilizzato in questo lavoro di tesi appartenente alla prima tecnica è l'SRP (Steered Response Power). Nei prossimi capitoli verrà analizzato e spiegato come l'algoritmo SRP è una tecnica di beamforming nella quale il filtraggio temporale applicato risulta essere il calcolo della cross-correlazione per versioni del segnale shiftate nel tempo.

### **2) Metodi basati sul sottospazio**

Questi metodi utilizzano le proprietà algebriche della matrice di covarianza dei dati registrati che mostrano come, una volta eseguita la sua decomposizione auto vettoriale, possano essere generati due sottospazi (del segnale e del rumore) tra loro ortogonali contenenti tutte le informazioni relative alle caratteristiche del campo sonoro presente e delle sorgenti, consentendo così di eseguire una stima dei parametri di quest'ultime [4].

Già Pisarenko nel 1973 [5] aveva introdotto una metodologia di analisi basata sull'analisi dei sottospazi del segnale e del rumore ma è con l'introduzione dell'algoritmo MUSIC (Multiple Signal

Classification) nel 1986 [6] che si è visto nascere un interesse crescente degli approcci basati sul metodo del sottospazio portando alla luce molteplici tipologie di tecniche che, differenziandosi per le loro caratteristiche, chiedono dunque un'ulteriore classificazione.

Come osservato da H. Wang and M. Kaveh in [4] gli algoritmi appartenenti a tale classe possono essere sotto classificati in:

**2.1) Metodi a banda stretta:** la variabilità della risposta dell'array in funzione della frequenza non viene considerata. Questi metodi, come osservato in [7] sono ulteriormente suddivisibili in

2.1a) Metodi a banda stretta basati sul sottospazio del rumore: una volta eseguita l'auto decomposizione della matrice di covarianza dei dati, questi metodi eseguono la stima dei parametri della sorgente utilizzando solo le informazioni contenute nel sottospazio del rumore e non risulta necessario ricreare il sottospazio del segnale. Tra gli algoritmi utilizzati in questo lavoro di tesi il MUSIC entra a far parte di questa classe.

2.1b) Metodi a banda stretta basati sul sottospazio del rumore e del segnale: una volta eseguita l'auto decomposizione della matrice di covarianza dei dati, questi metodi eseguono la stima dei parametri della sorgente utilizzando sia le informazioni contenute nel sottospazio del segnale sia quelle contenute nel sottospazio del rumore.

2.1c) Metodi a banda stretta basati sul sottospazio del segnale: una volta eseguita l'auto decomposizione della matrice di covarianza dei dati questi metodi eseguono la stima del numero e delle posizioni delle sorgenti presenti utilizzando solo le informazioni contenute nel sottospazio del segnale e, pertanto, non risulta necessario ricreare il sottospazio del rumore.

**2.2) Metodi a banda larga:** considerano la variabilità della risposta dell'array in funzione della frequenza. Come osservato in [18] questi sono ulteriormente suddivisibili in:

2.2a) Metodi a banda larga incoerenti: mediante la trasformata di Fourier suddividono l'uscita dell'array in tante bande strette di frequenza a ciascuna delle quali applicano un metodo a banda stretta per poi ottenere la stima finale mediando i risultati ottenuti dalle singole bande.

2.2b) Metodi a banda larga coerenti: una volta eseguita la decomposizione del segnale in tante bande strette combinano opportunamente le matrici di covarianza relative alle singole

bande per ottenere una matrice di covarianza unica (detta universale) dalla quale è possibile eseguire la stima dei parametri delle sorgenti applicando un metodo a banda stretta. Gli algoritmi CSSM (Coherent Signal Subspace Method) e il WAVES (Weighted Average of Signal Subspaces), utilizzati in questo lavoro di tesi, appartengono a questa categoria.

L'ultimo algoritmo utilizzato in questo lavoro di tesi, il TOPS (Test of Orthogonality of Projected Subspace), rientra nella categoria dei metodi basati sul sottospazio a banda larga. Prestando delle caratteristiche intermedie tra i metodi coerenti e incoerenti e operando in maniera diversa da entrambi, non può essere ritenuto appartenente a nessuna delle due classi [18].

### **3) Metodi parametrici**

In letteratura è presente una terza categoria di tecniche di stima dei parametri delle sorgenti: metodi parametrici. In questo lavoro di tesi nessun algoritmo appartenente a questa metodologia verrà utilizzato.

In virtù della classificazione sopra proposta nella tabella 1 è riportata la suddivisione degli algoritmi utilizzati.

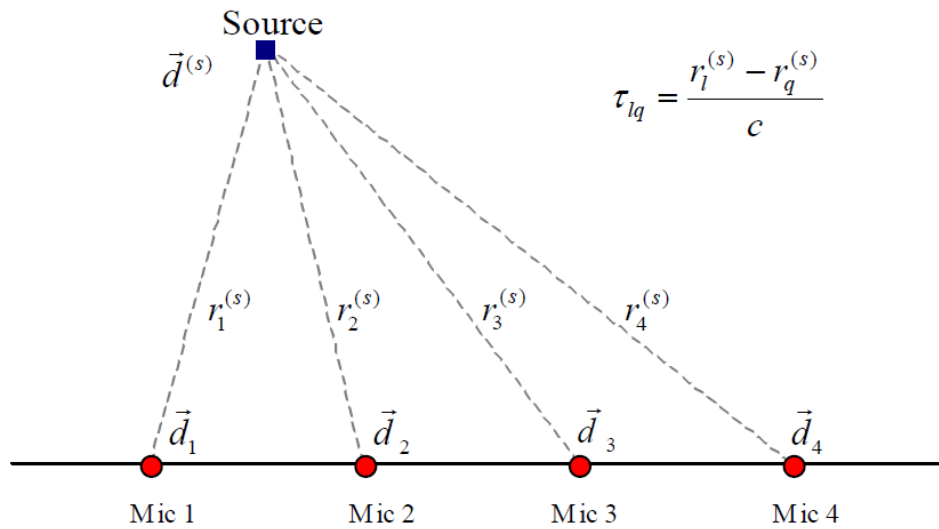
Tecniche di beamforming (SRP)		
Metodi basati sul sottospazio del segnale e del rumore	a banda stretta: non considerano la variabilità della risposta dell'array con la frequenza	segnale rumore (MUSIC) segnale e rumore
	a banda larga (TOPS): considerano la variabilità della risposta dell'array con la frequenza	coerenti (WAVES, MUSIC) incoerenti

**Tabella 1:** classificazione degli algoritmi usati nel seguente lavoro di tesi

#### 1.2.1.1 Steered Response Power (SRP)

Il metodo di localizzazione della sorgente sonora SRP consta di due fasi: una prima fase di discretizzazione e una seconda fase di ricerca. La prima fase prevede di discretizzare lo spazio di ricerca mediante una griglia spaziale e di associare a ciascuna posizione della griglia un vettore contenente tutti i ritardi relativi tra le coppie uniche dei microfoni per la sorgente che si ipotizza essere collocata in una posizione dello spazio. La figura 2 mette in evidenza come, conoscendo la

configurazione dell'array di sensori si possa associare a ciascuna posizione spaziale nella quale la sorgente potrebbe potenzialmente trovarsi, un insieme di ritardi temporali.



**Figura 2:** calcolo del ritardo temporale tra i microfoni  $l$  e  $q$ , che distano dalla sorgente  $r_l^{(s)}$  e  $r_q^{(s)}$ , per un'onda che si muove con velocità  $c$ .

La seconda fase di ricerca prevede dapprima di calcolare la funzione di cross-correlazione per i ritardi fisicamente ammissibili associati a ciascuna coppia di microfoni. In seguito, sommando tale funzione su tutte le coppie uniche di microfoni, si ottiene la funzione SRP dalla cui massimizzazione si deduce il vettore dei ritardi associato alla posizione reale della sorgente e, di conseguenza, la sua stima della posizione tra quelle presenti nella griglia [8].

Nel lavoro di Di Biase [8] viene riportata una chiara analisi del modo d'operare dell'algorithm SRP (descritto in precedenza). Inoltre, in [8] vengono validate le performance dell'algorithm su un array planare costituito da 8 microfoni in una sala conferenze mediamente riverberante e in presenza di un elevato rapporto segnale rumore mostrando un errore medio nella localizzazione della sorgente sonora di 10 gradi.

Successivamente il principale aspetto sul quale si è incentrata la letteratura ha riguardato la più grossa limitazione dell'algorithm SRP, costituita da un elevato onere computazionale. Si trovano dunque molti lavori che introducono metodologie finalizzate a rendere tale algorithm più leggero così da poter essere applicato in real time. Un contributo importante in questo ambito è stato dato da J. P. Dmochowski e J. Benesty che in [9] hanno proposto una generalizzazione del metodo attraverso l'introduzione di un parametro che permette di ridurre l'insieme dei ritardi sui quali si va a calcolare la mappa dell'SRP, considerando come validi solo quelli che presentano un valore elevato

della cross correlazione. Il metodo proposto in [9] è stato poi confrontato, per diversi valori del parametro di riduzione dello spazio di ricerca  $p$ , con l'algoritmo tradizionale mediante sperimentazioni che hanno mostrato che per un valore di  $p=8$  si riesce a ridurre del 93% l'onere computazionale sacrificando l'accuratezza solo per lo 0,07% in termini di errore quadratico medio valutato sulle stime non anomale (stime che si scostano dal valore reale della direzione di arrivo per meno di 5 gradi) della localizzazione della sorgente. Le figure 3 e 4 riportano i risultati della sperimentazione.

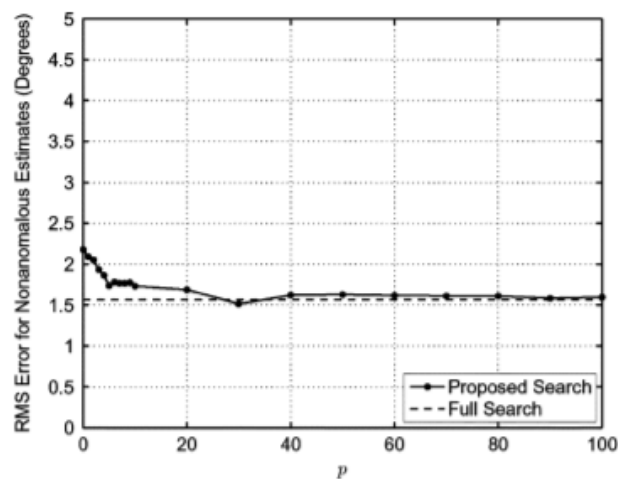


Figura 3: errore quadratico medio per le stime non anomale in funzione di  $p$  per l'algoritmo proposto e quello tradizionale

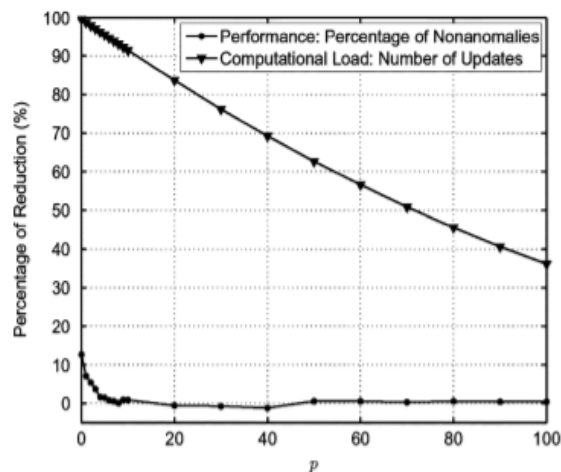


Figura 4: percentuale di riduzione dell'onere computazionale in funzione di  $p$  per l'algoritmo proposto

Va evidenziato come le sperimentazioni condotte in [9] abbiano visto un pre-processamento del segnale audio finalizzato a rimuovere dagli audio registrati la parte temporale non contenente il segnale vocale che si desiderava localizzare, un simile pre-processamento è stato condotto anche in questo lavoro di tesi.

### 1.2.1.2 Multiple Signal Classification (MUSIC)

L'algoritmo MUSIC venne formalizzato per la prima volta nel lavoro di Schmidt del 1986 [6] dove il suo modo di operare viene dettagliatamente descritto. Raccolti i dati in uscita dall'array di sensori si calcola la matrice di covarianza di questi ultimi e si passa poi alla decomposizione auto vettoriale di questa matrice. Tale passaggio risulta fondamentale in quanto dalla molteplicità del minimo autovalore verrà stimato il numero D delle sorgenti presenti mentre dagli M-D (con M numero dei sensori) auto vettori associati agli M-D autovalori più grandi si potrà generare il sottospazio del rumore sulla base del quale, sfruttando la relazione di ortogonalità col sottospazio del segnale, si potrà scrivere lo spettro spaziale del MUSIC i cui D massimi saranno indicativi della stima della direzione di arrivo delle sorgenti.

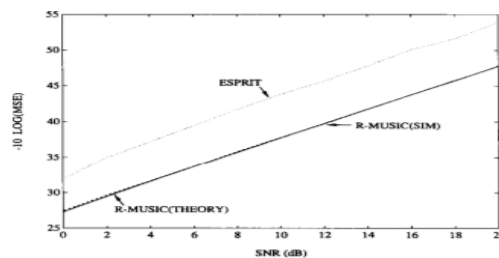
In letteratura molti autori si sono soffermati sull'analisi delle performance del metodo. Un lavoro interessante che è andato ad analizzare questo aspetto è stato quello condotto da Z. I. Khan, M.D. Kamal e K. Othman in [10] dove, su un array lineare di 15 sensori, sono state eseguite simulazioni per valutare l'accuratezza dell'algoritmo (valutata come differenza tra valore stimato e reale diviso valore reale) in funzione del numero di onde incidenti (da 1 a 5) ottenendo i risultati riportati in figura 5.

Number of Incident Signals	DOA	MUSIC	% of Error
1	20°	20.00°	0%
2	-15°	-14.00°	7%
	20°	20.00°	0%
3	-15°	-14.00°	7%
	10°	10.00°	0%
	30°	30.00°	0%
4	-25°	-26.00°	4%
	-15°	-16.00°	7%
	20°	20.00°	0%
	70°	70.00°	0%
5	-25°	-24.00°	4%
	-15°	-14.00°	7%
	20°	20.00°	0%
	50°	50.00°	0%
	70°	70.00°	0%

*Figura 5: errore di localizzazione del MUSIC in funzione del numero di onde d'onda incidenti*

Molteplici sono stati inoltre gli autori che hanno proposto modifiche nell'algoritmo per renderlo più performante in determinate condizioni [11]-[12]-[13]. In [11] vengono valutate le performance di

una variante dell'algoritmo MUSIC applicabile solo ad array lineari con sensori uniformemente distanziati nominata Root-Music nella quale la localizzazione è eseguita calcolando gli zeri di un polinomio costruito sulla base degli autovettori che generano il sottospazio del rumore. Per validare l'efficacia dell'algoritmo è stata dunque condotta sia un'analisi teorica che una simulazione nelle quali si è confrontato l'errore quadratico medio dell'algoritmo introdotto col già scientificamente noto ESPERIT [14], il tutto in funzione del rapporto segnale rumore. Tale analisi ha portato ai risultati mostrati in figura 6 dove si può notare la migliore accuratezza del Root-MUSIC in ogni condizione di rapporto segnale rumore.



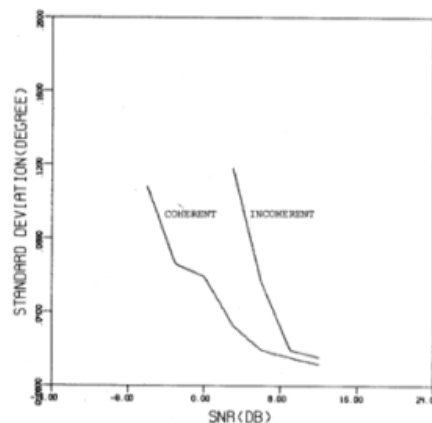
*Figura 6: errore quadratico medio in funzione del rapporto segnale rumore per il Root-MUSIC (simulato e teorico) e l'ESPRIT*

1.2.1.3 Coherent Signal Subspace Method (CSSM) e Weighted Average of Signal Subspaces (WAVES)  
 Nella letteratura del processamento dei segnali audio registrati da un array di sensori il termine CSSM assume un duplice significato: una metodologia di analisi dei segnali a banda larga e un preciso algoritmo.

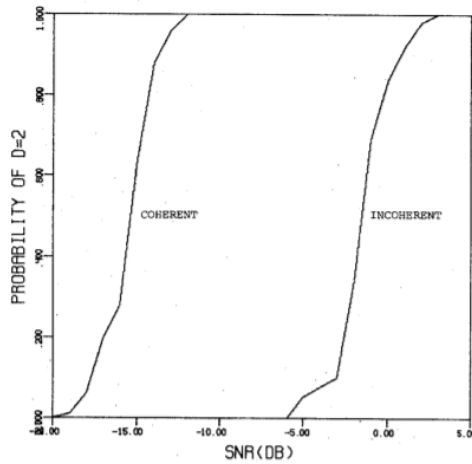
Come già accennato nella classificazione, si può affermare che appartengono alla classe CSSM gli algoritmi che, attraverso la trasformata di Fourier discreta, vanno dapprima a scomporre un segnale a banda larga in tante bande strette di frequenza per poi calcolare la matrice di covarianza spaziale di ogni singola banda, per poi ricondurre tali matrici ad un'unica frequenza di riferimento mediante un'opportuna matrice di trasformazione (la cui costruzione richiede una stima preliminare delle direzioni di arrivo delle onde incidenti) e infine calcolare, dalla media delle matrici di covarianza delle singole bande trasformate, un'unica matrice di covarianza (detta universale) che soddisfa il modello del segnale a banda stretta. In questo modo si possono ricavare i sottospazi del segnale e del rumore associati alla matrice di covarianza universale ed eseguire la stima dei parametri delle sorgenti applicando un metodo a banda stretta [4].



Questa metodologia di analisi dei segnali a banda larga introdotta da Wang e Kaveh in [4] è stata poi soggetta, nel loro stesso lavoro, ad una validazione sperimentale per la quale è stato applicato un preciso algoritmo, che prende appunto il nome di CSSM, che vede la matrice di trasformazione generata a partire da una stima delle direzioni di arrivo delle sorgenti ottenuta tramite il periodogramma spaziale [15] ed eseguire la stima finale dei parametri delle sorgenti attraverso l'algoritmo MUSIC. L'analisi sperimentale eseguita in [4] ha utilizzato un array lineare di 16 sensori uniformemente distribuiti per confrontare la deviazione standard della stima e la probabilità di riconoscimento, in funzione del rapporto segnale rumore, di due distinte sorgenti a banda larga distanziate angularmente di 3 gradi, tra l'algoritmo proposto e un algoritmo appartenente alla classe dei metodi del sottospazio a banda stretta incoerente introdotto da M. Wax e T. Kailath in [16] ottenendo i risultati riportati nelle figure 7 e 8. Possiamo osservare come la metodologia introdotta, presentando sempre una deviazione standard minore, risulti più accurata del metodo incoerente presentato in [16] e che rispetto a quest'ultimo è in grado di operare in condizioni in cui il rapporto segnale rumore anche molto basso.



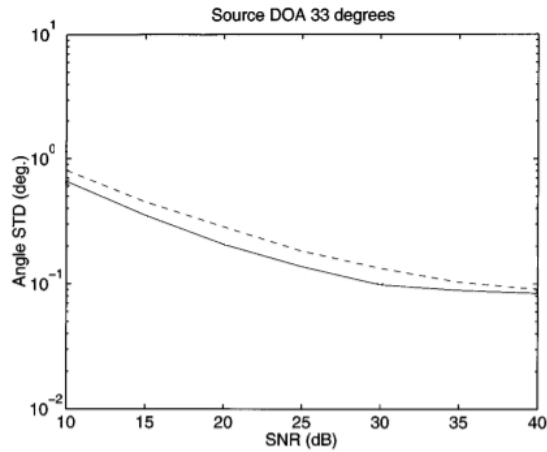
*Figura 7: deviazione standard della stima in funzione del rapporto segnale rumore*



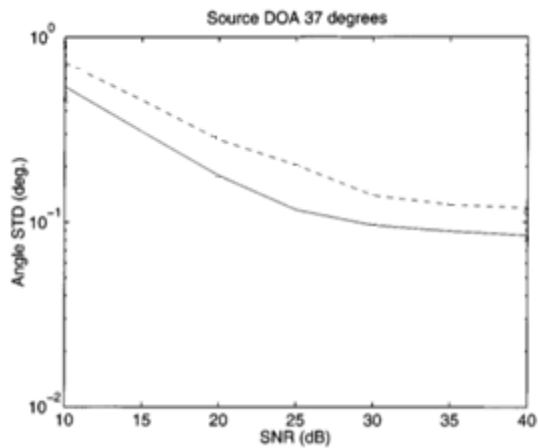
*Figura 8: probabilità di rilevazione delle due sorgenti distinte in funzione del rapporto segnale rumore*

Un importante contributo alla letteratura dei metodi basati sul sottospazio a banda larga coerenti è stato fornito da E. D. Di Claudio e R. Parisi che in [17] hanno proposto e valutato l'algoritmo WAVES (Weighted Average of Signal Subspaces). Come mostrato in [17] questo algoritmo, rispetto al CSSM, aggiunge un'ulteriore matrice costruita sulla base degli autovalori delle matrici di covarianza delle singole bande strette. Questa matrice viene quindi utilizzata, assieme alla matrice di trasformazione già presente nel CSSM, nel processo di media che porta ad ottenere la matrice di covarianza universale. Tale procedura consente di ponderare le singole matrici di covarianza delle varie bande strette sulla base dell'entità del segnale e del rumore presenti in ciascuna di esse per ottenere un algoritmo più robusto, ossia in grado di fornire stima più accurate in un ampio range di rapporto segnale rumore.

Tale conclusione è stata validata dalle simulazioni Montecarlo condotte in [17] che hanno visto il WAVES e il BI-CSSM (Beamforming Invariance Coherent Signal Subspace Method, variante del CSSM proposta e validata in [18]) implementato su un array lineare uniforme composto da 8 sensori per valutare la deviazione standard delle stime di due sorgenti, posizionate rispettivamente a 33 e 37 gradi, in funzione del rapporto segnale-rumore (figure 9 e 10).



**Figura 9:** deviazione standard delle stime della sorgente a 33 gradi eseguite col WAVES (linea continua) e col BICCSM (linea tratteggiata)



**Figura 10:** deviazione standard delle stime della sorgente a 37 gradi eseguite col WAVES (linea continua) e col BICCSM (linea tratteggiata)

#### 1.2.1.4 Test of Orthogonality of Projected Subspace (TOPS)

Questo algoritmo compare in letteratura nel 2006 a seguito del lavoro di Y. S. Yoon, L. M. Kaplan e J. H. McClellan [19]. In [19] viene mostrato il modo in cui opera il TOPS. Utilizza la matrice di trasformazione già introdotta in [17] e [18] per operare sui sottospazi del segnale e del rumore delle singole bande una trasformazione dipendente dagli angoli di arrivo delle sorgenti (per ora solo ipotizzati) andando ad eseguire la stima delle direzioni di arrivo sulla base dei valori per cui la trasformazione rende tali sottospazi ortogonali. Quest'ultima condizione si traduce nella perdita di rango di un'opportuna matrice generata a partire dalla matrice di trasformazione e dalle matrici del sottospazio del segnale e del rumore delle singole bande strette.

Sempre in [19] sono state condotte simulazioni per valutare le performance del TOPS e confrontarle con quelle dei metodi a banda larga già presenti, sia coerenti che incoerenti.

In [19] sono quindi stati implementati il TOPS, il CSSM, il WAVES e l' IMUSIC (metodo incoerente introdotto in [16]) su un array lineare di 10 sensori per localizzare due sorgenti a banda larga posizionate rispettivamente a  $33^\circ$  e  $37^\circ$  ed è stato valutato l'errore quadratico medio delle stime in funzione del rapporto segnale rumore (figure 11 e 12) dal cui andamento si può concludere che il TOPS si comporta meglio degli altri metodi nei range di rapporto segnale rumore intermedi (da 0 a 10 dB).

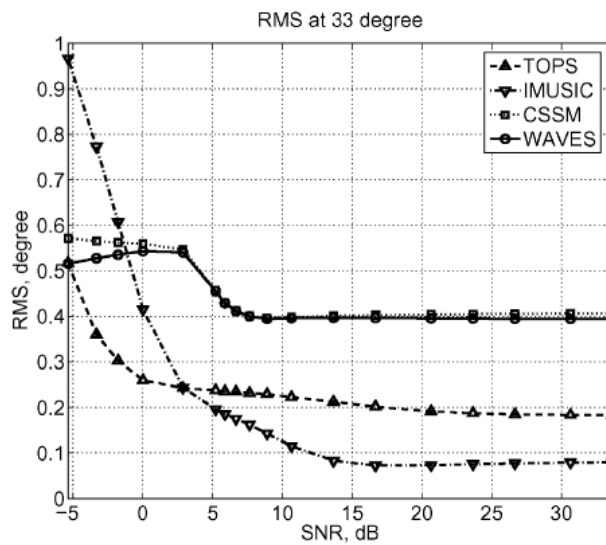


Figura 11: errore quadratico medio delle stime della sorgente a 33 gradi per gli algoritmi riportati in legenda

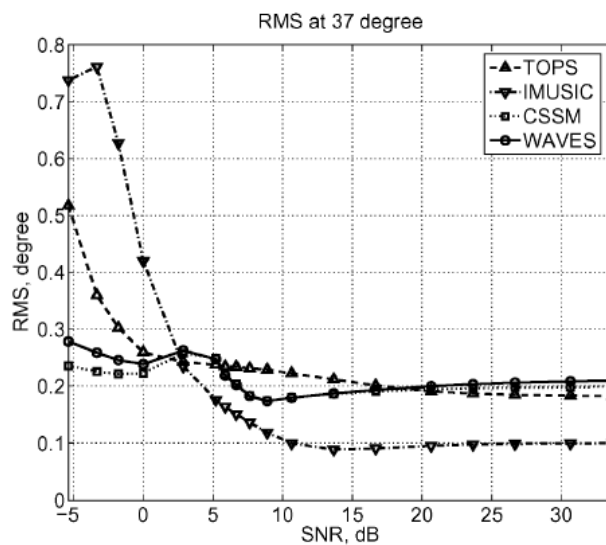


Figura 12: errore quadratico medio delle stime della sorgente a 37 gradi per gli algoritmi riportati in legenda

#### 1.2.1.5 SRP, MUSIC, CSSM, WAVES, TOPS a confronto

In virtù della ricerca bibliografica eseguita si è realizzata la tabella 2 che riporta le performance degli algoritmi di stima dei parametri delle sorgenti sonore utilizzati in questo lavoro di tesi in funzione delle seguenti caratteristiche:

-robustezza: rappresenta la capacità dell'algoritmo di operare in condizioni ambientali differenti in termini di rumore ambientale e riverbero

-risoluzione: rappresenta la capacità dell'algoritmo di distinguere due sorgenti ravvicinate

-versatilità array: rappresenta la capacità dell'algoritmo di operare sui dati ricevuti da tipologie di array differenti

-onere computazionale: rappresenta l'ammontare dei calcoli richiesti all'algoritmo per produrre la stima dei parametri delle sorgenti sonore

	Robustezza	Risoluzione	Versatilità array	Onere computazionale
SRP	Alta	Bassa	Bassa	Alto
MUSIC	Bassa	Alta	Alta	Basso
CSSM	Medio-bassa	Medio-alta	Media	Medio
WAVES	Media	Medio-alta	Media	Medio-alto
TOPS	Media	Medio-alta	Media	Medio-alto

*Tabella 2: confronto caratteristiche algoritmi utilizzati nel seguente lavoro di tesi*

#### 1.2.1.6 Nostro contributo alla letteratura

Dato il background letterario sulla localizzazione delle sorgenti sonore attraverso il processamento dei dati raccolti da un array di sensori il seguente lavoro di tesi fornisce un utile confronto dell'accuratezza dei 5 algoritmi utilizzati in tre diversi scenari di prova a tre diverse distanze array-sorgente: SRP, MUSIC, CSSM, WAVES, TOPS.

Al meglio delle conoscenze dell'autore, in letteratura vi è una mancanza del tema trattato dovuta soprattutto all'interesse dei ricercatori della validazione della metodologia di analisi da loro proposta. Si può quindi affermare che il presente lavoro fornisce un utile confronto oggettivo dell'incertezza di misura dei 5 algoritmi utilizzati nelle condizioni di prova descritte nella sezione 2.1.

## 1.2.2 Array signal processing applicato ai robot

Gli algoritmi di stima dei parametri delle sorgenti sonore costituiscono un utile strumento per migliorare la capacità percettiva dell'ambiente circostante un robot [20].

In analogia col sistema percettivo umano la capacità di un robot di percepire l'ambiente circostante viene sviluppata grazie alla presenza di un sistema di visione costituito da telecamere e di un sistema di ascolto costituito da un array di microfoni.

Una soluzione convenzionalmente adottata per consentire al robot di localizzare una sorgente sonora è quella del cosiddetto "stop-perceive-act", la quale prevede dapprima l'arresto del robot, in seguito la registrazione degli audio tramite l'array di microfoni e infine la localizzazione della sorgente sonora attraverso degli algoritmi di stima dei parametri delle sorgenti sugli audio registrati.

Il seguente lavoro di tesi è basato sul principio dello stop-perceive-act.

Questo tipo di soluzione coincide con quella indagata nella letteratura della localizzazione delle sorgenti sonore tramite un array di microfoni, dove le condizioni valutate considerano un array di microfoni statico, ossia fermo. Tuttavia, questa soluzione che prevede l'array di microfoni fermo applicata a un robot può apportare due notevoli svantaggi [20]:

- l'arresto del movimento del robot potrebbe comprometterne la capacità di svolgimento di un generico compito
- non si possono sfruttare gli eventuali vantaggi che il movimento del robot potrebbe fornire all'acquisizione degli audio consentendo così di ottenere stime più accurate.

Per non incorrere in questi svantaggi occorre riuscire ad applicare il processamento dei segnali in un contesto dinamico ed è questa la peculiarità del processamento dei segnali audio ricevuti da un array di sensori utilizzando un robot. Risulta pertanto fondamentale la conoscenza del movimento fornita dal sistema di controllo dei motori al robot e, da questo, l'implementazione di una funzione di trasformazione dei dati ricevuti dall'array di microfoni in relazione al movimento dell'array e alla sua risposta direzionale che gli consenta di operare in condizioni dinamiche [21].

Come precedentemente introdotto negli svantaggi del metodo "stop-perceive-act", un interessante vantaggio dell'implementazione degli algoritmi di stima dei parametri delle sorgenti sonore nei robot è la possibilità di sfruttare il movimento del robot per migliorare le performance della stima. Sotto questo punto di vista interessanti lavori sono stati condotti da V. Tourbabin, H. Barfuss, B.

Rafaely e W. Kellerman [22] così come da E. Vincent, A. Sini e F. Charpillat [23]. In [22] è stato mostrato come il movimento di un robot che vede implementati due array di sensori, il primo sulla testa ed il secondo sulle braccia, consenta di ampliare il range di frequenza del segnale in ingresso nel quale si ottengono le stime dai valori di varianza minore. Viene infatti mostrato come aumentando la velocità angolare dell'array sulla testa del robot, i cui dati vengono processati con l'algoritmo SSD (Space-Domain Distance, [24]), vengano migliorati i valori di varianza delle stime ottenute elaborando segnali a frequenze elevate e come muovendo le braccia del robot per allontanare i microfoni che costituiscono il secondo array, i cui dati vengono processati con l'algoritmo MUSIC, vengano migliorati i valori di varianza delle stime ottenute elaborando segnali a frequenze basse. I due movimenti possono dunque essere integrati in maniera complementare per ampliare il range di frequenza in cui il robot è in grado di fornire stime accurate.

In [23] invece è stata proposta una strategia di controllo del movimento del robot per una più rapida localizzazione della sorgente basata sulla stima dell'entropia delle posizioni relative di una griglia discretizzata. Sempre in [23], mediante analisi sperimentali condotte implementando l'algoritmo MUSIC sul sistema di audizione robot HARK [25], si è visto che tale strategia consente di abbassare l'errore medio di un fattore 4 rispetto a quello che si avrebbe in caso di movimento casuale del robot.

Va infine osservato che il problema principale relativo alla localizzazione delle sorgenti sonore da parte di un robot che si sta muovendo è dovuto alla presenza del rumore generato dai motori del robot che risultano essere molto vicini al sistema uditivo. Questo è il motivo per cui nel campo del processamento dei segnali audio da parte del robot molteplici lavori sono stati eseguiti al fine di trovare soluzioni in grado di sopprimere il rumore interno [26], [27], [28].

In particolare, in [26] gli autori notano come la conoscenza della struttura meccanica del robot, così come di una qualunque macchina, possa essere sfruttata per determinare quali siano e che relazioni presentino tra loro le principali componenti di frequenza del rumore generato dal movimento in funzione delle velocità e accelerazioni angolari dei motori. Gli autori propongono quindi una metodologia che a partire dalla conoscenza del movimento del robot, in termini di velocità e accelerazione dei motori, stima le componenti di frequenza del rumore interno generato dal movimento e va ad applicare al segnale ricevuto dall'array di sensori dei filtri che rimuovono il contributo delle componenti che ricadono negli intorno di queste frequenze. In [26] la procedura proposta è stata poi soggetta ad una conferma sperimentale che ha valutato la capacità del robot

nel riconoscere le parole contenute in un segnale audio che riproduce la voce umana in presenza di un movimento del braccio del robot. L'applicazione della procedura proposta ha consentito di passare da un "word error rate" (definita come percentuale di parole riconosciute correttamente) del 29.9% a un word error rate del 59,9 % confermando così l'efficacia della procedura nella riduzione del rumore interno.

Va infine evidenziato come l'importanza del sistema uditivo del robot diventa preponderante rispetto a quella del sistema visivo nelle situazioni in cui il soggetto, di cui si vuole rilevare la presenza, risulta mascherato dalla presenza di ostacoli. Un utile lavoro nella localizzazione di oggetti in questa situazione è stato fornito da J. Even, Y. Morales, N. Kallakuri, C. Ishi e N. Hagita in [29]. In questo lavoro l'algoritmo SRP viene utilizzato per implementare una procedura nella quale, una volta fornita al robot la conoscenza dell'ambiente circostante e della sua posizione in esso, si possa localizzare una sorgente sonora mascherata al sistema di visione stimando le direzioni di arrivo delle onde riflesse per poi ricostruire il loro percorso e valutarne il punto di intersezione.



# Capitolo 2

In questo capitolo andremo a descrivere le procedure utilizzate per:

- registrare gli audio dal robot Misty II
- elaborare gli audio con gli algoritmi SRP, MUSIC, CSSM, WAVES, TOPS implementati tramite il linguaggio di programmazione Python
- analizzare i risultati ottenuti al fine di valutare l'incertezza di misura

## 2.1 Registrazione degli audio utilizzando il Robot Misty II

### 2.1.1 Materiali

In questa sezione andiamo a descrivere quali sono i materiali che ci hanno consentito di realizzare il banco di prova tramite il quale si sono condotte le registrazioni degli audio sui quali sono stati implementati gli algoritmi di localizzazione della sorgente sonora introdotti nel capitolo 1.

#### 2.1.1.1 Misty II

La possibilità di avere un robot domestico di facile programmazione, che costituisca una piattaforma open source, sulla quale possa operare anche chi non ha mai lavorato con un robot, rende potenzialmente disponibile un'elevatissima quantità di tempo impiegata nello sviluppo di tale tecnologia e mette le basi per una forte innovazione nel campo della robotica.

Con questo fine l'azienda americana Misty Robotics ha rilasciato nel 2016 il robot Misty I. Nel 2019, una versione potenziata di Misty I, Misty II (figura 13), è stata commercializzata.



Figura 13: Misty II

Misty II è un robot dotato di due sistemi operativi (Windows IoT Core o Android 8), tre processori (Qualcomm SDA 820 SOM, Qualcomm SDA 820 PRO SOM, Qualcomm 410 SOM) e sedici sensori di vario tipo. Grazie ai due sistemi operativi e alla vasta rete di sensori, Misty II costituisce una piattaforma programmabile che offre agli sviluppatori enormi capacità di progettazione [30].

I sensori implementati sul robot possono essere classificati in base alle funzionalità che consentono di assolvere in:

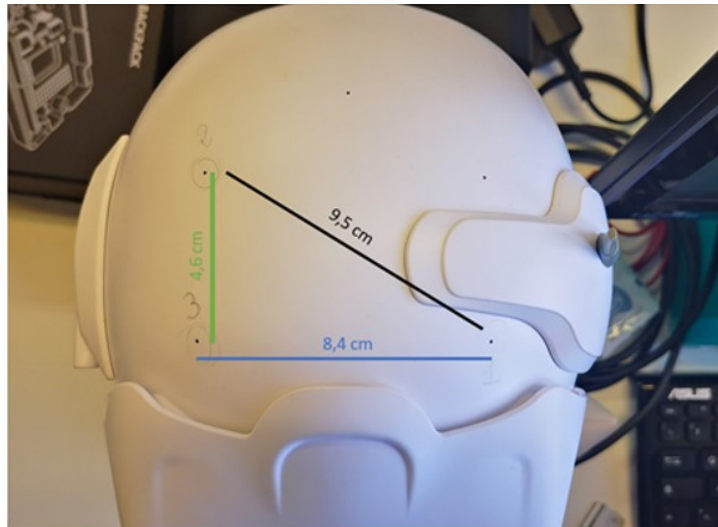
- telecamere: costituiscono il sistema di visione del robot e consentono di registrare video e immagini
- microfoni: costituiscono il sistema di ascolto del robot e consentono di registrare audio
- altoparlanti: consentono al robot di riprodurre segnali audio
- sensori tattili: consentono al robot di percepire il tocco umano
- sensori di distanza: consentono al robot di percepire la presenza di ostacoli così che, integrati con il sistema di controllo dei motori del robot, gli consentano di eseguire le traiettorie di movimento più opportune

Considerando il contesto casalingo è stato Misty II il robot che si è scelto di utilizzare in questo lavoro di tesi per localizzare l'utente nell'ambiente domestico.

#### 2.1.1.1.1 Array di microfoni

L'array di microfoni è collocato sulla testa di Misty II, la quale è posta ad un'altezza di 35 cm da terra.

Si tratta di un array costituito da 3 microfoni omnidirezionali a campo lontano, ossia microfoni la cui risposta è indipendente dalla direzione di arrivo dell'onda e progettati per ricevere segnali acustici provenienti da una distanza tale per cui le onde incidenti siano approssimabili come onde piane. La geometria dell'array è riportata in figura 14. I microfoni sono immersi in un materiale isolante per minimizzare l'acquisizione del rumore causato dai processori contenuti nello stesso case dei microfoni.



*Figura 14: geometria dell'array di microfoni presente sulla testa di Misty II*

#### 2.1.1.2 Casse altoparlanti

Due diverse casse (figura 15) sono state utilizzate per riprodurre rispettivamente una voce umana e un rumore domestico così da poter riuscire a simulare le reali condizioni di funzionamento del robot in un ambiente domestico: la voce umana che si è utilizzata è rimasta invariata in modo che l'input desiderato sia costante, mentre il rumore domestico utilizzato è quello di una aspirapolvere. Questo suono è stato estrapolato dal database dei suoni domestici "Everyday Life Sounds Database: Telemonitoring of Elderly or Disabled" (<https://www.degruyter.com/document/doi/10.1515/jisys-2014-0110/html>) [31].

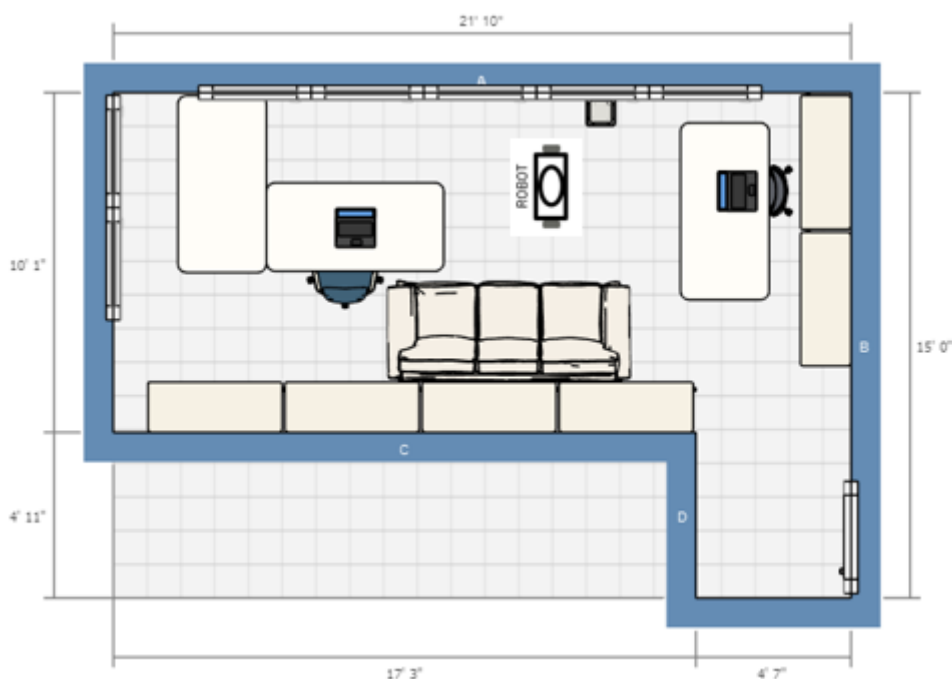


*Figura 15: casse altoparlanti utilizzate per riprodurre la voce umana e il rumore domestico*

## 2.1.2 Testing room

Per valutare l'accuratezza degli algoritmi SRP, MUSIC, CSSM, WAVES, TOPS sia in condizioni reali che in condizioni controllate, caratterizzate dall'assenza non solo del rumore di sottofondo ma anche del riverbero, le registrazioni sono state condotte in due diversi ambienti:

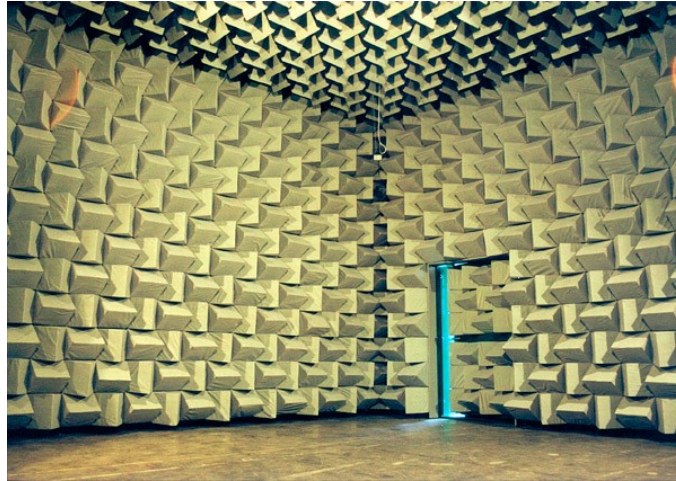
1) Stanza nell'Università Politecnica delle Marche (Q185\_012) la cui configurazione è riportata in figura (figura 16).



*Figura 16: ambiente di prova reale – stanza a quota 185 nell'Università Politecnica delle Marche*

2) Camera semi anecoica. Le camere anecoiche e semi anecoiche sono laboratori acustici per l'esecuzione di misurazioni fonometriche di precisione in condizioni ottimali di assenza di rumore di sottofondo e riverbero ottenute grazie agli elevati valori di isolamento dall'ambiente esterno e agli elevati coefficienti di fonoassorbimento delle pareti esterne. Le camere semi anecoiche (figura 17) si differenziano da quelle anecoiche in quanto sono dotate di pavimento piano e riflettente, mentre in quelle anecoiche anche il pavimento è realizzato in maniera tale da risultare fonoassorbente.

Si è scelto di utilizzare la camera semi anecoica piuttosto che quella anecoica in quanto il robot potrebbe doversi trovare ad operare in una condizione che si avvicina a quella definita da tale ambiente ma non in una condizione che si avvicina a quella definita dalla camera anecoica in quanto la riflessione delle onde sonore sul pavimento sarà sempre presente.



*Figura 17: camera semi anecoica*

## **2.1.3 Protocollo di misura**

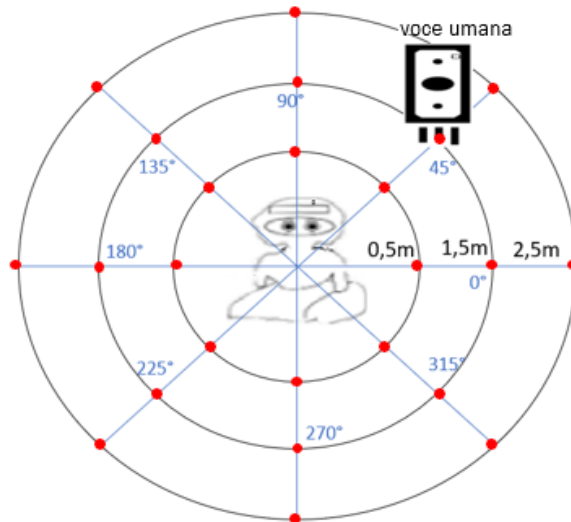
Sia nell'ambiente controllato che in quello reale sono state condotte tre diverse tipologie di test, descritte di seguito.

### 2.1.3.1 Protocollo di misura nell'ambiente reale

#### Test 1

In questa tipologia di test sono state eseguite 96 registrazioni della durata di 60 secondi utilizzando solo la cassa sulla quale l'audio che simula la voce umana è stato riprodotto: una volta posizionato il robot al centro della stanza per ognuna delle seguenti posizioni angolari relative cassa-robot,  $0^\circ$  -  $45^\circ$  -  $90^\circ$  -  $135^\circ$  -  $180^\circ$  -  $225^\circ$  -  $270^\circ$  -  $315^\circ$ , sono state eseguite 4 registrazioni con robot e cassa distanti 0,5m, 4 registrazioni con robot e cassa distanti 1,5m e 4 registrazioni con robot e cassa distanti 2,5m (figura 18).

Ogni registrazione è stata condotta attivando prima i microfoni posizionati sulla testa del robot, l'attivazione dei microfoni è stata programmata per avere una durata di 60 secondi. Durante la registrazione del suono, un led blu posto sulla testa del robot indica la corretta attivazione dei sensori. In contemporanea, sulla cassa predisposta a una determinata distanza, è stato trasmesso l'audio di 3,5s contenente la voce umana.

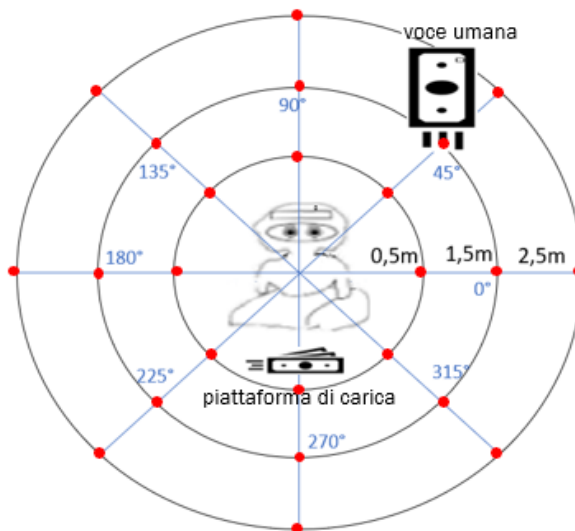


*Figura 18: configurazione delle registrazioni del test 1 nell'ambiente reale*

### Test 2

Anche in questo test è stata utilizzata solo la cassa che riproduce la voce umana, tuttavia il robot è stato messo sotto carica. Questo scenario è stato testato dal momento che il robot in carica genera rumore.

Per questo test il numero delle registrazioni eseguite, la loro durata e le posizioni relative cassa-robot valutate risultano invariate rispetto al test 1 (figura 19).



*Figura 19: configurazione delle registrazioni del test 2 nell'ambiente reale*

### Test 3

In questa tipologia di test si sono invece utilizzate entrambe le casse così da riprodurre sia la voce umana che il rumore domestico dell'aspirapolvere, la traccia audio contenente il rumore domestico

è stata attivata prima di eseguire la registrazione ed ha una durata tale da essere presente in tutti i 60 secondi della registrazione. La cassa che simula il rumore domestico è stata mantenuta ad una distanza e posizione angolare fissa. Anche ora il robot è stato tenuto sotto carica.

Le restanti condizioni di prova (numero di registrazioni, durata registrazioni, posizioni relative cassa che simula il rumore-robot) risultano invece invariate rispetto al test 1 (figura 20).

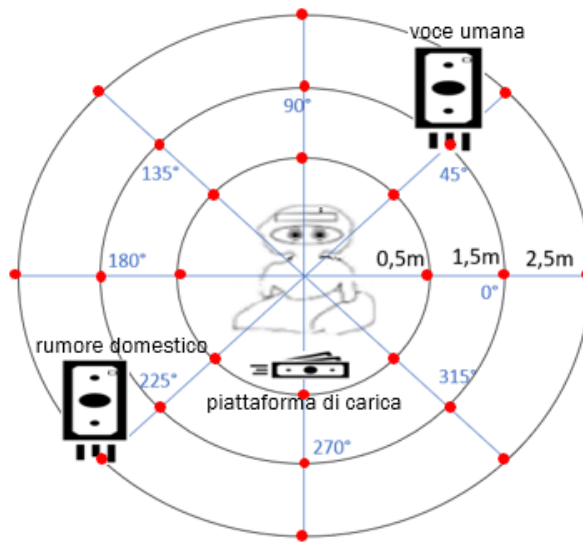


Figura 20: configurazione delle registrazioni del test 3 nell'ambiente reale

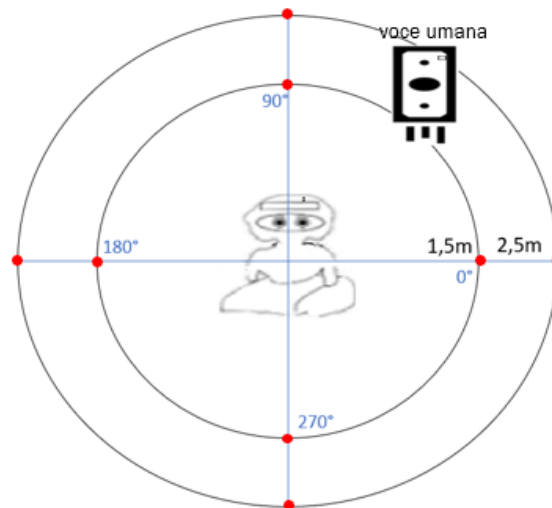
### 2.1.3.2 Protocollo di misura nell'ambiente controllato

In questa sezione verrà valutato il protocollo di misura utilizzato nell'ambiente controllato.

Condurre le registrazioni in questo ambiente di prova ci ha consentito di valutare anche le performance degli algoritmi introdotti nella sezione 1.1 in un ambiente controllato.

#### Test 1

In questa tipologia di test sono state eseguite 8 registrazioni della durata di 60 secondi utilizzando una sola cassa, quella che simula la voce umana: una volta posizionato il robot al centro della camera semi anecoica per ognuna delle seguenti posizioni angolari relative cassa-robot, 0° - 90° - 180° - 270° è stata eseguita una registrazione con robot e cassa distanti 1,5m e una registrazione con robot e cassa distanti 2,5m.

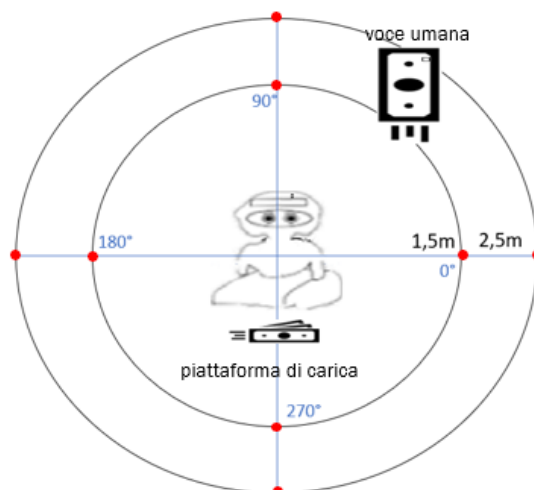


*Figura 21: configurazione delle registrazioni del test 1 condotte nell'ambiente controllato*

## Test 2

Anche in questo test è stata utilizzata solo la cassa che riproduce la voce umana, tuttavia il robot è stato messo sotto carica.

Per questo test il numero delle registrazioni eseguite, la loro durata e le posizioni relative cassa-robot valutate risultano invariate rispetto al test 1 condotto nell'ambiente controllato.



*Figura 22: configurazione delle registrazioni del test 2 nell'ambiente controllato*



### Test 3

In questa tipologia di test si sono invece utilizzate entrambe le casse così da simulare sia la voce umana che il rumore domestico della aspirapolvere. Quest'ultimo è stato trasmesso per tutta la durata della registrazione. La cassa è stata collocata ad una distanza e posizione angolare fissa. Durante la durata di questi test il robot è stato posto sotto carica.

Per questo test il numero delle registrazioni eseguite, la loro durata e le posizioni relative cassa-robot valutate risultano invariate rispetto al test 1 condotto nell'ambiente controllato.

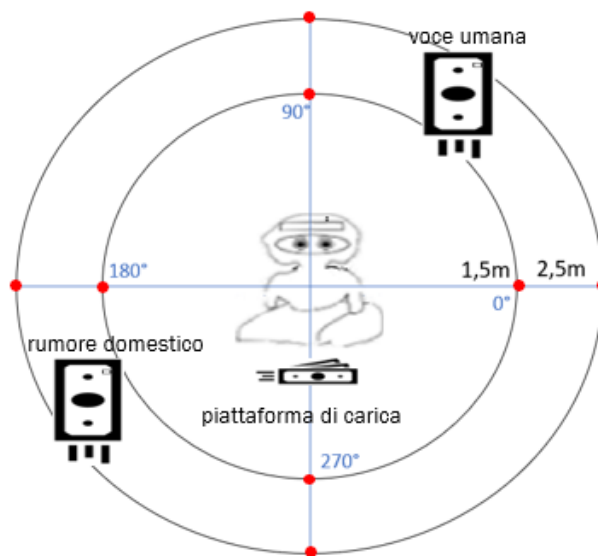


Figura 23: configurazione delle registrazioni del test 3 nell'ambiente controllato

## 2.2 Elaborazioni e valutazione dell'incertezza

Per valutare l'incertezza degli algoritmi di localizzazione delle sorgenti sonore introdotti in letteratura sotto le diverse condizioni di prova riportate nel paragrafo precedente si sono utilizzate le quattro diverse catene algoritmiche descritte nel seguito.

### 2.2.1 Prima catena algoritmica

Al fine di valutare se un pre-processamento dei segnali audio, volto a conservare il segnale utile (voce umana) e ridurre il più possibile la componente del rumore, consenta di migliorare le performance degli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES in un ambiente reale gli algoritmi sono stati implementati secondo quattro diverse catene algoritmiche.

In figura 24 è riportato lo schema a blocchi della prima catena algoritmica, la quale non ha previsto nessun pre-processamento dei segnali audio. Nelle sezioni seguenti verranno evidenziati gli input e gli output di ciascun blocco.

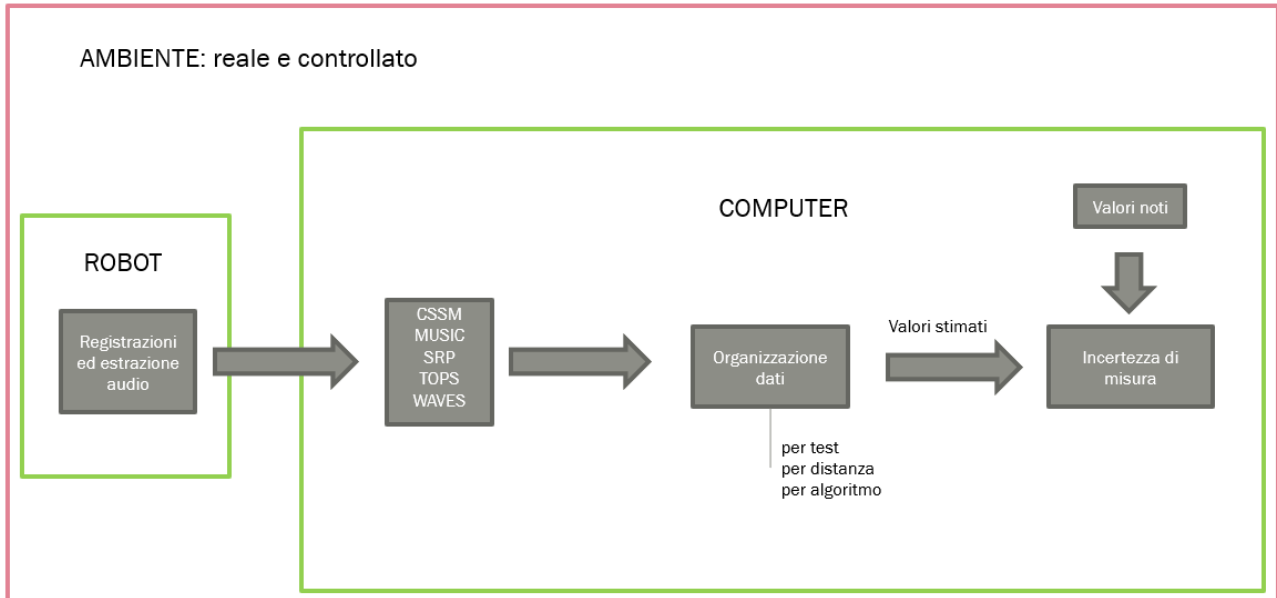


Figura 24: schema funzionale della prima catena algoritmica

#### 2.2.1.1 Prima catena algoritmica in ambiente reale

Grazie all'ausilio di un'interfaccia interattiva, l'estrazione degli audio dal robot al PC è stata eseguita in maniera automatica: quando il robot ha terminato di eseguire una registrazione il file audio registrato è stato automaticamente salvato sul computer con un nome tale da tenere traccia del valore reale della posizione della sorgente sonora rispetto al robot e della tipologia di test a cui questo file audio si riferisce.

Mediante il linguaggio di programmazione Python sono dunque stati implementati i 5 algoritmi di localizzazione della sorgente sonora (SRP, MUSIC, CSSM, WAVES e TOPS) dai quali si è ottenuta una stima della posizione angolare della sorgente rispetto al robot per ogni audio che è stato registrato.

Le stime così ottenute sono state raggruppate in 45 file, ciascuno contenente il valore delle stime per ogni algoritmo relative alle diverse distanze ad ogni tipologia di test eseguito. Si sono quindi ottenuti 15 diversi file, associati alle 3 distanze relative robot - cassa che simula la voce umana e ai 5 diversi algoritmi utilizzati, per ciascuno dei 3 test eseguiti nella stanza.

Una volta fatto questo si sono confrontati i 32 valori delle stime ottenute ad ogni distanza e ad ogni tipologia di test con i rispettivi valori noti per costruire la retta di taratura (figura 25) e si è valutata

l'incertezza di misura delle stime relative a ciascun algoritmo alle diverse distanze e alle diverse condizioni di prova.

L'incertezza di misura è stata valutata tramite la confidenza statistica con fattore di copertura 2, la quale rappresenta lo scostamento dal valore noto entro il quale cadono il 96% delle stime espresso in termini percentuali rispetto al valore medio dei valori noti su cui questa è valutata [32].

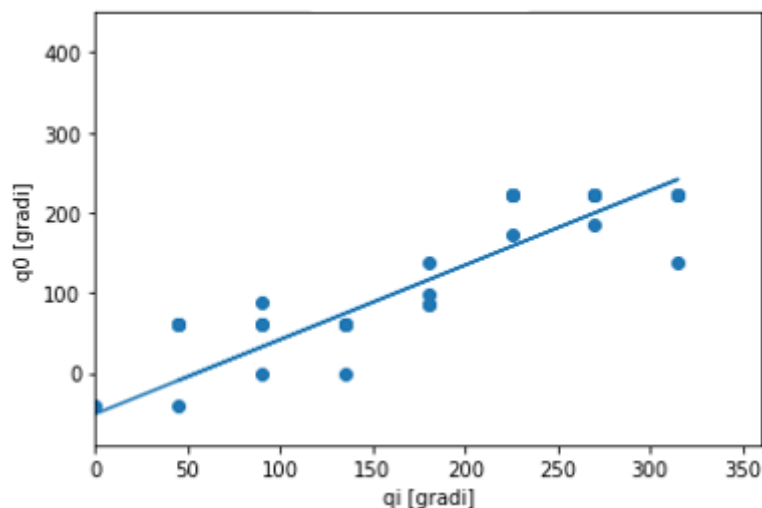
Le formule utilizzate per il calcolo della retta di taratura e della confidenza statistica con fattore di copertura due sono riportate nel seguito.

Indicando con  $q_0$  i valori delle posizioni della sorgente stimati da un determinato algoritmo e con  $q_i$  i valori noti delle posizioni della sorgente a cui i  $q_0$  sono associati, il coefficiente angolare  $m$  e l'intercetta  $b$  della retta di taratura sono dati da:

$$m = \frac{N \sum q_i q_0 - (\sum q_i)(\sum q_0)}{N \sum q_i^2 - (\sum q_i)^2}$$

$$b = \frac{(\sum q_0)(\sum q_i^2) - (\sum q_i q_0)(\sum q_i)}{N \sum q_i^2 - (\sum q_i)^2}$$

La retta di taratura risulta così determinata



**Figura 25:** retta di taratura relativa alle stime ottenute dall'elaborazione delle registrazioni del test 3 nell'ambiente reale alla distanza di 0,5m tramite l'algoritmo MUSIC implementato secondo la prima catena algoritmica

Una volta individuati il coefficiente angolare e l'intercetta della retta di taratura si è valutata la confidenza statistica con fattore di copertura 2 la quale sarà data da

$$\frac{2 * 100 * s_{q_i}}{(q_i)_{medio}}$$

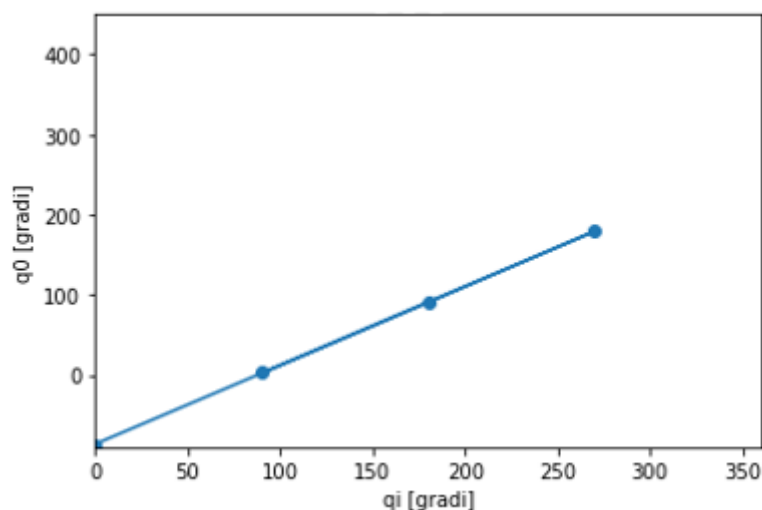
dove  $N$  è il numero totale dei valori di  $q_i$  considerati,  $s_{q_0}^2 = \frac{1}{N} \sum (mq_i + b - q_0)^2$ ,  $s_{q_i} = \sqrt{\frac{s_{q_0}^2}{m^2}}$  e  $(q_i)_{medio} = \frac{\sum q_i}{N}$ .

Al termine di tale procedura si ottengono quindi 45 valori della confidenza statistica con fattore di copertura 2 relativi alle diverse distanze e condizioni di prova per ciascun algoritmo.

### 2.2.1.2 Prima catena algoritmica in ambiente controllato

La differenza dell'implementazione di questa catena algoritmica nei dati raccolti nell'ambiente controllato sta nell'organizzazione dei dati e nella valutazione dell'incertezza.

Avendo eseguito nell'ambiente controllato un minor numero di registrazione le stime ottenute da ciascun audio sono state raggruppate diversamente. Le stime sono state quindi raggruppate in 30 file, ciascuno contenente il valore delle stime per ogni algoritmo relative alle diverse distanze (ora due piuttosto che tre) ad ogni tipologia di test eseguito. Poiché nella camera semi anecoica le registrazioni associate a ciascun test sono state eseguite per 4 posizioni angolari della sorgente (piuttosto che 8) e poiché in ciascuna di tali posizioni è stato registrato un solo audio (piuttosto che 4) ciascuno di questi file contiene 4 valori delle stime della posizione angolare della sorgente (piuttosto che 32). E' sulla base di questi 4 valori che è stata valutata la retta di taratura e la confidenza statistica con fattore di copertura 2 relativa a ciascun algoritmo per ogni distanza e condizione di prova testata nell'ambiente controllato.



**Figura 26:** retta di taratura relativa alle stime ottenute dall'elaborazione delle registrazioni del test 3 nell'ambiente controllato alla distanza di 1,5m tramite l'algoritmo SRP implementato secondo la prima catena algoritmica

## 2.2.2 Seconda catena algoritmica

Per valutare se un pre-processamento del segnale audio atto a eliminare la zona temporale in cui non risulta presente la voce umana consente di migliorare le performance degli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES operanti in un ambiente reale tali algoritmi sono stati implementati anche secondo la catena algoritmica descritta in questa sezione.

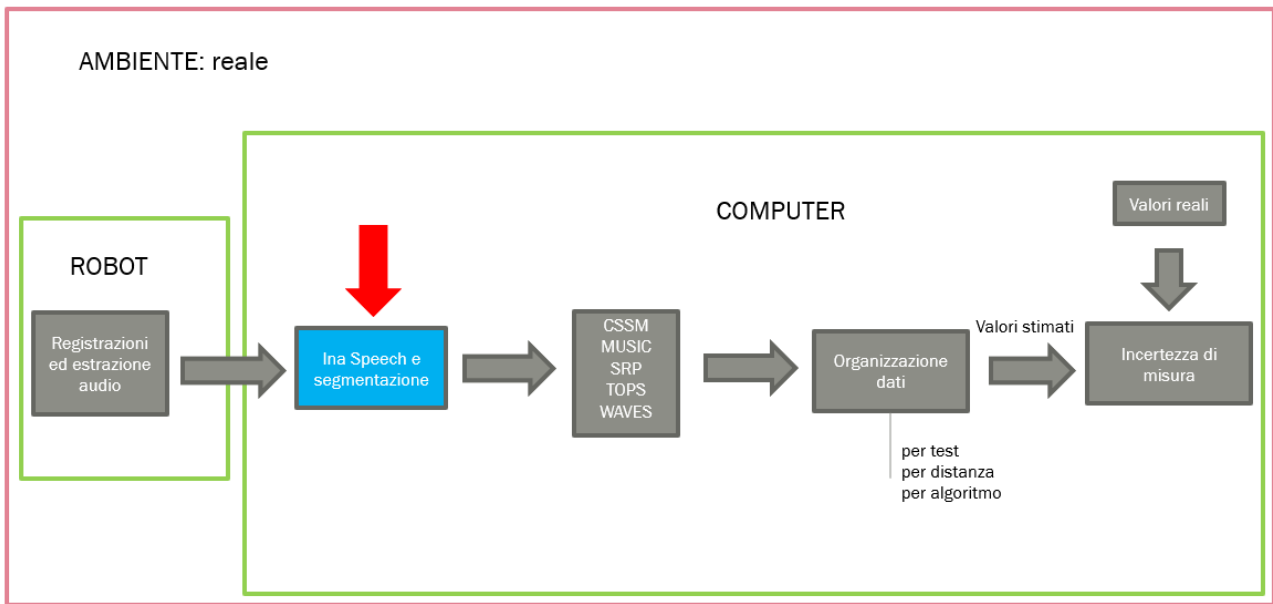


Figura 27: schema funzionale della seconda catena algoritmica

In tale procedura è stato introdotto un ulteriore algoritmo che prende il nome di Ina Speech. Questo algoritmo riceve in ingresso un file audio ed elaborandolo fornisce in output l'informazione relativa a quale tra le seguenti caratteristiche prevale in ciascun istante temporale dell'audio: rumore, musica, voce umana maschile, voce umana femminile (figura 28).

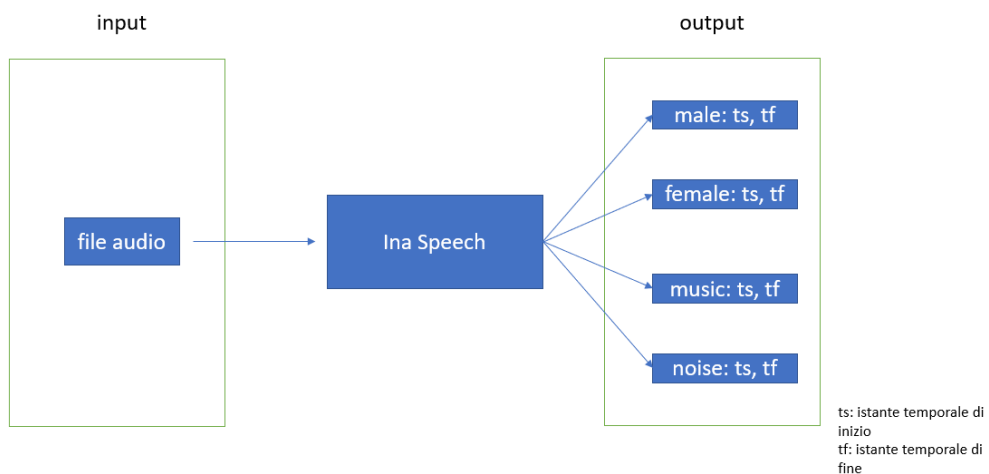
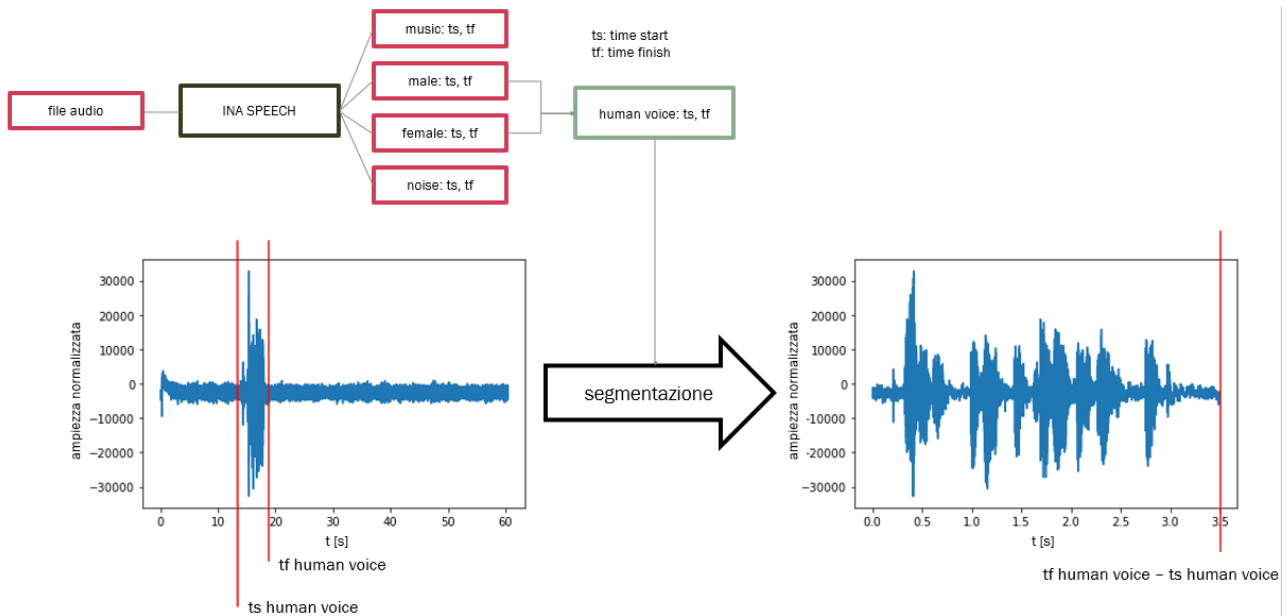


Figura 28: input e output di Ina Speech

Elaborati tutti gli audio registrati tramite questo algoritmo, le informazioni relative agli istanti temporali di inizio e di fine della presenza della voce umana ottenute sono state utilizzate per segmentare (troncare temporalmente) gli audio così da rimuovere la parte in cui non è presente il segnale relativo alla voce umana (figura 29).



*Figura 29: segmentazione dei segnali audio sulla base delle informazioni provenienti da Ina Speech*

Gli audio segmentati sono stati dati in ingresso agli algoritmi di localizzazione delle sorgenti sonore.

In questa catena algoritmica abbiamo previsto la possibilità di fallimento di Ina Speech nel rilevamento della voce umana. In queste condizioni gli audio che si sono dati in ingresso agli algoritmi di localizzazione della sorgente sonora sono gli audio interi.

Per quanto riguarda invece le registrazioni e l'estrazione degli audio, l'organizzazione dei dati e la valutazione dell'incertezza di misura è stata adottata la stessa procedura descritta nella sezione 2.2.1

### 2.2.3 Terza catena algoritmica

Per valutare se un pre processamento del segnale audio atto a eliminare le componenti della frequenza in cui non risulta presente la voce umana consenta di migliorare le performance degli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES operanti in un ambiente reale tali algoritmi sono stati implementati anche secondo la catena algoritmica descritta in questa sezione.

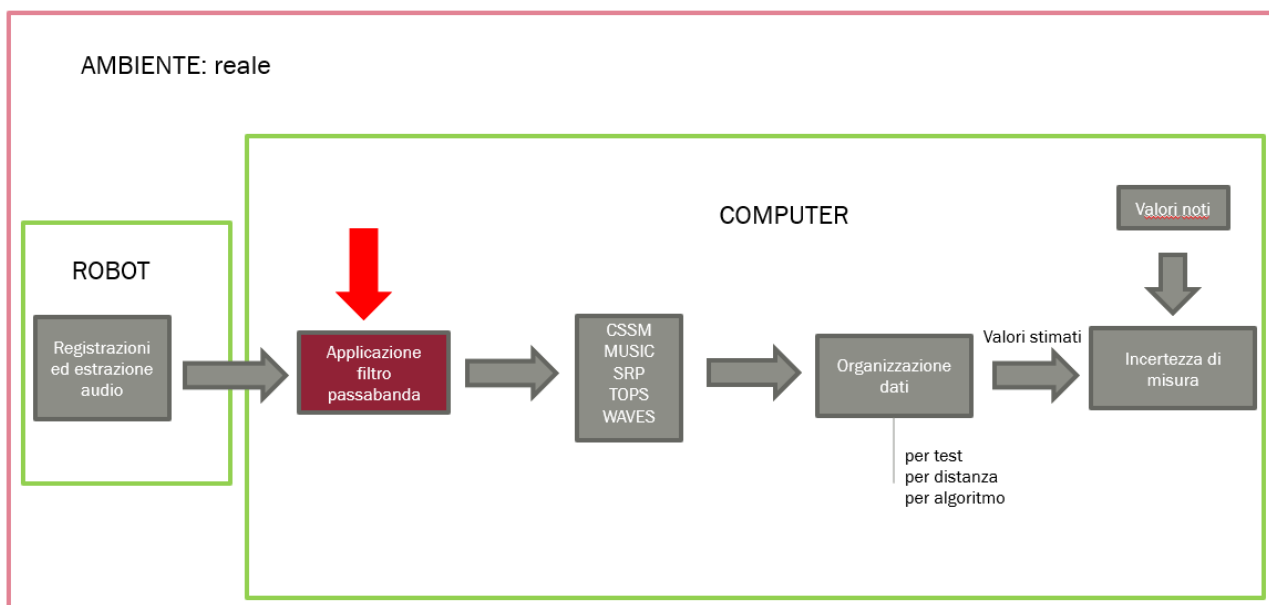


Fig 29: schema funzionale della terza catena algoritmica

In tale procedura gli audio che sono stati dati in ingresso agli algoritmi di localizzazione della sorgente sonora hanno subito un pre processamento che ha visto l'applicazione di un filtro passabanda atto a conservare solo le frequenze contenute nel segnale audio che simula la voce umana (400 – 2000 Hz) eliminando tutte le altre. Il filtro che si è scelto di usare è un filtro Butterworth del secondo ordine (figura 30).

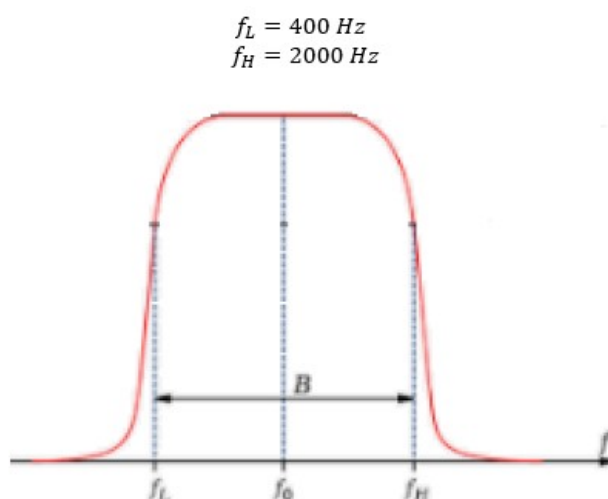
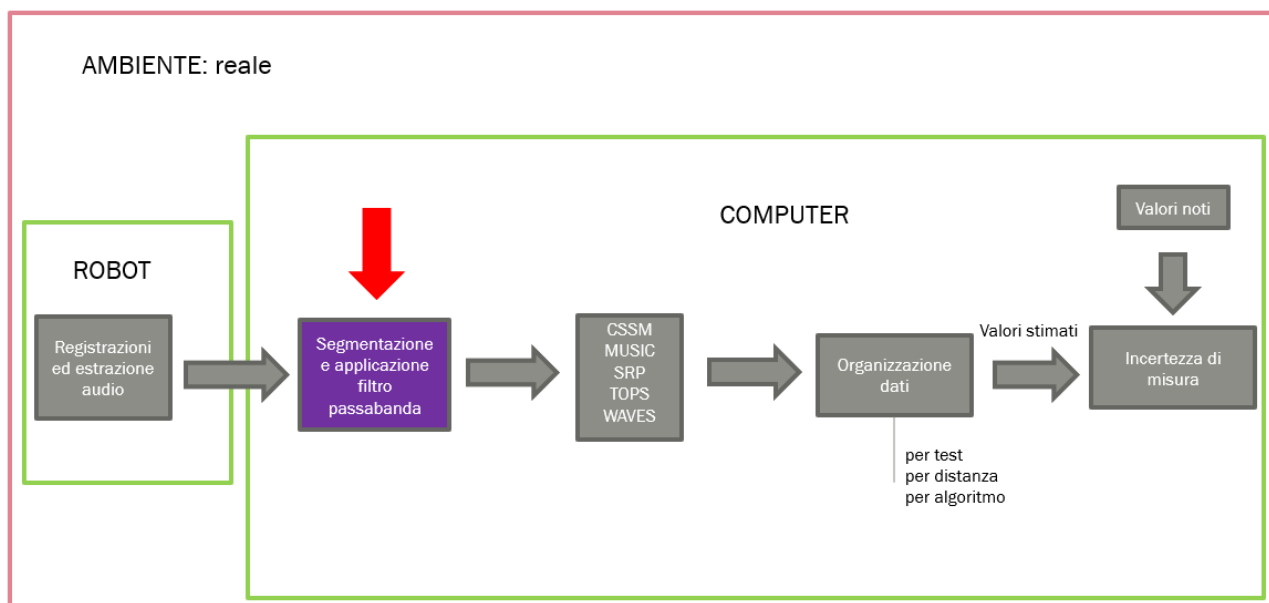


Figura 30: filtro butterworth del secondo ordine con frequenza di taglio pari a 400 e 2000 Hz

Le restanti parti della catena algoritmica sono invece rimaste invariate rispetto a quello descritte nella sezione 2.2.1.

## 2.2.4 Quarta catena algoritmica

Per valutare se un pre processamento del segnale audio atto a eliminare sia la zona temporale che le componenti di frequenza in cui non risulta presente la voce umana consenta di migliorare le performance degli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES operanti in un ambiente reale, tali algoritmi sono stati implementati anche secondo la catena algoritmica descritta in questa sezione.



*Figura 31: schema funzionale della quarta catena algoritmica*

Tale catena algoritmica prevede di pre processare i segnali audio forniti in ingresso agli algoritmi di localizzazione della sorgente sia tramite la segmentazione proposta nella sezione 2.2.2 che tramite l'applicazione del filtro passabanda proposta nella sezione 2.2.3.



# Capitolo 3

## 3.1 Risultati

In questo capitolo vengono riportati i risultati relativi alla stima della posizione angolare della sorgente sonora valutata secondo le modalità descritte nel capitolo 2.

Nel capitolo 2 è stato presentato come:

- Nell'ambiente reale abbiamo eseguito 32 registrazioni (4 registrazioni per ognuno degli 8 angoli in cui è stata posizionata la cassa) per ciascuna distanza (0,5m 1,5m 2,5m) e per ciascuna condizione di prova (test1 test2 test3). Ogni registrazione è stata elaborata in 20 diversi modi (5 algoritmi per 4 catene algoritmiche) per ottenere una stima della posizione angolare della sorgente. Per ognuna delle 20 tipologie di elaborazioni è stata calcolata la confidenza statistica con fattore di copertura due a ciascuna delle tre distanze e a ciascuno dei tre test confrontando i valori reali e le stime ottenute dalle 32 registrazioni a una data distanza in un dato test. Questo valore nel seguito verrà indicato come  $a_{i,j,k,z}$ , dove  $i$  è il pedice identificativo dell'algoritmo (CSSM, MUSIC, SRP, TOPS, WAVES),  $j$  è il pedice identificativo della catena algoritmica con cui l'algoritmo è stato implementato (prima, seconda, terza, quarta) e  $k$  e  $z$  sono rispettivamente i pedici identificativi della tipologia di test (test1, test2, test3) e della distanza( 0,5m, 1,5m, 2,5m).

- Nell'ambiente controllato abbiamo eseguito 4 registrazioni (una registrazione per ognuno dei 4 angoli in cui è stata posizionata la sorgente) per ciascuna distanza (1,5m 2,5m) e per ciascuna condizione di prova (test1 test2 test3). Ogni registrazione è stata elaborata in 5 diversi modi (5 algoritmi implementati con una sola catena algoritmica) per ottenere una stima della posizione angolare della sorgente. Per ognuna delle 5 tipologie di elaborazioni è stata calcolata la confidenza statistica con fattore di copertura due a ciascuna delle due distanze e a ciascuno dei tre test confrontando i valori reali e le stime ottenute dalle 4 registrazioni a una data distanza in un dato test.

### 3.1.1 Risultati ottenuti dall'implementazione della prima catena algoritmica

#### 3.1.1.1 Risultati ottenuti per la prima catena algoritmica nell'ambiente reale

Le incertezze di misura valutate tramite la confidenza statistica con fattore di copertura 2 ottenute dall'implementazione degli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES implementati secondo la catena algoritmica descritta nella sezione 2.2.1.1 per gli audio registrati nelle condizioni di prova descritte nella sezione 2.1.3.1 sono riportati nella tabella 3.

	TEST 1			TEST 2			TEST 3		
	0,5m	1,5m	2,5m	0,5m	1,5m	2,5m	0,5m	1,5m	2,5m
$C_{CSSM}[\%]$	7,9	9,1	20,3	7,6	14,3	8,2	9,9	7,6	11,6
$C_{MUSIC}[\%]$	9,4	15,3	10,3	10,3	31,3	7,7	7,4	3,5	9,8
$C_{SRP}[\%]$	17,1	13,0	8,6	13,5	8,5	11,2	3,3	3,7	7,1
$C_{TOPS}[\%]$	15,8	14,5	13,1	20,7	15,2	12,7	13,8	9,4	13,0
$C_{WAVES}[\%]$	16,7	13,6	13,2	10,4	13,9	11,3	13,9	15,6	9,1

*Tabella 3: confidenza statistica con fattore di copertura 2 ottenuta implementando i 5 algoritmi secondo la prima catena algoritmica sulle registrazioni condotte nell'ambiente reale*

Come mostrato in tabella 3 possiamo notare che implementando gli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES sugli audio registrati nell'ambiente reale secondo la prima catena algoritmica nelle condizioni del test 1 a 0,5m e 1,5 l'incertezza minore, valutata tramite la confidenza statistica con fattore di copertura due, è fornita dal CSSM con valori di  $a_{i,j,k,z}$  rispettivamente pari al 7,9% e 9,1%, nelle condizioni del test 1 a 2,5m l'incertezza minore è fornita dall' SRP con  $a_{i,j,k,z}$  uguale al 8,6%; nel test 2 invece alla distanza di 0,5m l'incertezza minore è stata fornita dal CSSM con  $a_{i,j,k,z}$  pari al 7,6%, alla distanza di 1,5m dall' SRP con  $a_{i,j,k,z}$  uguale al 8,5% e alla distanza di 2,5m dal MUSIC con  $a_{i,j,k,z}$  uguale al 7,7%; nelle condizioni del test 3 invece alle distanze di 0,5m e 2,5m a fornire l'incertezza minore è stato l'SRP con valori di  $a_{i,j,k,z}$  rispettivamente pari al 3,3% e 7,1%, mentre alla distanza di 1,5m è il CSSM che ha fornito l'incertezza minore con  $a_{i,j,k,z}$  uguale al 3,5%.

#### 3.1.1.2 Risultati ottenuti per la prima catena algoritmica nell'ambiente controllato

Le incertezze valutate tramite la confidenza statistica con fattore di copertura 2 delle stime ottenute dall'implementazione degli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES implementati secondo la catena algoritmica descritta nella sezione 2.2.1.1 per gli audio registrati nelle condizioni di prova descritte nella sezione 2.1.3.1 sono riportati nella tabella 4.

	TEST 1		TEST 2		TEST 3	
	1,5m	2,5m	1,5m	2,5m	1,5m	2,5m
C <sub>CSSM</sub> [%]	5,9	7,6	7,0	7,0	16,9	10,7
C <sub>MUSIC</sub> [%]	6,3	12,9	2,5	18,5	6,6	6,6
C <sub>SRP</sub> [%]	1,9	0,8	6,7	5,7	2,3	2,4
C <sub>TOPS</sub> [%]	25,6	4,0	26,5	4,3	8,6	5,6
C <sub>WAVES</sub> [%]	8,5	6,4	10,2	6,4	22,4	21,9

*Tabella 4: confidenza statistica con fattore di copertura 2 ottenuta implementando i 5 algoritmi secondo la prima catena algoritmica sulle registrazioni condotte nell'ambiente controllato*

Come mostrato in tabella 4 possiamo notare che implementando gli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES sugli audio registrati nell'ambiente controllato secondo la prima catena algoritmica nelle condizioni del test1 e del test3, per tutte le distanze, l'incertezza minore è fornita dall'SRP con valori della confidenza statistica con fattore di copertura 2 rispettivamente pari al 1,9%, 0,8%, 2,3% e 2,4%; nelle condizioni del test2 invece alla distanza di 1,5m l'incertezza minore è fornita dal MUSIC con confidenza statistica con fattore di copertura due uguale al 2,5% e alla distanza di 2,5m dal TOPS con confidenza statistica con fattore di copertura due uguale al 4,3%.

### 3.1.2 Risultati ottenuti dall'implementazione della seconda catena algoritmica

#### 3.1.2.1 Risultati ottenuti per la seconda catena algoritmica nell'ambiente reale

Le incertezze di misura valutate tramite la confidenza statistica con fattore di copertura 2 ottenute dall'implementazione degli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES implementati secondo la catena algoritmica descritta nella sezione 2.2.2 per gli audio registrati nelle condizioni di prova descritte nella sezione 2.1.3.1 sono riportati nella tabella 5.

	TEST 1			TEST 2			TEST 3		
	0,5m	1,5m	2,5m	0,5m	1,5m	2,5m	0,5m	1,5m	2,5m
C <sub>CSSM</sub> [%]	10,0	9,0	10,6	10,1	15,0	11,0	10,3	7,6	11,6
C <sub>MUSIC</sub> [%]	15,5	13,8	13,4	9,6	19,7	6,5	7,4	3,5	9,8
C <sub>SRP</sub> [%]	5,3	5,5	7,0	5,4	7,6	10,7	3,4	3,7	7,1
C <sub>TOPS</sub> [%]	17,4	11,6	13,3	18,0	13,7	16,3	16,4	9,4	13,0
C <sub>WAVES</sub> [%]	15,3	10,0	10,8	7,6	15,4	10,6	17,7	15,6	9,1

*Tabella 5: confidenza statistica con fattore di copertura 2 ottenuta implementando i 5 algoritmi secondo la seconda catena algoritmica sulle registrazioni condotte nell'ambiente reale*

Come mostrato dalla tabella 5 possiamo notare che implementando gli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES sugli audio registrati nell'ambiente reale secondo la seconda catena algoritmica nelle condizioni del test1 l'incertezza minore valutata tramite la confidenza statistica con fattore di

copertura due è fornita, per tutte le distanze, dall' SRP con valori di  $a_{i,j,k,z}$  rispettivamente pari al 5,3%, 5,5% e 7,0%; nelle condizioni del test2 invece alle distanze di 0,5m e 1,5m è ancora l' SRP a fornire l'incertezza minore con  $a_{i,j,k,z}$  rispettivamente pari al 5,4% e 7,6% mentre alla distanza di 2,5m il risultato migliore è dato dal MUSIC con  $a_{i,j,k,z}$  uguale a 6,5%; nelle condizioni del test3 infine l' SRP mostra ancora l'incertezza minore alle distanze di 0,5m e 2,5m con  $a_{i,j,k,z}$  rispettivamente uguali al 3,4% e 7,1% ma alla distanza di 1,5m è il MUSIC ad esibire l'incertezza minore con un  $a_{i,j,k,z}$  di 3,5%.

### 3.1.3 Risultati ottenuti dall'implementazione della terza catena algoritmica

3.1.3.1 Risultati ottenuti per la terza catena algoritmica nell'ambiente reale

Le incertezze di misura valutate tramite la confidenza statistica con fattore di copertura 2 ottenute dall'implementazione degli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES implementati secondo la catena algoritmica descritta nella sezione 2.2.3 per gli audio registrati nelle condizioni di prova descritte nella sezione 2.1.3.1 sono riportati nella tabella 6.

	TEST 1			TEST 2			TEST 3		
	0,5m	1,5m	2,5m	0,5m	1,5m	2,5m	0,5m	1,5m	2,5m
$C_{CSSM}[\%]$	8,0	3,3	4,8	3,8	13,8	5,7	7,4	5,5	9,8
$C_{MUSIC}[\%]$	13,5	8,8	9,6	10,7	13,3	10,9	7,2	6,8	10,2
$C_{SRP}[\%]$	14,6	11,2	7,0	11,2	10,0	11,3	5,1	5,0	7,7
$C_{TOPS}[\%]$	9,2	10,3	11,1	14,9	18,8	9,4	18,5	9,3	17,5
$C_{WAVES}[\%]$	12,6	4,9	9,1	11,3	16,1	13,5	10,1	5,9	7,8

*Tabella 6: confidenza statistica con fattore di copertura 2 ottenuta implementando i 5 algoritmi secondo la terza catena algoritmica sulle registrazioni condotte nell'ambiente reale*

Come mostrato da questa tabella possiamo notare che implementando gli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES sugli audio registrati nell'ambiente reale secondo la terza algoritmica nelle condizioni del test1 l'incertezza minore valutata tramite la confidenza statistica con fattore di copertura due è fornita, per tutte le distanze, dal CSSM con valori di  $a_{i,j,k,z}$  rispettivamente pari al 8,0%, 3,3% e 4,8%; nelle condizioni del test2 invece il CSSM fornisce l'incertezza minore alle distanze di 0,5m e 2,5m, dove presenta rispettivamente  $a_{i,j,k,z}$  uguale al 3,8% e 5,7%, mentre alla distanza di 1,5m l'algoritmo che esibisce la minore incertezza risulta essere l' SRP con  $a_{i,j,k,z}$  pari al 10%; infine

nelle condizioni del test3 l'incertezza minore è fornita, a tutte le distanze, dall'SRP con valori di  $a_{i,j,k,z}$  rispettivamente pari al 5,1%, 5,0% e 7,7%.

### 3.1.4 Risultati ottenuti dall'implementazione della quarta catena algoritmica

3.1.4.1 Risultati ottenuti per la quarta catena algoritmica nell'ambiente reale

Le incertezze di misura valutate tramite la confidenza statistica con fattore di copertura 2 ottenute dall'implementazione degli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES implementati secondo la catena algoritmica descritta nella sezione 2.2.4 per gli audio registrati nelle condizioni di prova descritte nella sezione 2.1.3.1 sono riportati nella tabella 7.

	TEST 1			TEST 2			TEST 3		
	0,5m	1,5m	2,5m	0,5m	1,5m	2,5m	0,5m	1,5m	2,5m
$C_{CSSM}[\%]$	7,9	3,4	4,8	5,5	14,8	4,8	8,3	5,5	9,8
$C_{MUSIC}[\%]$	14,4	7,7	9,6	9,8	13,9	10,1	8,6	6,8	10,2
$C_{SRP}[\%]$	5,4	5,6	9,2	5,5	13,4	9,1	5,1	5,0	7,7
$C_{TOPS}[\%]$	9,4	10,2	10,9	10,5	19,8	9,8	15,9	9,3	17,5
$C_{WAVES}[\%]$	10,6	9,0	8,1	12,5	22,4	12,5	10,7	5,9	7,8

*Tabella 7: confidenza statistica con fattore di copertura 2 ottenuta implementando i 5 algoritmi secondo la prima catena algoritmica sulle registrazioni condotte nell'ambiente reale*

Come mostrato dalla tabella 7 possiamo notare che implementando gli algoritmi CSSM, MUSIC, SRP, TOPS, WAVES sugli audio registrati nell'ambiente reale secondo la quarta catena algoritmica nelle condizioni del test1 alla distanza di 0,5m è l'SRP a fornire l'incertezza minore con  $a_{i,j,k,z}$  uguale al 5,4% mentre a 1,5 e 2,5 è il CSSM con  $a_{i,j,k,z}$  rispettivamente uguali al 3,4% e 4,8%; nelle condizioni del test alla distanza di 0,5m il CSSM e l'SRP presentano la minore incertezza con  $a_{i,j,k,z}$  uguale al 5,5%, alla distanza di 1,5m è l'SRP con  $a_{i,j,k,z}$  pari al 13,4% a fornire l'incertezza minore e a quella di 2,5 nuovamente il CSSM con  $a_{i,j,k,z}$  uguale al 4,8%; nelle condizioni del test3 invece è l'SRP che fornisce l'incertezza minore a tutte e tre le distanze con valori di  $a_{i,j,k,z}$  rispettivamente pari al 5,1%, 5,0% e 7,7%.

### 3.1.5 Risultati relativi all'accuratezza dell'algoritmo Ina Speech

Considerando l'accuratezza di Ina Speech pari alla percentuale di successo di rilevamento della presenza della voce umana, ossia il numero di audio in cui l'algoritmo ha riconosciuto la presenza

della voce umana diviso il numero di audio totali, si è riscontrato che relativamente alle registrazioni condotte nell'ambiente reale: nelle condizioni imposte dal test1 Ina Speech ha esibito un'accuratezza del 96,8%, in quelle del test2 del 86,5%, in quelle del test3 del 9,3%.

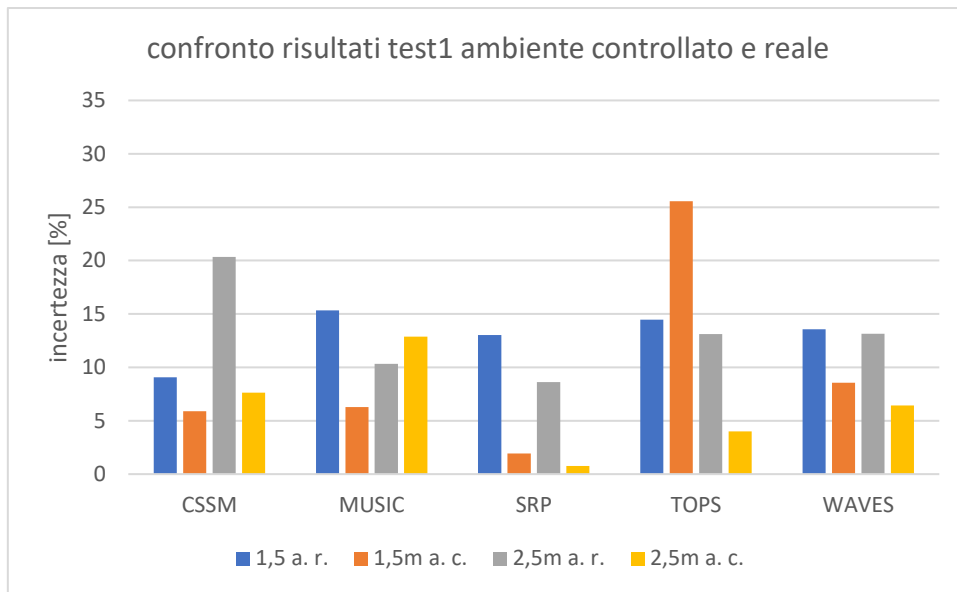
## **3.2 Analisi dei dati ottenuti**

In questa sezione dai risultati precedentemente riportati viene presentato:

- il comportamento degli algoritmi implementati nei due diversi ambienti di prova
- l'influenza delle catene algoritmiche sull'incertezza dei vari algoritmi ottenuta nell'ambiente reale
- il comportamento degli algoritmi implementati secondo le quattro diverse catene algoritmiche alle varie distanze nell'ambiente reale
- il comportamento degli algoritmi implementati secondo le quattro diverse catene algoritmiche nelle diverse condizioni di prova nell'ambiente reale.

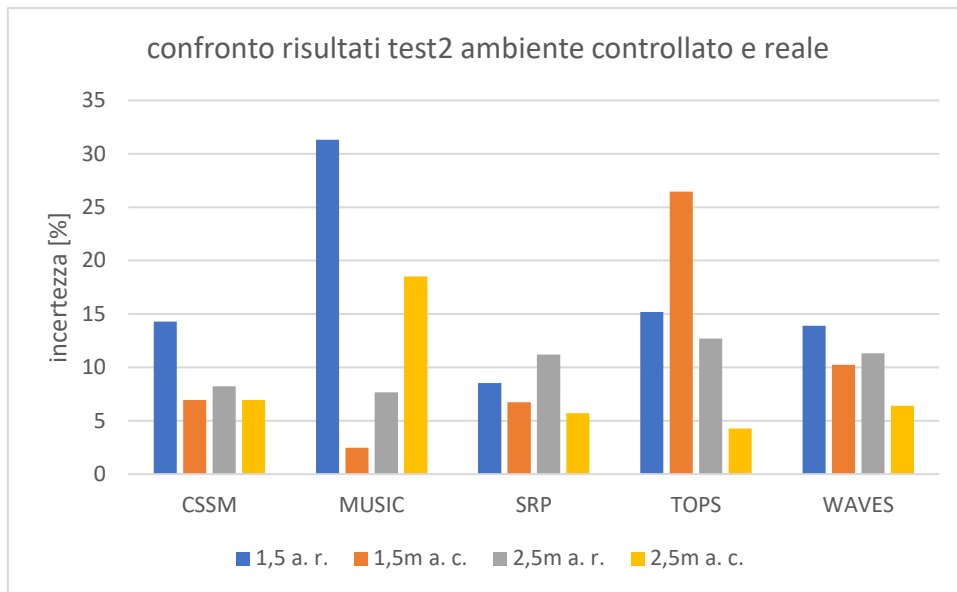
### **3.2.1 Analisi dei dati ottenuti: confronto degli algoritmi nei due ambienti**

Per confrontare il comportamento degli algoritmi nell'ambiente controllato e in quello reale si sono realizzate la figura 32, la figura 33 e la figura 34 nelle quali viene rappresentata la confidenza statistica con fattore di copertura due (sull'asse y) dei cinque algoritmi implementati secondo la prima catena algoritmica, ottenuta per le condizioni di prova comuni dei test condotti in ambiente controllato e reale in relazione alle distanze (sull'asse x) di 1,5m e 2,5m; test 1, test 2 e test3.



**Figura 32:** confidenza statistica con fattore di copertura due dei 5 algoritmi implementati secondo la prima catena algoritmica sulle registrazioni eseguite nell'ambiente controllato e reale alle distanze di 1,5m e 2,5m nelle condizioni di prova del test 1

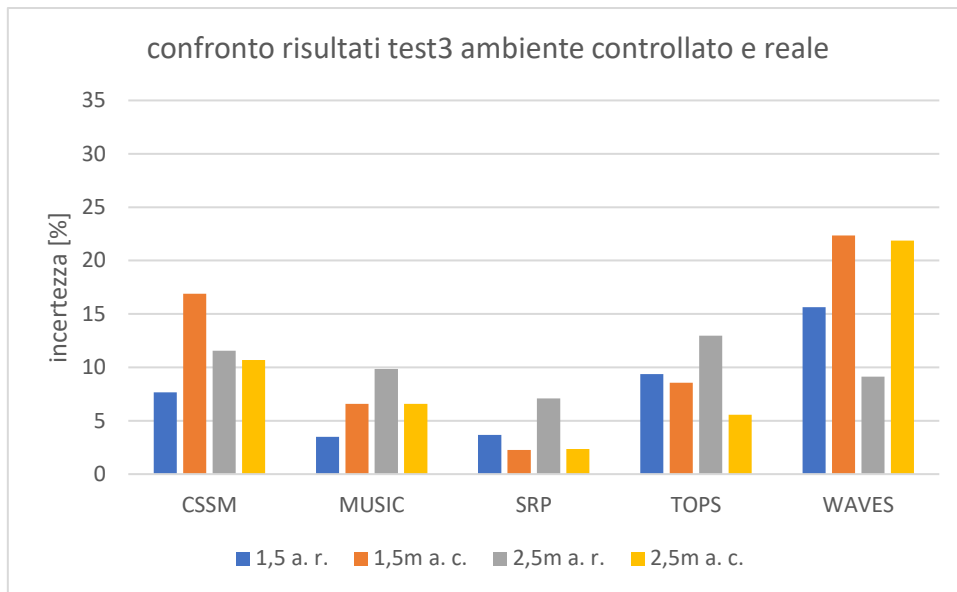
Nelle condizioni del test1 possiamo osservare: una minore incertezza del CSSM nell'elaborazione delle registrazioni condotte nell'ambiente controllato, vi è infatti una riduzione della confidenza statistica con fattore di copertura due del 35% alla distanza di 1,5m e del 62% alla distanza di 2,5m; una minore incertezza del MUSIC nell'ambiente controllato alla distanza di 1,5m (riduzione del 59%) ma maggiore alla distanza di 2,5m dove la confidenza statistica con fattore di copertura due aumenta del 25%; una minore incertezza dell' SRP nell'ambiente controllato sia alla distanza di 1,5m che a quella di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due dell' 85% e 91%; un miglior comportamento del TOPS nell'ambiente controllato alla distanza di 2,5m con una riduzione del 69% ma peggiore alla distanza di 1,5m con un aumento del 76%; una minore incertezza del WAVES sia alla distanza di 1,5m che di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due del 37% e 51%.



**Figura 33:** confidenza statistica con fattore di copertura due dei 5 algoritmi implementati secondo la prima catena algoritmica sulle registrazioni eseguite nell'ambiente controllato e reale alle distanze di 1,5m e 2,5m nelle condizioni di prova del test 2

Nelle condizioni del test2 possiamo osservare: una minore incertezza del CSSM nell'elaborazione delle registrazioni condotte nell'ambiente controllato, vi è infatti una riduzione della confidenza statistica con fattore di copertura due del 51% alla distanza di 1,5m e del 15% alla distanza di 2,5m; una minore incertezza del MUSIC nell'ambiente controllato alla distanza di 1,5m (riduzione del 92%) ma maggiore alla distanza di 2,5m dove la confidenza statistica con fattore di copertura due aumenta del 141%; una minore incertezza del SRP nell'ambiente controllato sia alla distanza di 1,5m che a quella di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due del 21% e 48%; una minore incertezza del TOPS nell'ambiente controllato alla distanza di 2,5m con una riduzione del 66% ma maggiore alla distanza di 1,5m con un aumento del 74%; una minore incertezza del WAVES sia alla distanza di 1,5m che di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due del 26% e 43%.





**Figura 34:** confidenza statistica con fattore di copertura due dei 5 algoritmi implementati secondo la prima catena algoritmica sulle registrazioni eseguite nell'ambiente controllato e reale alle distanze di 1,5m e 2,5m nelle condizioni di prova del test 3

Nelle condizioni del test3 possiamo osservare: relativamente al CSSM nell'elaborazione delle registrazioni condotte nell'ambiente controllato alla distanza di 2,5m una riduzione della confidenza statistica con fattore di copertura due del 7% e alla distanza di 1,5m un aumento del 120%; una minore incertezza del MUSIC nell'elaborazione delle registrazioni condotte nell'ambiente controllato alla distanza di 2,5m con una riduzione della confidenza statistica con fattore di copertura due del 33% ma maggiore alla distanza di 1,5m con un aumento del 88%; un una minore incertezza del SRP nell'ambiente controllato sia alla distanza di 1,5m che a quella di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due del 38% e 66%; una minore incertezza del SRP nell'ambiente controllato sia alla distanza di 1,5m che a quella di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due del 8% e 57%; una minore incertezza del WAVES nell'ambiente controllato sia alla distanza di 1,5m che a quella di 2,5m esibendo rispettivamente un aumento della confidenza statistica con fattore di copertura due del 42% e 139%.

### 3.2.2 Analisi dei dati ottenuti: confronto catene algoritmiche

Per valutare l'influenza delle diverse catene algoritmiche sull'incertezza relativa agli algoritmi utilizzati si è realizzata la tabella 8 che riporta, per ciascun algoritmo (sulle righe) e per ciascuna catena algoritmica con cui l'algoritmo è stato implementato (colonne), il valore della confidenza statistica con fattore di copertura due mediato sulle tre diverse distanze e sulle tre diverse condizioni di prova testate nell'ambiente reale, valore che indichiamo con  $B_{i,j}$ .

Essendo  $a_{i,j,k,z}$  la confidenza statistica con fattore di copertura due relativa alle registrazioni del test  $k$  ( $k = \text{test1, test2, test3}$ ) alla distanza  $z$  ( $z = 0,5m, 1,5m, 2,5m$ ) elaborate con l'algoritmo  $i$  ( $i = \text{CSSM, MUSIC, SRP, TOPS, WAVES}$ ) implementato secondo la catena algoritmica  $j$  ( $j = \text{prima catena algoritmica, seconda catena algoritmica, terza catena algoritmica, quarta catena algoritmica}$ ) abbiamo che

$$B_{i,j} = \frac{\sum_k \sum_z a_{i,j,k,z}}{9}$$

	prima c. a.	seconda c. a.	terza c. a.	quarta c. a.
$C_{\text{CSSM}}[\%]$	10,73	10,59	6,89	7,21
$C_{\text{MUSIC}}[\%]$	11,67	11,02	10,10	10,11
$C_{\text{SRP}}[\%]$	9,57	6,17	9,24	7,33
$C_{\text{TOPS}}[\%]$	14,23	14,33	13,24	12,59
$C_{\text{WAVES}}[\%]$	13,08	12,45	10,14	11,07

*Tabella 8: valori della confidenza statistica con fattore di copertura due mediati sui test condotti e sulle distanze*

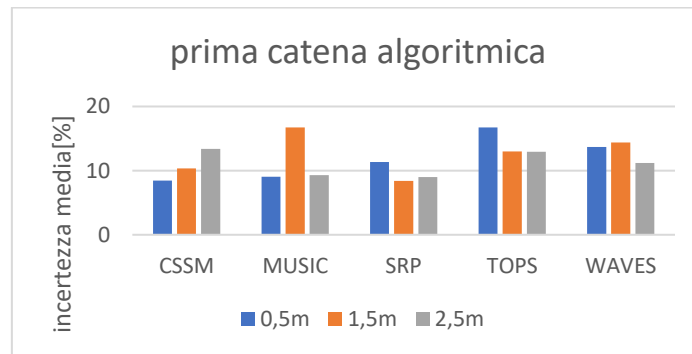
Dalla tabella 8 possiamo osservare che la terza catena algoritmica è quella che consente agli algoritmi CSSM, MUSIC, WAVES di fornire il valore più basso della confidenza statistica con fattore di copertura due mediata sulle 9 condizioni di prova testate nell'ambiente controllato ( $B_{i,j}$ ) con valori rispettivamente pari al 6,89%, 10,10%, 10,14%; per quanto riguarda l'implementazione dell'SRP è la seconda catena algoritmica a fornire il  $B_{i,j}$  più basso con un valore del 6,17%; per quanto riguarda l'implementazione del TOPS è invece la quarta catena algoritmica a fornire il  $B_{i,j}$  più basso con un valore del 12,59%.

### 3.2.3 Analisi dei dati ottenuti: confronto distanze

Le figure 35, 36, 37 e 38 riportano il comportamento degli algoritmi implementati secondo le quattro diverse catene algoritmiche alle varie distanze testate riportando la media dei valori della confidenza statistica con fattore di copertura due nelle 3 condizioni di prova definite dai 3 test. Tale valore lo indichiamo con  $C_{i,j,z}$

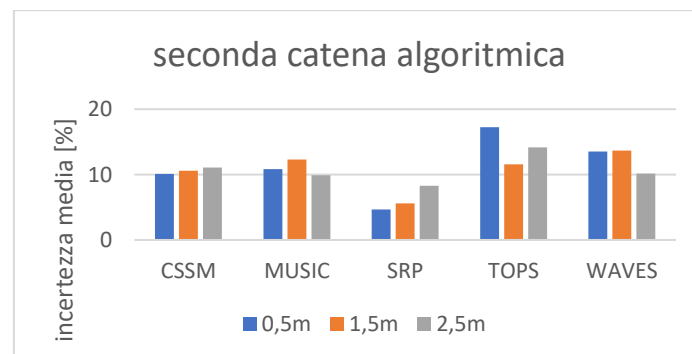
$$C_{i,j,z} = \frac{\sum_k a_{i,j,k,z}}{3}$$

Per evidenziare il comportamento degli algoritmi in funzione della distanza della sorgente per ognuna delle quattro diverse catene algoritmica tali 4 diverse figure, una per catena algoritmica, sono state organizzate riportando in ordinata l'incertezza media valutata tramite  $C_{i,j,z}$  e in ascissa, per ciascun algoritmo, le 3 distanze.



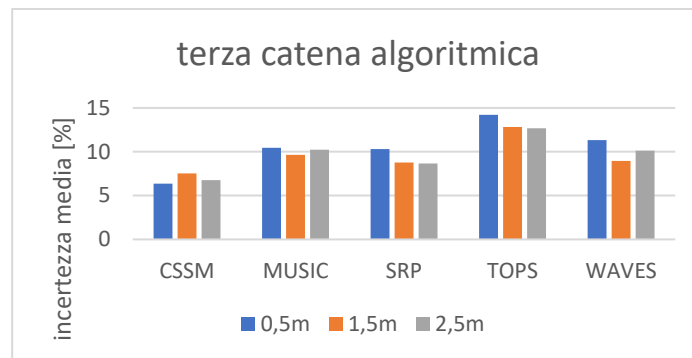
*Figura 35: valori della confidenza statistica con fattore di copertura due ottenuta dall'implementazione della prima catena algoritmica mediata sui test condotti*

Dalla figura 35 possiamo osservare che, a seguito dell'implementazione della prima catena algoritmica: il CSSM vede un valore di  $C_{i,j,z}$  che aumenta all'aumentare della distanza della sorgente; il TOPS vede un valore di  $C_{i,j,z}$  che diminuisce all'aumentare della distanza; il MUSIC, l'SRP e il WAVES hanno invece un comportamento irregolare in funzione della distanza, in particolare il MUSIC vede un valore di  $C_{i,j,z}$  pari al 16% alla distanza di 1,5m a fronte di un valore di 9% alle altre due distanze; l'SRP vede un valore di 9% alle distanze di 1,5m e 2,5m a fronte di un valore del 11,5% a 0,5m; il WAVES vede un valore del 13% a 0,5m, 14% a 1,5m e 11% a 2,5m.



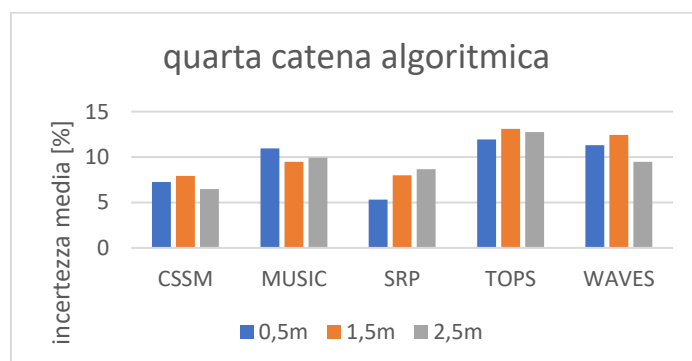
*Figura 36: valori della confidenza statistica con fattore di copertura due ottenuta dall'implementazione della seconda catena algoritmica mediata sui test condotti*

Dalla figura 36 possiamo osservare che, a seguito dell'implementazione della seconda catena algoritmica: il CSSM e l'SRP vedono un valore di  $C_{i,j,z}$  che aumenta all'aumentare della distanza della sorgente; il MUSIC, il TOPS e il WAVES hanno invece un comportamento irregolare in funzione della distanza, in particolare il MUSIC vede un valore di  $C_{i,j,z}$  pari al 12% alla distanza di 1,5m a fronte di un valore del 11% a 0,5m e 10 a 2,5m; TOPS vede il valore minore di  $C_{i,j,z}$  a 1,5m con 11% mentre mostra un valore del 17% a 0,5m e 14 a 2,5m; il WAVES vede il valore minore di  $C_{i,j,z}$  a 2,5m con il 10% mentre mostra un valore del 13% alle altre due distanze.



**Figura 37:** valori della confidenza statistica con fattore di copertura due ottenuta dall'implementazione della seconda catena algoritmica mediata sui test condotti

Dalla figura 37 possiamo osservare che, a seguito dell'implementazione della terza catena algoritmica: il TOPS e l'SRP vedono un valore di  $C_{i,j,z}$  che diminuisce all'aumentare della distanza della sorgente; il CSSM, il MUSIC e il WAVES hanno invece un comportamento irregolare in funzione della distanza, in particolare il CSSM vede un valore di  $C_{i,j,z}$  pari al 7,5% alla distanza di 1,5m a fronte di un valore del 6% a 0,5m e del 7% a 2,5m; il MUSIC vede il valore minore di  $C_{i,j,z}$  a 1,5m con il 9,5% mentre mostra un valore del 10,5% a 0,5m e del 10% a 2,5m; il WAVES vede il valore minore di  $C_{i,j,z}$  a 1,5m con il 10% mentre mostra un valore del 13% alle altre due distanze.



**Figura 38:** valori della confidenza statistica con fattore di copertura due ottenuta dall'implementazione della quarta catena algoritmica mediata sui test condotti

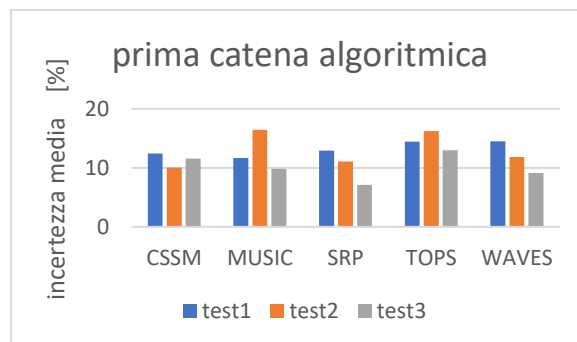
Dalla figura 37 possiamo osservare che, a seguito dell'implementazione della quarta catena algoritmica: l' SRP vede un valore di  $C_{i,j,z}$  che aumenta all'aumentare della distanza della sorgente passando dal 5% a 0,5m al 8% a 1,5m al 9% a 2,5m; il CSSM mostra il valore più alto a 1,5m con 8% mentre ha un valori del 7% a 0,5m e del 6% a 2,5m; il MUSIC mostra il valore più basso a 0,5m con 8% mentre ha un valori del 7% a 0,5m e del 6% a 2,5m; il TOPS vede un valore pari al 13% sia a 1,5m che a 2,5m mentre riporta un valore del 12% a 0,5m; il WAVES mostra il valore più alto a 1,5m con il 12% mentre ha un valore del 11% a 0,5m e 9 a 2,5m.

### 3.2.3 Analisi dei dati ottenuti: confronto test

Per valutare il comportamento degli algoritmi implementati secondo le quattro diverse catene algoritmiche nelle tre diverse condizioni di prova testate si sono realizzate la figura 39, la figura 40, la figura 41 e la figura 42 che riportano il valore della confidenza statistica con fattore di copertura due mediata, per ciascun algoritmo-catena algoritmica-test, nelle 3 distanze. Tale valore lo indichiamo con  $D_{i,j,k}$

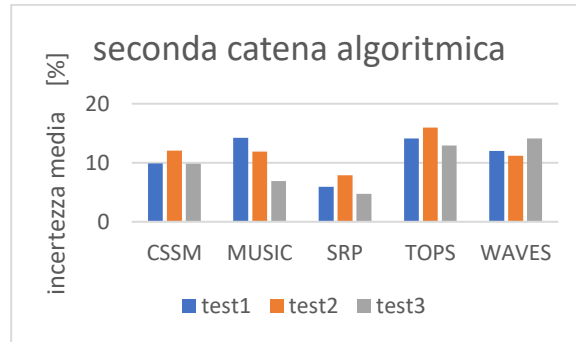
$$D_{i,j,k} = \frac{\sum_z a_{i,j,k,z}}{3}$$

Al fine di facilitare il confronto tra i valori di  $D_{i,j,k}$  per i vari algoritmi, implementati secondo le quattro diverse catene algoritmiche, nelle tre diverse condizioni di prova testate le 4 figure, una per catena algoritmica, sono state organizzate in maniera tale da riportare sull'asse delle ordinate il valore di  $D_{i,j,k}$  e su quello delle ascisse la tipologia di test e l'algoritmo a cui tali  $D_{i,j,k}$  si riferiscono.



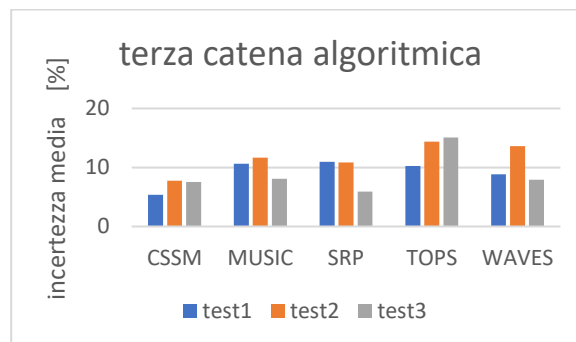
**Figura 39:** valori della confidenza statistica con fattore di copertura due mediati sulle distanze relativo all'implementazione della prima catena algoritmica

Dalla figura 39 possiamo osservare che, relativamente all'implementazione della prima catena algoritmica: il CSSM fornisce il valore di  $D_{i,j,k}$  più basso nelle condizioni del test2 con il 10%; il MUSIC, l'SRP, il TOPS e il WAVES in quelle del test3 rispettivamente con valori di  $D_{i,j,k}$  uguali al 10%,7%,12%,9%.



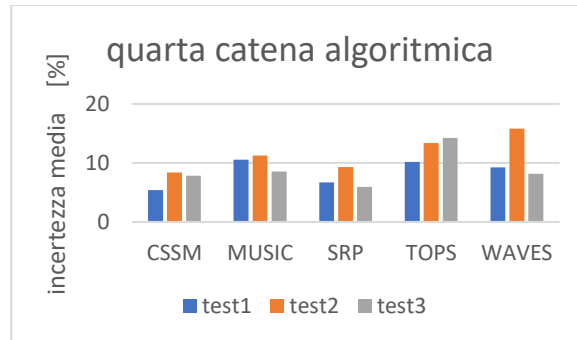
**Figura 40:** valori della confidenza statistica con fattore di copertura due mediati sulle distanze relativo all'implementazione della seconda catena algoritmica

Dalla figura 40 possiamo osservare che, relativamente all'implementazione della seconda catena algoritmica: il MUSIC, l'SRP e il TOPS forniscono il valore di  $D_{i,j,k}$  più basso nelle condizioni del test3 con valori rispettivamente pari al 7,4% e al 13%; il CSSM fornisce lo stesso valore di  $D_{i,j,k}$ , pari al 9,8% nelle condizioni del test 1 e 3, valore che risulta più basso rispetto a quello ottenuto nelle condizioni del test2; il WAVES invece presenta il  $D_{i,j,k}$  più basso nelle condizioni del test2 con un valore pari al 11%.



**Figura 41:** valori della confidenza statistica con fattore di copertura due mediati sulle distanze relativo all'implementazione della terza catena algoritmica

Dalla figura 41 possiamo osservare che, relativamente all'implementazione della terza catena algoritmica: il MUSIC, l'SRP e il WAVES forniscono il valore di  $D_{i,j,k}$  più basso nelle condizioni del test3 con valori rispettivamente pari al 8,6% e al 8%; il CSSM e il TOPS forniscono il valore di  $D_{i,j,k}$  più basso nelle condizioni del test3 con valori rispettivamente pari al 5% e al 10%.



**Figura 42:** valori della confidenza statistica con fattore di copertura due mediati sulle distanze relativo all'implementazione della quarta catena algoritmica

Dalla figura 42 possiamo osservare che, relativamente all'implementazione della quarta catena algoritmica: il MUSIC, l'SRP e il WAVES forniscono il valore di  $D_{i,j,k}$  più basso nelle condizioni del test3 con valori rispettivamente pari al 8,5%, 6% e al 8%; il CSSM e il TOPS forniscono il valore di  $D_{i,j,k}$  più basso nelle condizioni del test3 con valori rispettivamente pari al 5% e al 10%.

# Capitolo 4

## 4.1 Conclusioni

### 4.1.1 Prestazioni ottimali al variare dell'algoritmo e della catena algoritmica utilizzati

Con riferimento alla tabella 8 si può affermare che l'applicazione del filtro passa-basso consente di ridurre l'incertezza media, valutata tramite  $B_{i,j}$ , del CSSM del 35% e che la segmentazione consente di ridurre l'incertezza media, valutata tramite  $B_{i,j}$ , dell' SRP del 35%. In tutte le situazioni testate, il pre-processamento sugli audio, sia filtraggio sia la segmentazione e sia l'applicazione di entrambi, ha portato a una riduzione dell'incertezza media, valutata tramite  $B_{i,j}$ , di tutti gli algoritmi.

Sempre dalla tabella 8 si deduce che la combinazione algoritmo-catena algoritmica che fornisce i risultati migliori è l' SRP segmentato (implementato con la seconda catena algoritmica) con un'incertezza media, valutata tramite  $B_{i,j}$ , pari al 6,17%, il secondo risultato migliore è invece stato fornito dal CSSM filtrato (implementato con la terza catena algoritmica) con un'incertezza media pari al 6,9%. Occorre tuttavia notare come i risultati del CSSM si portino su livelli di incertezza media maggiori di quelli dell' SRP quando a questo non è applicato il filtraggio (per l' SRP e per il CSSM abbiamo  $B_{i,j}$  rispettivamente pari al 10,73% e 9,57% nella prima catena algoritmica, 10,59% e 6,17% nella seconda, 6,89% e 9,84% nella terza, al 7,21% e 7,33% nella quarta) e ciò ne costituisce un notevole svantaggio dato che in molte applicazioni non si conosce la frequenza del segnale da localizzare e, di conseguenza, non si può realizzare un filtro apposito. Per quanto riguarda l' SRP invece si possono notare un valore di  $B_{i,j}$  pari al 9,57% nella prima catena algoritmica, al 9,24% nella terza e al 7,33% nella quarta.

Se ne deduce quindi che l'algoritmo che presenta la minore incertezza nel localizzare la voce umana in un ambiente domestico nelle condizioni imposte in questo lavoro è, con un valore di  $B_{i,j}$  mediato sulle 4 catene algoritmiche pari al 8,1% è l' SRP e che questo esibisce un comportamento ancora migliore eseguendo la segmentazione portando a un  $B_{i,j}$  uguale al 6,17%. Ciò risulta confermato non solo dall'incertezza media ma anche dall'analisi delle occorrenze, ossia dal numero di volte che una determinata combinazione algoritmo-catena algoritmica ha fornito risultato dall'incertezza minore per le diverse combinazioni distanza-tipologia di test, facendo riferimento alla tabella 3, alla tabella 5, alla tabella 6 e alla tabella 7 vediamo infatti che sulle 9 condizioni di prova testate nell'ambiente



reale (3 distanze per 3 test) l' SRP segmentato fornisce il risultato migliore 4 volte e il CSSM filtrato fornisce il risultato migliore 3 volte.

Va inoltre evidenziato come l' SRP fornisca i risultati migliori anche relativamente alle registrazioni condotte nell'ambiente controllato: 4 volte su 6.

#### **4.1.2 Confronto prestazioni nei due diversi ambienti di prova**

Analizzando i risultati della figura 32 e della figura 33 possiamo affermare che nell'ambiente controllato si ha un miglioramento del comportamento dei 5 algoritmi implementati secondo la prima catena nelle condizioni imposte dal test1 che dal test2. Per quanto riguarda invece il confronto dei risultati tra l'ambiente controllato e quello reale nelle condizioni imposte dal test 3, con riferimento alla figura 34 possiamo invece affermare che in tale condizione di prova l'incertezza ottenuta dagli algoritmi non è risultata sensibile all'ambiente di registrazione. In particolare per quanto riguarda le condizioni imposte nel test 1 (figura 32) abbiamo: una minore incertezza del CSSM nell'elaborazione delle registrazioni condotte nell'ambiente controllato, vi è infatti una riduzione della confidenza statistica con fattore di copertura due del 35% alla distanza di 1,5m e del 62% alla distanza di 2,5m; una minore incertezza del MUSIC nell'ambiente controllato alla distanza di 1,5m (riduzione del 59%) ma peggiore alla distanza di 2,5m dove la confidenza statistica con fattore di copertura due aumenta del 25%; una minore incertezza dell' SRP nell'ambiente controllato sia alla distanza di 1,5m che a quella di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due dell' 85% e 91%; un miglior comportamento del TOPS nell'ambiente controllato alla distanza di 2,5m con una riduzione del 69% ma peggiore alla distanza di 1,5m con un aumento del 76%; una minore incertezza del WAVES sia alla distanza di 1,5m che di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due del 37% e 51%. Per quanto riguarda le condizioni imposte dal test 2 abbiamo (figura 33): una minore incertezza del CSSM nell'elaborazione delle registrazioni condotte nell'ambiente controllato, vi è infatti una riduzione della confidenza statistica con fattore di copertura due del 51% alla distanza di 1,5m e del 15% alla distanza di 2,5m; una minore incertezza del MUSIC nell'ambiente controllato alla distanza di 1,5m (riduzione del 92%) ma maggiore alla distanza di 2,5m dove la confidenza statistica con fattore di copertura due aumenta del 141%; una minore incertezza del SRP nell'ambiente controllato sia alla distanza di 1,5m che a quella di 2,5m esibendo rispettivamente

una riduzione della confidenza statistica con fattore di copertura due dell' 21% e 48%; una minore incertezza del TOPS nell'ambiente controllato alla distanza di 2,5m con una riduzione del 66% ma maggiore alla distanza di 1,5m con un aumento del 74%; una minore incertezza del WAVES sia alla distanza di 1,5m che di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due del 26% e 43%. Per quanto riguarda le condizioni del test 3 (figura 34) abbiamo: relativamente al CSSM nell'elaborazione delle registrazioni condotte nell'ambiente controllato alla distanza di 2,5m una riduzione della confidenza statistica con fattore di copertura due del 7% e alla distanza di 1,5m un aumento del 120%; una minore incertezza del MUSIC nell'elaborazione delle registrazioni condotte nell'ambiente controllato alla distanza di 2,5m con una riduzione della confidenza statistica con fattore di copertura due del 33% ma maggiore alla distanza di 1,5m con un aumento del 88%; un una minore incertezza del SRP nell'ambiente controllato sia alla distanza di 1,5m che a quella di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due dell' 38% e 66%; una minore incertezza del SRP nell'ambiente controllato sia alla distanza di 1,5m che a quella di 2,5m esibendo rispettivamente una riduzione della confidenza statistica con fattore di copertura due dell' 8% e 57%; una minore incertezza del WAVES nell'ambiente controllato sia alla distanza di 1,5m che a quella di 2,5m esibendo rispettivamente un aumento della confidenza statistica con fattore di copertura due dell' 42% e 139%.

### **4.1.3 Prestazioni ottimali al variare delle condizioni di prova**

Per quanto riguarda il comportamento degli algoritmi nelle diverse condizioni di prova testate possiamo affermare che sia il MUSIC che l'SRP, indipendentemente dalla catena algoritmica con cui sono stati implementati, hanno mostrato i risultati migliori nel test3: analizzando infatti la figura 39, la figura 40, la figura 41 e la figura 42 notiamo che per tutte le catene algoritmiche con cui è stato implementato il MUSIC ha esibito il valore di  $D_{i,j,k}$  minore nelle condizioni del test3 e che anche l'SRP, per tutte le catene algoritmiche con cui è stato implementato, ha visto nel test3 il valore di  $D_{i,j,k}$  minore.

### **4.1.4 Prestazioni ottimali al variare della distanza**

Per quanto riguarda infine il comportamento degli algoritmi in funzione delle distanze e indipendentemente dalla catena algoritmica adottata possiamo affermare che, confrontando i

valori riportati nella figura 35, nella figura 36, nella figura 37 e nella figura 38 il WAVES fornisce, tranne che quando implementato con la terza catena algoritmica, il valore di  $C_{i,j,z}$  più basso a 2,5m.

## 4.2 Sviluppi futuri

Per una più dettagliata analisi un utile proseguimento di questo lavoro potrebbe consistere nello sviluppare un'ulteriore condizione di prova nella quale la cassa altoparlante che riproduce la voce umana presenta un moto circolare uniforme con centro nell'array di microfoni individuando così, nell'arco temporale della durata della registrazione, un settore angolare ben definito. Si potrebbe quindi valutare l'efficacia dell'algoritmo proposto nel localizzare una sorgente in questa situazione valutando il numero di volte in cui la stima angolare fornita dall'algoritmo ricade nel settore angolare individuato dal moto della cassa nell'arco temporale a cui la registrazione, sulla quale viene eseguita la stima, si riferisce.

Risulterebbe quindi utile verificare se delle 20 combinazioni algoritmo-catena algoritmica valutate in questo lavoro di tesi l'SRP segmentato presenti i risultati migliori anche in questa situazione.

# 5 Bibliografia

- [1] M.S Barlett. "Smoothing Periodograms from Time Series with Continuous Spectra." *Nature*, 161:686-687,1948.
- [2] G. Bienvenu e L. Kopp. "Adaptivity to Background Noise Spatial Coherence for High Resolution Passive Methods." In *Int. Conf. Of Acoust., Speech and Signal Processing*, pages 307-310, 1980.
- [3] H. Krim e M. Viberg "Two Decades of Array Signal Processing Research." *IEEE*, 1996, pp. 67-94
- [4] H. Wang and M. Kaveh. "Coherent Signal-Subspace Processing for the Detection and Estimation of Angle of Multiple Wide-Band Sources", *IEEE*, 1985.
- [5] V.F Pisarenko. "The Retrieval of Harmonics from a Covariance Function" *Geophys. J. Roy. Astron. Soc.*, 33:347-366,1973
- [6] R.O. Schmidt: "Multiple Emitter Location and Signal Parameter Estimation", *IEEE*, 1986
- [7] M. Devendra, K. Manjunathanchari. "DOA Estimation of a System Using MUSIC Method" *IEEE*, 2015 p. 309
- [8] J. H. Di Biase. "A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays"
- [9] J. P. Dmochowski, J. Benesty. "A Generalized Steered Response Power Method for Computationally Viable Source Localization", *IEEE*, 2007
- [10] Z.I. Khan, M. MD. Kamal, N. Hamzah, K. Othman "Analysis of Performance for Multiple Signal Classification (MUSIC) in Estimating Direction of Arrival", *IEEE*, 2008, pp. 524-528.
- [11] B.D. Rao, K. V. S. Hari "Performance Analysis of Root-Music", *IEEE*, 1989
- [12] Bo, L. I. (2011). Estimation for Signal's DOA by MMUSIC Algorithm. *Journal of Computational Information Systems*, 7(15), 5461–5468
- [13] Zeng, W. J., So, H. C., & Huang, L. (2013). lp-MUSIC: Robust direction-of-arrival estimator for impulsive noise environments. *IEEE Transactions on Signal Processing*, 5, 99.
- [14] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech. Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989
- [15] M.J. Hinich, "Processing spatially aliased arrays," *J. Acoust. Soc. Am.* , vol. 64, no. 3, pp. 792-794, 1978
- [16] M. Wax and T. Kailath, "Determining the number of signals by information theoretical criteria," in *Proc. ASSP Spectrum Estimation Workshop 11*, Tampa, FL, 1983, pp. 192-196.
- [17] E. D. Di Claudio and R. Parisi, "WAVES: Weighted Average of Signal Subspace for Robust Wideband Direction Finding", *IEE*, 2001, pp. 2179-2191.

- [18] Ta Sung Lee, “ Efficient Wideband Source Localization Using Beamforming Invariance Technique”, IEEE, 1994, pp. 1376-1386.
- [19] Y. S. Yoon, L. M. Kaplan, J. H. McClellan “TOPS: New DOA Estimator for Wideband Signals”, IEEE, 2006, pp. 1977-1989.
- [20] K. Nakadai, H. G. Okuno, T. Laurens, H. Kitano “Humanoid Active Audition System”, IEEE, 2000.
- [21] K. Nakamura, K. Nakadai, F. Asano, Y. Hasegawa, and H. Tsujino “Intelligent sound source localization for dynamic environments”, IEEE, 2009, pp. 664-669.
- [22] V. Tourbabin, H. Barfuss, B. Rafaely e W. Kellerman “Enhanced robot audition by dynamic acoustic sensing in moving humanoids”, IEEE, 2015, pp 5625-5629.
- [23] E. Vincent, A. Sini e F. Charpillat “AUDIO SOURCE LOCALIZATION BY OPTIMAL CONTROL OF A MOBILE ROBOT”, IEEE, 2015, pp. 5630-5634.
- [24] V. Tourbabin and B. Rafaely, “Speaker localization by humanoid robots in reverberant environments,” in IEEE 28th Convention of Electrical Electronics Engineers in Israel (IEEEI), Dec. 2014, pp. 1–5.
- [25] K. Nakadai, T. Takahashi, H. G. Okuno, H. Nakajima, Y. Hasegawa, and H. Tsujino, “Design and implementation of robot audition system ‘HARK’ — open source software for listening to three simultaneous speakers,” *Adv. Robot.*, vol. 24, no. 5–6, pp. 739–761, 2010.
- [26] A. Schmidt, H W. Lollmann, W. Kellerman “A novel ego-noise suppression algorithm for acoustic signal enhancement in autonomous systems”, IEEE, 2018, pp. 6593-6587.
- [27] N. Mae, Y. Mitsui, S. Makino, D. Kitamura, N. Ono, T. Yamada, and H. Saruwatari, “Sound source localization using binaural difference for hose-shaped rescue robot,” *Proc. APSIPA 2017*.
- [28] N. Mae, Y. Mitsui, S. Makino, D. Kitamura, N. Ono, T. Yamada, and H. Saruwatari, “Sound source localization using binaural difference for hose-shaped rescue robot,” *Proc. APSIPA 2017*.
- [29] J. Even, Y. Morales, N. Kallakuri, C. Ishi e N. Hagita, “Audio ray tracing for position estimation of entities in blind regions”, IEEE, 2014, pp. 1920-1924.
- [30] I. Ciuffreda, “Analysis and development of robot on board sensors network to localize people in indoor living environments”, Master’s Degree’s Thesis in Biomedical Engineering, 2020.
- [31] Leila Abdoune and Mohamed Fezari “Everyday Life Sounds Database: Telemonitoring of Elderly or Disabled”, *J. Intell. Syst.* 2016; 25(1): 71–84
- [32] E. Doebelin, “Strumenti e metodi di misura” ed. Mc Graw Hills, 2004, pp. 103-106