



**Ελληνικό Μεσογειακό Πανεπιστήμιο - Hellenic Mediterranean University**

Σχολή Μουσικής και Οπτοακουστικών Τεχνολογιών - School of Music and Optoacoustic Technologies

Τμήμα Μουσικής Τεχνολογίας και Ακουστικής - Department of Music Technology & Acoustics

**M.Sc. – Τεχνολογίες Ήχου και Μουσικής - Sound and Music Technologies**

---

**Τεχνικές αυτόματης μίξης κατά την ηχογράφηση και  
αναπαραγωγή χωρικά εκτεταμένης ηχητικής σκηνής.**

Automatic mixing techniques for capturing and reproduction of a spatially extended sound scene.

---

**Διπλωματική Εργασία**

Σπυρίδων Ρουμελιώτης

**Επιβλέπων καθηγητής:**

Νικόλαος Στεφανάκης

Ρέθυμνο, Φεβρουάριος 2023

Ελληνικό Μεσογειακό Πανεπιστήμιο - Hellenic Mediterranean University

Σχολή Μουσικής και Οπτοακουστικών Τεχνολογιών - School of Music and Optoacoustic Technologies

Τμήμα Μουσικής Τεχνολογίας και Ακουστικής - Department of Music Technology & Acoustics

74133, Ρέθυμνο, Κρήτη

Phone: +30 28310 21902

<https://www.hmu.gr/>

[info@hmu.gr](mailto:info@hmu.gr)



**Ελληνικό Μεσογειακό Πανεπιστήμιο - Hellenic Mediterranean University**

Σχολή Μουσικής και Οπτοακουστικών Τεχνολογιών - School of Music and Optoacoustic Technologies

Τμήμα Μουσικής Τεχνολογίας και Ακουστικής - Department of Music Technology & Acoustics

**Τεχνικές αυτόματης μίξης κατά την ηχογράφηση και αναπαραγωγή χωρικά εκτεταμένης ηχητικής σκηνής.**

Automatic mixing techniques for capturing and reproduction of a spatially extended sound scene.

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

Σπύρος Ρουμελιώτης, ΑΜ: mta19

**Επιβλέπων καθηγητής:**

Νικόλαος Στεφανάκης

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 28/04/2023.

Νικόλαος Στεφανάκης, Επίκουρος καθηγητής, Τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Ελληνικό Μεσογειακό Πανεπιστήμιο (ΕΛΜΕΠΑ)

Ρέθυμνο, Φεβρουάριος 2023

## **ΣΠΥΡΙΔΩΝ ΡΟΥΜΕΛΙΩΤΗΣ**

Πτυχιούχος Μηχανικός Δομικών Έργων Τ.Ε., Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

Copyright © Σπυρίδων Ρουμελιώτης, 2023.

Με επιφύλαξη παντός δικαιώματος. All rights reserved. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Ελληνικού Μεσογειακού Πανεπιστημίου.

## **Ευχαριστίες**

Ένα μεγάλο ευχαριστώ στον επιβλέποντα καθηγητή μου Νικόλαο Στεφανάκη, για την εμπιστοσύνη που μου έδειξε, τη στήριξη του σε όλη τη διάρκεια της συνεργασίας μας καθώς και για την πολύτιμη συμβολή του σε όλη τη διαδικασία. Επίσης, θα ήθελα να ευχαριστήσω θερμά την οικογένεια μου και φυσικά τη σύντροφο μου Έυη, που είναι δίπλα μου και στηρίζουν πάντα όλες μου τις επιλογές.

## Περίληψη

Κατά την ηχητική κάλυψη ή ηχογράφιση εκτεταμένης χωρικής περιοχής, ενδείκνυται η χρήση πολλαπλών μικροφώνων τα οποία είναι κατανεμημένα στο χώρο. Η μίξη όμως και επεξεργασία των ηχητικών καταγραφών για τη δημιουργία του τελικού ηχητικού υλικού αποτελεί σημαντική πρόκληση, καθότι πρέπει να αντιμετωπιστούν προβλήματα όπως ο θόρυβος και η αντήχηση, οι φασματικοί χρωματισμοί και οι ανεπιθύμητες διακυμάνσεις της ηχητικής έντασης.

Στην εργασία αυτή υλοποιήθηκαν τεχνικές πολυκάναλης επεξεργασίας σήματος με στόχο την αυτοματοποίηση της μίξης για την παραγωγή του τελικού ηχητικού υλικού με όσο το δυνατόν λιγότερη παρέμβαση από τον χρήστη. Οι τεχνικές υλοποιήθηκαν σε συνθήκες προσομοίωσης με χρήση σήματος φωνής ενώ για την αξιολόγησή τους αξιοποιήθηκε μετρική βασισμένη στο κριτήριο του Perceptual Evaluation of Speech Quality (PESQ). Μέσα από ένα σύνολο από πρώτα πειράματα, αναδεικνύεται καταρχάς η εγκυρότητα του PESQ ως ένα αντικειμενικό κριτήριο για την ποσοτικοποίηση της υποβάθμισης της ποιότητας του σήματος φωνής σε ένα μεγάλο εύρος από συνθήκες αντήχησης και θορύβου.

Ένα δεύτερο σύνολο πειραμάτων υλοποιήθηκε με σκοπό τη σύγκριση διαφορετικών τεχνικών ηχητικής μίξης σε ένα περιβάλλον που προσομοιώνει θεατρική παράσταση. Τα αποτελέσματα αναδουκνείουν την υπεροχή της γνωστής τεχνικής *delay and sum* έναντι άλλων τεχνικών που βασίζονται σε αξιοποίηση ενός υποσυνόλου των διαθέσιμων μικροφώνων. Επιπλέον τεχνικές που αναδεικνύονται - όχι μόνο λόγω της επίδοσής τους αλλά και λόγω της ευκολής υλοποίησής τους - είναι μια προσέγγιση που αξιοποιεί ζεύγος μικροφώνων βασισμένο στη μέτρηση του *Magnitude Squared Coherence (MSC)* αλλά και η επιλογή του κοντινότερου στην πηγή μικροφώνου.

### Λέξεις Κλειδιά

*automix, beamforming, spotforming, differential microphone, maximum component preservation, delay and sum, magnitude squared coherence,*

## **Abstract**

When capturing a spatially extended sound scene, it is advisable to use multiple microphones distributed in space. However, the mixing and processing of the multiple audio channels is a challenging process due to problems such as noise, reverberation, spectral coloration and unwanted power variations. In this thesis, multichannel signal processing techniques are examined as the means to automate the mixing process, so as to produce an acoustic representation of the sound scene with as little intervention from the user as possible.

The techniques are implemented inside a simulated environment using dry speech signals as input, while the evaluation of the overall performance is based on the well-known metric of Perceptual Evaluation of Speech Quality (PESQ). A first series of experiments is deployed with the goal to verify the appropriateness of PESQ for measuring quality degradation in the audio signal across a wide range of noise and reverberation conditions. In a second round of experiments, the different mixing techniques are put into test in a scenario that simulates theatrical performance. The results indicate that the well-known delay-and-sum technique achieves the best performance in terms of PESQ, by exploiting all the available audio channels. However, two additional approaches that provide competitive performance and that are quite interesting due to the fact that they can be easily implemented are; an approach that utilizes a pair of microphones based on a measure of the magnitude squared coherence (MSC) and a much simpler approach that activates only a single microphone, the one that is closest to the sound source.

### **Keywords**

*automix, beamforming, spotforming, differential microphone, maximum component preservation, delay and sum, magnitude squared coherence,*

# ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

<b>1. Εισαγωγή</b>	<b>1</b>
1.1 Ορισμός προβλήματος	1
1.2 Ιστορική αναδρομή	2
1.2.1 <i>Αυτόματη μίξη (Dan Dugan)</i>	5
1.3 Υπάρχουσα μεθοδολογία	7
1.4 Σκοπός εργασίας	7
1.5 Δομή Διπλωματικής	8
<b>2. Βασικά Εργαλεία</b>	<b>10</b>
2.1. Pyroom	10
2.1.1. <i>Pyroomacoustics</i>	11
2.1.2. <i>Πυρήνας της Pyroomacoustics</i>	11
2.1.3. <i>Υπολογισμός κρουστικής απόκρισης δωματίου</i>	13
2.2. Perceptual Evaluation of Speech Quality (PESQ)	14
2.2.1. <i>Αξιολόγηση εργαλείων Pyroomacoustics &amp; PESQ</i>	15
2.2.2. <i>Συμπεράσματα</i>	17
2.3. Διαφορικές συστοιχίες μικροφώνων	18
2.4. Εκτίμηση Χρονικής Διαφοράς Αφιξης	21
<b>3. Μέθοδοι</b>	<b>24</b>
3.1. Εισαγωγή	24
3.2. Maximum Component Preservation	24
3.3. Delay & Sum	25
3.4. Spotforming	27
3.5. Magnitude Squared Coherence (MSC)	
<b>4. Υλοποίηση</b>	<b>32</b>
4.1. Περιγραφή πειραματικής διαδικασίας	32
1.1.1. <i>Μεταβλητές</i>	32
1.1.2. <i>Αποτελέσματα</i>	35
1.1.2.1. <i>Μελέτη ακραίων περιπτώσεων με βάση την απόσταση</i>	37
1.1.2.2. <i>Συνολική επίδοση των μεθόδων</i>	45
<b>5. Συμπεράσματα και Μελλοντική Εργασία</b>	<b>51</b>
5.1. Σύνοψη & Συμπεράσματα	51
5.2. Μελλοντικές Εφαρμογές & Βελτιώσεις	52
<b>Αναφορές</b>	<b>53</b>



## ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

<b>Εικόνα 1.1:</b> Παράδειγμα θεατρικής παράστασης και θέσης ηθοποιών σε σχέση με τα μικρόφωνα	2
<b>Εικόνα 1.2:</b> Ηχολήπτης (FOH) σε θεατρική παράσταση	2
<b>Εικόνα 1.3:</b> Ζώνη ομιλίας όταν $EAD=2Ds$	5
<b>Εικόνα 2.1:</b> Αρχικές δοκιμές σε φυσικό χώρο	10
<b>Εικόνα 2.2:</b> α) Παράδειγμα κώδικα Pyroomacoustics β) Παράδειγμα προσομοίωσης δωματίου με μια πηγή και μια συστοιχία 4 μικροφώνων στην ίδια ευθεία	12
<b>Εικόνα 2.3:</b> (Πάνω) Αποτύπωση κρουστικής απόκρισης σε δωμάτιο με $RT60: 0,6s$ στο 1ο μικρόφωνο της συστοιχίας (Κάτω) Το σήμα στο 1ο μικρόφωνο της συστοιχίας	13
<b>Εικόνα 2.4:</b> Η μέθοδος εικονικών μοντέλων, οι εικονικές πηγές αντιστοιχούν σε πολλαπλές διαδρομές διάδοσης του ήχου.	14
<b>Εικόνα 2.5:</b> Γράφημα που αποτυπώνει την πτώση της βαθμολογίας PESQ όσο αυξάνεται ο χρόνος αντήχησης $RT60$ και το SNR	16
<b>Εικόνα 2.6:</b> Γράφημα που αποτυπώνουν πως επιδρά η χρονική καθυστέρηση στη βαθμολογία PESQ	17
<b>Εικόνα 2.7:</b> Διάγραμμα που δείχνει τη διαφορετική βαθμολογία ανάλογα με το πολικό διάγραμμα του μικροφώνου	17
<b>Εικόνα 2.8:</b> Συλλογή όλων των ανακλώμενων ήχων, εκτός από τον απευθείας ήχο, σε ένα αμφιθέατρο.	18
<b>Εικόνα 2.9:</b> Δομή μιας διαφορικής συστοιχίας	20
<b>Εικόνα 2.10:</b> Πολικά διαγράμματα α) Καρδιοειδές, β) Υπερκαρδιοειδές, γ) Figure of 8	
<b>Εικόνα 2.11:</b> Χρονική καθυστέρηση ανάμεσα σε δύο μικρόφωνα απομακρυσμένης ηχητικής πηγής σε γωνία $\theta$ .	23
<b>Εικόνα 3.1:</b> (Αριστερά) Τυπική delay-and-sum διαμόρφωση δέσμης. (Δεξιά) Υλοποίηση του διαμορφωτή δέσμης στο πεδίο της συχνότητας, όπου τα σήματα $x_m(n)$ είναι οι μιγαδικοί συντελεστές Fourier συγκεκριμένης συχνότητας και τα βάρη $w_m$ είναι μιγαδικά.	27
<b>Εικόνα 3.2:</b> Spot-forming	29
<b>Εικόνα 4.1:</b> Απεικόνιση δωματίου που χρησιμοποιήθηκε κατά την πειραματική διαδικασία αριστερά με τέσσερα μικρόφωνα δεξιά με οκτώ, ενώ φαίνονται στις εικόνες και οι περισσότερες από τις θέσεις πηγής που μελετήθηκαν.	35
<b>Εικόνα 4.2:</b> Παράδειγμα μέρους πίνακα αποτελεσμάτων στο excel για τέσσερα μικρόφωνα, $RT60 = 60s$ και $SNR = 60$	35
<b>Εικόνα 4.3:</b> Θέσεις πηγών και μικροφώνων, τέσσερα μικρόφωνα (αριστερά), οκτώ μικρόφωνα (δεξιά)	39

## **ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ**

<b>Πίνακας 1:</b> Τεχνικές που εξετάστηκαν κατά τη διεξαγωγή των πειραμάτων	<b>37</b>
<b>Πίνακας 2:</b> Αναλυτικά αποτελέσματα για θέση πηγής [4.4 , 7.5 , 1] και 4 και 8 μικρόφωνα	<b>42</b>
<b>Πίνακας 3:</b> Αναλυτικά αποτελέσματα για θέση πηγής [3, 9, 1] και 4 και 8 μικρόφωνα	<b>45</b>
<b>Πίνακας 4:</b> Αποτελέσματα ΜΟ	<b>57</b>
<b>Πίνακας 5:</b> Αποτελέσματα μέσω των όρων από τιμές με και χωρίς συγχρονισμό για πειράματα με συστοιχία 4 μικροφώνων	<b>59</b>

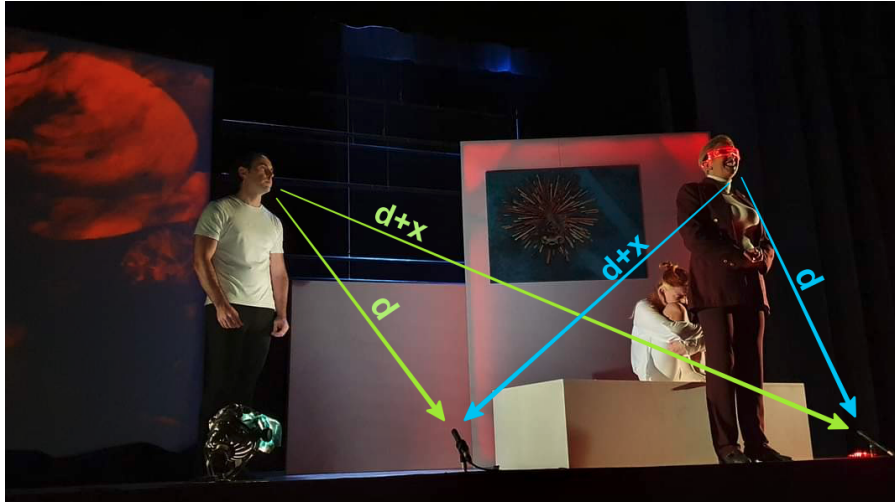
# 1. Εισαγωγή

---

## 1.1 Ορισμός προβλήματος

Η χωρική επεξεργασία ήχου γίνεται όλο και πιο σημαντική, καθώς η ποικιλία των εμπορικών εφαρμογών για πολυκάναλο ήχο αυξάνεται συνεχώς. Ένα συνηθισμένο σενάριο είναι η χρήση οικιακών συστημάτων ψυχαγωγίας για την ακρόαση πολυκάναλης μουσικής ή την παρακολούθηση ταινιών με ήχο surround. Επιπλέον, υπάρχει αυξανόμενο ενδιαφέρον για τη διάθεση χωρικού ήχου για συστήματα τηλεδιασκέψεων υψηλής ποιότητας. Είναι σαφές ότι η τελευταία εφαρμογή απαιτεί όχι μόνο μια ρεαλιστική αναπαραγωγή του χωρικού ήχου, αλλά και κατάλληλες μεθόδους για την καταγραφή και την ενίσχυση του (Thiergart et al., 2011).

Ένας ηχολήπτης που εργάζεται σε θεατρικές παραστάσεις με πολλούς ηθοποιούς οι οποίοι βρίσκονται ταυτόχρονα πάνω στη σκηνή, καλείται να αντιμετωπίσει αρκετές προκλήσεις. Με μία από αυτές καταπιάνεται η παρούσα διπλωματική εργασία. Σε μια σκηνή συνήθως τοποθετούνται κατευθυντικά μικρόφωνα κατα μήκος της σκηνής ή και σε άλλα σημεία εφόσον αυτό είναι απαραίτητο. Το αποτέλεσμα αυτής της συνθήκης είναι ότι όταν ένας ηθοποιός ξεκινήσει να μιλάει το πιθανότερο είναι ότι θα βρίσκεται πιο κοντά σε ένα από αυτά τα μικρόφωνα (Εικόνα 1.1), επομένως ο ηχολήπτης καλείται να γνωρίζει σε μεγάλο βαθμό το πότε μιλάει ο κάθε ηθοποιός και σε ποια θέση, ώστε να ενισχύει την ένταση του κοντινότερου σε αυτόν μικροφώνου και ταυτόχρονα να χαμηλώνει τα υπόλοιπα (Εικόνα 1.2). Αν δεν γίνει αυτό θα αρχίσουν να παρουσιάζονται στη μίξη προβλήματα φάσης, comp filtering, ανεπιθύμητη ήχοι κ.α. Στην παρούσα εργασία λοιπόν γίνεται μια προσπάθεια για τη δημιουργία ενός αλγόριθμου αυτόματης μίξης, με διαφορετική προσέγγιση από αυτές τις οποίες βασίζονται τα ήδη υπάρχοντα συστήματα αυτόματης μίξης.



*Εικόνα 1.1: Παράδειγμα θεατρικής παράστασης και θέσης ηθοποιών σε σχέση με τα μικρόφωνα*



*Εικόνα 1.2: Ηχολήπτης (FOH) σε θεατρική παράσταση*

## 1.2 Ιστορική αναδρομή

Είναι σύνηθες για έναν μηχανικό ήχου να σκέφτεται αρκετά τις τεχνικές λεπτομέρειες όταν προσεγγίζει ένα νέο πρόβλημα. Να ξεκινά με την απόφαση για το πού θα τοποθετηθούν τα ηχεία και τα μικρόφωνα και ποιος εξοπλισμός θα είναι ο καλύτερος για τη δουλειά. Στις περισσότερες περιπτώσεις, αυτή η προσέγγιση είναι απολύτως έγκυρη. Υπάρχει βέβαια πάντα ο κίνδυνος η ενασχόλησή με τον εξοπλισμό και τις προδιαγραφές να αποπροσανατολίσει από

τον πραγματικό σκοπό των προσπαθειών δηλαδή την ευκολότερη και πιο ικανοποιητική ενίσχυση του ήχου. Οι συγκεκριμένες προσεγγίσεις έγιναν στο παρελθόν από τον μηχανικό Dan Dugan και αποτελούν μέχρι σήμερα ένα από τα πιο διαδεδομένα εργαλεία αυτόματης μίξης. Όλα εξαρτώνται φυσικά από πολλούς παράγοντες, όπως από τις ακουστικές ιδιότητες του περιβάλλοντος, αλλά και από την ικανότητα του ατόμου ως ομιλητή. Οι μεταβλητές που επηρεάζουν την ομιλία ενός ατόμου μπορούν να καταγραφούν ως εξής:

1. η προσπάθεια
2. η φωνητική ικανότητα
3. η ακουστική οξύτητα
4. ο θόρυβος του περιβάλλοντος
5. η αντήχηση.

Τα σημεία 1 - 3 είναι ανθρώπινες μεταβλητές, τα 5 και 6 είναι περιβαλλοντικές μεταβλητές. Όλες οι μεταβλητές αλλάζουν συνεχώς, επομένως είναι δύσκολο να μετρηθούν. Εάν όμως στις μεταβλητές αποδοθούν τυπικά εύρη τιμών, θα καταστεί δυνατός ο σχεδιασμός χώρων στους οποίους η ομιλία θα είναι σχετικά εύκολη για σχεδόν όλους τους ανθρώπους, όπως ακριβώς σχεδιάζεται μια πόρτα ώστε να είναι αρκετά ψηλή για να περνούν οι άνθρωποι χωρίς να χτυπούν το κεφάλι τους. Ένα στοιχείο που δυσκολεύει τα πράγματα στον τομέα της ηχητικής ενίσχυσης, είναι η έλλειψη μιας άμεσης μέτρησης της αποτελεσματικότητας της επικοινωνίας που μεταδίδεται μέσω ενός συστήματος (Dugan, 1975).

Η ανάγκη για συναντήσεις με μεγάλο αριθμό ατόμων εισάγει ένα πρόβλημα επικοινωνίας. Ο μεγαλύτερος αριθμός που μπορεί να δημιουργήσει αμοιβαία επικαλυπτόμενες ζώνες ομιλίας 3 μέτρων είναι περίπου δώδεκα άτομα. Αυτή η φυσική συνθήκη θέτει όρια στην αποτελεσματικότητα των βασικών δραστηριοτήτων της δημοκρατίας, της θρησκείας και της τέχνης. Εκείνοι με τις πιο δυνατές φωνές δεν είναι απαραίτητα εκείνοι των οποίων η ομιλία ωφελεί περισσότερο τους γείτονές τους. Η ηχητική ενίσχυση δημιουργήθηκε προκειμένου να διευρύνει αυτά τα όρια επεκτείνοντας το χωρικό εύρος της ομιλίας. Αν και η επαφή με τα μάτια και το πρόσωπο χάνεται σε μεγαλύτερες αποστάσεις, η διατήρηση του λεκτικού καναλιού μεταφέρει τις περισσότερες πληροφορίες πέρα από τα φυσικά όρια. Ένα μικρόφωνο σε ένα δωμάτιο μπορεί να μεταφέρει αποτελεσματικά όλους όσους βρίσκονται εντός του διαγράμματος ακτινοβολίας του μεγαφώνου στη ζώνη ομιλίας του ομιλητή. Θα μπορούσε

αυτό να θεωρηθεί ως μια ακουστική στρέβλωση χώρου, τα τμήματα λήψης των ζωνών επικοινωνίας όλων των ακροατών προβάλλονται σε έναν συμπυκνωμένο χώρο γύρω από το μικρόφωνο. Έτσι, υπάρχει μια ειδική ζώνη ομιλίας γύρω από ένα μικρόφωνο, και θα μπορούσε να οριστεί με παρόμοιο τρόπο:

Η ενεργός ζώνη ομιλίας που σχετίζεται με ένα μικρόφωνο σε ένα σύστημα ενίσχυσης είναι η περιοχή εντός της οποίας ένας ομιλητής που παράγει λιγότερο από 65 dB (μετρούμενο στα 0,9 m) μπορεί να μιλήσει σε όλους τους συγκεκριμένους ακροατές με λιγότερο από 15% απωλειών ( $AL_{cons}$ )<sup>1</sup>.

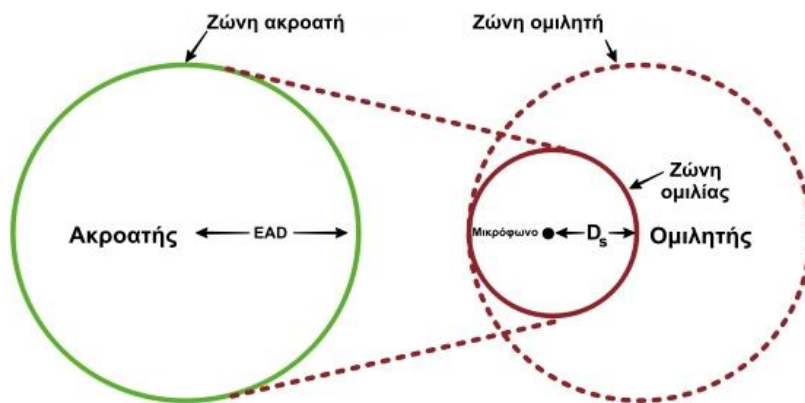
Και αυτός ο ορισμός απαιτεί έναν συμπληρωματικό ορισμό:

Η παθητική ζώνη ομιλίας που σχετίζεται με ένα σύστημα ηχητικής ενίσχυσης είναι η περιοχή που περιλαμβάνει όλους τους καθορισμένους ακροατές στην ομιλία των ενεργών συμμετεχόντων στο διάλογο. Οι πραγματικές αποστάσεις του ομιλητή από το μικρόφωνο διατηρούν μια σταθερή αναλογικότητα προς τις αντίστοιχες εικονικές αποστάσεις από τα αυτιά των ακροατών. Λόγω της ακουστικής ανατροφοδότησης, το αυτί του ακροατή μπορεί να τοποθετηθεί (εικονικά) μόνο ακριβώς στην απόσταση του μικροφώνου, στο ραδιόφωνο, στην ηχογράφιση ή σε συστήματα που διανέμουν τον ήχο επιλεκτικά. Για τους σκοπούς της παρούσας εργασίας η συζήτηση θα περιοριστεί μόνο στα συστήματα εκείνα που διανέμουν την ενισχυμένη ηχητική στάθμη ομοιόμορφα σε όλο το χώρο, τόσο στην περιοχή του ακροατή όσο και στην περιοχή του μικροφώνου. Αυτός είναι ο μόνος τύπος συστήματος που επιτρέπει άνετο διάλογο σε συνεδριάσεις, εκκλησίες, αίθουσες δικαστηρίων, συμβούλια, νομοθετικά σώματα κ.λπ.

Το μέγεθος της συρρίκνωσης ή της σμίκρυνσης της ζώνης γύρω από το μικρόφωνο εξαρτάται από την ποιότητα του ηχητικού συστήματος, ιδίως από το πλάτος και τη συχνότητα φάσης του. Ο λόγος  $EAD/D_s$  (Εικόνα 1.3), είναι ένας χρήσιμος δείκτης του πόσο "ζεστό" είναι ένα μικρόφωνο ηχοσυστήματος και ίσως είναι πιο "οργανικός" από το ακουστικό κέρδος που συνήθως καθορίζεται. Το  $EAD/D_s$ , περιγράφει άμεσα τη φυσική αναλογία μεταξύ των ζωνών ομιλίας του ακροατή και του μικροφώνου και είναι μια σταθερά για μια δεδομένη ρύθμιση ενισχυτή (Dugan, 1975).

---

<sup>1</sup> <https://www.itwissen.info/en/articulation-loss-of-consonants-audibility-ALcons-120021.html#gsc.tab=0> , <https://en.wikipedia.org/wiki/Consonant>



Εικόνα 1.3: Ζώνη ομιλίας όταν  $EAD=2D_s$

### 1.2.1 Αυτόματη μίξη (Dan Dugan)

Το όνειρο ενός αποτελεσματικού συστήματος φωνητικής ενίσχυσης επιδιώκεται από τους δημιουργικούς ηχολήπτες από τότε που αναγνωρίστηκε το πρόβλημα. Το εμπόδιο στο οποίο απέτυχαν οι περισσότερες απόπειρες είναι η μέθοδος με την οποία ανιχνεύεται η ανάγκη ενεργοποίησης ενός μικροφώνου. Εξαιρετικές επιδόσεις έχουν επιτευχθεί σε συστήματα στα οποία η ανίχνευση γινόταν με άλλα μέσα εκτός από ακουστικά, όπως διακόπτες push-to-talk. Ο έλεγχος της έντασης για την αποφυγή της ανατροφοδότησης (feedback noise) σε αυτά τα συστήματα έχει επιτευχθεί είτε με λογική που αποτρέπει να είναι ταυτόχρονα ανοιχτά περισσότερα από ένα μικρό αριθμό μικροφώνων, είτε με κύριο ρυθμιστή έντασης που ανταποκρίνεται στον αριθμό των ανοικτών μικροφώνων, είτε με συνδυασμό αυτών των μεθόδων. Το γενικό μειονέκτημα των συστημάτων αυτών έγκειται στην ποσότητα του ειδικού εξοπλισμού που απαιτείται για κάθε εργασία.

Τα συστήματα που λειτουργούν με φωνή δοκιμάζονται συχνά αλλά σπάνια είναι αποτελεσματικά στην αύξηση της έντασης ενός ηχητικού συστήματος. Αυτό οφείλεται στην ανεπάρκεια της διαδικασίας που χρησιμοποιείται για την ανίχνευση των ενεργών μικροφώνων, gating με σταθερό κατώφλι. Οι συσκευές VOX<sup>2</sup> (Voice-operated switch) μπορούν να είναι αποτελεσματικές στη μείωση της πρόσληψης του περιβαλλοντικού θορύβου σε ένα πολυφωνικό σύστημα καθώς η στάθμη θορύβου είναι πάντα μικρότερη από την

<sup>2</sup> [https://en.wikipedia.org/wiki/Voice-operated\\_switch](https://en.wikipedia.org/wiki/Voice-operated_switch)

ελάχιστη στάθμη ομιλίας. Αυτή η συνθήκη δεν υπάρχει στις περισσότερες αίθουσες συνεδριάσεων. Είναι αδύνατο να βρεθεί μια ικανοποιητική ρύθμιση για τα χειριστήρια VOX στις περισσότερες εφαρμογές. Εάν το κατώφλι VOX ενός καναλιού μικροφώνου ρυθμιστεί σε αρκετά χαμηλή στάθμη για να ανιχνεύει τα σήματα ελάχιστης στάθμης που εμφανίζονται, όπως ένας αδύναμος ομιλητής που είναι λίγο εκτός μικροφώνου, τότε οποιοσδήποτε δυνατός ήχος στην αίθουσα, όπως ένας δυνατός ομιλητής σε άλλο μικρόφωνο, μουσική ή χειροκρότημα θα ενεργοποιήσει το μικρόφωνο. Εάν, για να αποφευχθεί αυτή η λανθασμένη ενεργοποίηση, το κατώφλι VOX μιας εισόδου έχει ρυθμιστεί σε αρκετά υψηλό επίπεδο ώστε να απορρίπτει το θόρυβο και τους ήχους του μεγαφώνου, τότε θα είναι πολύ αναίσθητο για να πιάσει την απαλή ομιλία. Το ακουστικό αποτέλεσμα θα είναι ενοχλητικό (διακοπτόμενη ομιλία).

Έχουν κατασκευαστεί πειραματικά και πρακτικά συστήματα στα οποία χρησιμοποιείται είτε VOX είτε χειροκίνητο gating των καναλιών εισόδου και ένας αυτόματος εξασθενητής μειώνει το κύριο κέρδος σύμφωνα με τον αριθμό των εισόδων που ενεργοποιούνται κάθε φορά. Ένα σύστημα με έναν τέτοιο αυτόματο εξασθενητή θα παρέχει τη μέγιστη δυνατή ενίσχυση για τον αριθμό των μικροφώνων που είναι ενεργοποιημένα ανά πάσα στιγμή. Εάν ένας αυτόματος κύριος ρυθμιστής κέρδους αυτού του τύπου συνδυαζόταν με έναν δυναμικό επεξεργαστή τύπου gate, που λειτουργεί καλά με τη φωνή, θα μπορούσε να δημιουργηθεί ένα πλήρως αυτόματο σύστημα μίξης (Dugan, 1975).

Οι gate based αυτόματοι μίκτες και όχι αυτοί που χρησιμοποιούν την αύξηση κέρδους (gain), παρουσιάζουν μια χαρακτηριστική χρονική περιβάλλουσα στο σήμα που χρησιμοποιείται ως πύλη (gate). Η εντολή gate on ακολουθείται από την ενεργοποίηση της εισόδου ή, καλύτερα, από μια δυναμική αύξησης κέρδους σταθερού ρυθμού, συνήθως 3 έως 15 msec. Η πύλη διατηρείται ανοιχτή όσο διαρκεί η εντολή. Όταν τελειώνει η εντολή, μπορεί να υπάρχει χρόνος αναμονής που προορίζεται να κρατήσει την πύλη ανοιχτή κατά τη διάρκεια των σύντομων παύσεων στην ομιλία που εκτελείται, συνήθως 0,2 έως 0,4 sec. Μετά το χρόνο αναμονής (εάν υπάρχει), η ενίσχυση είτε μεταβαίνει στο χαμηλό της επίπεδο, είτε, σε πιο ήπια συστήματα, μειώνεται, συνήθως σε διάστημα 0,3 δευτερολέπτων. Η προσοχή στους χρόνους επαναφοράς, συγκράτησης και απόσβεσης μπορεί να κάνει πολλά για να προσεγγίσει έναν ανεκτό ήχο σε έναν αυτόματο μίκτη. Οι μίκτες που διαθέτουν μια λειτουργία επέκτασης παρά μια απλή πύλη παράγουν πολύ καλύτερο ήχο όταν λειτουργούν κοντά στο κατώφλι (Dugan, 1992).



### 1.3 Υπάρχουσα μεθοδολογία

Η τεχνολογία μείωσης του θορύβου είναι ένα από τα πιο σημαντικά ζητήματα για τα συστήματα προφορικού διαλόγου. Καθώς αυξάνεται η απόσταση μεταξύ ενός χρήστη και ενός μικροφώνου, ο λόγος σήματος προς θόρυβο (SNR) μειώνεται. Έχουν προταθεί πολλά είδη μεθόδων μείωσης του θορύβου όπως για καταγραφή ομιλίας με χρήση αλγορίθμων αποθορυβοποίησης, μέθοδοι επεξεργασίας σήματος για την ενίσχυση της ομιλίας, εξαγωγή χαρακτηριστικών με θόρυβο και στατιστική μοντελοποίηση με χρήση δεδομένων ομιλίας με θόρυβο. Οι τεχνολογίες για καταγραφή ομιλίας με χρήση αλγορίθμων αποθορυβοποίησης, είναι εύκολο να χρησιμοποιηθούν σε συστήματα προφορικού διαλόγου, επειδή δεν απαιτείται καμία τροποποίηση στο σύστημα, χρειάζεται μόνο να παρεμβληθεί στη ροή του σήματος ένας αλγόριθμος αποθορυβοποίησης. Οι μέθοδοι συστοιχίας μικροφώνων χρησιμοποιούνται συχνά για την καταγραφή ομιλίας χωρίς θόρυβο. Σε αυτά τα συστήματα χρησιμοποιούνται πολλά μικρόφωνα και η κατεύθυνση των πηγών βρίσκεται με την επεξεργασία των πολλών σημάτων ομιλίας που καταγράφονται από κάθε μικρόφωνο (η τεχνική ονομάζεται "σηματισμός δέσμης"). Όλα τα σήματα θορύβου που βρίσκονται εκτός της δέσμης μπορούν να αποσβεστούν. Επιπλέον, ο ομιλητής μπορεί να κινείται κατά την ομιλία του, επειδή η κατεύθυνση της δέσμης μπορεί να αλλάξει εύκολα αλλάζοντας ορισμένες παραμέτρους στον υπολογισμό.

### 1.4 Σκοπός εργασίας

Όπως προκύπτει από το προηγούμενο κεφάλαιο ο Dan Dugan ήταν από τους πρώτους που ξεκίνησε να πειραματίζεται με τα συστήματα αυτόματης μίξης.

Τα συστήματα του Dan Dugan παραμένουν μέχρι και σήμερα αρκετά διαδεδομένα και τα συναντά κανείς σε συνεδριακούς χώρους, σε τηλεοπτικά στούντιο αλλά πλέον με την είσοδο των ψηφιακών μικτών στις περισσότερες ψηφιακές κονσόλες σαν έξτρα λειτουργία. Ωστόσο το automix του Dan Dugan θεωρείται ένα gate based automix σύστημα. Στην παρούσα εργασία η ιδέα για τη δημιουργία ενός συστήματος αυτόματης μίξης μπορεί να προήλθε από τον Dan Dugan, όμως μετά την παρότρυνση του επόπτη μου κ. Νίκου Στεφανάκη ο οποίος ερευνητικά ασχολείται με συστοιχίες μικροφώνων, πρότεινε να διερευνηθεί η δυνατότητα

δημιουργίας ενός αλγορίθμου αυτόματης μίξης, με διαφορετική προσέγγιση από εκείνη του gate based, που θα υλοποιείται στο πεδίο του χρόνου συχνότητας και θα μπορούσε να έχει καλά αποτελέσματα σε θεατρικές παραστάσεις, όπου οι ηχητικές πηγές (ηθοποιοί) δεν είναι σταθερές. Η παρούσα εργασία λοιπόν σκοπό έχει τη δημιουργία ενός συστήματος μίξης χρησιμοποιώντας κατανεμημένες διατάξεις μικροφώνων και ένα συνδυασμό μεθόδων και εργαλείων με σκοπό την βελτιστοποίηση της αποδοσης.

## **1.5 Δομή Διπλωματικής**

Η παρουσίαση της εργασίας ξεκινάει με ένα εισαγωγικό κεφάλαιο, στο οποίο γίνεται μια εισαγωγή στο θέμα, καθώς και μια σχετική ιστορική αναδρομή. Στη συνέχεια, η εργασία χωρίζεται σε δύο βασικά μέρη, στο Θεωρητικό και στο Πειραματικό.

Το θεωρητικό μέρος αναφέρονται και αναλύονται τα βασικά εργαλεία και οι μέθοδοι που χρησιμοποιήθηκαν για τις ανάγκες τις εκπόμενες των πειραμάτων.

Έπειτα ακολουθεί το πειραματικό μέρος, στο οποίο γίνεται η περιγραφή του πειράματος, η παρουσίαση των αποτελεσμάτων αυτών καθώς και τα συμπεράσματα που προέκυψαν.

---

**Μέρος Α΄**  
**Βασικά Εργαλεία - Μέθοδοι**

---

## 2. Βασικά Εργαλεία

---

### 2.1. Pyroom

Τα πρώτα πειράματα ξεκίνησαν να διεξάγονται σε πραγματικές συνθήκες, σε φυσικό χώρο, με πραγματικά μικρόφωνα (Εικόνα 2.1). Αυτό που πολύ σύντομα παρατηρήθηκε ήταν η δυσκολία σε σχέση με τη εξαγωγή σταθερών συμπερασμάτων διότι οι αστάθμητοι παράγοντες ήταν αρκετοί και δημιουργούσαν αρκετά προβλήματα. Πέρα από τους ασταθμητους παραγοντες όμως, και οι τεχνικές δυσκολίες, όπως για παράδειγμα το γεγονός ότι δεν υπήρχε το απόλυτο ταίριασμα στα μικρόφωνα ή ότι ακουστική του χώρου ήταν ακατάλληλη, ήταν ζητήματα τα οποία έπρεπε να αντιμετωπιστούν. Ένα ακόμα σημαντικό μειονέκτημα, των αρχικών πειραμάτων ήταν ο περιορισμός των επιλογων και των σεναρίων που θα μπορούσαν να εξεταστούν.



*Εικόνα 2.1: Αρχικές δοκιμές σε φυσικό χώρο*

Απο το ξεκίνημα του πειραματικού μέρους της παρουσιας εργασίας δημιουργήθηκε η ανάγκη για αλλαγή προσανατολισμού σε σχέση με τις δοκιμές. Η δημιουργία προσομοίωσης δωματίων και μικροφώνων θεωρήθηκε μια πολυ καλη λύση, όπου μετα απο έρευνα, η προσομοίωση δωματίου με χρήση της Python βιβλιοθήκης “Pyroomacoustics” έδειχνε να είναι μια αρκετά καλή επιλογή. Ξεκίνησαν οι δοκιμές και

το στήσιμο των προσομοιωτών “Pyroomacoustics” το οποίο εν τέλει αποτελεί ένα από τα βασικά εργαλεία των πειραμάτων.

### **2.1.1. Pyroomacoustics**

Το pyroomacoustics είναι ένα πακέτο λογισμικού που ως στόχο έχει την προσομοίωση της ακουστικής κλειστών χώρων στο πεδίο του χρόνου, με έμφαση κυρίως στον υπολογισμό της κρουστικής απόκρισης (Scheibler, Bezzam, Dokmanic 2018).

Το περιεχόμενο του πακέτου χωρίζεται σε 3 βασικές κατηγορίες:

- κατασκευή διαφορετικών σεναρίων προσομοίωσης που περιλαμβάνουν πολλαπλές πηγές ήχου και μικρόφωνα σε 2D και 3D χώρους
- Παραγωγή κρουστικών αποκρίσεων χώρων και προσομοίωση της διάδοσης μεταξύ πηγών και μικροφώνων
- Απαραίτητες βιβλιοθηκες για Beamforming, direction finding, adaptive filtering

Με τη βοήθεια του εργαλείου αυτού υπάρχει η δυνατότητα μιας αρκετά ρεαλιστικής απεικόνισης χώρων σε ένα ελεγχόμενο περιβάλλον, δυνατότητα για πολλές επαναλήψεις, με παραμέτρους που μπορούν να μετατρέψουν τον προσομοιωτή σε κάτι που μπορεί σε καλό βαθμό να αντικαταστήσει μια διεξαγωγή πειραμάτων σε φυσικό χώρο. Η προσομοίωση βασίζεται στο μοντέλο εικονικής πηγής (Image Source Method - ISM), στο οποίο προσομοιώνονται οι ανακλάσεις στους τοίχους με εικονικές πηγές που αναπαράγουν τον ίδιο ήχο της πηγής με την απαραίτητη καθυστέρηση, εξασθένηση και φιλτράρισμα, δημιουργώντας έτσι μια κρουστική απόκριση του χώρου που έχει δημιουργηθεί για μεγάλο αριθμό πηγών (Allen, J. B., & Berkley 1979).

### **2.1.2. Πυρήνας της Pyroomacoustics**

Εκμεταλλούμενη τα αντικειμενοστραφή χαρακτηριστικά της Python η “Pyroomacoustics” δημιουργεί ένα “Application Programming Interface (API)” για την προσομοίωση της ακουστικής χώρου, οι 3 βασικές κλάσεις είναι η κλάση δωμάτιο (Room), η ηχητική πηγή (Sound Source) και η συστοιχία μικροφώνων (Microphone Array). Δημιουργείται ένα

σενάριο προσομοίωσης ορίζοντας ένα δωμάτιο, θέσεις πηγών (ηχείων) και μικροφώνων και από το χρήστη εισάγεται και το ηχητικό σήμα που θα τροφοδοτηθεί στα ηχεία. Η μέθοδος ISM εφαρμόζεται με βάση τη γεωμετρία του δωματίου για την εύρεση όλων των ειδώλων των πηγών και από την πρωταρχική πηγή και τα είδωλά της προκύπτουν τα χαρακτηριστικά της κρουστικής απόκρισης του εκάστοτε χώρου. Το τελικό ηχητικό σήμα σε κάθε μικρόφωνο υπολογίζεται μέσω της συνέλιξης του ακατέργαστου ηχητικού σήματος με την κρουστική απόκριση που αφορά το κάθε μικρόφωνο. Ιδανικά, ένα ανηχοϊκό (dry) ηχητικό σήμα που ορίζεται ως διέγερση, θα παράγει στην έξοδο ένα επεξεργασμένο (wet) ηχητικό σήμα το οποίο θα περιέχει τα χαρακτηριστικά της αντήχησης του εικονικού δωματίου. Ένα απλο παράδειγμα του κώδικα και του αποτελέσματος που προκύπτει φαίνεται στην Εικόνα 2.2.

```

%% The desired reverberation time and dimensions of the room
rt60_tgt = 0.2 # seconds
room_dim = [7.5, 5, 3] # meters

# import a mono wavfile as the source signal
# the sampling frequency should match that of the room
audio, fs = librosa.load("cmdmic.wav")
noise, fs = librosa.load("noise.wav")

# We invert Sabine's formula to obtain the parameters for the ISM simulator
e_absorption, max_order = pra.inverse_sabine(rt60_tgt, room_dim)

%% Create the room
m = pra.make_materials(
    ceiling="hard_surface",
    floor="concrete_floor",
    east="brickwork",
    west="brickwork",
    north="brickwork",
    south="brickwork",)
h = pra.Material(energy_absorption="hard_surface")
room = pra.ShoeBox(
    room_dim, fs=16000, material=pra.Material(e_absorption), max_order=max_order
)

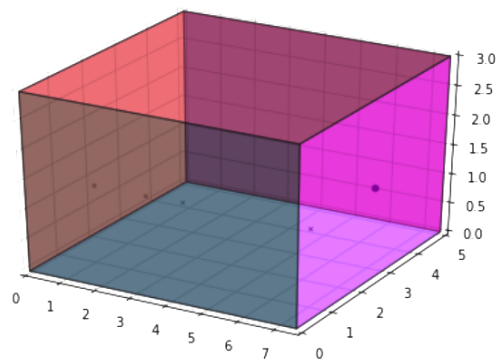
%% Source & Mics
# place the source in the room
room.add_source([1, 1, 1.28], signal=audio, delay=0.5)
room.add_source([1, 4, 1.28], signal=noise, delay=0.5)

# define the locations of the microphones
mic_locs = np.c_[
    [7, 1, 1.28], # mic 1
    [7, 2, 1.28], # mic 2
    [7, 3, 1.28], # mic 3
    [7, 4, 1.28], # mic 4
]

# finally place the array in the room
room.add_microphone_array(mic_locs)
room.plot(img_order=0)
plot.show()
%% Run the simulation (this will also build the RIR automatically)
room.simulate()

```

a) Κώδικας

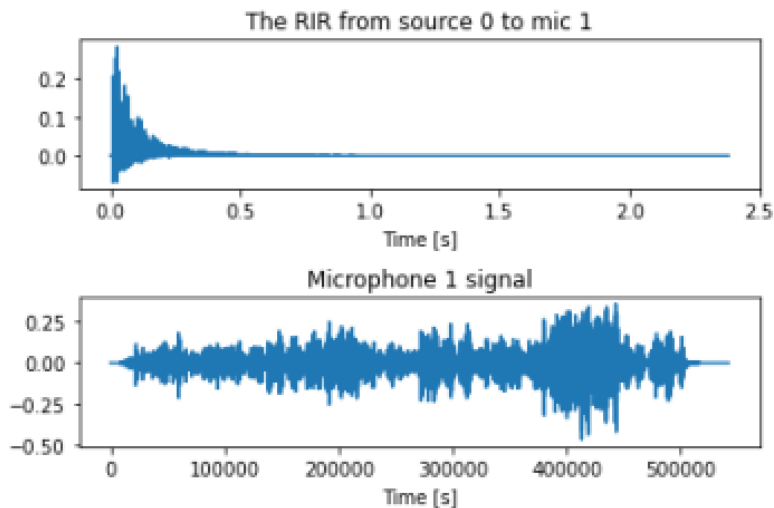


b) Δωμάτιο

**Εικόνα 2.2:** a) Παράδειγμα κώδικα Pyroomacoustics b) Παράδειγμα προσομοίωσης δωματίου με μια πηγή και μια συστοιχία 4 μικροφώνων στην ίδια ευθεία

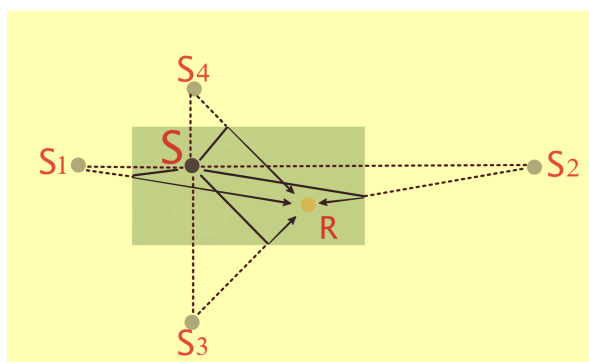
### 2.1.3. Υπολογισμός κρουστικής απόκρισης δωματίου

Όπως αναφέρθηκε παραπάνω η γεννήτρια κρουστικής απόκρισης βασίζεται στα μοντέλα εικονικών πηγών και εξετάζει δύο περιπτώσεις: ορθογώνια δωμάτια (shoebox) και αυθαίρετα πολυεδρικά δωμάτια. Για την περίπτωση των ορθογώνιων δωματίων - η οποία αποτελεί τη μελέτη περίπτωσης της παρούσας εργασίας- το “Pyroomacoustics” χρησιμοποιεί τον αλγόριθμο των Allen and Berkley [2], για τον υπολογισμό της κρουστικής απόκρισης (Εικόνα 2.3). Η περίπτωση των πολυεδρικών δωματίων είναι αρκετά πιο σύνθετη καθώς ο αριθμός των εικονικών πηγών αυξάνεται εκθετικά και δεν ελέγχεται η ορατότητα των πηγών. Επομένως ο υπολογισμός της κρουστικής απόκρισης γίνεται με διαφορετικό τρόπο (Schimmel, Müller, Dillier 2009).



*Εικόνα 2.3: (Πάνω) Αποτύπωση κρουστικής απόκρισης σε δωμάτιο με  $RT60: 0,6s$  στο 1ο μικρόφωνο της συστοιχίας (Κάτω) Το σήμα στο 1ο μικρόφωνο της συστοιχίας*

Στη μέθοδο εικονικής πηγής, η ηχητική πηγή καθρεφτίζεται στις επιφάνειες του δωματίου με σκοπό να δημιουργηθούν οι εικονικές πηγές, όπως απεικονίζεται στη Εικόνα 2.4. Μια ευθεία γραμμή από μια εικονική πηγή στον δέκτη αντιστοιχεί σε μια διαδρομή διάδοσης του ήχου στο δωμάτιο που περιλαμβάνει μία ή περισσότερες κατοπτρικές ανακλάσεις. Γνωρίζοντας το μήκος της γραμμής και την απορρόφηση των επιφανειών που τέμνει, μπορεί να υπολογιστεί η κρουστική απόκριση του δωματίου, από τη συμβολή της αντίστοιχης διαδρομής διάδοσης. Η μέθοδος της πηγής εικόνας εγγυάται την ακριβή εύρεση όλων των διαδρομών διάδοσης σε ένα δωμάτιο (Allen & Berkley 1979).



*Εικόνα 2.4: Η μέθοδος εικονικών μοντέλων, οι εικονικές πηγές αντιστοιχούν σε πολλαπλές διαδρομές διάδοσης του ήχου.*

Στο παραπάνω σχήμα διακρίνεται η πηγή (**S**), ο δέκτης (**R**) καθώς και οι εικονικές πηγές (**S1**, **S2**, **S3**, **S4**). Οι δευτερεύουσες πηγές ορίζουν ουσιαστικά μια καθυστέρηση και μια εξασθένηση η οποία λαμβάνεται υπόψη για τον υπολογισμό της κρουστικής απόκρισης.

## 2.2. Perceptual Evaluation of Speech Quality (PESQ)

Αφου έγιναν αρκετές δοκιμές στους προσομοιωτές δωματίων και μικροφώνων και άρχισαν τα εξαγόμενα αρχεία να τροφοδοτούν τον κεντρικό κώδικα με τις μεθόδους αυτόματης μίξης, δημιουργήθηκε η ανάγκη για ένα τρόπο αξιολόγησης των αποτελεσμάτων, ένας τρόπος βαθμολόγησης ο οποίος θα βοηθούσε στην εξαγωγή συμπερασμάτων για την αποτελεσματικότητα της έρευνας. Το δεύτερο πολύ βασικό εργαλείο, που χρησιμοποιήθηκε και πάνω στο οποίο βασίστηκε σε μεγάλο βαθμό η εξέλιξη της διπλωματικής είναι η μέθοδος μέτρησης PESQ (Perceptual Evaluation of Speech Quality).

Η μέθοδος μέτρησης "Perceptual Evaluation of Speech Quality" (PESQ), η οποία δημοσιεύθηκε από τη Διεθνή Ένωση Τηλεπικοινωνιών το 2001, που ουσιαστικά εξυπηρετεί στην αξιολόγηση των σημάτων ομιλίας των κινητών τηλεφώνων και επιτρέπει τη μέτρηση τους χρησιμοποιώντας μεθόδους ψυχοακουστικής κωδικοποίησης υψηλής συμπίεσης (Rix et al 2001). Το PESQ χρησιμοποιεί έναν αλγόριθμο που επιτρέπει την αξιολόγηση αυτών των σημάτων συγκρίνοντάς τα με σήματα αναφοράς. Αναπτύχθηκε με τη χρήση μεγάλου αριθμού ηχογραφήσεων που περιείχαν προτάσεις από διάφορους ομιλητές σε διάφορες γλώσσες. Οι ηχογραφήσεις έγιναν με τη χρήση διαφόρων κωδικοποιητών ομιλίας με διαφορετικά επίπεδα



ποιότητας και με τυπικές διαταραχές μετάδοσης δικτύου. Σε μια σειρά δοκιμών ακρόασης, ένας επαρκής αριθμός ακροατών ταξινόμησε αυτά τα παραδείγματα σε μια κλίμακα ποιότητας ομιλίας που κυμαίνεται από το 1 (κακή) έως το 5 (εξαιρετική).

Ο στόχος στην ανάπτυξη του PESQ ήταν μια μέθοδος για τον προσδιορισμό μιας αντικειμενικής μέτρησης που σχετίζεται με τα αποτελέσματα των δοκιμών ακρόασης, με βάση τη σύγκριση του αρχικού, μη υποβαθμισμένου σήματος ομιλίας (το σήμα αναφοράς) με το υποβαθμισμένο σήμα (το μετρούμενο σήμα) (Rix et al 2001). Η χρήση της PESQ μετρικής στην παρούσα εργασία είναι προφανή. Ορίζουμε ένα καθαρό (χωρίς θόρυβο) ανηγικό ηχητικό σήμα φωνής ως το σήμα αναφοράς, και το αποτέλεσμα που παίρνουμε από τις διαφορετικές τεχνικές ηχητικής μίξης το ορίζουμε ως το μετρούμενο σήμα. Με αυτό τον τρόπο, η μετρική PESQ θα μας δώσει ένα αντικειμενικό κριτήριο για το βαθμό που υποβαθμίζεται η ποιότητα του πρωταρχικού σήματος λόγω του θορύβου και της αντήχησης που περνάει μέσα σε αυτό.

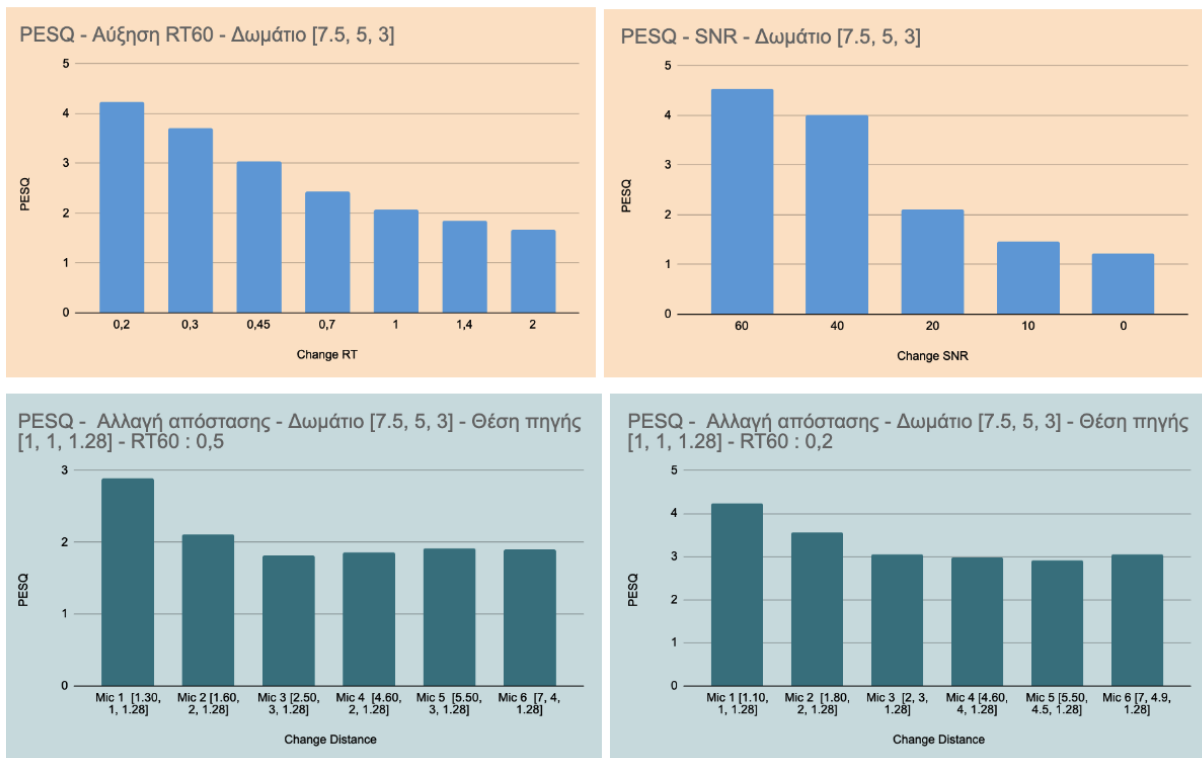
### **2.2.1. Αξιολόγηση εργαλείων Pyroomacoustics & PESQ**

Τα δύο εργαλεία, ο προσομοιωτής (Pyroomacoustics) και η αξιολόγηση σήματος φωνής (PESQ) αποτελούν καθοριστικής σημασίας εργαλεία καθώς μας επιτρέπουν να αξιολογήσουμε την αποτελεσματικότητα των μεθόδων που αναπτύσσονται. Για την διεξαγωγή λοιπόν σωστών συμπερασμάτων, κρίθηκε αναγκαία η διεξαγωγή βοηθητικών πειραμάτων που θα επέτρεπαν να επαληθευτεί ότι ο συνδυασμός των δύο αυτών εργαλείων οδηγεί σε λογικά αποτελέσματα.

Σχεδιάστηκαν λοιπόν κάποια πειράματα με σκοπό την επαλήθευση του PESQ και της ορθής λειτουργίας αυτού, σε διάφορες μεταβολές συνθηκών. Οι μεταβολές αυτές ήταν οι εξής:

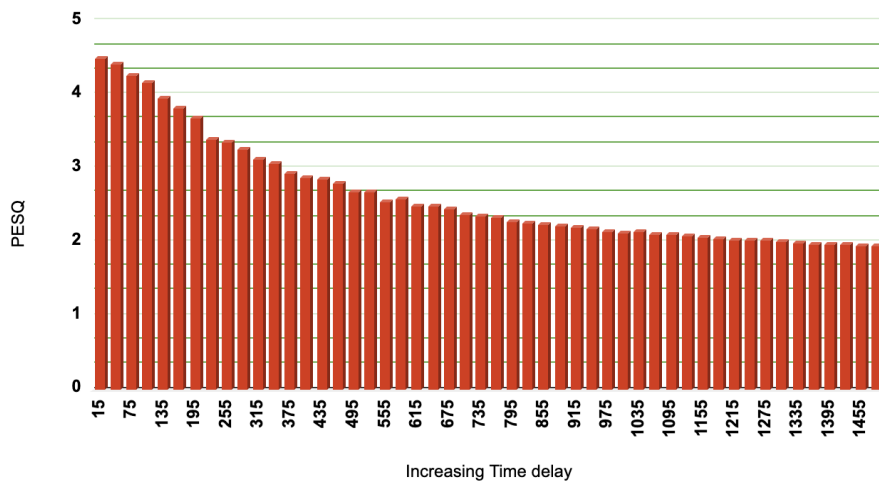
- Αύξηση του χρόνου αντήχησης (RT60) του δωματίου
- Αύξηση του λόγου σήματος - θορύβου (SNR)
- Σταδιακή αύξηση απόστασης πηγής - μικροφώνου
- Σταδιακή αύξηση καθυστέρησης (Delay)
- Διαφορετικά πολικά διαγράμματα μικροφώνων (θα γίνει αναφορά σε επόμενη ενότητα)

Στις εικόνες που ακολουθούν παρατίθενται αναλυτικά τα αποτελέσματα των πειραμάτων.

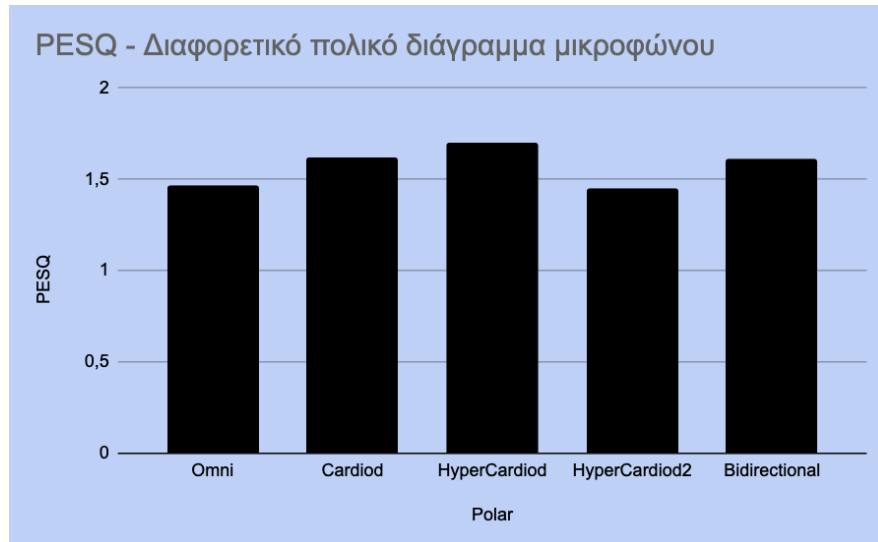


**Εικόνα 2.5:** Γράφημα που αποτυπώνει την μεταβολή του PESQ συναρτήσει του χρόνου αντήχησης RT60 (πάνω αριστερά), του SNR (πάνω δεξιά) και της απόστασης από την πηγή σε ένα δωμάτιο με RT60=0.5 s (κάτω αριστερά) και με RT60=0.2s (κάτω δεξιά).

### Increasing Time delay (PESQ)



**Εικόνα 2.6:** Γράφημα που αποτυπώνουν πως επιδρά η χρονική καθυστέρηση στη βαθμολογία PESQ



**Εικόνα 2.7:** Διάγραμμα που δείχνει τη διαφορετική βαθμολογία ανάλογα με το πολικό διάγραμμα του μικροφώνου

### 2.2.2. Συμπεράσματα

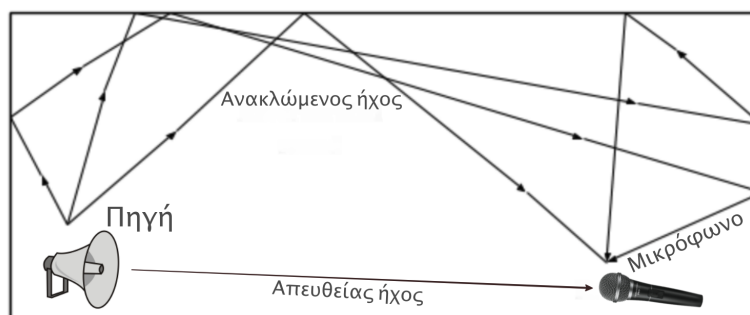
Απο τα παραπάνω πειράματα<sup>3</sup> (test) προκύπτουν τα εξής:

Η βαθμολογία δείχνει σταθερή πτώση όσο αυξάνεται ο χρόνος αντήχησης του δωματίου (Εικόνα 2.5 πάνω αριστερά).

Σε ένα κλειστό χώρο, όπως είναι ένα αμφιθέατρο ή μία αίθουσα θεάτρου, αντήχηση ονομάζεται η άθροιση όλων των ανακλώμενων ήχων από τις επιφάνειες του χώρου.

Επομένως στα μικρόφωνα εκτός απο τον απευθείας ήχο φτάνει και ο ανακλώμενος ήχος (Εικόνα 2.8)

(<http://hyperphysics.phy-astr.gsu.edu/hbase/acoustic/reverb.html>)



**Εικόνα 2.8:** Συλλογή όλων των ανακλώμενων ήχων, εκτός από τον απευθείας ήχο, σε ένα αμφιθέατρο.

<sup>3</sup> Τα αποτελέσματα αναλυτικά μπορείτε να τα δείτε Εδώ: [PESQ Tests](#)

Το αποτέλεσμα είναι ότι καθώς το μικρόφωνο δεν μπορεί να διαχωρήσει τον απευθείας ήχο από την αντήχηση, το τελικό ηχητικό σήμα εμφανίζεται σημαντικά αλλοιωμένο σε σχέση με τον “καθαρό” ήχο που εκπέμπει η πηγή. Επομένως αυτό δείχνει ότι οι δείκτες PESQ ανταποκρίνονται καλά.

Αντίστοιχα το επόμενο τεστ αφορά στην αξιολόγηση του PESQ σχετικά με διαφορετικές τιμές λόγου σήματος - θορύβου (SNR). Παρατηρείται λοιπόν ότι όσο μειώνεται η τιμή της μεταβλητής SNR μειώνεται αντίστοιχα αισθητά και η βαθμολογία του PESQ (Εικόνα 2.5 πάνω δεξιά).

Στη συνέχεια εξετάζεται το πώς επηρεάζει τη βαθμολογία η απομάκρυνση του μικροφώνου από την πηγή, η αλλαγή δηλαδή του λόγου απευθείας ήχου προς ανακλώμενο (Direct to Reverberant Ratio - DRR), όπου ενώ ο ανακλώμενος ήχος παραμένει σταθερός ο απευθείας ήχος ελαττώνεται όσο η πηγή απομακρύνεται. Για αυτό, σε ένα δωμάτιο με διαστάσεις [7.5, 5, 3] έγιναν δοκιμές σε πολλές διαφορετικές θέσεις μικροφώνων και σε διαφορετικούς χρόνους αντήχησης. Παρατηρείται λοιπόν και εδώ η πτώση της μετρικής όσο ο λόγος απευθείας σήματος προς αντήχηση ελαττώνεται (Εικόνα 2.5 κάτω).

Τέλος στην Εικόνα 2.7 αποτυπώνεται η βαθμολογία PESQ σε σχέση με τα διαφορετικά πολικά διαγράμματα των μικροφώνων. (Τι συνθήκες έχουμε εδώ? αντήχηση, θόρυβος?) Η χειρότερη τιμή PESQ προκύπτει για το παντοκατευθυντικό μικρόφωνο και αυτό οφείλεται καθαρά στο ότι δεν υπάρχει επιλεκτικότητα ως προς τη γωνία. Απο την άλλη, ότι το υπερκαρδιοειδές και το καρδιοειδές μικρόφωνο όταν είναι στραμμένα προς την πηγή ενδιαφέροντος βελτιώνουν σημαντικά την ποιότητα του σήματος, ενισχύοντας τον απευθείας ήχο συγκριτικά με τις ανακλώμενες συνιστώσες του και κατά συνέπεια οδηγούν σε αύξηση του PESQ.

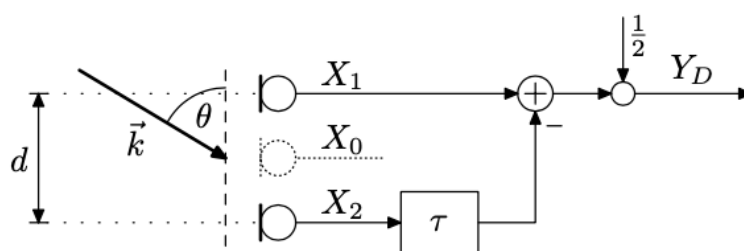
### **2.3. Διαφορικές συστοιχίες μικροφώνων**

Όπως φαίνεται παραπάνω στο κεφάλαιο με τα PESQ τεστ, ένα από αυτά τα τεστ αφορά στο πώς επηρεάζονται τα αποτελέσματα ανάλογα με το πολικό διάγραμμα των μικροφώνων που χρησιμοποιούνται. Σε πραγματικές συνθήκες σε ανάλογο στήσιμο με αυτό που εξετάζεται στα πειράματα που διενεργήθηκαν, συνήθως ένας ηχολήπτης θα επέλεγε είτε καρδιοειδή μικρόφωνα ή ίσως υπερκαρδιοειδή.

Στις προσομοιώσεις που πραγματοποιήθηκαν τα μικρόφωνα του Pyroom είναι πανκατευθυντικά (omni) επομένως, έπρεπε να ληφθεί υπόψιν και αυτός ο παράγοντας και να ελεγχθεί το κατά πόσο επηρεάζεται το αποτέλεσμα ανάλογα με την κατευθυντικότητα των μικροφώνων. Αυτό που προκύπτει σαν συμπέρασμα από τα τεστ είναι ότι επηρεάζεται οπότε, το πολικό διαγράμμα ήταν μια ακόμα παράμετρος που εντάχθηκε στα πειράματα. Για να επιτευχθεί η μετατροπή του πολικού διαγράμματος των πανκατευθυντικών μικροφώνων χρησιμοποιήθηκε η τεχνική των διαφορικών συστοιχιών μικροφώνων, όπου τα ηχητικά σήματα από δύο ή περισσότερα πανκατευθυντικά μικρόφωνα επιδέχονται επεξεργασία ώστε να τροποποιηθεί η γωνιακή απόκριση της όλης συστοιχίας.

Στην παρούσα διπλωματική εργασία, η κατευθυντική συμπεριφορά τροποποιείται από την εφαρμογή τεχνικών επεξεργασίας σήματος πάνω σε δυο omni μικροφώνων πολύ κοντά το ένα στο άλλο, έτσι ώστε ο άξονας μέγιστης απόκρισης της συστοιχίας να είναι στραμμένος προς την ηχητική σκηνή. Στην περίπτωση που εξετάζεται σε αυτά τα πειράματα τα μικρόφωνα αποτελούν μέρος της προσομοίωσης επομένως υπάρχει το πλεονέκτημα ότι οι αποκρίσεις των δύο μικροφώνων είναι πανομοιότυπες και ιδανικές.

Αξιοποιώντας τα ηχητικά σήματα από δύο omni μικρόφωνα τοποθετημένα σε μικρή απόσταση μεταξύ τους ( $d$ ) όπως φαίνεται στο Σχήμα 2.9, η κατευθυντική συμπεριφορά του διαφορικού μικροφώνου είναι δυνατόν να τροποποιείται ανάλογα με τη χρονική καθυστέρηση που εφαρμόζεται μεταξύ των αισθητήρων.



**Εικόνα 2.9:** Δομή μιας διαφορικής συστοιχίας

Η συστοιχία διεγείρεται από επίπεδα κύματα με κυματάρημο  $k$  ( $k=\omega/c$ ). Εξαιτίας της ακτινικής συμμετρίας, η χρονική διαφορά άφιξης  $\tau_A$  μεταξύ των δύο σημάτων  $X_1(\omega)$  και  $X_2(\omega)$  μπορούν να συσχετιστούν με τη γωνία πρόσπτωσης  $\theta$  μέσω της σχέσης  $kd = \omega\tau_A$ .

Έστω ότι στο κέντρο της συστοιχίας το σήμα  $X_0(\omega)$  λαμβάνεται από ένα εικονικό μικρόφωνο, το επίπεδο κύμα προσπίπτει με γωνία  $\theta$  και  $\kappa = \frac{2\pi}{\lambda}$  οπότε τα σήματα των μικροφώνων  $X_1(\omega)$  και  $X_2(\omega)$  μπορούν να εκφραστούν από τη σχέση,

$$X_1(\omega) = X_0(\omega) \cdot \exp\left\{j\frac{\kappa d}{2}\cos\theta\right\}, \quad (2.1)$$

$$X_2(\omega) = X_0(\omega) \cdot \exp\left\{-j\frac{\kappa d}{2}\cos\theta\right\}. \quad (2.2)$$

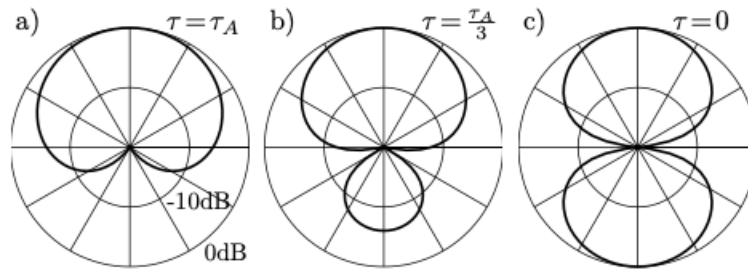
Τα σήματα των δύο μικροφώνων συνδυάζονται με τον παρακάτω τρόπο,

$$Y_D(\omega) = \frac{1}{2} (X_1(\omega) - X_2(\omega) \cdot \exp\{j\omega\tau\}). \quad (2.3)$$

και η γωνιακή απόκριση της συστοιχίας  $H_D$  δίνεται από τη Σχέση

$$H_D(\omega, \theta) = j \cdot \exp\left\{-j\frac{\omega\tau}{2}\right\} \cdot \sin\left(\frac{\kappa d}{2}\left(\frac{\tau}{\tau_A} + \cos\theta\right)\right) \quad (2.4)$$

όπου η κατεύθυνση της επιθυμητής ηχητικής πηγής είναι τυπικά στη γωνία  $\theta = 0$ .



**Εικόνα 2.10 :** Πολικά διαγράμματα α) Καρδιοειδές, β) Υπερκαρδιοειδές, γ) Figure of 8

Στη βιβλιογραφία συνήθως θεωρούνται πολύ μικρές αποστάσεις  $\kappa d \ll 1$  που δικαιολογούν την προσέγγιση  $\sin a \approx a$ . Σε αυτή την περίπτωση υπάρχει η ιδανική κατευθυντική απόκριση  $H_D$

$$H_D(\omega, \theta) \approx H_D(\omega, \theta) = j \cdot \frac{\kappa d}{2} \cdot \left(\frac{\tau}{\tau_A} + \cos\theta\right). \quad (2.5)$$

Σε αυτή την αναπαράσταση, τα βασικά χαρακτηριστικά των διαφορετικών συστοιχιών είναι:

- Το σχήμα του  $H_D(\theta)$  καθορίζεται από το  $\frac{\tau}{\tau_A} + \cos\theta$  το οποίο δεν εξαρτάται από τη συχνότητα. Γνωστά σχήματα όπως το figure of 8 ( $\frac{\tau}{\tau_A} = 0$ ), το καρδιοειδές ( $\frac{\tau}{\tau_A} = 1$ ) και το υπερκαρδιοειδές ( $\frac{\tau}{\tau_A} = \frac{1}{3}$ ) (Εικόνα 2.10).
- Λόγω της αφαίρεσης του σήματος (Εικόνα 2.9) προκύπτει μετατόπιση φάσης  $\frac{\pi}{2}$ .
- Η αφαίρεση που λαμβάνει χώρα στην Σχέση (2.3) έχει ως αποτέλεσμα την εξασθένηση των χαμηλών συχνοτήτων. Για αυτό το λόγο, για τη δημιουργία του τελικού σήματος η έξοδος της συστοιχίας  $Y(\omega)$  πολλαπλασιάζεται από ένα συντελεστή  $W_{eq}(\omega)$  ο οποίος εξαρτάται από τη συχνότητα και ορίζεται ως εξής:

$$W_{eq}(\omega) = \left\{ \frac{1}{\sin\left(\frac{\pi \cdot \omega}{2 \cdot \omega_c}\right)}, 0 < \omega < \omega_c \right\} \quad (2.6)$$

## 2.4. Εκτίμηση Χρονικής Διαφοράς Άφιξης (ΧΔΑ)

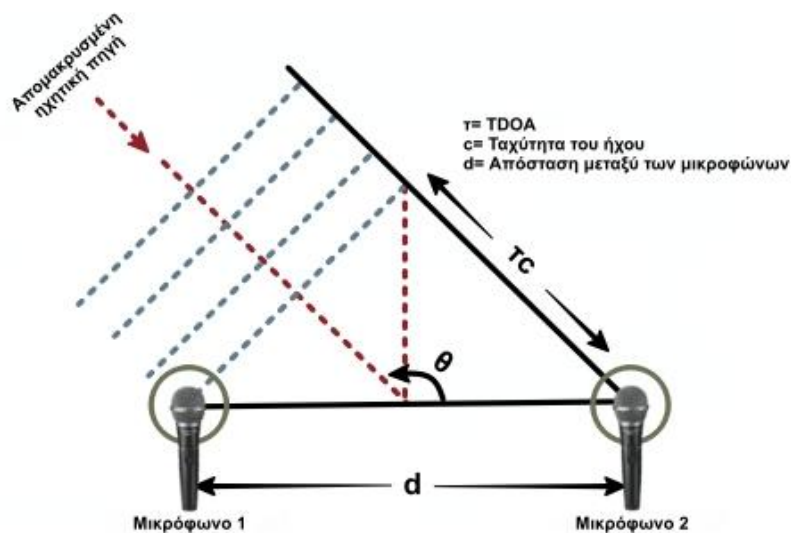
Η εκτίμηση του ΧΔΑ (Time Difference Of Arrival - TDOA) σε ένα σύστημα με πολλά μικρόφωνα είναι απαραίτητη προϋπόθεση για εφαρμογές που εστιάζουν στο πρόβλημα του εντοπισμού της θέσης της πηγής. Αυτό επιτυγχάνεται με την εκτίμηση της διαφοράς χρόνου άφιξης μεταξύ ζευγών αισθητήρων και κατά κύριο λόγο εφαρμόζεται σε μικρόφωνα που είναι σχετικά κοντά το ένα στο άλλο. (Blandin et al 2012). Ο εντοπισμός θέσης (Sound Source Localization - SSL), έχει αρκετές εφαρμογές σε πολλούς τομείς, αρκετοι απο τους οποίους αφορούν στην βελτίωση της ομιλίας, αλλά και στον εντοπισμό της θέσης μιας ή περισσότερων ακουστικών πηγών.

Η εκτίμηση των ΧΔΑ είναι επίσης απαραίτητη προϋπόθεση για την εφαρμογή τεχνικών σχηματισμού λοβού (beamforming) που με τη σειρά τους αποτελούν τη βάση για ένα πλήθος προσεγγίσεων που στοχεύουν στην βελτίωση της ποιότητας ηχητικού σήματος φωνής ή μουσικής.

Οι προσεγγίσεις με βάση την TDOA βασίζονται στην εκτίμηση της σχετικής διαφοράς χρόνου άφιξης μεταξύ ζευγαριών μικροφώνων και γίνονται με χρήση της πράξης της

ετεροσυσχέτισης (Cross Correlation - CC). Συνήθως για να βελτιωθεί η προσέγγιση της CC κάθε σήμα φιλτράρεται εκ των προτέρων και εφαρμόζεται μετά η CC και σε αυτή την περίπτωση έχουμε την Generalized Cross Correlation (GCC) (Hosseini et al 2017).

Η μέθοδο που εκτιμούν την κατεύθυνση των πηγών με χρήση της GCC είναι βασισμένες στην διάδοση της διαδρομής του επίπεδου κύματος (Εικόνα 2.11). Σε αυτό το μοντέλο τα σήματα που λαμβάνονται, ουσιαστικά αποτελούν την καθυστερημένη και εξασθενημένη εκδοχή του αρχικού σήματος, το οποίο προέρχεται από μια σημειακή πηγή η οποία έχει αλλοιωθεί από έναν λευκό γκαουσιανό θόρυβο, ο οποίος είναι ασυσχέτιστος με το σήμα της πηγής.



*Εικόνα 2.11: Χρονική καθυστέρηση ανάμεσα σε δύο μικρόφωνα απομακρυσμένης ηχητικής πηγής σε γωνία  $\theta$ .*

Η εκτίμηση της χρονικής καθυστέρησης με τη μέθοδο GCC μπορεί να υλοποιηθεί στο πεδίο της συχνότητας ως

$$R_{x_1 x_2}^g(\tau) = \int_{-\infty}^{\infty} X_1(\omega) X_2^*(\omega) \psi(\omega) e^{j2\pi\omega\tau} d\omega, \quad (2.7)$$

όπου  $X_1(\omega)$  και  $X_2(\omega)$  είναι οι συντελεστές Fourier των δύο μικροφωνικών σημάτων στη συχνότητα  $\omega$  και  $\psi(\omega)$  είναι ένας συχνοτικά μεταβαλλόμενος συντελεστής. Η χρονική καθυστέρηση προκύπτει ως η τιμή του  $\tau$  όπου η Σχέση (2.7) μεγιστοποιείται, δηλαδή



$$\tau' = \operatorname{argmax} \left\{ R_{x_1 x_2}^g(\tau) \right\}. \quad (2.8)$$

Το 1973, προτάθηκε η μέθοδος PHAT ή αλλιώς Cross-Power Spectrum Phase (CSP). Αποτελούσε μια λύση για την εκτίμηση της χρονικής καθυστέρησης, η οποία εξάγει πληροφορίες χρονικής καθυστέρησης από τη φάση του εγκάρσιου κύματος. Ουσιαστικά, ο πολλαπλασιασμός μεταξύ των μιγαδικών τιμών  $X_1(\omega)X_2^*(\omega)$ , θα επιστρέψει μιγαδική τιμή, η οποία εμπεριέχει τη διαφορά φάσης μεταξύ των δύο καναλιών αλλά και το μέτρο τους. Στη μέθοδο PHAT, η συνάρτηση  $\psi(\omega)$  παίρνει τη μορφή

$$\psi_g^{PHAT}(\omega) = \frac{1}{\sqrt{|X_1(\omega)X_2^*(\omega)|}} \quad (2.9)$$

που έχει ως αποτέλεσμα την εξουδετέρωση των διακυμάνσεων λόγω του πλάτους μεταξύ διαφορετικών συχνοτήτων και την αξιοποίηση μόνο της φάσης στον υπολογισμό. Με αυτό τον τρόπο, όλες οι συχνότητες ανάλυσης λαμβάνονται υπόψιν με την ίδια βαρύτητα στον υπολογισμό. Η μέθοδος PHAT παρουσιάζει καλά αποτελέσματα σε συνθήκες υψηλού SNR και σε συνθήκες μέτριας έως υψηλής αντήχησης (Hosseini et al 2017), και επομένως υιοθετήθηκε και για τον υπολογισμό του TDOA μεταξύ ζευγών μικροφώνων στην παρούσα διπλωματική.

## 3. Μέθοδοι

---

### 3.1. Εισαγωγή

Σε όλα τα παρακάτω υποκεφάλαια, όπου παρουσιάζονται οι βασικές μέθοδοι που χρησιμοποιήθηκαν για τις ανάγκες των πειραμάτων, πραγματοποιείται η μεταφορά από το πεδίο του χρόνου στο πεδίο του χρόνου συχνότητας (ΧΣ) με χρήση του short-time Fourier transform (STFT)

$$x[n] \rightarrow X(\tau, \omega), \quad (3.1)$$

όπου  $n$  ο δείκτης δείγματος στο πεδίο του χρόνου,  $\tau$  ο δείκτης χρονικού παραθύρου και  $\omega$  ο δείκτης συχνότητας.

### 3.2. Maximum Component Preservation (MCP)

Στην παρούσα διπλωματική εργασία προτείνεται και αξιολογείται για πρώτη φορά μια καινοτόμα προσέγγιση στην σύλληψη του ήχου που την ονομάσαμε Maximum Component Preservation (MCP). Η προτεινόμενη τεχνική εμπνέεται από την προσέγγιση του Dugan που αναλύθηκε σε προηγούμενη ενότητα, ωστόσο διαφέρει σημαντικά από αυτήν καθώς υλοποιείται στο πεδίο του χρόνου συχνότητας.

Η υλοποίηση της MCP είναι πολύ απλή και μαθηματικά μπορεί να διατυπωθεί ως

$$Y(\tau, \omega) = X_q(\tau, \omega) \quad (3.2)$$

όπου

$$q = \operatorname{argmax} \{ |X_1(\tau, \omega)|, |X_2(\tau, \omega)|, \dots, |X_M(\tau, \omega)| \} \quad (3.3)$$

είναι ο δείκτης του μικροφώνου με την μεγαλύτερη ενέργεια στο συγκεκριμένο σημείο ΧΣ και  $|X_m(\tau, \omega)|$  είναι το μέτρο του  $m$ -ιστού μικροφώνου στο σημείο  $(\tau, \omega)$ . Το σήμα εξόδου  $Y(\tau, \omega)$  προσδιορίζεται με βάση το πλάτος και τη φάση του μικροφώνου με τη μεγαλύτερη ενέργεια, ωστόσο για το ίδιο χρονικό πλαίσιο  $\tau$ , επιτρέπεται να υπάρχουν παραπάνω από ένα

μικρόφωνα που συμμετέχουν στη διαμόρφωση του τελικού σήματος. Η προσέγγιση είναι σε συμφωνία με τη διαίσθησή ότι

1) η επιλογή του καναλιού με τη μέγιστη ενέργεια θα αναδεικνύει τα συστατικά του απευθείας ήχου, τα οποία και λογικά θα εμφανίζονται ενισχυμένα στα μικρόφωνα που είναι κοντά στη πηγή

2) η επιλογή του καναλιού με τη μέγιστη ενέργεια θα αναδεικνύει ενδεχομένως και συστατικά από τις πρώτες ανακλάσεις οι οποίες γνωρίζουμε ότι συμβάλλουν θετικά στην καταληπτότητα του ήχου

3) εξασφαλίζοντας ότι για κάθε συχνότητα  $\omega$  υπάρχει συνεισφορά από ένα μόνο κανάλι, αποφεύγονται φαινόμενα φασματικού χρωματισμού τύπου κτένας και ακυρώσεις ήχου οι οποίες ενδεχομένως να επιδρουν αρνητικά στην ποιότητα του σήματος εξόδου.

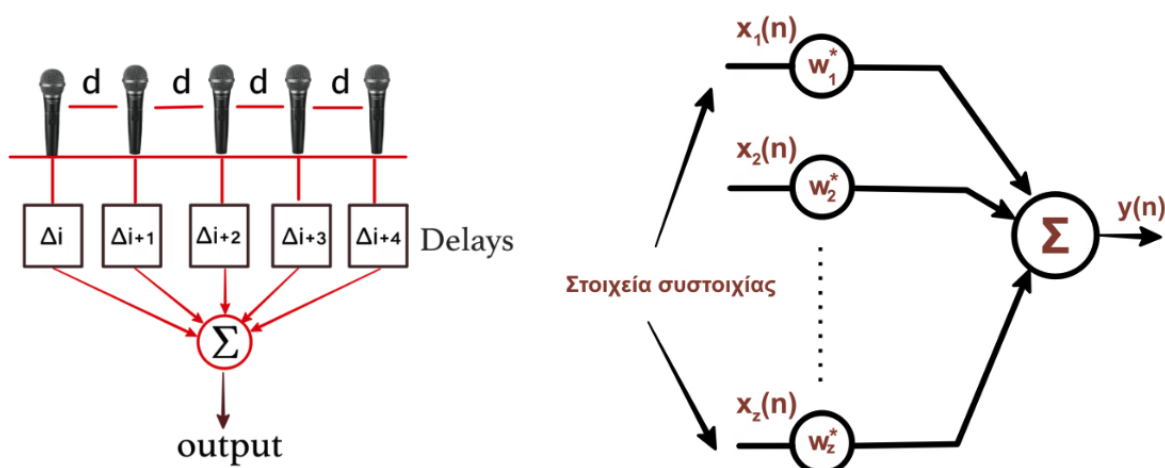
Η τεχνική αυτή αξιολογείται σε επόμενη ενότητα.

### 3.3. Delay & Sum

Τα συστήματα που έχουν σχεδιαστεί για τη λήψη σημάτων χωρικής διάδοσης συχνά αντιμετωπίζουν την παρουσία σημάτων παρεμβολής. Εάν το επιθυμητό σήμα και οι παρεμβολές καταλαμβάνουν την ίδια ζώνη συχνοτήτων, τότε το απλό φιλτράρισμα δεν μπορεί να χρησιμοποιηθεί για το διαχωρισμό του σήματος από την παρεμβολή. Ωστόσο, τα επιθυμητά σήματα και τα σήματα παρεμβολής συχνά προέρχονται από διαφορετικές χωρικές θέσεις. Αυτός ο χωρικός διαχωρισμός μπορεί να αξιοποιηθεί για το διαχωρισμό του σήματος από τις παρεμβολές με τη χρήση ενός χωρικού φίλτρου στο δέκτη (Veen & Buckley, 1999).

Ο διαμορφωτής δέσμης είναι ένα χωρικό φίλτρο που χρησιμοποιείται σε συστοιχίες αισθητήρων, για κατευθυνόμενη μετάδοση ή λήψη σήματος, προκειμένου να ενισχύσει το πλάτος ενός συνεκτικού μετώπου κύματος σε σχέση με το θόρυβο υποβάθρου και τις παρεμβολές. Ο στόχος της διαμόρφωσης δέσμης είναι η άθροιση πολλαπλών στοιχείων για την επίτευξη μιας στενότερης απόκρισης (στενοτερου λοβού) προς μια επιθυμητή κατεύθυνση, η βελτίωση σε σύγκριση με την πανκατευθυντική λήψη/μετάδοση είναι γνωστή ως κατευθυντικότητα της συστοιχίας.

Η μέθοδος Delay & Sum είναι από τις πιο απλές προσεγγίσεις στο σχηματισμό ακουστικού λοβού με συστοιχίες μικροφώνων (Εικόνα 3.1). Η αρχή λειτουργίας του Delay & Sum βασίζεται στην επιβολή κατάλληλων χρονικών καθυστερήσεων (ή φασικών ολισθήσεων) στο σήμα κάθε στοιχείου λήψης με σκοπό το σχηματισμό του επιθυμητού λοβού προς κάποια συγκεκριμένη κατεύθυνση. Έτσι για κάθε επιθυμητή κατεύθυνση του λοβού χρησιμοποιείται διαφορετική τιμή χρονικής καθυστέρησης. Η καθυστέρηση που χρησιμοποιείται για κάθε στοιχείο αισθητήρα καθορίζεται από τη γεωμετρία της συστοιχίας.



**Εικόνα 3.1:** (Αριστερά) Τυπική delay-and-sum διαμόρφωση δέσμης. (Δεξιά) Υλοποίηση του διαμορφωτή δέσμης στο πεδίο της συχνότητας, όπου τα σήματα  $x_m(n)$  είναι οι μιγαδικοί συντελεστές Fourier συγκεκριμένης συχνότητας και τα βάρη  $w_m$  είναι μιγαδικά.

Γενικά, ο διαμορφωτής δέσμης μπορεί να υλοποιηθεί και στο πεδίο της συχνότητας, όπως φαίνεται στην Εικόνα 3.1 (Δεξιά). Σε αυτήν την περίπτωση η έξοδος  $y(\tau, \omega)$  δίνεται από τον γραμμικό συνδυασμό στο πεδίο του χρόνου - συχνότητας των δεδομένων ως εξής:

$$y(\tau, \omega) = \sum_{l=1}^M w_l^*(\omega) x_l(\tau, \omega) \quad (3.4)$$

όπου  $M$  είναι το πλήθος των μικροφώνων,  $\omega$  η συχνότητα και το  $(*)$  υποδηλώνει μιγαδική συζυγία. Εάν το σήμα σε κάθε αισθητήρα θεωρηθεί ως είσοδος, τότε ένας διαμορφωτής δέσμης αντιπροσωπεύει ένα σύστημα πολλαπλών εισόδων και μιας εξόδου.

Για της ανάγκες τις παρούσας εργασίας, η μέθοδος delay and sum υλοποιείται στο πεδίο του χρόνου ως εξής:

$$y(n) = \sum_{l=1}^M x_l[n - \tau_l] \quad (3.5)$$

όπου  $\tau_l$  η χρονική καθυστέρηση η οποία προσδιορίζεται είτε μέσω της TDOA τεχνικής που αναφέρθηκε παραπάνω, ή μέσω της γεωμετρίας του εκάστοτε πειράματος.

### 3.4. Spotforming

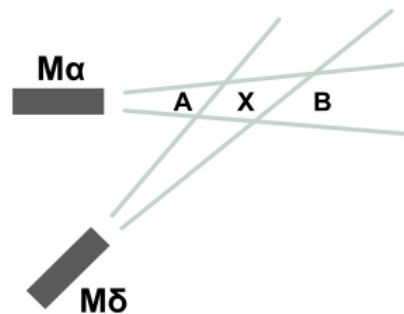
Οι μέθοδοι συστοιχίας μικροφώνων είναι μερικές από τις πιο αποτελεσματικές μεθόδους για την καταγραφή χωρίς θόρυβο. Ωστόσο, έχουν μειονεκτήματα όπως ότι, απαιτούνται πολλά μικρόφωνα, απαιτείται μεγάλη υπολογιστική ισχύς και αρκετές από τις μεθόδους είναι ευάλωτες σε αλλαγές στο ακουστικό περιβάλλον (Suzuki & Honjo, 2015).

Ένας άλλος τρόπος για την για την καταπίεση (ή καταπολέμηση) του θορύβου χρησιμοποιεί κατευθυντικά μικρόφωνα. Ειδικά, ένα μικρόφωνο τύπου shotgun είναι εξαιρετικά κατευθυντικό και εύκολο στη χρήση. Βέβαια να σημειωθεί εδώ ότι σε σύγκριση με μια συστοιχία μικροφώνων όπου μπορούν να εφαρμοστούν τεχνικές όπως η TDOA ή DOA, με μικρόφωνο τύπου shotgun είναι δύσκολο να εντοπιστεί η κατεύθυνση του χρήστη. Επίσης τα shotgun -εξαιτίας του πολικού τους διαγράμματος- δεν μπορούν να καταστείλουν έναν θόρυβο που βρίσκεται πίσω από τον ομιλητή για παράδειγμα, εάν μια τηλεόραση βρίσκεται πίσω από τον ομιλητή, τότε το σήμα θα έχει θόρυβο, πράγμα που προκαλεί επιδείνωση της απόδοσης.

Για την επίλυση αυτού του προβλήματος, προτείνεται μια μέθοδος "σχηματισμού σημείου" με τη χρήση δύο μικροφώνων τύπου shotgun. Με τη μέθοδο αυτή η δέσμες εστιάζουν σε μια περιοχή ενδιαφέροντος. Τυχόν θόρυβοι πίσω από το στοχευόμενο ηχείο μπορούν να κατασταλούν με την προσαρμογή της θέσης του "σημείου" στο στόμα του ομιλητή. Πρόκειται για τη μέθοδος spotforming η οποία παρόλο που δεν χρησιμοποιείται στην παρούσα εργασία αξίζει να αναφερθεί.

Η βασική ιδέα αυτής της μεθόδου είναι πολύ απλή. Ένα μικρόφωνο τύπου shotgun έχει μια μεγάλη και στενή δέσμη κατευθυντικότητας. Δύο μικρόφωνα τύπου shotgun Ma και Mb τοποθετούνται παράλληλα και στη συνέχεια περιστρέφεται το ένα, ώστε οι δέσμες κατευθυντικότητάς τους να τέμνονται (Εικόνα 3.3). Όταν μια ηχητική πηγή βρίσκεται στο

σημείο διασταύρωσης (X) των δύο ακτίνων κατευθυντικότητας, τότε και τα δύο μικρόφωνα θα την καταγράψουν. Από την άλλη πλευρά, όταν μια ηχητική πηγή βρίσκεται σε οποιαδήποτε άλλη περιοχή (A,B), τότε τουλάχιστον ένα μικρόφωνο δεν θα την καταγράψει. Επομένως, το ηχητικό σήμα που βρίσκεται στο σημείο (X) μπορεί να ενισχυθεί με την πρόσθεση των δύο καταγεγραμμένων σημάτων, ενώ τα ηχητικά σήματα που βρίσκονται σε οποιαδήποτε άλλα σημεία δε θα ενισχυθούν. Φυσικά, τα δύο καταγεγραμμένα σήματα πρέπει να προστίθενται με κατάλληλο χρόνο καθυστέρησης, επειδή οι αποστάσεις μεταξύ της ηχητικής πηγής και των μικροφώνων είναι διαφορετικές (delay and sum). Να σημειωθεί ότι ένα από τα πιο σημαντικά ζητήματα είναι η σωστή ρύθμιση των δύο μικροφώνων. (Suzuki & Honjo, 2015).



*Εικόνα 3.2: Spot-forming*

### 3.5. Magnitude Squared Coherence (MSC)

Πρόκειται για την αξιοποίηση μιας στατιστικής μέτρησης των σημάτων η οποία προκύπτει από την σύγκριση δύο καναλιών ήχου δηλαδή στην περίπτωση που εξετάζεται εδώ, δύο μικροφωνικών σημάτων. Η χρήση της MSC τεχνικής έχει προταθεί για την αποδόμηση του ήχου στις συνιστώσες του απευθείας και του διάχυτου ήχου (Stefanakis 2015). Λόγω της σημασίας που έχουν αυτές οι δύο συνιστώσες στην χωρική αντίληψη του ήχου, η MSC τεχνική παίζει σημαντικό ρόλο σε κάποιες τεχνικές αναπαραγωγής χωρικού ήχου (Thiergart 2011).

Η υλοποίηση της τεχνικής στο πεδίου του ΧΣ για δύο κανάλια ήχου με δείκτες  $n$  και  $m$  μπορεί να γραφεί ως εξής:

$$MSC_{nm}(\tau, \omega) = \sqrt{\frac{|E\{X_n(\tau, \omega)X_m^*(\tau, \omega)\}|^2}{E\{|X_n(\tau, \omega)|^2\}E\{|X_m(\tau, \omega)|^2\}}}. \quad (3.6)$$

όπου το  $E\{\}$  συμβολίζει προσδοκώμενη τιμή. Από τη Σχέση (3.6) θα προκύψει ένας πραγματικός αριθμός όπου η μεγαλύτερη τιμή που μπορεί να πάρει είναι το 1 και η μικρότερη το 0 ( $0 < MSC < 1$ ).

Πρακτικά, φασικές εξαρτήσεις ανάμεσα στα κανάλια  $n$  και  $m$  θα οδηγούν σε υψηλές τιμές του MSC, ενώ για σήματα που είναι στατιστικά ασυσχέτιστα μεταξύ τους, το αποτέλεσμα θα τείνει στο 0. Αυτό είναι πολύ σημαντικό διότι, σε ένα σημείο του ΧΣ όπου επικρατεί ο απευθείας ήχος αναμένονται ισχυρές φασικές εξαρτήσεις, ενώ σε σημεία ΧΣ όπου επικρατεί ο θόρυβος περιβάλλοντος ή η αντήχηση, οι διαφορές φάσης είναι θεωρητικά τυχαίες επομένως το MSC θα δίνει μικρές τιμές. Επομένως αυτή η μέτρηση σαν εργαλείο είναι πολύ χρήσιμη διότι μπορεί να αναδείξει τον απευθείας ήχο χρησιμοποιώντας μόνο την συνεισφορά από τα ζεύγη καναλιών που εμφανίζουν ισχυρές συσχετίσεις.

Επειδή ακριβώς η μέθοδος λειτουργεί με τη σύγκριση δύο καναλιών, όσα μικρόφωνα και να χρησιμοποιούνται στο εκάστοτε πείραμα, θα πρέπει να γίνει η επιλογή ζευγαριών με κάποιο συγκεκριμένο τρόπο. Στα πειράματα που πραγματοποιήθηκαν για την εργασία επιλέχθηκε τα ζεύγη να διαμορφώνονται με βάση τη γειτνίαση, δηλαδή κάθε ένα μικρόφωνο συμμετέχει στον υπολογισμό της MSC μέσα από δύο ζεύγη, ένα με το γειτονικό του αριστερόστροφα και ένα με το γειτονικό του δεξιόστροφα. Με τον τρόπο αυτό διαμορφώνονται  $M$  μόνο ζευγάρια (όσο και το πλήθος των μικροφώνων) και αποφεύγεται ο υπολογισμός της MSC για όλα τα πιθανά ζεύγη μικροφώνων (δηλ.  $M(M-1)$  ζεύγη), κάτι που θα αύξανε την πολυπλοκότητα της διεργασίας. Αν τώρα συμβολίσουμε ως  $MSC_m(\tau, \omega)$ ,  $m=1,2,\dots,M$  τις τιμές της MSC για το  $m$ -ιοστό ζεύγος και το σημείο ΧΣ  $(\tau, \omega)$ , το σήμα στην έξοδο υπολογίζεται με βάση την παρακάτω σχέση:

$$Y(\tau, \omega) = MSC_q \cdot X_q(\tau, \omega) \quad (3.7)$$

όπου

$$q = \text{argmax}\{MSC_1(\tau, \omega), MSC_2(\tau, \omega), \dots, MSC_M(\tau, \omega)\} \quad (3.8)$$

είναι ο δείκτης του ζεύγους με τη μεγαλύτερη τιμή MSC για το συγκεκριμένο  $(\tau, \omega)$  σημείο του ΧΣ.

Ο υπολογισμός των προσδοκώμενων τιμών στη Σχέση (3.5) γίνεται μέσω των σχέσεων:

$$\begin{aligned} E \left\{ \left| X_n(\tau, \omega) \right|^2 \right\} &= \Psi_{nn}(\tau, \omega) \\ E \left\{ X_n(\tau, \omega) X_m^*(\tau, \omega) \right\} &= \Psi_{nm}(\tau, \omega) \end{aligned} \quad (3.9)$$

όπου

$$\mathbf{\Psi}(\tau, \omega) = E \{ \mathbf{X}(\tau, \omega) \mathbf{X}^H(\tau, \omega) \} \quad (3.10)$$

είναι ο  $M \times M$  μιγαδικός πίνακας συνδιακύμανσης του σήματος και

$$\mathbf{X}(\tau, \omega) = \left[ X_1(\tau, \omega), X_2(\tau, \omega), \dots, X_M(\tau, \omega) \right]^T \quad (3.11)$$

είναι το διάνυσμα με όλα τα σήματα μικροφώνων για το σημείο  $(\tau, \omega)$  του ΧΣ.

Στην πράξη, ο υπολογισμός της προσδοκώμενης τιμής για το  $\mathbf{\Psi}$  στη Σχέση (3.10) γίνεται λαμβάνοντας υπόψιν την τελευταία τιμή αλλά και παρελθοντικές τιμές και υπολογίζεται μέσα από την αναδρομική σχέση:

$$\mathbf{\Psi}(\tau, \omega) = (1 - \alpha) \mathbf{\Psi}(\tau, \omega) + \alpha \mathbf{X}(\tau, \omega) \mathbf{X}^H(\tau, \omega) \quad (3.12)$$

όπου  $\alpha$  είναι ένας πραγματικός συντελεστής ( $0 < \alpha < 1$ ).



---

**Μέρος Β΄**  
**Πειραματικό Μέρος**

---

## 4. Υλοποίηση

---

### 4.1. Περιγραφή πειραματικής διαδικασίας

Όπως ήδη έχει αναφερθεί στο Μέρος Α', η πειραματική διαδικασία πραγματοποιήθηκε σε συνθήκες προσομοίωσης. Στην προσομοίωση αυτή επιλέχθηκαν συνθήκες οι οποίες θα μπορούσαν να πλησιάσουν τις συνθήκες μιας θεατρικής παράστασης σε ένα κλειστό θέατρο. Οι διαστάσεις του δωματίου και τα χαρακτηριστικά του είναι κάποιες από αυτές. Εκτός όμως από το δωμάτιο σε μια θεατρική παράσταση υπάρχουν και άλλες ιδιαιτερότητες όπως το γεγονός ότι, οι ηθοποιοί διαρκώς κινούνται πάνω στη σκηνή και αυτό έχει σαν αποτέλεσμα να αλλάζουν οι εντάσεις και η σχέση απόστασης από το κάθε μικρόφωνο της εκάστοτε συστοιχίας. Αυτός ήταν ένας λόγος που επιλέχθηκε να εξεταστούν διαφορετικές θέσεις πηγής, με σκοπό την εξαγωγή όσο το δυνατόν πιο ρεαλιστικών συμπερασμάτων. Σε αυτή τη λογική επιλέχθηκαν 11 διαφορετικές θέσεις πάνω στη σκηνή (Εικόνα 4.1) με αποτέλεσμα να εξεταστεί το πώς ανταποκρίνονται οι μέθοδοι σε κάθε μια από αυτές. Ακόμα, επειδή ο κάθε ηθοποιός έχει άλλη ένταση και ηχοχρωμα στη φωνή του, επιλέχθηκαν κάποιες ηχογραφήσεις όπου εννεα διαφορετικοί άνθρωποι, διαφορετικού φύλου (πέντε γυναίκες, τέσσερις άντρες), διαβάζουν ένα κείμενο. Αυτές οι ηχογραφήσεις συγχωνεύτηκαν σε ένα αρχείο στο οποίο επιλέχθηκε μια πρόταση από τον κάθε άνθρωπο, επομένως το συγκεκριμένο ηχητικό αρχείο αποτέλεσε το ηχητικό σήμα που εισήχθη στον αλγόριθμό και αντιμετωπίστηκε σαν ηχητική πηγή. Με αυτό τον τρόπο αξιολογώντας έναν αριθμό διαφορετικών φωνών και ηχοχρωμάτων.

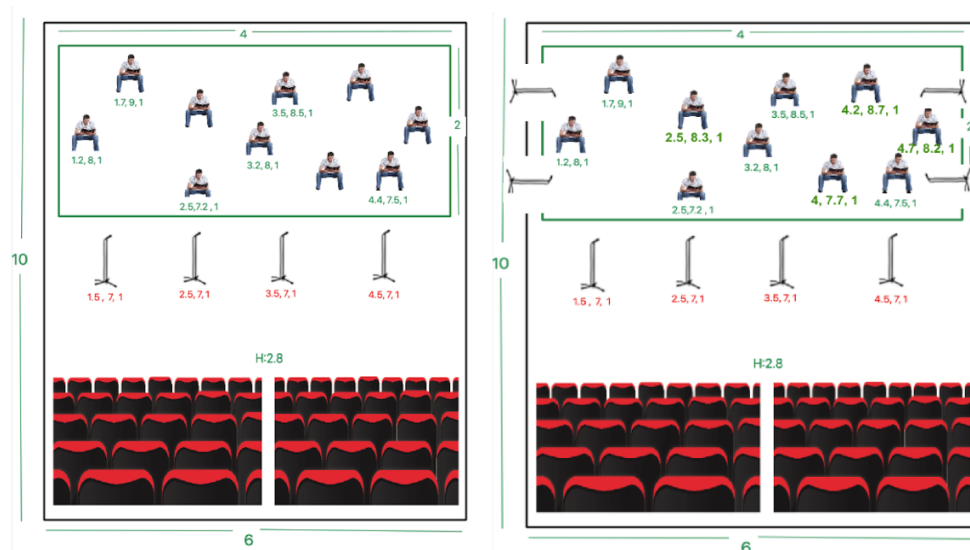
#### 1.1.1. Μεταβλητές

Σε αυτή την προσομοίωση πραγματοποιούνται σενάρια στα οποία μετά από αρκετά πειράματα επικεντρώθηκε η έρευνα με σκοπό την εξαγωγή συμπερασμάτων, με σκοπό να καλύπτουν ένα αρκετά μεγάλο εύρος συνθηκών. Ιδανικά αναμενόταν ένα καθολικό συμπέρασμα για να αναδειχθεί κάποια μέθοδος ή έστω συνδυασμός μεθόδων που να

λειτουργεί σε όλες τις περιπτώσεις, όμως αποδείχθηκε ότι αυτό είναι κάτι δύσκολο. Όλα τα αποτελέσματα που εξήχθησαν αποτελούν στην ουσία έναν συνδυασμό τεσσάρων μεθόδων, τα οποία προκύπτουν συναρτήσει μιας σειράς παραμέτρων που μεταβάλλονται.

Αφού ολοκληρώθηκαν οι δοκιμές των βασικών εργαλείων με σκοπό να εξασφαλιστεί η αξιοπιστία τους (Κεφ. 2.2.1) έγινε η επιλογή των συνθηκών στις οποίες θα πραγματοποιούνταν οι προσομοιώσεις. Οι βασικές και σταθερές μεταβλητές των πειραμάτων είναι το δωμάτιο και οι θέσεις των μικροφώνων. Πραγματοποιήθηκαν δύο ξεχωριστά πειράματα, το ένα αφορούσε σε συστοιχία τεσσάρων μικροφώνων και το δεύτερο σε συστοιχία οκτώ μικροφώνων (Εικόνα 4.1). Στη συνέχεια, επειδή η παρούσα εργασία καταπιάνεται με την αυτόματη μίξη σε σενάριο θεατρικής παράστασης, όπου η πηγή δεν είναι σταθερή -καθώς ο ηθοποιός μετακινείται πάνω στη σκηνή μεταβάλλοντας διαρκώς τη θέση του- επιλέχθηκαν κάποιες θέσεις πάνω στη σκηνή όπως φαίνεται επίσης στην Εικόνα 4.1, στην προσπάθεια να καλυφθούν όσο το δυνατόν περισσότερα πιθανά σημεία και περιπτώσεις. Η επιλογή της εκάστοτε θέσης έχει να κάνει με την απόσταση από ένα μικρόφωνο ή ένα ζευγάρι μικροφώνων, για αυτό επιλέχθηκαν απομακρυσμένα σημεία, αρκετά κοντινά, αλλά και σημεία ανάμεσα σε ζεύγη μικροφώνων.

Οι διαστάσεις της αίθουσας είναι 10 x 6 m και υψος  $H=2.8$  m ενώ τα μικρόφωνα τοποθετήθηκαν με 1 m απόσταση μεταξύ τους καθώς και 1 m απόσταση από το έδαφος.



**Εικόνα 4.1:** Απεικόνιση δωματίου που χρησιμοποιήθηκε κατά την πειραματική διαδικασία με τέσσερα μικρόφωνα (αριστερά) και με οκτώ μικρόφωνα (δεξιά), ενώ στις εικόνες φαίνονται και οι περισσότερες από τις θέσεις πηγής που μελετήθηκαν.

Οι θέσεις της πηγής αναλυτικά βρίσκονται στις εξής θέσεις: [1.2 , 8, 1], [1.7 , 9, 1], [2.5, 7.2, 1], [2.5 , 8.3 , 1], [3 , 9 , 1], [3.2 , 8 , 1], [3.5 , 8.5 , 1], [4 , 7.7 , 1], [4.2 , 8.7 , 1], [4.4 , 7.5 , 1], [4.7 , 8.2 , 1].

Εκτός από τη θέση της πηγής μια άλλη μεταβαλλόμενη παράμετρος είναι ο χρόνος αντήχησης (Reverberation Time -  $RT_{60}$ ). Οι τιμές  $RT_{60}$  που εξετάστηκαν είναι η τιμές  $RT_{60} = 0.3, 0.6$  και  $1.1s$ . Ο λόγος σήματος προς θόρυβο (Signal to noise Ratio - SNR) είναι μια ακόμη μεταβαλλόμενη παράμετρος με τιμές  $SNR = 10, 20$  και  $60$  dB.

Τα παραπάνω αποτελούν τις βασικές παραμέτρους σε σχέση με το στήσιμο του πειράματος δηλαδή ο αλγόριθμος στήθηκε ώστε να τρέχει τη διαδικασία για ένα δωμάτιο μια συστοιχία μικροφώνων (τέσσερα ή οκτώ), μια τιμή  $RT_{60}$  και μια τιμή SNR. Ο αλγόριθμος λοιπόν χρειάστηκε να τρέξει συνολικά 18 πειράματα <sup>4</sup> και με αυτή τη λογική παρουσιάζονται και τα αποτελέσματα αυτών (Εικόνα 4.2).

PESQ	difMicOn:True synchronizeON:True useGroundTruthTDOA:True		difMicOn:True synchronizeON:False useGroundTruthTDOA:False		difMicOn:False synchronizeON:True useGroundTruthTDOA:False		difMicOn:False synchronizeON:False useGroundTruthTDOA:False	
Source [1.2 , 8, 1]								
Automix		1,908	1,785	1,886	1,712	1,723	1,626	
Mix		2,071	1,739	1,976	1,766	1,863	1,549	
Refmic		2,062	2,062	2,062	1,742	1,742	1,742	
MSC		1,875	1,772	1,878	1,750	1,742	1,638	
Source		4,549	4,549	4,549	4,549	4,549	4,549	
TheClosestMic		2,062	2,062	2,062	1,742	1,742	1,742	
Source [1.7 , 9, 1]								
Automix		1,777	1,688	1,717	1,582	1,609	1,566	
Mix		1,952	1,752	1,865	1,663	1,709	1,606	
Refmic		1,783	1,783	1,783	1,597	1,596	1,603	
MSC		1,894	1,804	1,832	1,685	1,695	1,678	
Source		4,549	4,549	4,549	4,549	4,549	4,549	
TheClosestMic		1,783	1,783	1,783	1,597	1,596	1,603	
Source [2.5 , 7.2 , 1]								
Automix		3,058	3,014	3,056	2,657	2,657	2,594	
Mix		2,943	2,743	2,939	2,612	2,612	2,339	
Refmic		1,790	1,790	1,790	1,791	1,791	1,791	
MSC		1,900	1,777	1,897	2,134	2,134	1,938	
Source		4,549	4,549	4,549	4,549	4,549	4,549	
TheClosestMic		3,426	3,426	3,426	3,021	3,021	3,021	

**Εικόνα 4.2:** Μέρος του πίνακα αποτελεσμάτων που παράχθηκε στο excel, για τέσσερα μικρόφωνα,  $RT_{60} = 60s$  και  $SNR = 60$  dB (Οι πίνακες excel θα παραδοθούν με την εργασία)

Σε κάθε ξεχωριστό πείραμα, εφόσον οριστεί χρόνος αντήχησης και SNR εξετάζονται όλες οι μέθοδοι που έχουν επιλεγεί καθώς και κάποιες ακόμα παράμετροι οι οποίες επηρεάζουν την βαθμολογία PESQ, με αποτέλεσμα την εξαγωγή συγκεκριμένων συμπερασμάτων, με βάση συγκεκριμένες συνθήκες. Αυτό έχει σαν αποτέλεσμα μια ασφαλή εξαγωγή συμπερασμάτων όπως, σε ποιες συνθήκες επικρατεί ποιά μέθοδος και που δεν έχει τόσο καλή απόδοση, είτε

<sup>4</sup> 4mic , 8mic

ποια απο τις μεθόδους παρουσιάζει μεγαλύτερη σταθερότητα ανεξάρτητα από τις εκάστοτε συνθήκες.

Οι παράμετροι που μεταβάλλονται σε δεύτερο επίπεδο στον κώδικα είναι η θέση της πηγής η κατευθυντική συμπεριφορά των μικροφώνων και ο συγχρονισμός ή μη των μικροφώνων. Με τον όρο συγχρονισμό, εννοείται χρονική μετατόπιση των σημάτων των μικροφώνων ώστε ο απευθείας ήχος από την πηγή να φτάνει ταυτόχρονα σε όλα τα μικρόφωνα. Ο συγχρονισμός ουσιαστικά υλοποιείται με τον υπολογισμό των ΧΔΑ, οι οποίες υπολογίζονται είτε με βάση τη γεωμετρία (πραγματικές ΧΔΑ ή groundtruth TDOAs) - είτε με βάση την gcc-phat τεχνική (εκτιμητές ΧΔΑ ή estimated TDOAs). Για παράδειγμα στα αποτελέσματα τυχαίνει να υπάρχει περίπτωση κατά την οποία με ενεργοποιημένο το συγχρονισμό υπολογίζοντας τις αποστάσεις μέσω ground truth και χρησιμοποιούνται κατευθυντικά μικρόφωνα, η Delay & Sum να δίνει καλύτερα αποτελέσματα, στην περίπτωση όμως που δεν ισχύει κάτι από τα παραπάνω καλύτερα αποτελέσματα επιστρέφει η μέθοδος MSC.

Επομένως με αυτό το τρόπο η παρούσα εργασία μπορεί να προτείνει μια μέθοδο για την μίξη σαν γενικό συμπέρασμα μέσα από έναν συγκεκριμένο αριθμό πειραμάτων, επίσης μπορεί να προτείνει παραπάνω μεθόδους ανάλογα κάθε φορά τις συνθήκες που επικρατούν.

### **1.1.2. Αποτελέσματα**

Όλα τα αποτελέσματα από τα πειράματα μαζί με τον αλγόριθμο, επισυνάπτονται με την παρούσα εργασία ωστόσο στο συγκεκριμένο κεφάλαιο θα παρουσιαστούν πίνακες με κάποια συγκεκριμένα παραδείγματα.

Στον Πίνακα 1 φαίνονται αναλυτικά όλες οι τεχνικές οι μέθοδοι και οι συνδυασμοι αυτών, που μελετήθηκαν και αξιολογήθηκαν στις πειραματικές διαδικασίες με σκοπό την εξαγωγή συμπερασμάτων. Ειδικότερα οι συνδυασμοί παραμέτρων όπως φαίνεται και στον Πίνακα 1 αλλά και στα αποτελέσματα που παρουσιάζονται πιο κάτω, μπορεί να είναι καθοριστικής σημασίας.

Όλες οι μέθοδοι αλλά και μεταβλητές εξετάζονται σε δύο περιπτώσεις, με κατευθυντικά και παντοκατευθυντικά μικρόφωνα. Όπως φάνηκε και στα τεστ (κεφ 2.1.1) το αποτέλεσμα επηρεάζεται αισθητά απο το αν τα μικρόφωνα είναι η όχι κατευθυντικά, έχει λοιπόν ενδιαφέρον να εξεταστεί το κατα πόσο αυτη η διαφορά μπορεί να επηρεάσει την εκάστοτε

μέθοδο αλλά και πόσο επηρεάζεται από το παραμέτρους όπως ο συγχρονισμός. Σε ένα κατευθυντικό μικρόφωνο σε αντίθεση με ένα omnidirectional, υπάρχει η δυνατότητα να επιλεγεί ο προσανατολισμός του κάθε μικροφώνου με αποτέλεσμα να μεγιστοποιείται η πιθανότητα η ηχητική πηγή να είναι πάνω στον άξονα του μικροφώνου. Για παράδειγμα, στα πειράματα που χρησιμοποιούνται οκτώ μικρόφωνα (Εικόνα 4.3), όσα βρίσκονται στη Νότια πλευρά της σκηνής είναι προσανατολισμένα ώστε ο άξονας τους να κοιτάζει το Βορρά, ενώ όλα τα μικρόφωνα στη Δυτική πλευρά της σκηνής κοιτάζουν προς Ανατολή. Εκτός από την δυνατότητα όμως καλύτερου εντοπισμού της πηγής με ένα κατευθυντικό μικρόφωνο σημαντικό ρόλο παίζει και το γεγονός ότι, ένα κατευθυντικό μικρόφωνο έχει συγκεκριμένη γωνιακή απόκριση, με αποτέλεσμα να είναι λιγότερο εκτεθειμένο σε ανεπιθύμητους ήχους του περιβάλλοντος που βρίσκονται στην off-axis περιοχή του.

Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	✓	✓	✓	✓	✓	✓
MSC	✓	✓	✓	✓	✓	✓
Delay & Sum			✓	✓	✓	✓
MIX	✓	✓				
Κοντινότερο μικρόφωνο	✓	✓				

*Πίνακας 1: Τεχνικές που εξετάστηκαν κατά τη διεξαγωγή των πειραμάτων*

Σε σχέση με το συγχρονισμό συμβαίνει το εξής: Η περίπτωση “χωρίς συγχρονισμό” υπονοεί ουσιαστικά την αξιοποίηση των μικροφωνικών σημάτων απευθείας χωρίς κάποια διαδικασία χρονικής μετατόπισης, όπως δηλαδή αυτά τα σήματα εισέρχονται στην κονσόλα μίξης. Τα ηχητικά σήματα σε αυτήν την περίπτωση να μην είναι συγχρονισμένα με βάση το ρολόι της συσκευής, δε λαμβάνεται όμως υπόψιν η ΧΔΑ. Επομένως εδώ δεν υπάρχει μέριμνα ώστε να αποφευχθούν προβλήματα λόγω διαφοράς φάσης, μια συνθήκη η οποία είναι πολύ κοντα στην πραγματικότητα καθώς είναι πολύ δύσκολο να πραγματοποιηθεί συγχρονισμός μικροφώνου σε συνθήκες που η πηγή κινείται διαρκώς. Ωστόσο, εξετάζεται επίσης η περίπτωση, τα κανάλια εισόδου να συγχρονίζονται και με βάση τα groundtruth ή τα εκτιμητέα TDOAs, έτσι ώστε όλα τα μικροφωνικά σήματα να βρίσκονται σε συμφωνία φάσης. Τα

groundtruth TDOAs υπολογίζονται με βάση την γεωμετρία της διάταξης θεωρώντας γνωστές τις θέσεις όλων των μικροφώνων αλλά και της ηχητική πηγής. Από την άλλη τα εκτιμητέα TDOAs υπολογίζονται βάσει της σχετικής διαφοράς χρόνου άφιξης μεταξύ ζευγαριών μικροφώνων με τη χρήση της χρήσης της πράξης της ετεροσυσχέτισης (Κεφ. 2.4).

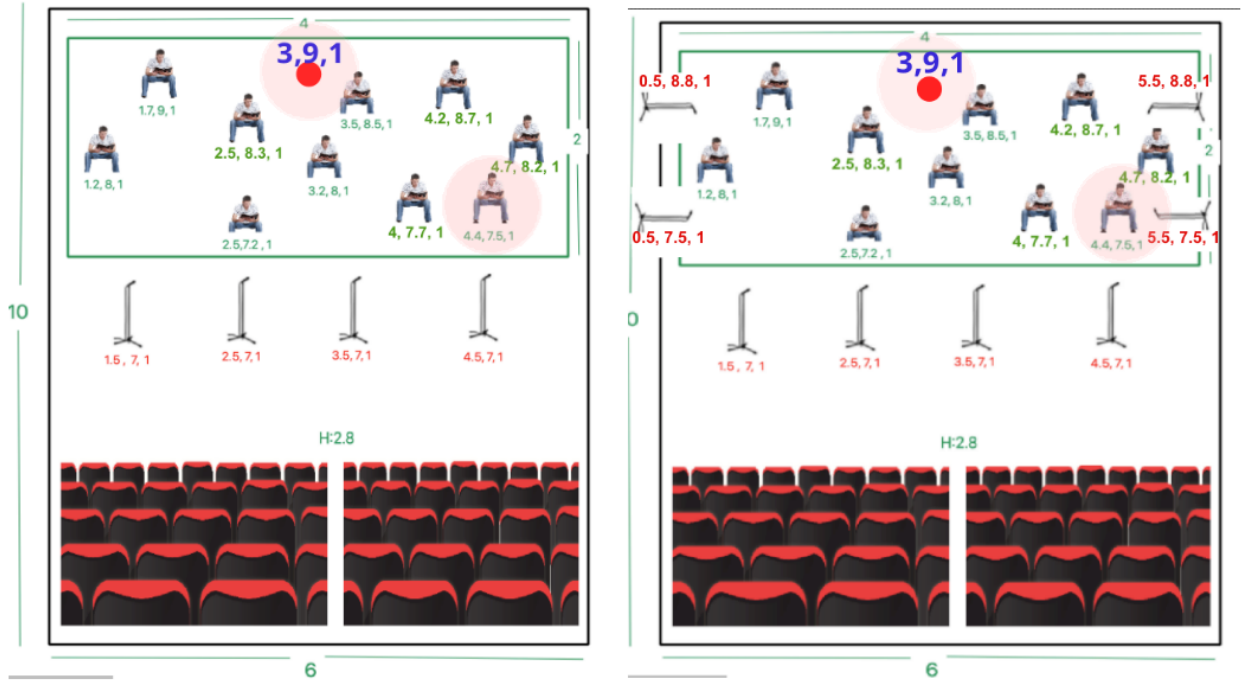
Ο συγχρονισμός με TDOA είναι απαραίτητη προϋπόθεση για την εφαρμογή της delay and sum τεχνικής. Από την άλλη, όταν τα σήματα μιξάρονται χωρίς TDOA συγχρονισμό παρουσιάζεται μία ακόμα περίπτωση στην οποία αναφέρεται με τον όρο MIX η οποία στην ουσία είναι η απλή πρόσθεση των σημάτων των μικροφώνων.

Τέλος, παρουσιάζεται η βαθμολογία του του σήματος του μικροφώνου που βρίσκεται πιο κοντά κάθε φορά στην πηγή, για την περίπτωση όπου δεν είναι ενεργοποιημένος ο συγχρονισμός. Αυτό στα αποτελέσματα φαίνεται σαν “κοντινότερο μικρόφωνο” και υπολογίζεται επίσης με τη γεωμετρία εφόσον είναι γνωστή η θέση της πηγής και του μικροφώνου.

#### **1.1.2.1. Μελέτη ακραίων περιπτώσεων με βάση την απόσταση**

Τα πρώτα αποτελέσματα που παρουσιάζονται σε αυτό το κεφάλαιο είναι τα εξής, επιλέχθηκαν δύο ακραίες θέσεις πηγής, μια πολύ κοντά σε ένα μικρόφωνο και η πιο απομακρυσμένη με σκοπο να εξεταστεί το πως ανταποκρίνονται οι μέθοδοι σε αυτές και στις εκάστοτε ιδιαίτερες συνθήκες. Στη συνέχεια παρουσιάζονται πίνακες που εμφανίζουν τον μέσο όρο των βαθμολογιών όλων των θέσεων με σκοπό να προκύψει ένα γενικό συμπέρασμα.

Οι θέσεις είναι 11 στο σύνολο (Εικόνα 4.3) και οι θέσεις που παρουσιάζονται είναι η θέση [4.4 , 7.5. 1] και η [3 , 9 , 1].



Εικόνα 4.3: Θέσεις πηγών και μικροφώνων, τέσσερα μικρόφωνα (αριστερά), οκτώ μικρόφωνα (δεξιά)

4 mic		RT <sub>60</sub> = 0.3s, SNR = 10				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,698	<b>1,997</b>	1,750	2,016	1,749	2,023
MSC	<b>1,737</b>	1,845	1,814	1,894	1,811	1,906
Delay & Sum			<b>1,976</b>	<b>2,187</b>	<b>1,970</b>	<b>2,198</b>
MIX	1,620	1,822				
Κοντινότερο μικρόφωνο	<b>1,974</b>	<b>2,314</b>				

4 mic		RT <sub>60</sub> = 0.3s, SNR = 20				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	<b>2,237</b>	<b>2,646</b>	2,388	2,740	2,389	2,752
MSC	2,084	2,198	2,250	2,340	2,249	2,332
Delay & Sum			<b>2,724</b>	<b>2,996</b>	<b>2,730</b>	<b>3,006</b>
MIX	2,151	2,460				
Κοντινότερο μικρόφωνο	<b>2,699</b>	<b>3,110</b>				

4 mic		RT <sub>60</sub> = 0.3s, SNR = 60				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	<b>2,641</b>	<b>3,164</b>	2,899	3,350	2,898	3,351
MSC	2,318	2,463	2,580	2,680	2,580	2,680
Delay & Sum			<b>3,294</b>	<b>3,619</b>	<b>3,294</b>	<b>3,619</b>
MIX	2,563	2,896				
Κοντινότερο μικρόφωνο	<b>3,234</b>	<b>3,699</b>				

4 mic		RT <sub>60</sub> = 0.6s, SNR = 10				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,439	<b>1,635</b>	1,476	1,662	1,475	1,653
MSC	<b>1,483</b>	1,556	1,525	1,588	1,518	1,583
Delay & Sum			<b>1,605</b>	<b>1,774</b>	<b>1,604</b>	<b>1,724</b>
MIX	1,390	1,519				
Κοντινότερο μικρόφωνο	<b>1,596</b>	<b>1,865</b>				



4 mic		RT <sub>60</sub> = 0.6s, SNR = 20				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,744	2,053	1,826	2,117	1,826	2,118
MSC	1,719	1,818	1,814	1,876	1,812	1,876
Delay & Sum			2,042	2,308	2,041	2,309
MIX	1,675	1,878				
Κοντινότερο μικροφώνο	2,007	2,394				

4 mic		RT <sub>60</sub> = 0.6s, SNR = 60				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,896	2,246	2,017	2,331	2,017	2,331
MSC	1,806	1,918	1,928	1,995	1,929	1,995
Delay & Sum			2,266	2,559	2,266	2,559
MIX	1,837	2,061				
Κοντινότερο μικροφώνο	2,208	2,630				

4 mic		RT <sub>60</sub> = 1.1s, SNR = 10				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,318	1,463	1,333	1,477	1,333	1,470
MSC	1,326	1,391	1,357	1,429	1,347	1,418
Delay & Sum			1,421	1,552	1,382	1,506
MIX	1,299	1,385				
Κοντινότερο μικροφώνο	1,418	1,624				

4 mic		RT <sub>60</sub> = 1.1s, SNR = 20				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,487	1,713	1,539	1,748	1,531	1,740
MSC	1,494	1,582	1,550	1,631	1,519	1,591
Delay & Sum			1,658	1,951	1,657	1,956
MIX	1,658	1,954				
Κοντινότερο μικροφώνο	1,468	1,595				

4 mic		RT <sub>60</sub> = 1.1s, SNR = 60				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,559	1,807	1,626	1,852	1,623	1,852
MSC	1,541	1,640	1,616	1,703	1,606	1,704
Delay & Sum			1,775	1,980	1,765	1,980
MIX	1,549	1,685				
Κοντινότερο μικροφώνο	1,757	2,069				

8 mic		RT <sub>60</sub> = 0.3s, SNR = 10				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,638	1,913	1,705	1,966	1,664	1,930
MSC	1,672	1,792	1,798	1,938	1,684	1,803
Delay & Sum			1,979	2,229	1,721	1,929
MIX	1,522	1,674				
Κοντινότερο μικροφώνο	1,967	2,309				

8 mic		RT <sub>60</sub> = 0.3s, SNR = 20				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	<b>2,148</b>	<b>2,519</b>	2,298	2,665	2,207	2,543
MSC	1,980	2,110	2,243	2,375	2,007	2,125
Delay & Sum			<b>2,702</b>	<b>3,014</b>	<b>2,296</b>	<b>2,603</b>
MIX	1,964	2,200				
Κοντινότερο μικροφώνο	<b>2,700</b>	<b>3,117</b>				

8 mic		RT <sub>60</sub> = 0.3s, SNR = 60				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	<b>2,502</b>	<b>2,925</b>	2,748	3,207	2,537	2,810
MSC	2,158	2,297	2,514	2,638	2,258	2,233
Delay & Sum			<b>3,251</b>	<b>3,606</b>	<b>2,770</b>	<b>2,887</b>
MIX	2,332	2,582				
Κοντινότερο μικροφώνο	<b>3,233</b>	<b>3,699</b>				

8 mic		RT <sub>60</sub> = 0.6s, SNR = 10				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,394	<b>1,573</b>	1,439	1,621	1,417	1,595
MSC	<b>1,463</b>	1,543	1,539	1,640	1,466	1,542
Delay & Sum			<b>1,608</b>	<b>1,822</b>	<b>1,481</b>	<b>1,610</b>
MIX	1,358	1,441				
Κοντινότερο μικροφώνο	<b>1,596</b>	<b>1,876</b>				

8 mic		RT <sub>60</sub> = 0.6s, SNR = 20				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,654	<b>1,933</b>	1,749	2,019	1,692	1,964
MSC	<b>1,686</b>	1,781	1,803	1,919	1,720	1,755
Delay & Sum			<b>2,020</b>	<b>2,308</b>	<b>1,790</b>	<b>1,984</b>
MIX	1,584	1,712				
Κοντινότερο μικροφώνο	<b>2,006</b>	<b>2,401</b>				

8 mic		RT <sub>60</sub> = 0.6s, SNR = 60				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	<b>1,784</b>	<b>2,098</b>	1,903	2,228	1,855	2,138
MSC	1,773	1,873	1,910	2,030	1,759	1,832
Delay & Sum			<b>2,222</b>	<b>2,540</b>	<b>2,045</b>	<b>2,181</b>
MIX	1,712	1,855				
Κοντινότερο μικροφώνο	<b>2,208</b>	<b>2,630</b>				

8 mic		RT <sub>60</sub> = 1.1s, SNR = 10				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5 ,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,293	<b>1,411</b>	1,313	1,443	1,286	<b>1,416</b>
MSC	<b>1,318</b>	1,391	1,353	1,465	<b>1,339</b>	1,384
Delay & Sum			<b>1,427</b>	<b>1,582</b>	1,334	1,386
MIX	1,275	1,324				
Κοντινότερο μικροφώνο	<b>1,424</b>	<b>1,628</b>				

8 mic		RT <sub>60</sub> = 1.1s, SNR = 20				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5, 1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,444	<b>1,624</b>	1,476	1,684	1,444	<b>1,636</b>
MSC	<b>1,484</b>	1,547	1,533	1,667	<b>1,488</b>	1,558
Delay & Sum			<b>1,653</b>	<b>1,881</b>	1,418	1,585
MIX	1,431	1,489				
Κοντινότερο μικρόφωνο	<b>1,660</b>	<b>1,954</b>				

8 mic		RT <sub>60</sub> = 1.1s, SNR = 60				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [4.4, 7.5, 1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,507	<b>1,704</b>	1,541	1,770	<b>1,512</b>	<b>1,723</b>
MSC	<b>1,534</b>	1,607	1,573	1,719	1,509	1,633
Delay & Sum			<b>1,750</b>	<b>1,989</b>	1,503	1,713
MIX	1,488	1,559				
Κοντινότερο μικρόφωνο	<b>1,757</b>	<b>2,069</b>				

**Πίνακας 2:** Αναλυτικά αποτελέσματα για θέση πηγής [4.4, 7.5, 1] και 4 και 8 μικρόφωνα

Στα αποτελέσματα που παρουσιάζονται παραπάνω τα οποία αφορούν στη θέση πηγής [4.4, 7.5, 1] η οποία είναι ακριβώς μπροστά από το τελευταίο μικρόφωνο της οριζόντιας συστοιχίας παρατηρείται -όπως αναμενόταν- μια σταθερή βαθμολογική επικράτηση της μεθόδου closest mic. Στην περίπτωση που υπάρχει συγχρονισμός μικροφώνων η delay and sum ενώ φαίνεται να συμπεριφέρεται επίσης καλά με μια σταθερότητα ενώ στην περίπτωση που δεν υπάρχει συγχρονισμός η μέθοδος MCP παρουσιάζει επίσης καλά αποτελέσματα, κυρίως όμως στις περιπτώσεις όπου δεν υπάρχει πολύς θόρυβος. Σε περιπτώσεις χωρίς συγχρονισμό για SNR = 10 & 20 dB και για omni μικρόφωνα η μέθοδος MSC δείχνει επιστρέφει καλύτερα αποτελέσματα σε σχέση με το MIX και την MCP. Επίσης στο πιο ακραίο σενάριο για RT<sub>60</sub>= 1.1 s και SNR = 10 dB επικρατεί και στην περίπτωση του συγχρονισμού με εκτιμητέες ΧΔΑ σε omni μικρόφωνο.

Ακολουθούν τα αποτελέσματα για την πιο απομακρυσμένη θέση πηγής [3, 9, 1]:

4 mic		RT <sub>60</sub> = 0.3s, SNR = 10				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,444	1,460	1,451	1,517	1,453	1,515
MSC	<b>1,653</b>	<b>1,726</b>	<b>1,659</b>	1,754	<b>1,660</b>	1,750
Delay & Sum			1,631	<b>1,771</b>	1,632	<b>1,769</b>
MIX	1,548	1,580				
Κοντινότερο μικροφώνο	1,458	1,518				

4 mic		RT <sub>60</sub> = 0.3s, SNR = 20				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,748	1,881	1,802	1,966	1,803	1,967
MSC	<b>1,814</b>	<b>2,067</b>	1,925	<b>2,289</b>	1,918	2,067
Delay & Sum			<b>1,999</b>	2,074	<b>2,001</b>	<b>2,290</b>
MIX	1,802	2,012				
Κοντινότερο μικροφώνο	1,837	1,979				

4 mic		RT <sub>60</sub> = 0.3s, SNR = 60				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	2,090	2,249	2,093	2,402	2,092	2,402
MSC	2,085	2,296	2,072	2,288	2,072	2,288
Delay & Sum			<b>2,200</b>	<b>2,609</b>	<b>2,200</b>	<b>2,609</b>
MIX	<b>2,127</b>	<b>2,299</b>				
Κοντινότερο μικροφώνο	2,159	2,390				

4 mic		RT <sub>60</sub> = 0.6s, SNR = 10				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,353	1,382	1,327	1,376	1,328	1,385
MSC	<b>1,506</b>	<b>1,560</b>	<b>1,495</b>	<b>1,565</b>	<b>1,492</b>	<b>1,561</b>
Delay & Sum			1,401	1,510	1,397	1,513
MIX	1,425	1,469				
Κοντινότερο μικροφώνο	1,377	1,411				

4 mic		RT <sub>60</sub> = 0.6s, SNR = 20				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,547	1,625	1,513	1,628	1,511	1,617
MSC	<b>1,625</b>	<b>1,732</b>	<b>1,555</b>	1,728	<b>1,638</b>	1,725
Delay & Sum			1,554	<b>1,733</b>	1,558	<b>1,734</b>
MIX	1,612	1,704				
Κοντινότερο μικροφώνο	1,599	1,685				

4 mic		RT <sub>60</sub> = 0.6s, SNR = 60				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,680	1,776	1,614	1,764	1,611	1,764
MSC	1,666	1,783	<b>1,676</b>	1,799	<b>1,676</b>	1,799
Delay & Sum			1,610	<b>1,819</b>	1,610	<b>1,819</b>
MIX	<b>1,688</b>	<b>1,810</b>				
Κοντινότερο μικροφώνο	<b>1,702</b>	<b>1,827</b>				

4 mic		RT <sub>60</sub> = 1.1s, SNR = 10				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,288	1,306	1,261	1,298	1,284	1,304
MSC	<b>1,390</b>	<b>1,424</b>	<b>1,373</b>	<b>1,427</b>	<b>1,373</b>	<b>1,435</b>
Delay & Sum			1,310	1,370	1,361	1,366
MIX	1,339	1,367				
Κοντινότερο μικροφώνο	1,292	1,331				

4 mic		RT <sub>60</sub> = 1.1s, SNR = 20				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,406	1,447	1,367	1,444	1,365	1,441
MSC	1,448	<b>1,512</b>	<b>1,454</b>	<b>1,527</b>	<b>1,443</b>	<b>1,541</b>
Delay & Sum			1,391	1,478	1,391	1,477
MIX	<b>1,476</b>	1,505				
Κοντινότερο μικροφώνο	1,437	<b>1,521</b>				

4 mic		RT <sub>60</sub> = 1.1s, SNR = 60				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,473	1,524	1,439	1,502	1,430	1,502
MSC	1,449	1,536	<b>1,453</b>	<b>1,552</b>	<b>1,453</b>	<b>1,552</b>
Delay & Sum			1,410	1,511	1,410	1,511
MIX	<b>1,501</b>	<b>1,543</b>				
Κοντινότερο μικροφώνο	1,462	<b>1,567</b>				

8 mic		RT <sub>60</sub> = 0.3s, SNR = 10				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,424	1,447	1,475	1,507	1,464	1,510
MSC	<b>1,636</b>	<b>1,710</b>	1,655	1,752	1,649	1,751
Delay & Sum			<b>1,804</b>	<b>1,967</b>	<b>1,801</b>	<b>1,963</b>
MIX	1,572	1,613				
Κοντινότερο μικροφώνο	1,457	1,521				

8 mic		RT <sub>60</sub> = 0.3s, SNR = 20				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,760	1,848	1,841	1,966	1,841	1,964
MSC	1,928	<b>2,066</b>	1,931	2,092	1,938	2,091
Delay & Sum			<b>2,257</b>	<b>2,562</b>	<b>2,256</b>	<b>2,557</b>
MIX	<b>1,952</b>	2,035				
Κοντινότερο μικροφώνο	1,854	1,988				

8 mic		RT <sub>60</sub> = 0.3s, SNR = 60				
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]					
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	2,034	2,186	2,135	2,386	2,135	2,386
MSC	2,079	2,296	2,072	2,288	2,073	2,288
Delay & Sum			<b>2,493</b>	<b>2,914</b>	<b>2,493</b>	<b>2,914</b>
MIX	<b>2,144</b>	<b>2,312</b>				
Κοντινότερο μικροφώνο	<b>2,152</b>	<b>2,390</b>				

8 mic		RT <sub>60</sub> = 0.6s, SNR = 10					
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)			Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]						
	omni	cardioid	omni	cardioid	omni	cardioid	
MCP	1,325	1,354	1,333	1,387	1,316	1,382	
MSC	<b>1,481</b>	<b>1,538</b>	1,482	1,552	<b>1,478</b>	1,542	
Delay & Sum			<b>1,508</b>	<b>1,663</b>	1,365	<b>1,662</b>	
MIX	1,417	1,472					
Κοντινότερο μικρόφωνο	1,368	1,404					

8 mic		RT <sub>60</sub> = 0.6s, SNR = 20					
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)			Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]						
	omni	cardioid	omni	cardioid	omni	cardioid	
MCP	1,530	1,580	1,516	1,621	1,488	1,622	
MSC	<b>1,620</b>	<b>1,725</b>	1,629	1,727	<b>1,627</b>	1,738	
Delay & Sum			<b>1,699</b>	<b>1,945</b>	1,498	<b>1,948</b>	
MIX	1,589	1,705					
Κοντινότερο μικρόφωνο	1,590	1,693					

8 mic		RT <sub>60</sub> = 0.6s, SNR = 60					
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)			Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]						
	omni	cardioid	omni	cardioid	omni	cardioid	
MCP	1,612	1,709	1,618	1,761	1,579	1,761	
MSC	<b>1,668</b>	1,783	1,690	1,799	<b>1,676</b>	1,799	
Delay & Sum			<b>1,762</b>	<b>2,045</b>	1,535	<b>2,045</b>	
MIX	1,644	<b>1,790</b>					
Κοντινότερο μικρόφωνο	<b>1,705</b>	1,827					

8 mic		RT <sub>60</sub> = 1.1s, SNR = 10					
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)			Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]						
	omni	cardioid	omni	cardioid	omni	cardioid	
MCP	1,259	1,289	1,277	1,292	1,260	1,299	
MSC	<b>1,362</b>	<b>1,425</b>	1,355	1,428	<b>1,366</b>	1,422	
Delay & Sum			<b>1,365</b>	<b>1,482</b>	1,276	<b>1,480</b>	
MIX	1,343	1,381					
Κοντινότερο μικρόφωνο	1,291	1,352					

8 mic		RT <sub>60</sub> = 1.1s, SNR = 20					
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)			Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]						
	omni	cardioid	omni	cardioid	omni	cardioid	
MCP	1,388	1,422	1,370	1,416	1,366	1,373	
MSC	<b>1,438</b>	<b>1,514</b>	1,432	1,515	<b>1,443</b>	<b>1,518</b>	
Delay & Sum			<b>1,461</b>	<b>1,629</b>	1,369	1,397	
MIX	1,422	1,457					
Κοντινότερο μικρόφωνο	1,416	1,503					

8 mic		RT <sub>60</sub> = 1.1s, SNR = 60					
Μέθοδοι	Χωρίς συγχρονισμό	Με συγχρονισμό βάση TDOA (oracle)			Με συγχρονισμό βάση TDOA (εκτιμητέο)		
	Source [3,9,1]						
	omni	cardioid	omni	cardioid	omni	cardioid	
MCP	1,427	1,483	1,432	1,497	1,411	1,438	
MSC	<b>1,449</b>	<b>1,541</b>	1,453	1,552	<b>1,451</b>	<b>1,552</b>	
Delay & Sum			<b>1,487</b>	<b>1,672</b>	1,364	1,415	
MIX	1,431	1,505					
Κοντινότερο μικρόφωνο	<b>1,462</b>	<b>1,567</b>					

**Πίνακας 3:** Αναλυτικά αποτελέσματα για θέση πηγής [3, 9, 1] και 4 και 8 μικρόφωνα

Όπως φαίνεται στον Πίνακα 3 παρουσιάζονται τα αποτελέσματα για μια αρκετά απομακρυσμένη θέση πηγής. Η θέση [3, 9, 1] (Εικόνα 4.3) εκτός από το ότι είναι μια αρκετά απομακρυσμένη θέση από τα μικρόφωνα, έχει το χαρακτηριστικό ότι σχεδόν ισαπέχει από όλα τα μικρόφωνα. Η MSC δείχνει να επικρατεί με μεγάλη σταθερότητα στην περίπτωση που δεν υπάρχει συγχρονισμός ενώ αξιοσημείωτο είναι το γεγονός πως σε αρκετές

περιπτώσεις ξεπερνάει την τιμή του κοντινότερου μικροφώνου. Στην περίπτωση που έχουμε συγχρονισμό φαίνεται να έχει πιο σταθερή επικράτηση η μέθοδος Delay & Sum όμως αυτο συμβαίνει κυρίως στις περιπτώσεις που έχουμε καλές συνθήκες. Σε συνθήκες με πολύ θόρυβο, μεγάλο reverberation time και μη κατευθυντικό μικρόφωνο φαίνεται να αντιδρά πολύ καλά η MSC. Επίσης σε αυτή τη συνθήκη παρατηρείται η κακή επίδοση της MCP.

### 1.1.2.2. Συνολική επίδοση των μεθόδων

Στους πίνακες που ακολουθούν παρουσιάζονται τα αποτελέσματα των μέσων όρων των τιμών όλων το θέσεων πηγής που έχουμε επιλέξει, έντεκα στο σύνολο, για συστοιχίες με τέσσερα και οκτώ μικρόφωνα όπως φαίνεται και στην Εικόνα 4.3 και για κάθε μια από τις παραμέτρους που αλλάζουν, δηλαδή χρόνο αντήχησης, SNR και συγχρονισμό μικροφώνων. Πιο αναλυτικά διαστάσεις δωματίου [6, 10, 2.8] και θέσεις πηγών: [1.2, 8, 1], [1.7, 9, 1], [2.5, 7.2, 1], [2.5, 8.3, 1], [3, 9, 1], [3.2, 8, 1], [3.5, 8.5, 1], [4, 7.7, 1], [4.2, 8.7, 1], [4.4, 7.5, 1], [4.7, 8.2, 1].

4 mic			RT <sub>60</sub> = 0.3s, SNR = 10			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,519	1,636	1,570	1,691	1,570	1,693
MSC	<b>1,659</b>	<b>1,755</b>	1,732	1,830	1,732	1,832
Delay & Sum			<b>1,763</b>	<b>1,917</b>	<b>1,754</b>	<b>1,907</b>
MIX	1,542	1,654				
Κοντινότερο μικρόφωνο	1,639	<b>1,812</b>				

4 mic			RT <sub>60</sub> = 0.3s, SNR = 20			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,929	2,109	2,037	2,242	2,036	2,244
MSC	1,954	2,077	2,082	2,219	2,075	2,217
Delay & Sum			<b>2,337</b>	<b>2,580</b>	<b>2,323</b>	<b>2,582</b>
MIX	<b>1,966</b>	<b>2,156</b>				
Κοντινότερο μικρόφωνο	<b>2,142</b>	<b>2,426</b>				

4 mic			RT <sub>60</sub> = 0.3s, SNR = 60			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	2,250	2,509	<b>2,423</b>	<b>2,745</b>	<b>2,418</b>	<b>2,744</b>
MSC	2,107	2,280	<b>2,284</b>	<b>2,479</b>	<b>2,275</b>	<b>2,479</b>
Delay & Sum			<b>2,765</b>	<b>3,135</b>	<b>2,742</b>	<b>3,133</b>
MIX	<b>2,288</b>	<b>2,574</b>				
Κοντινότερο μικρόφωνο	<b>2,533</b>	<b>2,945</b>				

4 mic			RT <sub>60</sub> = 0.6s, SNR = 10			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,383	1,482	1,383	1,482	1,407	1,506
MSC	<b>1,471</b>	<b>1,560</b>	<b>1,471</b>	<b>1,560</b>	<b>1,508</b>	1,590
Delay & Sum			1,389	1,496	1,498	<b>1,616</b>
MIX	1,389	1,496				
Κοντινότερο μικρόφωνο	1,449	<b>1,592</b>				

4 mic			RT <sub>60</sub> = 0.6s, SNR = 20			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,615	1,779	1,656	1,824	1,648	1,819
MSC	<b>1,659</b>	1,767	1,713	1,815	1,711	1,814
Delay & Sum			<b>1,789</b>	<b>1,984</b>	<b>1,757</b>	<b>1,977</b>
MIX	1,612	<b>1,797</b>				
Κοντινότερο μικροφώνο	<b>1,718</b>	<b>1,952</b>				

4 mic			RT <sub>60</sub> = 0.6s, SNR = 60			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	<b>1,729</b>	1,928	1,781	1,987	1,775	1,974
MSC	1,716	1,839	1,781	1,897	1,777	1,890
Delay & Sum			<b>1,922</b>	<b>2,159</b>	<b>1,892</b>	<b>2,143</b>
MIX	1,722	<b>1,946</b>				
Κοντινότερο μικροφώνο	<b>1,851</b>	<b>2,132</b>				

4 mic			RT <sub>60</sub> = 1.1s, SNR = 10			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,299	1,376	1,385	1,310	1,385	1,385
MSC	<b>1,355</b>	<b>1,424</b>	1,443	<b>1,370</b>	1,438	1,443
Delay & Sum			<b>1,450</b>	1,345	<b>1,440</b>	<b>1,450</b>
MIX	1,307	1,372				
Κοντινότερο μικροφώνο	1,345	<b>1,451</b>				

4 mic			RT <sub>60</sub> = 1.1s, SNR = 20			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,442	1,546	1,457	1,562	1,452	1,561
MSC	<b>1,480</b>	<b>1,564</b>	1,508	1,598	<b>1,502</b>	1,586
Delay & Sum			<b>1,515</b>	<b>1,655</b>	1,494	<b>1,636</b>
MIX	1,437	1,543				
Κοντινότερο μικροφώνο	<b>1,502</b>	<b>1,662</b>				

4 mic			RT <sub>60</sub> = 1.1s, SNR = 60			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,503	1,613	1,520	1,635	1,517	1,632
MSC	<b>1,517</b>	1,603	1,547	1,640	1,548	1,627
Delay & Sum			<b>1,573</b>	<b>1,731</b>	<b>1,572</b>	<b>1,708</b>
MIX	1,495	<b>1,614</b>				
Κοντινότερο μικροφώνο	<b>1,572</b>	<b>1,739</b>				

8 mic			RT <sub>60</sub> = 0.3s, SNR = 10			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,518	1,609	1,589	1,693	1,553	1,669
MSC	<b>1,641</b>	<b>1,712</b>	1,757	1,849	1,691	1,806
Delay & Sum			<b>1,871</b>	<b>2,053</b>	<b>1,727</b>	<b>1,932</b>
MIX	1,499	1,572				
Κοντινότερο μικροφώνο	<b>1,680</b>	<b>1,845</b>				

8 mic			RT <sub>60</sub> = 0.3s, SNR = 20			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	<b>1,921</b>	<b>2,059</b>	2,072	2,256	2,012	2,171
MSC	1,915	2,013	2,125	2,271	2,029	2,163
Delay & Sum			<b>2,514</b>	<b>2,792</b>	<b>2,318</b>	<b>2,570</b>
MIX	1,902	2,039				
Κοντινότερο μικροφώνο	<b>2,206</b>	<b>2,477</b>				

8 mic			RT <sub>60</sub> = 0.3s, SNR = 60			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	<b>2,228</b>	2,419	2,459	2,747	2,374	2,607
MSC	2,049	2,182	2,331	2,547	2,224	2,379
Delay & Sum			<b>2,996</b>	<b>3,397</b>	<b>2,758</b>	<b>3,139</b>
MIX	2,209	<b>2,429</b>				
Κοντινότερο μικροφώνο	<b>2,597</b>	<b>2,987</b>				



8 mic			RT <sub>60</sub> = 0.6s, SNR = 10			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,380	1,450	1,414	1,507	1,388	1,482
MSC	<b>1,466</b>	<b>1,525</b>	1,523	1,612	<b>1,493</b>	1,569
Delay & Sum			<b>1,574</b>	<b>1,752</b>	1,459	<b>1,624</b>
MIX	1,361	1,418				
Κοντινότερο μικροφώνο	<b>1,470</b>	<b>1,605</b>				

8 mic			RT <sub>60</sub> = 0.6s, SNR = 20			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,602	1,723	1,663	1,819	1,620	1,775
MSC	<b>1,652</b>	<b>1,736</b>	1,742	1,866	1,701	1,800
Delay & Sum			<b>1,910</b>	<b>2,185</b>	<b>1,717</b>	<b>1,988</b>
MIX	1,568	1,672				
Κοντινότερο μικροφώνο	<b>1,748</b>	<b>1,959</b>				

8 mic			RT <sub>60</sub> = 0.6s, SNR = 60			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,708	<b>1,855</b>	1,790	1,983	1,739	1,920
MSC	<b>1,709</b>	1,806	1,818	1,957	1,764	1,872
Delay & Sum			<b>2,063</b>	<b>2,384</b>	<b>1,879</b>	<b>2,154</b>
MIX	1,667	1,800				
Κοντινότερο μικροφώνο	<b>1,880</b>	<b>2,134</b>				

8 mic			RT <sub>60</sub> = 1.1s, SNR = 10			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,304	1,351	1,312	1,382	1,299	1,366
MSC	<b>1,353</b>	<b>1,405</b>	1,374	1,449	<b>1,366</b>	1,427
Delay & Sum			<b>1,399</b>	<b>1,542</b>	1,328	<b>1,437</b>
MIX	1,282	1,322				
Κοντινότερο μικροφώνο	<b>1,355</b>	<b>1,455</b>				

8 mic			RT <sub>60</sub> = 1.1s, SNR = 20			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,434	1,510	1,451	1,560	1,438	1,526
MSC	<b>1,484</b>	<b>1,543</b>	1,518	1,614	<b>1,498</b>	1,576
Delay & Sum			<b>1,583</b>	<b>1,789</b>	1,468	<b>1,610</b>
MIX	1,408	1,466				
Κοντινότερο μικροφώνο	<b>1,520</b>	<b>1,658</b>				

8 mic			RT <sub>60</sub> = 1.1s, SNR = 60			
Μέθοδοι	Χωρίς συγχρονισμό		Με συγχρονισμό βάση TDOA (oracle)		Με συγχρονισμό βάση TDOA (εκτιμητέο)	
	omni	cardioid	omni	cardioid	omni	cardioid
MCP	1,480	1,571	1,510	1,629	1,495	1,591
MSC	<b>1,514</b>	<b>1,582</b>	1,557	1,661	<b>1,533</b>	1,626
Delay & Sum			<b>1,651</b>	<b>1,877</b>	1,527	<b>1,664</b>
MIX	1,450	1,523				
Κοντινότερο μικροφώνο	<b>1,582</b>	<b>1,736</b>				

*Πίνακας 4: Αποτελέσματα μέσωσ όρων από όλες τις θέσεις*

Στα αποτελέσματα που παρουσιάζονται για τις συστοιχίες τεσσάρων και οκτώ μικροφώνων παρατηρείται μια αισθητή επικράτηση της μεθόδου MSC στην περίπτωση που δεν είναι ενεργοποιημένος ο συγχρονισμός και μάλιστα επικρατεί σε όλες τις περιπτώσεις όταν χρησιμοποιούνται omni μικρόφωνα. Ακόμα, σε αρκετές περιπτώσεις σημειώνεται υψηλότερη βαθμολογία στην μέθοδο MSC με ενεργοποιημένο τον συγχρονισμό την περίπτωση εκτιμώμενων TDOAS και πάλι το πολικό διάγραμμα των μικροφώνων είναι omni. Έχει επίσης ενδιαφέρον το γεγονός ότι στην περίπτωση της οριζόντιας συστοιχίας των τεσσάρων μικροφώνων, που αποτελεί και μια συνηθισμένη διάταξη σε θεατρικές παραστάσεις η μέθοδος MSC επικρατεί και του κοντινότερου μικροφώνου.

Στην περίπτωση που ενεργοποιείται ο συγχρονισμός η επικρατέστερη μέθοδος φαίνεται να είναι η Delay & Sum η οποία δίνει σταθερά καλά αποτελέσματα, εκτός από κάποιες περιπτώσεις όπου πάλι σε δύσκολες συνθήκες υψηλού θορύβου και μη κατευθυντικών μικροφώνων κυρίως στην περίπτωση που χρησιμοποιούνται οκτώ μικρόφωνα φαίνεται να υπερισχύει η MSC. Και σε αυτά τα αποτελέσματα η MCP τεχνική φαίνεται να μην ανταποκρίνεται καθόλου καλά πέρα από ελάχιστες περιπτώσεις όπου δεν υπάρχει θόρυβος και αυτό είναι λογικό καθώς η τεχνική αυτή αναδεικνύει το μικρόφωνο με τη μεγαλύτερη ενέργεια στο συχνοτικό φάσμα κατ' επέκταση ενισχύεται και ο θόρυβος.

Ένα γενικό συμπέρασμα που θα μπορούσε να προκύψει από τα παραπάνω αποτελέσματα είναι ότι η MSC είναι μια μέθοδος που λειτουργεί καλύτερα σε “δύσκολες συνθήκες”. Στην περίπτωση που ενεργοποιείται ο συγχρονισμός υπερέχει η τεχνική delay & sum η οποία φαίνεται να επικρατεί στις περισσότερες περιπτώσεις. Άρα είναι μια μέθοδος που επίσης λειτουργεί καλά και παρουσιάζει μια σταθερότητα.

Σε αυτή τη φάση θα μπορούσε με ασφάλεια να παρατηρηθεί το γεγονός ότι μια αξιόπιστη μέθοδος αυτόματης μίξης είναι η MSC σε συνθήκες δύσκολες, ενώ στην περίπτωση που υπάρχει η δυνατότητα καθυστερήσεις των μικροφώνων και ελεγχόμενο περιβάλλον η μέθοδος delay & sum είναι πιο αποδοτική και αξιόπιστη.

Ωστόσο είναι σημαντικό να τονιστεί η βαρύτητα της επικράτησης της μεθόδου MSC στην περίπτωση που δεν γίνεται συγχρονισμός ακριβώς γιατί η συγκεκριμένη εργασία σκοπό έχει να μελετήσει αυτές τις μεθόδους σε ένα συγκεκριμένο πλαίσιο, δηλαδή σε μια θεατρική παράσταση. Σε αυτή τη συνθήκη με ηθοποιούς που κινούνται διαρκώς που περιστρέφουν το κεφάλι τους και γενικότερα δεν υπάρχει προβλέψιμη συμπεριφορά, είναι αμφίβολο κατά πόσο θα μπορεί να γίνει σωστά ο υπολογισμός των TDOAs που είναι απαραίτητος για την delay and sum τεχνική. Το ίδιο ισχύει και για την μέθοδο του κοντινότερου μικροφώνου η οποία επιστρέφει επίσης καλά αποτελέσματα όμως απαιτεί εντοπισμό θέσης πηγής. Ακόμα, μια θεατρική παράσταση διεξάγεται σε συνθήκες που μπορεί να είναι απρόβλεπτες με αστάθμητους παράγοντες όπως το θόρυβο που επικρατεί στη σκηνή η στο περιβάλλον ή τα χαρακτηριστικά της εκάστοτε αίθουσας. Λαμβάνοντας λοιπόν υπόψη όλα τα παραπάνω, τα αποτελέσματα της μεθόδου MSC είναι αρκετά ικανοποιητικά και θα μπορούσαν να αποτελέσουν πρόταση για περαιτέρω έρευνα.

Τέλος σημαντικό είναι να γίνει και μια γενικότερη σύγκριση των εργαλείων τα οποία φαίνεται να επηρεάζουν σημαντικά τα αποτελέσματα. Στο κεφάλαιο 2.2.1 όπου πραγματοποιήθηκε η

αξιολόγηση εργαλείων Pyroomacoustics & PESQ φάνηκε η καλύτερη βαθμολογία των καρδιοειδών μικροφώνων σε σχέση με τα omni, κάτι που παρατηρείται και στα τελικά πειράματα της εργασίας. Είναι ξεκάθαρη η βαθμολογική υπεροχή των cardioid σε κάθε συνθήκη και αυτό είναι κάτι που λαμβάνεται υπόψιν, καθώς όταν η συστοιχία αποτελείται από omni μικρόφωνα, αντιμετωπίζεται σαν μια δύσκολη συνθήκη στην αξιολόγηση των μεθόδων.

Η ίδια σύγκριση μπορεί να γίνει και για τις περιπτώσεις συγχρονισμού, να εξεταστεί δηλαδή η βαθμολογία σε ομάδες με σκοπο να διαπιστωθούν οι διαφορές μεταξύ περιπτώσεων που υπάρχει ή όχι συγχρονισμός καθώς και μεταξύ περιπτώσεων που ο συγχρονισμός γίνεται με βάση την πραγματική ΧΔΑ ή με βάση την εκτιμητέα ΧΔΑ. Χαρακτηριστικό δείγμα στον Πίνακα 5 από μέσους όρους τιμών με πραγματική ΧΔΑ και με εκτιμητέα ΧΔΑ καθώς και χωρίς συγχρονισμό.

Όπως λοιπόν παρατηρείται στα παραπάνω πειράματα, οι εκτιμητέες ΧΔΑ δεν συμπίπτουν πάντα με τις πραγματικές ΧΔΑ, γεγονός που κυρίως οφείλεται στην αντήχηση του δωματίου, ή στο θόρυβο. Τα αποτελέσματα όπως αναμενόταν δείχνουν ότι η χρήση των πραγματικών ΧΔΑ έναντι των εκτιμώμενων οδηγούν σε εμφανή βελτίωση του PESQ.

Δωμάτια	Χωρίς συγχρονισμό	Με συγχρονισμό βάση πραγματικά TDOA (oracle)	Με συγχρονισμό βάση εκτιμητέα TDOA
RT03_SNR10	1.592	<b>1.802</b>	1.730
RT03_SNR20	1.975	<b>2.338</b>	2.210
RT03_SNR60	2.253	<b>2.746</b>	2.580
RT06_SNR10	1.433	<b>1.564</b>	1.503
RT06_SNR20	1.659	<b>1.864</b>	1.767
RT06_SNR60	1.758	<b>1.999</b>	1.888
RT11_SNR10	1.336	<b>1.410</b>	1.371
RT11_SNR20	1.474	<b>1.586</b>	1.519
RT11_SNR60	1.520	<b>1.648</b>	1.573

*Πίνακας 5: Αποτελέσματα μέσω των όρων από τιμές με και χωρίς συγχρονισμό για πειράματα με συστοιχία 4 μικροφώνων.*

Αντίστοιχα, σημαντικές διαφορές παρατηρούνται και στη σύγκριση της περίπτωσης όπου ενεργοποιείται ο συγχρονισμός σε σχέση με την περίπτωση που δεν είναι ενεργοποιημένος. Όπως είναι αναμενόμενο, παρατηρείται μεγαλύτερη βαθμολογία με ενεργοποιημένο συγχρονισμό, ακόμα και αν οι ΧΔΑ είναι οι εκτιμώμενες. Αυτές οι γενικές συγκρίσεις είναι χρήσιμες ως τέτοιες, όμως βοηθούν αρκετά και στην αξιολόγηση των μεθόδων και την εξαγωγή συμπερασμάτων.

# 5. Συμπεράσματα και Μελλοντική Εργασία

---

## 5.1. Σύνοψη & Συμπεράσματα

Συνοψίζοντας, στα παραπάνω κεφάλαια παρουσιάστηκε μια εργασία που σκοπό έχει να προτείνει μια μέθοδο αυτόματης μίξης -με δυνατότητα εφαρμογής σε θεατρικές παραστάσεις- η οποία διαφοροποιείται από τις υπάρχουσες gate-based τεχνικές. Για να επιτευχθεί αυτό πραγματοποιήθηκαν πειραματικές διαδικασίες σε περιβάλλον προσομοίωσης και χρησιμοποιήθηκαν εργαλεία τα οποία προηγουμένως δοκιμαστήκαν, επίσης σε συνθήκες προσομοίωσης. Στις πειραματικές διαδικασίες αξιολογήθηκαν μια σειρά από μεθόδους και εργαλεία τα οποία εξετάστηκαν και συνδυαστικά και σε διάφορες συνθήκες και μεταβλητές.

Τα συμπεράσματα που προέκυψαν είναι πώς αφενός, δεν αναδεικνύεται καθολικά κάποια μέθοδος σαν η πιο αποτελεσματική όμως, μιας και τα πειράματα αποτελούν έναν συνδυασμό τεσσάρων μεθόδων με εναλλαγές μεταβλητών, προκύπτουν συγκεκριμένα συμπεράσματα για συγκεκριμένες συνθήκες. Αρχικά στην παρούσα διπλωματική έγινε η προσπάθεια να αξιολογηθεί μια καινοτόμα προσέγγιση η Maximum Component Preservation (MCP). Η συγκεκριμένη μέθοδος όπως έδειξαν τα πειράματα δεν ανταποκρίθηκε καλά επομένως έγινε η αναζήτηση για άλλες μεθόδους κατά τη διάρκεια αυτής της διαδικασίας. Οι μέθοδοι που επίσης ερευνήθηκαν και αξιολογήθηκαν η τεχνική delay and sum και η μέθοδος Magnitude Squared Coherence (MSC).

Η μέθοδος delay and sum παρουσιάζει μια σταθερότητα και πιο καλά αποτελέσματα σε σχέση με την MSC στις περιπτώσεις όπου υπάρχει εκτίμηση των ΧΔΑ από την πηγή σε όλα τα μικρόφωνα. Στις περιπτώσεις όπου δεν υπάρχει συγχρονισμός και γενικότερα όσο οι συνθήκες δεν είναι καλές, δηλαδή υψηλός χρόνος αντήχησης, υψηλός θόρυβος παντοκατευθυντικά μικρόφωνα, μεγάλη απόσταση πηγής από τα μικρόφωνα η μέθοδος MSC φαίνεται να έχει ενδιαφέρουσα συμπεριφορά. Επειδή εξ αρχής η πρόθεση αυτής της εργασίας ήταν να μελετήσει διαφορετικές προσεγγίσεις αυτόματης μίξης, σε περιβάλλον θεατρικής παράστασης, όπου οι συνθήκες δεν είναι προβλεψιμες και τεχνικές όπως η εκτίμηση των ΧΔΑ είναι δύσκολη υπόθεση, τα αποτελέσματα της μεθόδου MSC παρουσιάζουν ενδιαφέρον και θα μπορούσαν να αποτελέσουν πρόταση για περαιτέρω έρευνα.

## 5.2. Μελλοντικές Εφαρμογές & Βελτιώσεις

Λαμβάνοντας υπόψη τα συμπεράσματα που προέκυψαν από τα πειράματα, σε πρώτη φάση θα μπορούσαν να συνεχιστούν οι έρευνες σε συνθήκες προσομοίωσης με περισσότερα σενάρια και μεταβλητές να μεταβάλλονται όπως για παράδειγμα διαφορετικές διατάξεις μικροφώνων και αίθουσες. Υπάρχει η δυνατότητα για επιμέρους παρεμβάσεις, όπως για παράδειγμα το ότι η μέθοδος MSC λειτουργεί με τη σύγκριση δύο καναλιών, όσα μικρόφωνα και να χρησιμοποιούνται στο εκάστοτε πείραμα. Στα πειράματα που πραγματοποιήθηκαν για την εργασία επιλέχθηκε τα ζεύγη να διαμορφώνονται με βάση τη γειτνίαση, επομένως αυτή είναι μια συνθήκη που θα μπορούσε να διερευνηθεί όπως για παράδειγμα να επιλέγονται τα πιο απομακρυσμένα μικρόφωνα.

Φυσικά εκτός από όλες αυτές τις διαδικασίες ο αλγόριθμος που έχει δημιουργηθεί έχει τη δυνατότητα να μεταφερθεί σε πραγματικές συνθήκες. Θα μπορούσε να στηθεί μια πειραματική διαδικασία κατά την οποία, θα πραγματοποιούνταν μια θεατρική παράσταση σε φυσικό χώρο με ηθοποιούς έτσι ώστε να αξιολογηθούν τα μικροφωνικά σήματα. Συνδυαστικά, θα μπορούσε να λειτουργεί παράλληλα και ένα σύστημα αυτόματης μίξης, που υπάρχει σε ψηφιακές κονσόλες, με σκοπό να συγκριθεί με τις μεθόδους που μελετώνται στην παρούσα εργασία.

Στην περίπτωση που πραγματοποιηθούν τα παραπάνω και προκύψουν πιο ασφαλή συμπεράσματα θα μπορούσε η μέθοδος Magnitude Squared Coherence (MSC) να αξιοποιηθεί με επιτυχία σαν ένα βασικό εργαλείο, με σκοπό τη δημιουργία ενός καινοτόμου συστήματος αυτόματης μίξης και έτσι τα συστήματα αυτόματης μίξης, να πάψουν να αποτελούν τη χειρότερη επιλογή για τους ηχολήπτες θεατρικών παραστάσεων.

Κλείνοντας, η αρκετά καλή επίδοση του closest mic είναι κάτι που επίσης πρέπει να αναφερθεί. Βέβαια, στην προκειμένη περίπτωση ο τρόπος με τον οποίο υπολογίζεται, μέσω γεωμετρίας, προϋποθέτει ότι είναι γνωστή η θέση της εκάστοτε πηγή και μικροφώνου. Επομένως, βάση αποτελεσμάτων φαίνεται ότι θα άξιζε ενδεχομένως να υλοποιηθεί και να αξιολογηθεί μια μέθοδος που να εκτιμάει το closest mic με βάση ενεργειακά κριτήρια αλλά και μέσω TDOA.

# Αναφορές

---

- Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, 65(4), 943–950. <https://doi.org/10.1121/1.382599>
- Blandin, C., Ozerov, A., & Vincent, E. (2012). Multi-source TDOA estimation in reverberant audio using angular spectra and clustering. *Signal Processing*, 92(8), 1950–1960. <https://doi.org/10.1016/j.sigpro.2011.09.032>
- Buck, M., & Rößler, M. (2001). First order differential microphone arrays for automotive applications. *Proc. IWAENC*, 3–6.
- Dugan, D. (1975, May). Automatic microphone mixing. In *Audio Engineering Society Convention 51*. Audio Engineering Society.
- Dugan, D. (1992). Tutorial: Application of automatic mixing techniques to audio consoles. *SMPTE journal*, 101(1), 19-27.
- Hosseini, M. S., Rezaie, A. H., & Zanjireh, Y. (2017). Time difference of arrival estimation of sound source using cross correlation and modified maximum likelihood weighting function. *Scientia Iranica*, 24(6), 3268–3279. <https://doi.org/10.24200/sci.2017.4355>
- Rix, A.W., Beerends, J.G., Hollier, M.P. and Hekstra, A.P. (2001), Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In 2001 IEEE international conference on acoustics, speech, and signal processing. (Vol. 2, pp. 749-752).
- Schimmel, S. M., Müller, M. F., & Dillier, N. (2009). A fast and accurate “shoebox” room acoustics simulator. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (pp. 241–244).
- Stefanakis, N., & Mouchtaris, A. (2015). Foreground suppression for capturing and reproduction of crowded acoustic environments. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (Vol. 2015-August, pp. 51–55). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICASSP.2015.7177930>
- Stefanakis, N., Mastorakis, Y., Alexandridis, A., & Mouchtaris, A. (2019). Automating mixing of user-generated audio recordings from the same event. *AES*:

Journal of the Audio Engineering Society, 67(4), 201–212.  
<https://doi.org/10.17743/jaes.2019.0008>

- Suzuki, M., & Honjo, T. (2015, December). Spot-forming method by using two shotgun microphones. In 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA) (pp. 188-191). IEEE.
- Thiergart, O., Kallinger, M., Galdo, G. D., & Kuech, F. (2011). Parametric spatial sound processing using linear microphone arrays. In *Microelectronic Systems* (pp. 321-329). Springer, Berlin, Heidelberg.
- Van Veen, B. D., & Buckley, K. M. (1988). Beamforming: A versatile approach to spatial filtering. *IEEE assp magazine*, 5(2), 4-24.