

Υλοποίηση και υποκειμενική αξιολόγηση της Αμφιηχίας  
με σφαιρική διάταξη ηχείων

Άγγελος Ηλίας



Η διπλωματική εργασία υπέβαλε εκπλήρωση των απαιτήσεων για το

Μεταπτυχιακό δίπλωμα ειδίκευσης Επιστήμης Υπολογιστών

Πανεπιστήμιο Κρήτης

Σχολή Επιστημών και Μηχανικών

Τμήμα Επιστήμης Υπολογιστών

Βασιλικά Βουτών, Τ.Κ 700 13 Ηράκλειο, Κρήτη, Ελλάδα

Επόπτης καθηγητής: Παναγιώτης Τσακαλίδης

Πανεπιστήμιο Κρήτης  
Τμήμα Επιστήμης Υπολογιστών

Υλοποίηση και υποκειμενική αξιολόγηση της Αμφιηχίας με σφαιρική διάταξη ηχείων

Διπλωματική Εργασία του

Άγγελος Ηλίας

Η διπλωματική εργασία υπέβαλε εκπλήρωση των απαιτήσεων για το  
Μεταπτυχιακό δίπλωμα ειδίκευσης Επιστήμης Υπολογιστών  
Συμβουλευτική Επιτροπή

Συγγραφέας:

\_\_\_\_\_

Άγγελος Ηλίας

Έγκριση Επιτροπής:

\_\_\_\_\_

Παναγιώτης Τσακαλίδης  
Καθηγητής, Επόπτης Μεταπτυχιακής Εργασίας

\_\_\_\_\_

Ξενοφώντας Δημητρόπουλος  
Αναπλ. Καθηγητής, Μέλος Επιτροπής

\_\_\_\_\_

Νικόλαος Στεφανιάκης  
Επικ.Καθηγητής, Μέλος επιτροπής

\_\_\_\_\_

Ιωάννης Πανταζής  
Ερευνητής Γ, Μέλος επιτροπής

\_\_\_\_\_

Αντώνης Αργυρός  
Καθηγητής, Διευθυντής Μεταπτυχιακών Σπουδών

Ηράκλειο, Οκτώβριος 2019



## ΠΕΡΙΕΧΟΜΕΝΑ

<b>1</b>	<b>Εισαγωγή</b>	<b>3</b>
<b>2</b>	<b>Αμφιηχία</b>	<b>6</b>
2.1	Κωδικοποίηση 1 <sup>ης</sup> -τάξης Αμφιηχίας . . . . .	6
2.2	Αποκωδικοποίηση 1 <sup>ης</sup> -τάξης Αμφιηχίας . . . . .	9
2.3	Υλοποίηση Με Υψηλότερες Τάξεις . . . . .	10
2.3.1	Σφαιρικές Αρμονικές . . . . .	11
2.3.2	Κωδικοποίηση Υψηλότερης Τάξης . . . . .	12
<b>3</b>	<b>Παραμετρικές Μέθοδοι</b>	<b>15</b>
3.1	Σύλληψη και Αναπαραγωγή με την Μέθοδο DirAC . . . . .	15
3.1.1	Εκτίμηση Κατεύθυνσης Άφιξης . . . . .	15
3.1.2	Υπολογισμός Συντελεστή Διάχυσης . . . . .	16
3.1.3	Αναπαραγωγή . . . . .	16
3.2	Σύλληψη και Αναπαραγωγή με την Μέθοδο ImmACS . . . . .	18
<b>4</b>	<b>Υλοποίηση της Αμφιηχίας</b>	<b>22</b>
4.1	Με Χρήση Sennheizer . . . . .	22
4.1.1	Βαθμονόμηση Κάρτας Ήχου . . . . .	25
4.2	Με την Χρήση Κυκλικής Συστοιχίας Μικροφώνων . . . . .	28
4.2.1	Συστοιχία Διαφορικών Μικροφώνων 1ης Τάξης . . . . .	28
4.2.2	Βαθμονόμηση Συστοιχίας Μικροφώνων . . . . .	31
<b>5</b>	<b>Αξιολόγηση</b>	<b>33</b>
5.1	Πειραματική Διαδικασία . . . . .	33
5.1.1	Ηχογραφήσεις (Στατικές - Κινούμενες πηγές) . . . . .	33
5.1.2	Επεξεργασία για Αναπαραγωγή Στο Acoustic Lab . . . . .	36
5.2	Υποκειμενική Αξιολόγηση . . . . .	39
5.2.1	Ερωτηματολόγιο και Αποτελέσματα . . . . .	39
5.2.2	Αντικειμενική Αξιολόγηση . . . . .	42
<b>6</b>	<b>Συμπεράσματα</b>	<b>46</b>
<b>7</b>	<b>Παράρτημα</b>	<b>47</b>
7.1	VBAP . . . . .	47
7.2	Binaural Reproduction με HRTFs . . . . .	49

## ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

1	Σχηματική αναπαράσταση πολικών συντεταγμένων ενός πανκατευθυντικού μικροφώνου (πάνω κομμάτι) και τριών δικατευθυντικών ( άξονες $x, z, y$ ) στον τρισδιάστατο χώρο. Πηγή [1] . . . . .	8
2	Σχηματική αναπαράσταση πλατωνικών στερεών . . . . .	10
3	Σύστημα συντεταγμένων πάνω σε σφαίρα . . . . .	11
4	Σχηματική αναπαράσταση σφαιρικών αρμονικών $n = 4$ τάξεως στον 3-διάστατο χώρο . . . . .	12
5	Διάγραμμα μεθόδου DirAC για την ανακατασκευή ηχητικού γεγονότος σε μια διάταξη ηχείων . . . . .	17
6	Αναπαράσταση Ιστογράμματος γωνιών άφιξης για ένα σύνολο πολλαπλών πηγών. Οι γωνίες στις κορυφές του ιστογράμματος, αντιστοιχούν στο πλήθος των ενεργών πηγών. . . . .	18
7	Στάδιο ανάλυσης της μεθόδου ImmACS πάνω σε μια μικροφωνική διάταξη. . . . .	20
8	Στάδιο αναπαραγωγής πλήθους ηχητικών πηγών σε μια διάταξη ηχείων με την μέθοδο ImmACS. . . . .	21
9	Ambeo Sennheizer VR . . . . .	23
10	Ambeo plugin converter . . . . .	24
11	Αναπαράσταση διαγράμματος σύλληψης και επεξεργασίας του AMBEO. . . . .	24
12	Αναπαράσταση διαγράμματος αναπαραγωγής σήματος του μικροφώνου. . . . .	25
13	Τοποθέτηση ηχείου - μικροφώνου κατά την διαδικασία βαθμονόμησης της κάρτας ήχου. . . . .	26
14	Βάρη διόρθωσης πλάτους (a) και φάσης (b) συναρτήσει της συχνότητας για την κάρτα ήχου. . . . .	27
15	Μικροφωνική συστοιχία τεσσάρων μικροφώνων σε κυκλική διάταξη . . . . .	28
16	Δομή διαφορικής συστοιχίας 2 μικροφώνων . . . . .	29
17	Διαφορική συστοιχία 1ης τάξης : (a) 4-κάναλη κυκλική συστοιχία ακτίνας $q$ , (b) διάγραμμα κατευθυντικότητας της μικροφωνικής διάταξης $\alpha_{1,1} = 0$ . . . . .	30
18	Βάρη διόρθωσης πλάτους (a) και φάσης (b) συναρτήσει της συχνότητας για το σύστημα κάρτα ήχου – συστοιχία μικροφώνων. . . . .	32
19	Κάτοψη αίθουσας διεξαγωγής πειραμάτων . . . . .	33
20	Τοποθέτηση των δύο τύπων μικροφώνων κατά τη φάση της σύλληψης του ηχητικού πεδίου . . . . .	34
21	Διάγραμμα κινούμενων πηγών στον χώρο . . . . .	35
22	Περιβάλλον λογισμικού Reaper για την μετατροπή σημάτων από A-Format σε B-Format. . . . .	35
23	Διάταξη 12 ηχείων στο Acoustic Lab σε σχήμα σφαίρας. . . . .	36

24	Διάγραμμα τοποθέτησης ηχείων: (a) μεσαίο επίπεδο, (b) κάτω επίπεδο, (c) πάνω επίπεδο. . . . .	37
25	Ηχείο Monitor Genelec 8040b . . . . .	38
26	Αποστάσεις ηχείων της διάταξης . . . . .	38
27	Ποσοστά σωστών απαντήσεων για στατικές πηγές με χρήση του Shenheizer για 12 και 4 ηχεία. . . . .	40
28	Ποσοστά σωστών απαντήσεων για στατικές πηγές αναφορικά με το ηχητικό σήμα και τον τύπο μικροφώνου που έγινε η σύλληψη και επεξεργασία. . . . .	40
29	Συνολικά ποσοστά σωστών απαντήσεων για στατικές πηγές αναφορικά με τον τύπο μικροφώνου που έγινε η σύλληψη και επεξεργασία. . . . .	41
30	Ποσοστά σωστών απαντήσεων για κινούμενες πηγές αναφορικά με το ηχητικό σήμα και τον τύπο μικροφώνου που έγινε η σύλληψη και επεξεργασία. . . . .	41
31	Ποσοστά σωστών απαντήσεων για κινούμενες πηγές αναφορικά με το ηχητικό σήμα και τον τύπο μικροφώνου που έγινε η σύλληψη και επεξεργασία. . . . .	42
32	Κατανομή των εκτιμώμενων γωνιών άφιξης με χρήση της κυκλικής συστοιχίας για πηγή λευκού θορύβου στις 90°: (a) χωρίς βαθμονόμηση, (b) με βαθμονόμηση . . . . .	43
33	Κατανομή των εκτιμώμενων γωνιών άφιξης με χρήση του Shenheizer για πηγή λευκού θορύβου στις 90°. Οι εκτιμήσεις φαίνονται για ανάλυση μέχρι τα 6kHz στα (a) και (b) και για ανάλυση μέχρι τη συχνότητα Nyquist στα (c) και (d). Οι εκτιμήσεις κατά το αζιμούθιο απεικονίζονται στα (a) και (c) και η γωνία ανύψωσης στα (b) και (d). . . . .	44
34	Κατανομή εκτιμώμενων γωνιών άφιξης με την χρήση: (a) Shenheizer, (b) κυκλική συστοιχία . . . . .	44
35	Στερεοφωνική αναπαράσταση εικονικής πηγής με χρήση 2 ηχείων . . . . .	47
36	3D αναπαράσταση εικονικής πηγής με χρήση 3 ηχείων . . . . .	49
37	Διάταξη χρήστη και ηχητικών πηγών σε ακουστικό περιβάλλον για τον υπολογισμό HRTFs . . . . .	50

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

1	Πλατωνικά στερεά . . . . .	10
2	SN3D αρμονικές και Fuma συντελεστές για σήματα αμφιχίας 3ης τάξης	13
3	Μοντέλα μικροφώνων τα οποία αποτελούνται απο διαφορετικό πλήθος καρδιοειδών μικροφώνων. . . . .	22
4	Πολικά μοτίβα για διαφορετικές τιμές του $\alpha$ . . . . .	30

5	Τιμές των γωνιών και ακτίνας χωροθεσίας των ηχείων. . . . .	37
---	---	----

## Ευχαριστίες

Η παρούσα διπλωματική εργασία πραγματοποιήθηκε στο Τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης τον Οκτώβριο του 2019.

Έχοντας ολοκληρώσει την διπλωματική μου εργασία θα ήθελα να ευχαριστήσω θερμά:

Τον κ. Νίκο Στεφανάκη, Επίκουρο Καθηγητή του Ελληνικού Μεσογειακού Πανεπιστημίου, για την επίβλεψη της παρούσας εργασίας καθώς παρακολουθούσε την υλοποίηση της εργασίας παρεμβαίνοντας καθοριστικά με συμβουλές και προτάσεις.

Τον κ. Αθανάσιο Μουχτάρη, Αναπληρωτή Καθηγητή του Πανεπιστημίου Κρήτης που με εισήγαγε στο αντικείμενο της επεξεργασίας ήχου.

Για την πραγματοποίηση θα ήθελα να ευχαριστήσω θερμά όλους όσους συνέλαβαν και βοήθησαν σε όλη την διαδικασία.

Τέλος, θα ήθελα να απευθύνω τις ευχαριστίες μου στους γονείς μου καθώς και όλους τους φίλους μου που με στήριζαν το χρονικό διάστημα αυτό.



## Περίληψη

Στην παρούσα μεταπτυχιακή εργασία παρουσιάζονται και αξιολογούνται διαφορετικές προσεγγίσεις για την υλοποίηση της τεχνικής της Αμφιηχίας (Ambisonics), που είναι από τις πιο εδραιωμένες τεχνικές για τη σύλληψη και αναπαραγωγή ηχητικού πεδίου. Παρουσιάζεται το θεωρητικό υπόβαθρο για ανάλυση του ηχητικού πεδίου με βάση τις σφαιρικές αρμονικές πρώτης τάξης καθώς και για την ανασύνθεσή του με χρήση συστοιχίας ηχείων. Για τη σύλληψη του ηχητικού πεδίου χρησιμοποιήθηκαν δύο διαφορετικοί αισθητήρες: μια τετράεδρη διάταξη καρδιοειδών μικρόφωνων η οποία είναι και διαθέσιμη στο εμπόριο, και μια αυτοσχέδια διάταξη πανκατευθυντικών μικροφώνων κυκλικής γεωμετρίας. Για την αναπαραγωγή στήθηκε μια τρισδιάστατη διάταξη 12 ηχείων, σχεδιασμένη ώστε να αξιοποιεί τις ήδη υπάρχουσες υποδομές του Εργαστηρίου Ακουστικής του Ινστιτούτου Πληροφορικής του ΙΤΕ. Στην εργασία περιγράφονται οι τεχνικές που αναπτύχθηκαν για την υλοποίηση της Αμφιηχίας με βάση των υπάρχοντα εξοπλισμό και παρουσιάζονται τα αποτελέσματα από υποκειμενικά τεστ που έγιναν με σκοπό την επιβεβαίωση της ορθής λειτουργίας του συστήματος, αλλά και τη σύγκριση των δύο διαφορετικών τύπων αισθητήρων

## Abstract

This Master Thesis presents different approaches for implementing Ambisonics, one the most well established techniques for capturing and reproducing a sound field. The Thesis presents the basic theory behind first-order spherical harmonics for 3-dimensional sound field analysis and for sound field synthesis using a loudspeaker array. For sound field capturing, two different sensing devices were used: a commercially available tetrahedral microphone and a custom-made circular array of omnidirectional microphones. For reproduction, a three-dimensional configuration of 12 loudspeakers was built, designed in accordance to the specifications of the Acoustic Lab of the Institute of Computer Science of FORTH. The Thesis presents the techniques that were used for implementing Ambisonics in accordance with the limitations of the available equipment and also presents the results derived from listening tests that were made with the goal to verify the correctness of implementation and to compare the two different sensing devices.



## 1 Εισαγωγή

Ο χωρικός ήχος (spatial audio) είναι ένα επιστημονικό πεδίο που συνδυάζει γνώσεις από διάφορους τομείς όπως η ακουστική, η ηλεκτρακουστική, η ψυχοακουστική και η επεξεργασία σήματος. Είναι ένα ενεργό ερευνητικό πεδίο με εφαρμογές που εστιάζουν στη βελτίωση της εμπειρίας ακρόασης, μέσω της σύλληψη, αποθήκευσης, μετάδοσης και αναπαραγωγή χωρικής πληροφορίας που συνθέτει ένα πραγματικό ή συνθετικό ακουστικό γεγονός. Στις μέρες μας, παρατηρείται μια άνθηση όσον αφορά την εμπορευματοποίηση της τεχνολογίας που σχετίζεται με τον χωρικό ήχο. Παράγοντες που συντελούν σε αυτό είναι κατ' αρχάς η ολοένα και μεγαλύτερη διείσδυση της σχετικής τεχνολογίας σε εφαρμογές και υπηρεσίες ψυχαγωγίας όπως είναι ο κινηματογράφος, η μουσική, ή τηλεόραση και το ίντερνετ – εφαρμογές επικοινωνίας (πχ. τηλεδιάσκεψη) καθώς και η διείσδυση του ψηφιακού ήχου και των **home theatre** συστημάτων στο οικιακό περιβάλλον. Επιπρόσθετος λόγος για αυτήν την άνθηση είναι οι εξελίξεις στην τεχνολογία και τις εφαρμογές που σχετίζονται με την εικονική πραγματικότητα (virtual reality) και το 360 βίντεο, καθώς είναι προφανές ότι η εμπειρία του χρήστη βελτιώνεται σημαντικά όταν μαζί με το οπτικό περιεχόμενο υπάρχει και το αντίστοιχο ακουστικό ερέθισμα. Ένα παράδειγμα που καταδεικνύει την άνθηση των τεχνολογιών σχετικών με το χωρικό ήχο στις μέρες μας είναι η εμφάνιση στην αγορά πολλαπλών μικροφωνικών διατάξεων για τη σύλληψη 3D ήχου, καθώς και αντίστοιχου software (plugins) για την αποκωδικοποίηση της πολυκάναλης ηχογράφησης και για αναπαραγωγή από ηχεία ή από ακουστικά.

Η Αμφιηχία αποτελεί μια από τις πιο παλιές θεωρητικές προσεγγίσεις για σύλληψη και αναπαραγωγή τρισδιάστατου ήχου. Πριν ακόμα την ψηφιακή επανάσταση, όταν τα ηχητικά σήματα και συστήματα ήταν στην πλειοψηφία τους ακόμα αναλογικά, η Αμφιηχία αποτέλεσε την πρώτη λύση για κωδικοποίηση της χωρικής πληροφορίας μιας πραγματικής ηχητικής σκηνής και για τη μετέπειτα αναπαραγωγή της μέσα από διατάξεις πολλαπλών ηχείων. Επιπλέον, έθεσε τη βάση για τη μετέπειτα εξέλιξη της μεθόδου **Directional Audio Coding (DirAC)**, η οποία με τη σειρά της αποτέλεσε έναυσμα για την εμφάνιση πολλών άλλων παραμετρικών τεχνικών χωρικού ήχου.

Στο πεδίο των παραμετρικών τεχνικών χωρικού ήχου, το Εργαστήριο Επεξεργασίας Σήματος (Signal Processing Lab – SPL) του Ινστιτούτου Πληροφορικής του ΙΤΕ έχει σημαντική συνεισφορά. Μέσα από την πτυχιακή αυτή εργασία ωστόσο, είναι η πρώτη φορά που το εργαστήριο καταπιάνεται με τη μέθοδο της Αμφιηχίας. Αφορμή για την ενασχόληση με αυτή τη μέθοδο μέσα από αυτήν την πτυχιακή εργασία είναι

1. η πρόσφατη απόκτηση της μικροφωνικής διάταξης **AMBEO VR**, που αποτελεί προϊόν της εταιρείας **Shennheizer**, και είναι ιδανική για εφαρμογή της Αμφιηχίας

καθώς και για συγγενικές μεθόδους,

2. η πρόσφατη αναβάθμιση του εξοπλισμού στο Εργαστήριο Ακουστικής του SPL που επέτρεψε την τοποθέτηση ηχείων σε σφαιρική διάταξη, πάνω και κάτω από τον ακροατή, ώστε να είναι εφικτή και η αναπαραγωγή τρισδιάστατου ηχητικού πεδίου,
3. η επιθυμία για την περαιτέρω αξιοποίηση των μικροφωνικών διατάξεων κυκλικής γεωμετρίας που το SPL έχει ήδη στη διάθεσή του, εφαρμόζοντας πάνω σε αυτές μεθόδους ηχητικής σύλληψης που δεν είχαν δοκιμαστεί από το εργαστήριο στο παρελθόν,
4. η συνεργασία με επιστημονική ομάδα από το Εθνικό Μετσόβιο Πολυτεχνείο για έρευνα πάνω στην ακουστικοποίηση (auralization) κλειστών χώρων με χρήση της μεθόδου της Αμφιηχίας.

Η δομή της πτυχιακής είναι η παρακάτω:

Κεφάλαιο 2 : Γίνεται παρουσίαση της μεθόδου της Αμφιηχίας. Περιγράφεται η μέθοδος κωδικοποίησης και αποκωδικοποίησης 1ης τάξης για αναπαραγωγή με διατάξεις ηχείων που είναι τοποθετημένα πάνω σε κανονικά πολύγωνα. Επιπλέον γίνεται μια σύντομη περιγραφή της μεθόδου για υλοποίηση με υψηλότερες τάξεις.

Κεφάλαιο 3 : Σε αυτό το κεφάλαιο παρουσιάζονται δύο παραμετρικές τεχνικές απεικόνισης του ηχητικού πεδίου, η μέθοδος DirAC και η μέθοδος ImmACS.

Κεφάλαιο 4 : Περιγράφεται αναλυτικά ο τρόπος που έγινε η υλοποίηση της Αμφιηχίας στο Εργαστήριο Ακουστικής του SPL τόσο με την χρήση του Shennheizer AMBEO VR όσο και με χρήση κυκλική συστοιχίας μικροφώνων που το εργαστήριο είχε ήδη στη διάθεσή του. Επιπλέον περιγράφονται οι τεχνικές βαθμονόμησης που ακολουθήθηκαν αναφορικά και με τους δύο ακουστικούς αισθητήρες για βελτίωση της διεργασίας σύλληψης του ηχητικού πεδίου.

Κεφάλαιο 5 : Περιγράφεται η διαδικασία που ακολουθήθηκε για την αναπαραγωγή από σφαιρική διάταξη 12 ηχείων καθώς και ο τρόπος που δημιουργήθηκαν τα δεδομένα για την αξιολόγηση του συστήματος. Τέλος δίνεται περιγραφή των αποτελεσμάτων της υποκειμενικής αξιολόγησης (τεστ ακρόασης) καθώς και της αντικειμενικής αξιολόγησης του συστήματος

Κεφάλαιο 6 : Παρουσιάζονται τα συμπεράσματα που προέκυψαν από την όλη εργασία.



## 2 Αμφιηχία

Η αμφιηχία έχει τις ρίζες της στις τεχνικές εγγραφής των τετραφωνικών συστημάτων (quadraphony) και αργότερα προτάθηκε ως διάδοχος της στερεοφωνίας (Gerzon 1985, Lipshitz 1986). Η αμφιηχία ουσιαστικά στοχεύει στην υλοποίηση μιας τεχνικής γνώστης ως θεωρία ανακατασκευής μετώπου κύματος. Αυτό γίνεται αν λάβουμε υπόψιν την κατευθυντική πληροφορία ενός ηχητικού πεδίου το οποίο θέλουμε να ανακατασκευάσουμε. Σε συστήματα αναπαραγωγής (stereo, 5.1 η 7.1 surround), κάθε ηχογραφημένο σήμα αντιστοιχεί σε ένα ηχείο αναπαραγωγής. Στην μέθοδο της αμφιηχίας, κάθε κανάλι περιλαμβάνει επιπλέον πληροφορίες για τις ιδιότητες ενός ακουστικού πεδίου όπως είναι οι αυξομειώσεις και η κατευθυντικότητα του. Διάφορα στάδια υλοποίησης της στην θεωρία τόσο στην σύλληψή όσο και στην αναπαραγωγή περιλαμβάνονται στην μέθοδο αυτή.

1. Μηδενική-τάξη αμφιηχίας, περιλαμβάνει πληροφορία για τις αυξομειώσεις του ακουστικού πεδίου (ηχογράφηση με την χρήση ενός πανκατευθυντικού μικροφώνου). Το κανάλι αυτό συμβολίζεται ως  $W$ .
2. Πρώτη-τάξη αμφιηχίας, περιλαμβάνει πληροφορία για την προέλευση της κατεύθυνσης του ακουστικού πεδίου (ηχογράφηση με την χρήση τριών δικτευθυντικών μικροφώνων τοποθετημένα σε κάθε άξονα). Τα κανάλια αυτά συμβολίζονται με  $X$ ,  $Y$ ,  $Z$  και περιλαμβάνουν την πληροφορία της κατεύθυνσης του ηχητικού πεδίου πάνω σε κάθε άξονα
3. Υψηλότερες τάξεις αμφιηχίας, περιλαμβάνουν ακριβέστερη πληροφορία σχετικά με την κατευθυντικότητα ενός ακουστικού πεδίου συγκριτικά με χαμηλότερες τάξεις.

Στο παρόν κεφάλαιο θα δώσουμε περισσότερη έμφαση στην σύλληψή καθώς και την αναπαραγωγή  $1^{η}$  τάξης με αμφιηχία.

### 2.1 Κωδικοποίηση $1^{η}$ -τάξης Αμφιηχίας

Η αναπαράσταση ενός ηχητικού πεδίου με την μέθοδο της αμφιηχίας σε πρώτο στάδιο, απαιτεί την σύλληψή ενός ακουστικού γεγονότος με την χρήση πολλαπλών μικροφώνων τοποθετημένα στο χώρο. Για την καλύτερη κατανόηση της αμφιηχίας ας υποθέσουμε ότι το επιθυμητό ηχητικό πεδίο είναι αυτό ενός επίπεδου κύματος πλάτους  $A$  και κατεύθυνσης  $\mathbf{k}$ .

Η αυξομείωση της πίεσης μιας ηχητικής πηγής  $s(t)$  και κατεύθυνσης  $\mathbf{k} = k_x \mathbf{x} + k_y \mathbf{y} + k_z \mathbf{z}$ , όπου  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\mathbf{z}$  είναι τα μοναδιαία διανύσματα κατά τους άξονες  $-x$ ,  $-y$  και  $-z$  μπορεί να διατυπωθεί ως :

$$p(t, \mathbf{r}) = s\left(t + \mathbf{k} \cdot \frac{\mathbf{r}}{c}\right), \quad (1)$$

όπου  $c$  η ταχύτητα διάδοσης της ηχητικής πηγής και  $\mathbf{r}$  το διάνυσμα θέσης του μικροφώνου.

Για παράδειγμα, αν έχουμε ακουστική πηγή ημιτονοειδές σήμα  $s(t) = A \sin(\omega t)$  πλάτους  $A$  και γωνιακής συχνότητας  $\omega$ , τότε σύμφωνα με την παραπάνω σχέση η πίεση της ηχητικής πηγής θα πάρει την μορφή :

$$p(t, \mathbf{r}) = A \sin\left(\omega t + \omega \mathbf{k} \cdot \frac{\mathbf{r}}{c}\right), \quad (2)$$

Επιπλέον, το διάνυσμα ταχύτητας  $\mathbf{u}$  της ηχητικής πηγής είναι :

$$\mathbf{u}(t, \mathbf{r}) = (1/Z)s\left(t + \mathbf{k} \cdot \frac{\mathbf{r}}{c}\right)\mathbf{k}, \quad (3)$$

με  $Z = \rho_0 c$  να αποτελεί την ακουστική εμπέδηση με  $\rho_0$  η πυκνότητα του αέρα. Ακόμα το διάνυσμα της ταχύτητας μπορεί να αναλυθεί στις παρακάτω συνιστώσες:

$$u_x(t, \mathbf{r}) = (1/Z)s\left(t + \mathbf{k} \cdot \frac{\mathbf{r}}{c}\right) \cos \theta \cos \phi, \quad (4a)$$

$$u_y(t, \mathbf{r}) = (1/Z)s\left(t + \mathbf{k} \cdot \frac{\mathbf{r}}{c}\right) \sin \theta \cos \phi, \quad (4b)$$

$$u_z(t, \mathbf{r}) = (1/Z)s\left(t + \mathbf{k} \cdot \frac{\mathbf{r}}{c}\right) \sin \phi. \quad (4c)$$

με  $\theta$  να αποτελεί την γωνία του αζιμούθιου και  $\phi$  την γωνία ανύψωσης. Το διάνυσμα αυτό αποτελεί την κλίση της συνάρτησης πίεσης  $p(t, \mathbf{r})$ .

Με την χρήση ειδικών μικροφώνων μπορούμε να υπολογίσουμε απευθείας τις αυξομειώσεις καθώς και την κατεύθυνση πρόσπτωσης ενός ηχητικού σήματος ( σχέσεις (1), (3) ). Οι αυξομειώσεις της πίεσης  $p(t, \mathbf{r})$  σχετίζεται με την ηχογράφηση ενός πανκατευθυντικού μικροφώνου. Η ιδέα της κωδικοποίησης με την χρήση αμφιηχίας υλοποιείται χρησιμοποιώντας τέσσερα διαφορετικά κανάλια τα οποία ορίζονται ως :

$$W(t) = \frac{s(t)}{\sqrt{2}} \quad (5a)$$

$$X(t) = s(t) \cos \theta \cos \phi \quad (5b)$$

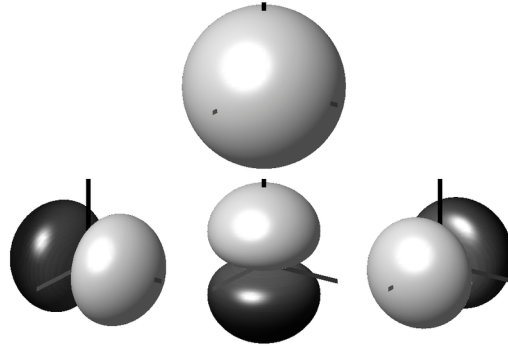
$$Y(t) = s(t) \sin \theta \cos \phi \quad (5c)$$

$$Z(t) = s(t) \sin \phi \quad (5d)$$

όπου  $s(t)$  αποτελεί ένα μονοφωνικό σήμα πηγής και  $\theta, \phi$  είναι η αζιμουθιανή γωνία και η γωνία ανύψωσης της πηγής.

Το κανάλι  $W(t)$  περιέχει πληροφορία σχετικά με την απευθείας μετάδοσή της ηχητικής

πηγής από οποιαδήποτε κατεύθυνση του χώρου, ενώ τα κανάλια  $X(t)$ ,  $Y(t)$  και  $Z(t)$  εμπεριέχεται η πληροφορία της κατεύθυνσης του ηχητικού σήματος πάνω στους άξονες που έχουν τοποθετηθεί τα μικρόφωνα  $x$ ,  $y$ ,  $z$ . Το σύνολο των καναλιών  $W, X, Y, Z$  συμβάλουν στην κωδικοποίηση πρώτης τάξης αμφιηχίας [2] για την ανακατασκευή ενός ηχητικού γεγονότος και η μέθοδος αυτή ονομάζεται *B – Format*.



Εικόνα 1: Σχηματική αναπαράσταση πολικών συντεταγμένων ενός πανκατευθυντικού μικροφώνου (πάνω κομμάτι) και τριών διχατευθυντικών ( άξονες  $x$ ,  $z$ ,  $y$  ) στον τρισδιάστατο χώρο. Πηγή [1]

Επιπρόσθετα, για ένα σύνολο πολλαπλών ηχητικών πηγών στον χώρο  $s_i$  κατεύθυνσης  $(\theta_i, \delta_i)$ , τα διαφορετικά κανάλια της αμφιηχίας μπορούν να πάρουν την μορφή [3]:

$$W(t) = \sum_i^N \frac{s_i(t)}{\sqrt{2}}, \quad (6a)$$

$$X(t) = \sum_i^N s_i(t) \cos \theta_i \cos \phi_i, \quad (6b)$$

$$Y(t) = \sum_i^N s_i(t) \sin \theta_i \cos \phi_i, \quad (6c)$$

$$Z(t) = \sum_i^N s_i(t) \sin \phi_i, \quad (6d)$$

με  $N$  να αποτελεί το σύνολο των ηχητικών πηγών και  $s_i$  το μονοφωνικό σήμα γωνίας  $(\theta_i, \phi_i)$  που θέλουμε να κωδικοποιήσουμε στις αντίστοιχες θέσεις. Τέλος πρέπει να σημειωθεί ότι οι παραπάνω εξισώσεις δείχνουν ότι για την σύνθεση αμφιηχίας απαιτείται κάθε μονοφωνικό σήμα πηγής να πολλαπλασιαστεί με την αντίστοιχη γωνία  $(\theta_i, \phi_i)$



για την χωρική αναπαράσταση του.

## 2.2 Αποκωδικοποίηση 1<sup>ης</sup>-τάξης Αμφιηχίας

Η αποκωδικοποίηση 1ης τάξης αμφιηχίας, αποτελεί το επόμενο στάδιο της διαδικασίας κωδικοποίησης για την διαδικασία αναπαραγωγής ενός ακουστικού πεδίου με ηχεία. Η έννοια 1η τάξη δηλώνει το πλήθος των καναλιών που θα γίνει η αποκωδικοποίηση σύμφωνα με την τάξη αμφιηχίας που χρησιμοποιήθηκε. Τα σήματα αναπαραγωγής της διάταξης, αποτελούν γραμμικό συνδυασμό σημάτων κωδικοποίησης με την προσθήκη επιπλέον βαρών τα οποία καθορίζουν το πόσο πίστα θα γίνει η αναπαραγωγή σε σχέση με το επιθυμητό αποτέλεσμα [3]. Επιπλέον ισχύει ο γενικός κανόνας ότι το πλήθος των ηχείων θα πρέπει να είναι μεγαλύτερο από τα κανάλια κωδικοποίησης.

Έτσι, για έναν δοθέντα αριθμό ηχείων  $L$ , η έξοδος μας εφαρμόζεται πάνω σε κάθε ηχείο  $i$  όπου το κάθε ηχητικό σήμα  $s_i$  αναπαράγεται από την αντίστοιχη κατεύθυνση προέλευσης του  $(\theta_i, \delta_i)$ . Η παραπάνω διαδικασία μπορεί να εκφραστεί ως:

$$s_i(t) = w_i W(t) + x_i X(t) + y_i Y(t) + z_i Z(t) \quad (7)$$

όπου το σύνολο των μεταβλητών  $w_i, x_i, y_i, z_i$  αποτελούν τιμές βαρών για κάθε ένα κανάλι. Οι τιμές των παραπάνω συντελεστών διαφοροποιούνται ανάλογα με τις θέσεις των ηχείων στον χώρο. Θεωρώντας ότι οι αποστάσεις των ηχείων από το κέντρο της συστοιχίας είναι ίσες, το πρώτο στάδιο για τον υπολογισμό των σημάτων εισόδου στα ηχεία είναι ο υπολογισμός των παρακάτω συντελεστών [4]:

$$w_i = \frac{\sqrt{2}g_0}{N} \quad (8a)$$

$$x_i = \frac{3g_1 \cos \theta_i \sin \phi_i}{N} \quad (8b)$$

$$y_i = \frac{3g_1 \sin \theta_i \cos \phi_i}{N} \quad (8c)$$

$$z_i = \frac{3g_1 \sin \phi_i}{N}, \quad (8d)$$

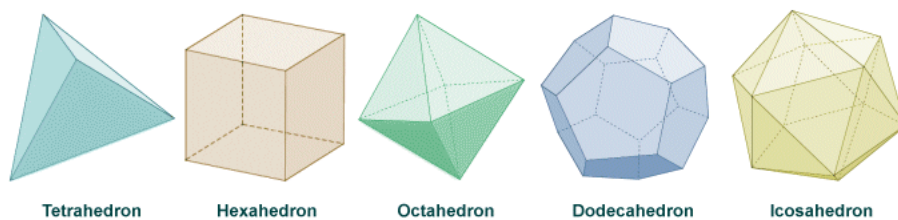
όπου  $\theta_i$  η γωνία του αζιμούθιου,  $\phi_i$  η γωνία ανύψωσης και  $N$  το πλήθος των ηχείων. Οι τιμές για τα  $g_0$  και  $g_1$  στις παραπάνω μεταβλητές, αντιστοιχούν στην τιμή ένα. Στην περίπτωση μη κανονικών διατάξεων οι τιμές αυτές διαφοροποιούνται.

Η αποκωδικοποίηση της αμφιηχίας είναι ανεξάρτητη του πλήθους των ηχείων που μπορούν να χρησιμοποιηθούν. Στην περίπτωση του B-Format που αναλύθηκε, το ελάχιστο πλήθος ηχείων που μπορούν να χρησιμοποιηθούν σε τρισδιάστατο χώρο (περιφωνία) είναι τέσσερα. Ακόμα, σε ένα οριζόντιο επίπεδο (χωρίς την χρήση του κατακόρυφου

άξονα  $Z$ ), ο ελάχιστος αριθμός των ηχείων είναι τρία. Σε υψηλότερες τάξεις αμφιηχίας απαιτείται μεγαλύτερος αριθμός ηχείων. Επιπλέον, είναι επιθυμητό για καλύτερη εικόνα του ηχητικού πεδίου η τοποθέτηση περισσότερων ηχείων έναντι του ελάχιστου αριθμού καθώς η ποιότητα του ήχου βελτιώνεται. Η συνηθέστερη τακτική για την τοποθέτηση των ηχείων στο χώρο είναι οι αποστάσεις και οι γωνίες να είναι ίσες μεταξύ τους καθώς για αυτές τις διατάξεις το σύστημα μας προσφέρει καλύτερα αποτελέσματα κατά την αναπαραγωγή. Είναι σημαντικό να σημειωθεί ότι η μέθοδος μας δίνει το επιθυμητό αποτέλεσμα όταν η θέση του ακροατή είναι στο κέντρο της διάταξης (*sweet spot*) και για συχνότητες άνω των 500 Hz. Η οποιαδήποτε μετακίνηση από το κέντρο προκαλεί φαινόμενα χωρικής αλλοίωσης κυρίως λόγω διαφοράς φάσης μεταξύ του σήματος αναπαραγωγής των ηχείων. Στον παρακάτω πίνακα παρουσιάζονται μερικές ιδιότητες των διατάξεων αυτών.

Πολύεδρο	Ακμές	Έδρες	Σχήμα
Τετράεδρο	4	4	Τριγωνικό
Εξάεδρο	8	6	κύβος
Οκτάεδρο	6	8	Τριγωνικό
Δωδεκάεδρο	20	12	Πεντάγωνο
Εικοσάεδρο	12	20	Τριγωνικό

Πίνακας 1: Πλατωνικά στερεά



Εικόνα 2: Σχηματική αναπαράσταση πλατωνικών στερεών

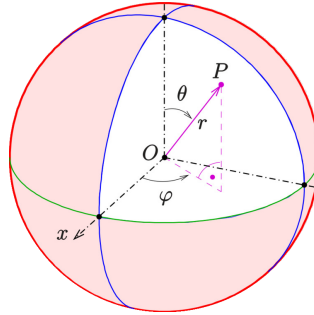
### 2.3 Υλοποίηση Με Υψηλότερες Τάξεις

Σε αυτήν την ενότητα θα γίνει μια σύντομη εισαγωγή σε υψηλότερες τάξεις αμφιηχίας. Οι τάξεις αυτές μας δίνουν καλύτερη ανάλυση της χωρικής πληροφορίας συγκριτικά με μικρότερες μεγέθους τάξεις. Ωστόσο όσο αυξάνεται το μέγεθος της τάξης τόσο μεγαλύτερο θα είναι και το πλήθος των καναλιών που θα χρησιμοποιηθούν. Αυτό απαιτεί συνολικά  $(N + 1)^2$  αριθμό καναλιών και για κάθε τάξη  $2N + 1$  κανάλια, όπου

$N$  ο αριθμός τάξης. Η αναπαράσταση αυτή γίνεται με την χρήση σφαιρικών αρμονικών, καθώς κάθε ηχητική πηγή μπορεί να περιγραφεί ως άθροισμα επίπεδων κυμάτων.

### 2.3.1 Σφαιρικές Αρμονικές

Ένας τρόπος για την αναπαράσταση ενός ακουστικού πεδίου, είναι η ανάλυση του με την χρήση σφαιρικών αρμονικών (Spherical Harmonics). Σε αυτή την περίπτωση, η ανάλυση των ηχητικών πηγών ως εξισώσεις κυμάτων γίνεται με την χρήση σφαιρικών συντεταγμένων  $(r, \theta, \phi)$ , όπου  $r$  η ακτίνα,  $\theta$  η γωνία πάνω στον οριζόντιο άξονα και  $\phi$  η γωνία του κατακόρυφου άξονα (σχήμα 3).



Εικόνα 3: Σύστημα συντεταγμένων πάνω σε σφαίρα

Η κυματική εξίσωση στο πεδίο του χρόνου ορίζεται ως :

$$\nabla^2 p(r, \theta, \phi, t) - \frac{1}{c^2} \frac{\partial^2 p(r, \theta, \phi, t)}{\partial t^2} = 0, \quad (9)$$

όπου  $c$  η ταχύτητα διάδοσης του ήχου και  $p(r, \theta, \phi, t)$  η συνάρτηση πεδίου.

Η λύση της παραπάνω εξίσωσης μπορεί να εκφραστεί ως άθροισμα επίπεδων κυμάτων σε οποιοδήποτε ακουστικό πεδίο. Ένα επίπεδο κύμα που μεταδίδεται κατά μία γωνία, μπορεί να αναλυθεί ως σειρά **Fourier-Bessel** με βάρη κατευθυνόμενες συναρτήσεις  $Y_{mn}^\sigma(\theta, \phi)$  οι οποίες καλούνται σφαιρικές αρμονικές και παραμετρικές συναρτήσεις **Bessel** πρώτου είδους  $J_m(kr)$  [5]. Έτσι θα έχουμε :

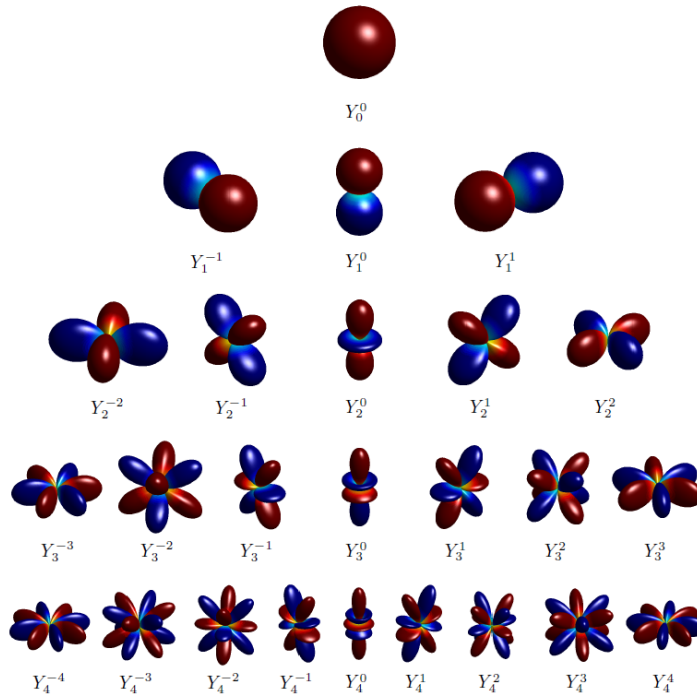
$$p(\vec{r}) = \sum_{m=0}^{\infty} (2m+1) j^m j_m(kr) \sum_{0 \leq n \leq \sigma = \pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \phi) \quad (10)$$

με  $m$  να αποτελεί τον βαθμό,  $n$  την τάξη,  $\sigma$  την ιδιοπεριστροφή (spin) και  $k = \frac{2\pi f}{c}$  το μήκος κύματος. Επιπλέον η μεταβλητή  $B_{mn}^\sigma$  αποτελεί όρους των σφαιρικών αρμονικών.

Οι σφαιρικές αρμονικές δίνονται από την παρακάτω εξίσωση :

$$Y_{mn}^{\sigma}(\theta, \phi) = N_{mn} P_{m|n|}(\sin \phi) \times \begin{cases} \cos(|n|\theta) & \text{αν } \sigma = +1 \\ \sin(|n|\theta) & \text{αν } \sigma = -1 \end{cases} \quad (11)$$

όπου  $P_{mn}$  αποτελεί πολυώνυμο Legendre και  $N_{mn}$  παράγοντας κανονικοποίησης.



Εικόνα 4: Σχηματική αναπαράσταση σφαιρικών αρμονικών  $n = 4$  τάξεως στον 3-διάστατο χώρο

### 2.3.2 Κωδικοποίηση Υψηλότερης Τάξης

Για την σύλληψη ενός ηχητικού σήματος η κωδικοποίηση με τη χρήση της μεθόδου **B-Format** μπορεί να υλοποιηθεί και σε υψηλότερες τάξεις ( $n > 1$ ). Η ακρίβεια της αναπαράστασης ενός ηχητικού γεγονότος, αυξάνεται όσο αυξάνεται και ο αριθμός της τάξης. Ο παρακάτω Πίνακας 2 περιλαμβάνει τους συντελεστές κανονικοποίησης **Furse-Malham** καθώς και τη μορφή καναλιών κωδικοποίησης για τάξεις ανώτερες του ένα.

τάξη	$B_{mn}^\sigma$	$\begin{pmatrix} \sigma \\ mn \end{pmatrix}$	$Y_{mn}^{\sigma(SN3D)}(\theta, \phi)$	συντελεστές Fuma
0	$W$	$\begin{pmatrix} 1 \\ 00 \end{pmatrix}$	1	$\frac{1}{\sqrt{2}}$
1	$X$	$\begin{pmatrix} 1 \\ 11 \end{pmatrix}$	$\cos \theta \cos \phi$	1
	$Y$	$\begin{pmatrix} -1 \\ 00 \end{pmatrix}$	$\sin \theta \cos \phi$	1
	$Z$	$\begin{pmatrix} 1 \\ 10 \end{pmatrix}$	$\sin \phi$	1
2	$R$	$\begin{pmatrix} 1 \\ 20 \end{pmatrix}$	$(3 \sin^2 \phi - 1)/2$	1
	$S$	$\begin{pmatrix} 1 \\ 21 \end{pmatrix}$	$(\frac{\sqrt{3}}{2}) \cos \theta \sin (2\phi)$	$\frac{2}{\sqrt{3}}$
	$T$	$\begin{pmatrix} -1 \\ 21 \end{pmatrix}$	$(\frac{\sqrt{3}}{2}) \sin \theta \sin (2\phi)$	$\frac{2}{\sqrt{3}}$
	$U$	$\begin{pmatrix} 1 \\ 22 \end{pmatrix}$	$(\frac{\sqrt{3}}{2}) \cos (2\theta) \cos^2 \phi$	$\frac{2}{\sqrt{3}}$
	$V$	$\begin{pmatrix} -1 \\ 22 \end{pmatrix}$	$(\frac{\sqrt{3}}{2}) \sin (2\theta) \cos^2 \phi$	$\frac{2}{\sqrt{3}}$
	$K$	$\begin{pmatrix} 1 \\ 30 \end{pmatrix}$	$\sin \phi (5 \sin^2 \phi - 3)/2$	1
3	$L$	$\begin{pmatrix} 1 \\ 31 \end{pmatrix}$	$(\frac{\sqrt{3}}{8}) \cos \theta \cos \phi (5 \sin^2 \phi - 1)$	$\frac{\sqrt{45}}{32}$
	$M$	$\begin{pmatrix} -1 \\ 31 \end{pmatrix}$	$(\frac{\sqrt{3}}{8}) \sin \theta \cos \phi (5 \sin^2 \phi - 1)$	$\frac{\sqrt{45}}{32}$
	$N$	$\begin{pmatrix} 1 \\ 32 \end{pmatrix}$	$(\frac{\sqrt{15}}{2}) \cos (2\theta) \sin \phi \cos^2 \phi$	$\frac{3}{\sqrt{5}}$
	$O$	$\begin{pmatrix} -1 \\ 32 \end{pmatrix}$	$(\frac{\sqrt{15}}{2}) \sin (2\theta) \sin \phi \cos^2 \phi$	$\frac{3}{\sqrt{5}}$
	$P$	$\begin{pmatrix} 1 \\ 33 \end{pmatrix}$	$(\frac{\sqrt{5}}{8}) \cos (3\theta) \cos^3 \phi$	$\frac{\sqrt{8}}{5}$
	$Q$	$\begin{pmatrix} -1 \\ 33 \end{pmatrix}$	$(\frac{\sqrt{5}}{8}) \sin (3\theta) \cos^3 \phi$	$\frac{\sqrt{8}}{5}$

Πίνακας 2: SN3D αρμονικές και Fuma συντελεστές για σήματα αμφιηχίας 3ης τάξης

Σύμφωνα με τα παραπάνω, οι συναρτήσεις 2ης τάξης κωδικοποίησης για ένα επίπεδο κύμα μπορούν να περιγραφούν ως:

$$\begin{cases} R^{(SN3D)} = S \cdot Y_{20}^{-1 (SN3D)}(\theta, \phi) = S(3 \sin^2(\phi) - 1)/2 \\ S^{(SN3D)} = S \cdot Y_{21}^{-1 (SN3D)}(\theta, \phi) = S \frac{\sqrt{3}}{2} \cos \theta \sin 2\phi \\ T^{(SN3D)} = S \cdot Y_{21}^{-1 (SN3D)}(\theta, \phi) = S \frac{\sqrt{3}}{2} \sin \theta \sin 2\phi \\ U^{(SN3D)} = S \cdot Y_{22}^{-1 (SN3D)}(\theta, \phi) = S \frac{\sqrt{3}}{2} \cos 2\theta \cos^2 \phi \\ V^{(SN3D)} = S \cdot Y_{22}^{-1 (SN3D)}(\theta, \phi) = S \frac{\sqrt{3}}{2} \sin 2\theta \cos^2 \phi \end{cases} \quad (12)$$

όπου  $S$  το προσπίπτον σήμα της πηγής. Με όμοιο τρόπο προκύπτουν οι συναρτήσεις κωδικοποίησης ανώτερης τάξης αμφιηχίας.

Ακόμα, οι τιμές των συντελεστών αυξάνονται ανάλογα με την τάξη κάτι το οποίο δημιουργεί προβλήματα στο δυναμικό εύρος μεταξύ των μικροφώνων της διάταξης. Σε διαφορετική περίπτωση όπου η πηγή βρίσκεται πολύ κοντά στην διάταξη των μικροφώνων, ένας επιπλέον παράγοντας θα πρέπει να ληφθεί υπόψιν καθώς τα ηχητικά κύματα δεν έχουν πλέον την μορφή επίπεδων κυμάτων. Έτσι, το ηχητικό πεδίο εξαρτάται από την συχνότητα των επιμέρους πηγών.

### 3 Παραμετρικές Μέθοδοι

Αξίζει σε αυτήν την ενότητα να γίνει αναφορά σε τεχνικές σύλληψης και αναπαραγωγής του ηχητικού πεδίου που είναι μεταγενέστερες της Αμφιηχίας. Οι τεχνικές αυτές ονομάζονται παραμετρικές με την έννοια ότι τα σήματα στα κανάλια αναπαραγωγής προκύπτουν συναρτήσει παραμέτρων που μεταβάλλονται δυναμικά και οι οποίοι προκύπτουν από διάφορα στάδια ανάλυσης του ηχητικού πεδίου. Τέτοιες παράμετροι όπως θα δούμε είναι η κατεύθυνση άφιξης, ο συντελεστής διάχυσης και ο αριθμός των ηχητικών πηγών. Σε τελικό στάδιο η αναπαραγωγή επιτυγχάνεται με την χρήση των παραπάνω παραμέτρων σε συνδυασμό με τεχνικές δημιουργίας ηχητικών ειδώλων (VBAP). Σε αυτήν την ενότητα θα αναλυθεί η διαδικασία αναπαράστασης ενός ακουστικού γεγονότος για ένα ηχητικό πεδίο με την χρήση των μεθόδων Directional Audio coding (DirAC) και Immersive Audio Communication System (ImmACS).

#### 3.1 Σύλληψη και Αναπαραγωγή με την Μέθοδο DirAC

##### 3.1.1 Εκτίμηση Κατεύθυνσης Άφιξης

Η μέθοδος DirAC, όπως και η αμφιηχία, βασίζεται σε σφαιρικές αρμονικές πρώτης τάξης, δηλαδή στα τέσσερα κανάλια  $W, X, Y$ , και  $Z$  [6]. Ένας τρόπος για να γίνει η ανάλυση στο πεδίο χρόνου συχνότητας είναι ο βραχύς μετασχηματισμός Fourier (STFT). Για ένα οποιοδήποτε κανάλι μικροφώνου, εφαρμόζουμε μετασχηματισμό Fourier (STFT) με την ταυτόχρονη χρήση παραθύρου (Hamming Window) για ένα χρονικό διάστημα π.χ 20 ms. Η αναπαράσταση του προσπίπτον ηχητικού σήματος, ορίζεται στο πεδίο της συχνότητας ως  $X(\tau, k)$  με  $\tau$  να αποτελεί τον δείκτη χρόνου και  $k$  ο δείκτης τιμών της συχνότητας

Το κωδικοποιημένο σήμα  $W$  μπορεί να γραφτεί και στο πεδίο της συχνότητας ως  $W(\tau, k)$ . Επιπλέον για τα υπόλοιπα κανάλια θα έχουμε:

$$\mathbf{V}(\tau, k) = [X(\tau, k), Y(\tau, k), Z(\tau, k)]^T, \quad (13)$$

τα οποία αποτελούν στοιχεία μετασχηματισμού STFT. Με την χρήση αυτής της τεχνικής, μπορούμε να ορίσουμε το διάνυσμα της έντασης ως :

$$\mathbf{I}(\tau, k) = \frac{1}{\sqrt{2\rho_0 c^2}} \text{Re}\{W^*(\tau, k)\mathbf{V}(\tau, k)\}, \quad (14)$$

όπου  $\rho_0$  η μέση τιμή πίεσης μιας επιφάνειας και  $c$  η ταχύτητα διάδοσης του ήχου,  $(\cdot)^*$  μιγαδικός αριθμός και  $\text{Re}\{\cdot\}$  το πραγματικό μέρος ενός μιγαδικού αριθμού. Τέλος το διάνυσμα της κατεύθυνσης του ηχητικού σήματος ορίζεται ως η αντίθετη ποσότητα του

διανύσματος της έντασης:

$$\mathbf{D}(\tau, k) = -\mathbf{I}(\tau, k). \quad (15)$$

Κατά αυτόν τον τρόπο, το διάνυσμα της κατεύθυνσης μας δίνει την πληροφορία που χρειαζόμαστε για την αναπαραγωγή της κατευθυντικής συνιστώσας καθώς η υλοποίηση του μπορεί να γίνει με την χρήση του διανύσματος  $\mathbf{p}$  που ορίζεται στην μέθοδο VBAP [7].

Παρακάτω μπορούμε να εκτιμήσουμε απευθείας την αζιμούθια και τη γωνία ανύψωσης  $\theta$  και  $\phi$  αντίστοιχα. Η γωνία του αζιμούθιου της κατεύθυνσης άφιξης μπορεί να υπολογιστεί ως :

$$\theta(\tau, k) = \arctan \frac{\operatorname{Re}\{W(\tau, k)Y(\tau, k)\}}{\operatorname{Re}\{W(\tau, k)X(\tau, k)\}} \quad (16)$$

και η γωνία ανύψωσης ως:

$$\phi(\tau, k) = \arccos \frac{\operatorname{Re}\{W(\tau, k)Z(\tau, k)\}}{\operatorname{Re}\{W(\tau, k)\sqrt{X(\tau, k)^2 + Y(\tau, k)^2 + Z(\tau, k)^2}\}} \quad (17)$$

### 3.1.2 Υπολογισμός Συντελεστή Διάχυσης

Ο συντελεστής διάχυσης είναι μια θετική παράμετρος μεταξύ 0 και 1 η οποία εκφράζει τη σχέση μεταξύ απευθείας και διάχυτου ήχου. Ο υπολογισμός γίνεται με την χρήση της ενέργειας του προσπίπτον σήματος σε συνδυασμό με την κατεύθυνση του [8]. Συγκριμένα η ενέργεια στο πεδίο της συχνότητας που αναλύουμε δίνεται από :

$$E(\tau, k) = \frac{1}{2\rho_0 c^2} \left[ |W(\tau, k)|^2 + \frac{1}{2} \|\mathbf{V}(\tau, k)\|^2 \right]. \quad (18)$$

Τα σήματα των μικροφώνων υποθέτουμε ότι είναι κανονικοποιημένα. Ο συντελεστής διάχυσης του ηχητικού πεδίου εν τέλει μπορεί να υπολογιστεί από την παρακάτω εξίσωση:

$$\Psi(\tau, k) = 1 - \frac{\|E(\mathbf{I}(\tau, k))\|}{cE\{E(\tau, k)\}}, \quad (19)$$

όπου  $E\{\cdot\}$  η προσδόκιμη τιμή. Από τις σχέσεις (18) και (14) ο συντελεστής διάχυσης μπορεί να εκφραστεί και ως:

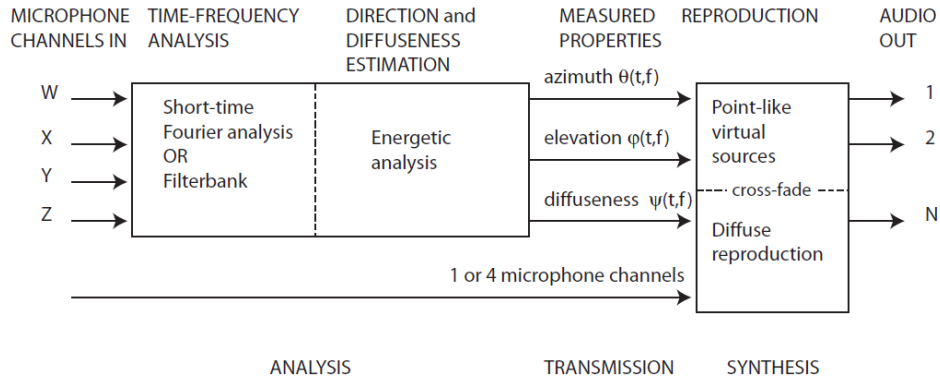
$$\Psi(\tau, k) = 1 - \frac{\operatorname{Re}\{W^*(\tau, k)\mathbf{V}(\tau, k)\}}{E\{|W(\tau, k)|^2 + \|\mathbf{V}(\tau, k)\|^2/2\}}. \quad (20)$$

### 3.1.3 Αναπαραγωγή

Όπως έχει ήδη αναφερθεί, οι δύο παράμετροι που εμπλέκονται στη μέθοδο DirAC είναι το διάνυσμα κατεύθυνσης του ήχου και ο συντελεστής διάχυσης του ηχητικού πεδίου,



τα οποία προκύπτουν ανεξάρτητα από το πλήθος των πηγών που μπορεί να περιέχονται σε αυτό.



Εικόνα 5: Διάγραμμα μεθόδου *DirAC* για την ανακατασκευή ηχητικού γεγονότος σε μια διάταξη ηχείων

Κατά την διαδικασία αναπαραγωγής στο κομμάτι που αφορά το κατευθυντικό τμήμα του σήματος των μικροφώνων, χρησιμοποιείται το διάνυσμα της έντασης  $\mathbf{I}(\tau, k)$  σε συνδυασμό με την μέθοδο *VBAP*. Με αυτόν τον τρόπο ενεργοποιείται ένα υποσύνολο ηχείων της διάταξης του οποίου το σήμα της κατευθυντικής συνιστώσας αναπαράγεται από τα ηχεία του υποσυνόλου. Οι συντελεστές κέρδους για κάθε ηχείο μπορούν να υπολογιστούν ως εξής :

$$g_l(\tau, k) = VBAP(\mathbf{I}(\tau, k), \mathbf{r}_l), \quad (21)$$

όπου  $\mathbf{r}_l$  το διάνυσμα συντεταγμένων του ηχείου με δείκτη  $l$ .

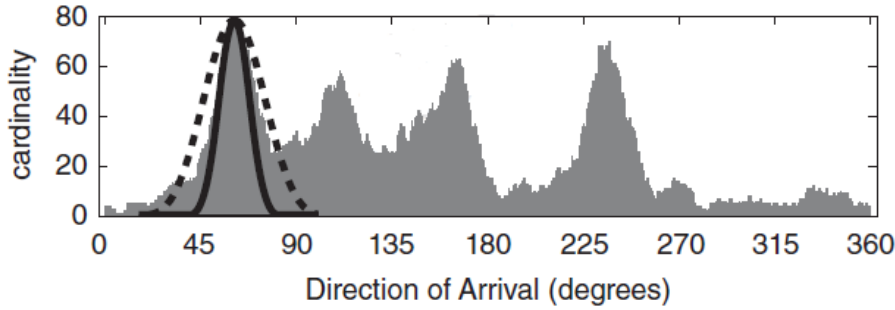
Τέλος, το σήμα των μικροφώνων που αφορά την διάχυτη συνιστώσα αναπαράγεται σε όλη ηχητική διάταξη πολλαπλασιάζοντας τις τιμές του σήματος των μικροφώνων με τον συντελεστή διάχυσης. Επιπλέον δίνεται μια τυχειότητα στην φάση στο σήμα των ηχείων για την αποφυγή φαινομένων χρωματισμού αλλά και τυχών αλλοιώσεων του σήματος κατά την διάρκεια αναπαραγωγής. Έτσι για κάθε ηχείο με δείκτη  $l$ , το σήμα που περιλαμβάνει την διάχυτη συνιστώσα προκύπτει από τον παρακάτω τύπο:

$$s_l^{dir}(\tau, k) = \sqrt{1 - \Psi(\tau, k)} g_l(\tau, k) w(\tau, k) \quad (22)$$

με  $w(\tau, k)$  να αποτελεί το σήμα ενός μικροφώνου της διάταξης.

### 3.2 Σύλληψη και Αναπαραγωγή με την Μέθοδο **ImmACS**

Στην τεχνική **ImmACS**, οι δύο βασικοί παράμετροι που εμπλέκονται είναι το πλήθος των ηχητικών πηγών και η γωνία της κάθε πηγής. Η διαδικασία αυτή πραγματοποιείται με την κατασκευή του ιστογράμματος (σχήμα 6) το οποίο περιλαμβάνει τη συλλογή των εκτιμώμενων γωνιών άφιξης  $\theta(\tau, k)$  για όλους τους δείκτες της συχνότητας σε ένα σύνολο από διαδοχικά χρονικά πλαίσια. Στην εργασία [9] η εκτίμηση της κατεύθυνσης άφιξης έγινε με τη μέθοδο **CICS** αλλά είναι προφανές ότι σύνολο των χρονικών πλαισίων περιλαμβάνει ένα ιστορικό από το οποίο γίνεται η συλλογή αυτή. Από τις κορυφές που εμφανίζονται στο ιστόγραμμα, γίνεται η εκτίμηση του πλήθους των πηγών και των γωνιών για την κάθε πηγή. Πχ, από το εικονιζόμενο διάγραμμα είναι προφανές ότι υπάρχει μία ηχητική πηγή περίπου στις 60 μοίρες καθώς και άλλες τρεις πηγές στις γωνίες  $105^\circ, 165^\circ$  και  $240^\circ$ .



Εικόνα 6: Αναπαράσταση Ιστογράμματος γωνιών άφιξης για ένα σύνολο πολλαπλών πηγών. Οι γωνίες στις κορυφές του ιστογράμματος, αντιστοιχούν στο πλήθος των ενεργών πηγών.

Αφού γίνει η εκτίμηση των γωνιών άφιξης καθώς και το πλήθος των ενεργών πηγών, γίνεται ο διαχωρισμός τους με τη χρήση τεχνικών **beamforming**, σε συνδυασμό με χρήση μεταφίλτρου (**postfilter**). Έτσι, αν  $X(\tau, k)$  το σήμα εισόδου του μικροφώνου για ένα χρονικό πλαίσιο  $\tau$  στο πεδίο της συχνότητας, οι τιμές του φίλτρου με την χρήση της μεθόδου **Superdirective Beamforming** θα είναι:

$$\mathbf{w}(k, \theta_s) = \frac{[\epsilon \mathbf{I} + \mathbf{\Gamma}(k)]^{-1} \mathbf{d}(k, \theta_s)}{\mathbf{d}(k, \theta_s)^H [\epsilon \mathbf{I} + \mathbf{\Gamma}(k)]^{-1} \mathbf{d}(k, \theta_s)} \quad (23)$$

όπου  $\mathbf{w}(k, \theta_s)$  αποτελεί διάνυσμα μήκους  $M \times 1$  μιγαδικών συντελεστών του φίλτρου,  $\theta_s$  η γωνία της  $s$  πηγής,  $\mathbf{d}(k, \theta_s)$  το διάνυσμα κατεύθυνσης στροφής της διάταξης,  $\mathbf{\Gamma}(k)$  ο  $M \times M$  πίνακας συνδυασμένης του θορύβου για σφαιρικά ιστροπικό ηχητικό πεδίο. Για κάθε μικροφώνο, το σήμα εισόδου πολλαπλασιάζεται με τους συντελεστές

beamforming  $\mathbf{w}(k, \theta_s)$  οι οποίοι συμβάλουν στον διαχωρισμό των επιμέρους σημάτων. Η έξοδος της παραπάνω διαδικασίας δίνεται απο την σχέση:

$$B_s(\tau, k) = \mathbf{w}^H(k, \theta_s)X(\tau, k), \quad s = 1, 2, 3, \dots, \hat{P} \quad (24)$$

με  $\hat{P}$  να αποτελεί το πλήθος των ενεργών πηγών.

Επιπλέον, μια ακόμη διαδικασία εφαρμόζεται με την χρήση δυαδικής μάσκας η οποία συμβάλει στον περαιτέρω διαχωρισμό των ηχητικών πηγών. Έτσι η δυαδική μάσκα για κάθε  $s$ -στο σήμα δίνεται ως :

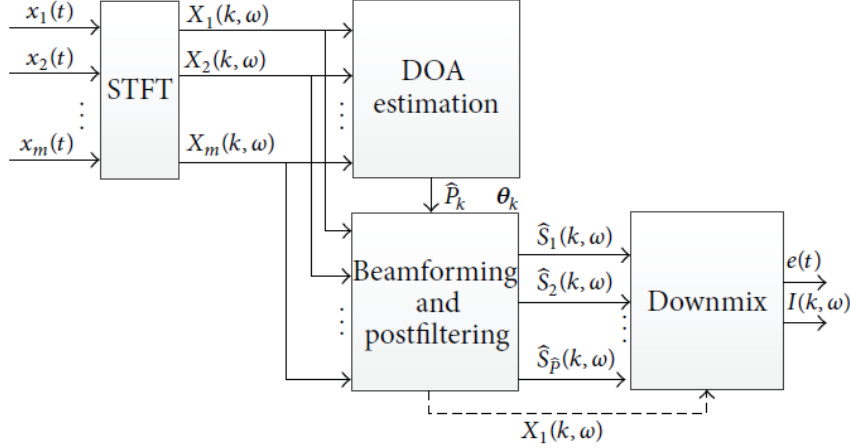
$$U_s(\tau, k) = \begin{cases} 1 & s = \operatorname{argmax}_p |B_p(\tau, k)|^2, p = 1, \dots, \hat{P} \\ 0 & \text{διαφορετικά} \end{cases} \quad (25)$$

Συνεπώς, από (24) και (25) κάθε διαχωριζόμενο ηχητικό σήμα μπορεί να γραφτεί ως:

$$\hat{S}_s(\tau, k) = U_s(\tau, k)B_s(\tau, k), \quad s = 1, 2, \dots, \hat{P} \quad (26)$$

Το πλήθος των διαχωριζόμενων σημάτων μπορούν να συνδυαστούν και να προκύψει ένα μονοφωνικό κανάλι εισόδου. Μια ακόμα υπόθεση που λαμβάνεται υπόψιν είναι η αρχή της ανεξαρτησίας μεταξύ των ηχητικών πηγών (WDO) [10]. Σύμφωνα με αυτήν για κάθε χρονικό πλαίσιο και για κάθε συχνοτική τιμή που αναλύεται, αντιστοιχίζεται μια μόνο πηγή. Η επιλογή της γίνεται ως το σήμα με την μεγαλύτερη τιμή της ενέργειας και οι υπόλοιπες πηγές δεν λαμβάνονται υπ'οψιν. Έτσι, αν προσθέσουμε τις επιμέρους διαχωριζόμενες πηγές, το σήμα του μονοφωνικού καναλιού που προκύπτει δίνεται από τον παρακάτω τύπο ως εξής:

$$E(\tau, k) = \sum_{s=1}^{\hat{P}} \hat{S}_s(\tau, k). \quad (27)$$



Εικόνα 7: Στάδιο ανάλυσης της μεθόδου ImmACS πάνω σε μια μικροφωνική διάταξη.

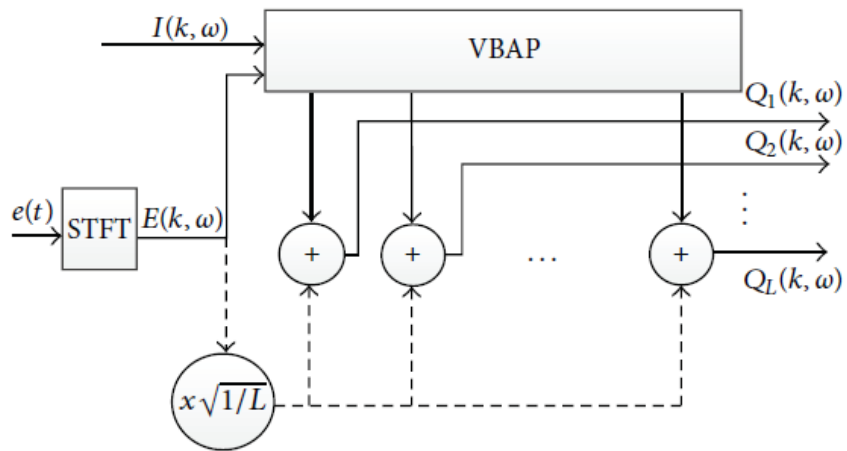
Για ένα οποιαδήποτε αριθμό σημάτων, γίνεται υπολογισμός της γωνίας άφιξης, του πλήθους ενεργών πηγών καθώς και ο διαχωρισμός τους (σχήμα 7). Τα σήματα αυτά μπορούν να αθροιστούν (Downmix) σε ένα ενιαίο κανάλι και κωδικοποιηθούν μέσω ενός συστήματος κωδικοποίησης/αποκωδικοποίησης για εφαρμογές πραγματικού χρόνου

Κατά την διαδικασία της σύνθεσης το φασματικό περιεχόμενο χωρίζεται σε κατευθυντική συνιστώσα και διάχυτη συνιστώσα. Για  $k \leq \omega_{cutoff}$ , το ηχητικό πεδίο θεωρείται κατευθυντικό και υπόκειται στην ανάλυση που έχει προηγηθεί, ενώ για  $k > \omega_{cutoff}$  το πεδίο θεωρείται διάχυτο. Η κατευθυντική συνιστώσα με την χρήση του  $I(k)$  και της VBAP συμβάλουν στην αναπαραγωγή του ηχητικού σήματος ενεργοποιώντας τον κατάλληλο αριθμό ηχείων της διάταξης. Τέλος, το τμήμα του σήματος που αφορά την διάχυτη συνιστώσα αναπαράγεται από το πλήθος των ηχείων της διάταξης για τιμές του  $k > \omega_{cutoff}$ .

Έτσι σε μια διάταξη ηχείων στο χώρο, το συνολικό σήμα αναπαραγωγής του κάθε ηχείου δίνεται απο την παρακάτω σχέση:

$$Q_l(\tau, k) = \begin{cases} g_l(\tau, k)E(\tau, k) & \text{Για } k \leq \omega_{cutoff} \\ \frac{1}{\sqrt{L}}E(\tau, k) & \text{Για } k > \omega_{cutoff} \end{cases} \quad (28)$$

όπου  $\omega_{cutoff}$  η συχνότητα κατωφλίου και  $g_l$  ο συντελεστής κέρδους για το  $l$ -στο ηχείο.



Εικόνα 8: Στάδιο αναπαραγωγής πλήθους ηχητικών πηγών σε μια διάταξη ηχείων με την μέθοδο ImmACS.

## 4 Υλοποίηση της Αμφιηχίας

### 4.1 Με Χρήση Sennheizer

Στις προηγούμενες ενότητες είδαμε πώς μπορούμε από ένα σύνολο μονοφωνικών καναλιών να ηχογραφήσουμε ένα ηχητικό γεγονός και να εξάγουμε από αυτό χωρική πληροφορία. Αυτό μπορεί να γίνει με την χρήση ειδικών συστοιχιών μικροφώνων καθώς και την χρήση της μεθόδου της αμφιηχίας. Μια συστοιχία μικροφώνων περιλαμβάνει ένα πλήθος από αισθητήρες οι οποίοι μπορεί να είναι σε μία διάταξη σφαίρας είτε ενός τετράεδρου. Οι αισθητήρες αυτοί μπορεί να είναι πανκατευθυντικοί ή καρδιοειδή. Στον παρακάτω πίνακα βλέπουμε μια λίστα από εμπορικά διαθέσιμα μικρόφωνα καθώς περιλαμβάνουν διαφορετικό αριθμό αισθητήρων.

Μικρόφωνο	Αριθμός αισθητήρων	Τάξη αμφιηχίας	Χρονιά
SoundField	4	1	1978
Tetramic	4	1	2007
Eigenmike	32	4	2008
<b>Ambeo</b>	<b>4</b>	<b>1</b>	<b>2016</b>
Zoom H2n	3	1 horizontal	2016
Zylia ZM-1	19	3	2017
Twirling720 Lite	4	1	2017
8ball	8	1 horizontal	2017

Πίνακας 3: Μοντέλα μικροφώνων τα οποία αποτελούνται από διαφορετικό πλήθος καρδιοειδών μικροφώνων.

Ανάλογα με το πλήθος των μικροφώνων μπορούμε να πετύχουμε υψηλότερες τάξεις αμφιηχίας. Η 1η τάξη αμφιηχίας μας δίνει 4 κανάλια κωδικοποίησης τα οποία αποτελούν :

- Το κανάλι  $W$  το οποίο είναι στην ουσία η πληροφορία που θα κατέγραφε ένα πανκατευθυντικό μικρόφωνο
- Τρία δικατευθυντικά μικρόφωνα για τα κανάλια  $X$ ,  $Y$ ,  $Z$

Το ελάχιστο πλήθος καναλιών που μπορεί να χρησιμοποιηθεί για μια αποτελεσματική αναπαράσταση ενός ηχητικού γεγονότος σε μια διάταξη ηχείων αντιστοιχεί σε  $(N + 1)^2$  κανάλια, με  $N$  να αποτελεί τον αριθμό της τάξης.

Στην παρούσα εργασία, έγινε η χρήση του Sennheizer Ambeo VR Mic το οποίο αποτελείται από 4 καρδιοειδή μικρόφωνα σε διάταξη τετράεδρου (εικόνα 9). Για κάθε τέτοιου είδους μικροφώνου, κατά τη σύλληψη ενός ακουστικού γεγονότος το σύνολο των μικροφωνικών

σημάτων αποτελούν το A-Format. Ο μετασχηματισμός των καναλιών της μορφής A-Format σε B-Format [11], γίνεται με τη χρήση του παρακάτω γραμμικού συστήματος των καναλιών του μικροφώνου:

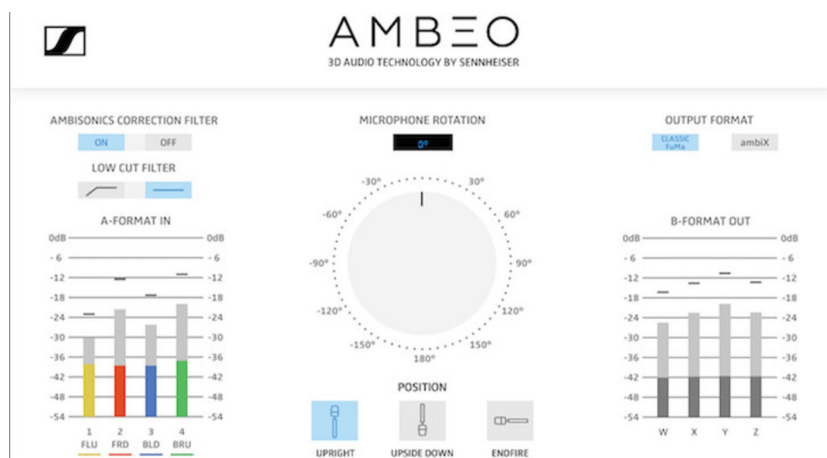
$$\begin{cases} W = FLU + FRD + BLD + BRU \\ X = FLU + FRD - BLD - BRU \\ Y = FLU - FRD + BLD - BRU \\ Z = FLU - FRD - BLD + BRU \end{cases} \quad (29)$$

όπου FLU, FRD, BLD και BRU αποτελούν σήματα του A-Format.



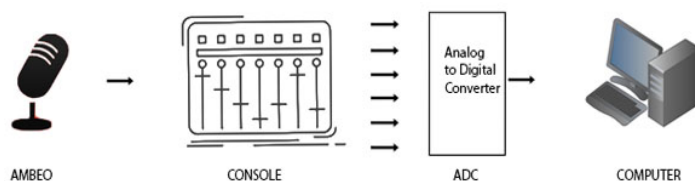
Εικόνα 9: Ambeo Sennheizer VR

Στη περίπτωση μας, έγινε χρήση ειδικού σχεδιασμένου εργαλείου λογισμικού Sennheizer AMBEO A-B format converter [12](εικόνα 12).



Εικόνα 10: Ambeo plugin converter

Η έξοδος του **B-format** περιλαμβάνει τα κωδικοποιημένα κανάλια *W*, *X*, *Y*, *Z*, όπου *W* αντιστοιχεί στην απόκριση μιας πανκατευθυντικής συνιστώσας στο κέντρο της διάταξης και τα *X*, *Y*, *Z* αντιπροσωπεύουν την κατανομή του πεδίου στους τρεις άξονες. Θα πρέπει να αναφερθεί ότι ο συνδυασμός των καναλιών του **B-Format** δεν περιορίζεται μόνο στην χρήση της αμφιηχίας, για παράδειγμα η μέθοδος **DiRAC** που είναι μια παραμετρική τεχνική βασίζεται στο **B-format** όπως αναφέρθηκε και σε προηγούμενη ενότητα.

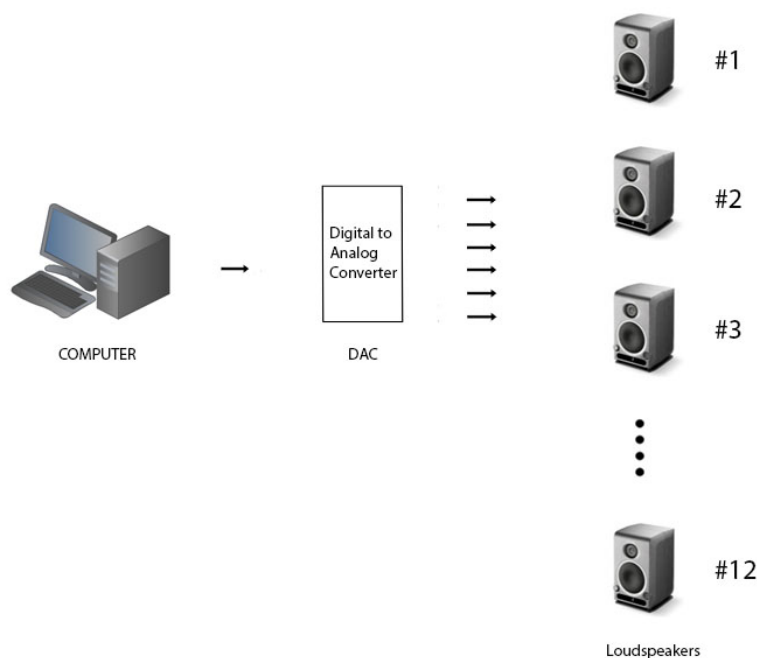


Εικόνα 11: Αναπαράσταση διαγράμματος σύλληψης και επεξεργασίας του AMBEO.

Το παραπάνω διάγραμμα απεικονίζει την διαδικασία σύλληψης και επεξεργασίας ενός ακουστικού γεγονότος με την χρήση του μικροφώνου **AMBEO**. Κατά την σύνδεση του μικροφώνου στα κανάλια των προενισχυτών, τα σήμα των καναλιών στέλνεται στις συσκευές ψηφιοποίησης και ύστερα στον υπολογιστή. Η κωδικοποίηση υλοποιείται μέσω υπολογιστή με την χρήση του λογισμικού **Reaper** και του **AMBEO plugin converter**. Η χρήση του **AMBEO plugin converter** διαφοροποιεί το σήμα κωδικοποίησης



του B-Format έναντι της (29) καθώς εφαρμόζεται μια διαδικασία κανονικοποίησης με την χρήση διορθωτικών καθώς και χαμηλοπερατών φίλτρων. Τέλος, για την αναπαραγωγή των σημάτων της ηχογράφησης το σήμα στέλνεται από τον ηλεκτρονικό υπολογιστή στους DA μετατροπείς και ύστερα στην διάταξη των ηχείων εφαρμόζοντας την διαδικασία αποκωδικοποίησης 1ης τάξης αμφιηχίας.



Εικόνα 12: Αναπαράσταση διαγράμματος αναπαραγωγής σήματος του μικροφώνου.

#### 4.1.1 Βαθμονόμηση Κάρτας Ήχου

Κάθε μικροφωνικό κανάλι του Shennheizer πρέπει να περάσει από προενισχυτή πριν το σήμα εισέλθει σε μετέπειτα στάδια επεξεργασίας. Επειδή τα ποτενσιόμετρα του ενισχυτή ρυθμίζονται χειρωνακτικά, είναι πολύ πιθανόν να μην υπόκεινται όλα τα κανάλια στο ίδιο κέρδος ενίσχυσης, γεγονός που μπορεί να επηρεάσει σημαντικά τα επόμενα στάδια επεξεργασίας. Κρίθηκε επομένως απαραίτητη μια διαδικασία βαθμονόμησης των προενισχυτών της κονσόλας που υποδέχονται τα A-format κανάλια του Shenheizer. Η βαθμονόμηση γίνεται με τον προσδιορισμό διορθωτικών συντελεστών  $\lambda(k)$  οι οποίοι αναλύονται στο πεδίο της συχνότητας. Τα τελικά σήματα των καναλιών του μικροφώνου που προκύπτουν με την χρήση των παραπάνω συντελεστών θα είναι:

$$\begin{cases} FLU'(\tau, k) = Re\{\lambda_1(k)\}FLU(\tau, k) \\ FRD'(\tau, k) = Re\{\lambda_2(k)\}FRD(\tau, k) \\ BLD'(\tau, k) = Re\{\lambda_3(k)\}BLD(\tau, k) \\ BRU'(\tau, k) = Re\{\lambda_4(k)\}BRU(\tau, k) \end{cases} \quad (30)$$

Έτσι, για την βαθμονόμηση της κάρτας ήχου ακολουθήσαμε την εξής διαδικασία: Αρχικά ένα ηχείο τοποθετήθηκε σε συγκεκριμένη θέση στο πάτωμα του δωματίου. Ένα μικρόφωνο (Shure SM93) χρησιμοποιήθηκε για την λήψη του ηχητικού σήματος. Η απόσταση μεταξύ μικροφώνου και ηχείου τοποθετήθηκε στο 1 μέτρο (σχήμα 13). Κατά την διεξαγωγή του πειράματος, το σήμα που παράχθηκε από το ηχείο ήταν λευκός θόρυβος διάρκειας 5 δευτερολέπτων. Η διαδικασία επαναλήφθηκε τέσσερις φορές, δηλαδή όσο είναι και το πλήθος των καναλιών της μικροφωνικής μας διάταξης για κάθε ένα ξεχωριστό κανάλι της κάρτας ήχου. Κατ' αυτόν τον τρόπο τα βάρη βαθμονόμησης που υπολογίστηκαν κρίθηκαν ικανά για να αντισταθμίσουν αποκλίσεις που οφείλονταν κυρίως στα ποτενσιόμετρα των προενισχυτών (εικόνα 14).



Εικόνα 13: Τοποθέτηση ηχείου - μικροφώνου κατά την διαδικασία βαθμονόμησης της κάρτας ήχου.

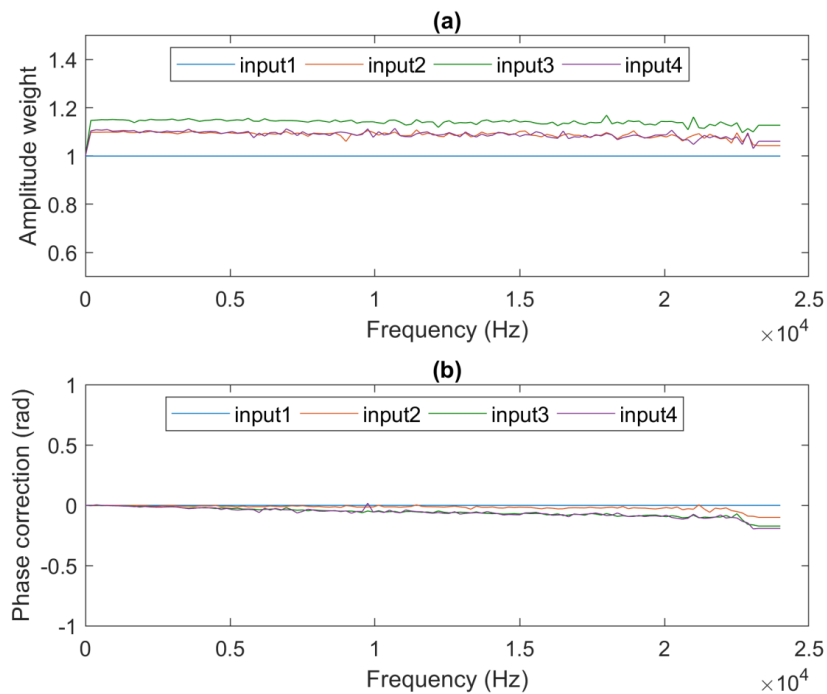
Η βαθμονόμηση έγινε λαμβάνοντας τη σχετική απόκριση μεταξύ κάθε καναλιού  $m = 2, 3, 4$  και του πρώτου καναλιού, μέσω της σχέσης :

$$\lambda_m(k) = \left\langle \frac{X_1(\tau, k)}{X_m(\tau, k)} \right\rangle \quad (31)$$

όπου το  $\langle \rangle$  συμβολίζει την προσδοκώμενη τιμή,  $\tau$  είναι το χρονικό πλαίσιο και  $k$  ο δείκτης συχνότητας.

Στο παρακάτω σχήμα φαίνονται τα βάρη διόρθωσης πλάτους και φάσης για την κάρτα ήχου. Οι τιμές που προκύπτουν σε κάθε περίπτωση είναι κανονικοποιημένες σε

σχέση το πρώτο κανάλι εισόδου. Και στις δύο περιπτώσεις παρατηρούνται πολύ μικρές αποκλίσεις στο πλάτος αλλά και στην φάση για κάθε είσοδο καναλιού. Με αυτόν τον τρόπο, οδηγούμαστε στο συμπέρασμα ότι με την χρήση μιας επαγγελματικής κάρτας ηχογράφησης οι αποκλίσεις που θα έχουν τα μικρόφωνα μεταξύ τους στην είσοδο της κάρτας μπορεί να είναι μικρές αλλά όχι αμελητέες.

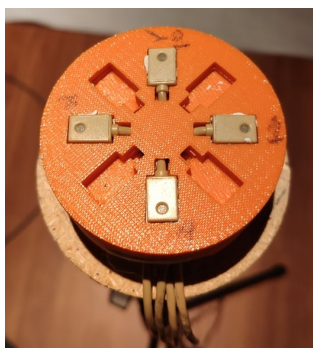


Εικόνα 14: Βάρη διόρθωσης πλάτους (a) και φάσης (b) συναρτήσει της συχνότητας για την κάρτα ήχου.

## 4.2 Με την Χρήση Κυκλικής Συστοιχίας Μικροφώνων

### 4.2.1 Συστοιχία Διαφορικών Μικροφώνων 1ης Τάξης

Το Εργαστήριο Επεξεργασίας Σήματος του ΙΤΕ διαθέτει τις δικές του κυκλικές συστοιχίες μικροφώνων και τις χρησιμοποιεί για ερευνητικούς σκοπούς (σχήμα 15). Η υλοποίηση της αμφιηχίας με κυκλική συστοιχία μικροφώνων είναι δυνατή εφαρμόζοντας τεχνικές διαφορικών μικροφώνων (DMA - Differential Microphone Arrays). Με χρήση της τεχνικής αυτής, μπορούμε σχετικά εύκολα να μετατρέψουμε το ηχητικό σήμα από τα πανκατευθυντικά μικρόφωνα στο **B-format** μορφή, που είναι απαραίτητη για την υλοποίηση της αμφιηχίας.



Εικόνα 15: Μικροφωνική συστοιχία τεσσάρων μικροφώνων σε κυκλική διάταξη

Αυτό πραγματοποιείται αν αφαιρέσουμε τα σήματα εξόδου της διάταξης ανά ζευγάρια μεταξύ τους. Το αποτέλεσμα της εξόδου μπορεί να χαρακτηριστεί ως μια διαφορική συστοιχία  $N$ -στης τάξης ανάλογα με τον αριθμό των ακουστικών αισθητήρων που χρησιμοποιείται. Για να είναι αποτελεσματική η κωδικοποίηση, η απόσταση μεταξύ των αισθητήρων πρέπει να είναι όσο το δυνατόν μικρότερη ώστε να αποφεύγονται φαινόμενα φασματικής αναδίπλωσης. Σε μια κυκλική διάταξη η μέγιστη τιμή της συχνότητας που μπορούμε να έχουμε για την αποφυγή τέτοιου είδους φαινομένων προκύπτει παρακάτω ως εξής:

$$f_{max} < \frac{c}{4r \sin \frac{\pi}{N}} \quad (32)$$

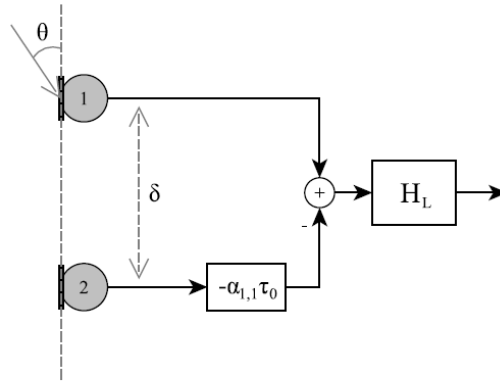
όπου  $r$  η ακτίνα της διάταξης,  $N$  ο αριθμός των μικροφώνων και  $c$  η ταχύτητα διάδοσης του ήχου.

Κατά την υλοποίηση 1ης τάξης διαφορικής συστοιχίας, απαιτείται η χρήση τουλάχιστον δύο μικροφώνων. Στην μέθοδο που υλοποιείται, γίνεται η χρήση τεσσάρων ακουστικών αισθητήρων ανά δύο ζεύγη μεταξύ τους. Κατά την σύλληψη ενός ακουστικού γεγονό-

τος, πραγματοποιείται κατάλληλη επεξεργασία με την χρήση κατάλληλου φίλτρου για την διεξαγωγή του σήματος κωδικοποίησης. Η τιμή της απόκρισης του συστήματος για τα σήματα των μικροφώνων αναλύεται στο πεδίο της συχνότητας. Έτσι, για σήματα δύο μικροφώνων η απόκριση  $h(\omega)$  του συστήματος [13] θα είναι :

$$h(\omega) = \frac{1}{1 - e^{-j\omega\tau_0(a_{1,1}-1)}} \begin{bmatrix} 1 \\ -e^{j\omega\tau_0 a_{1,1}} \end{bmatrix} \quad (33)$$

με  $\tau_0 = \frac{\delta}{c}$  να αποτελεί την χρονική καθυστέρηση του προσπίπτον σήματος μεταξύ των αισθητήρων,  $j = \sqrt{-1}$  η μιγαδική ποσότητα,  $\omega$  η γωνιακή συχνότητα και  $a_{1,1}$  η τιμή της κατευθυντικής απόκρισης.



Εικόνα 16: Δομή διαφορικής συστοιχίας 2 μικροφώνων

Επιπλέον απο την σχέση  $\tau_0 = \frac{\delta}{c}$  για  $\delta \ll \lambda$  και με την εφαρμογή της παρακάτω προσέγγισης  $e^x \simeq x + 1$ , η εξίσωση (33) μπορεί να αναλυθεί και ως:

$$h(\omega) = \frac{j}{(a_{1,1} - 1)\tau_0\omega} \begin{bmatrix} 1 \\ -e^{j\omega\tau_0 a_{1,1}} \end{bmatrix}. \quad (34)$$

Ακόμα, απο την (34), η απόκριση του συστήματος μπορεί να χωριστεί σε δύο μέρη εκ των οποίων περιλαμβάνουν, την καθυστέρηση του σήματος πρόσπτωσης στην διάταξη (σχήμα 16):

$$H_1(\omega) = 1 \quad (35a)$$

$$H_2(\omega) = -e^{-j\omega\tau_0 a_{1,1}} \quad (35b)$$

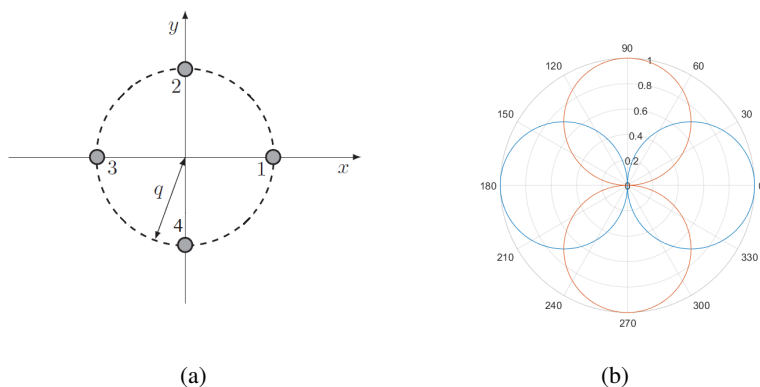
και την τιμή του φίλτρου ανακατασκευής του σήματος:

$$H_L(\omega) = \frac{j}{(\alpha_{1,1} - 1)\tau_0\omega}. \quad (36)$$

Το αποτέλεσμα της εξόδου περιλαμβάνει το κωδικοποιημένο σήμα της διάταξης. Κατά τον ίδιο τρόπο γίνεται η επεξεργασία και του δεύτερου καναλιού κωδικοποίησης για το άλλο ζεύγος μικροφώνων. Στον παρακάτω πίνακα απεικονίζονται διαφορετικές αποκρίσεις των μικροφώνων για διαφορετικές τιμές του  $\alpha_{1,1}$ .

Μοτίβο ακουστικής δέσμης	$\alpha_{1,1}$
Πανκατευθυντικό	1
Υποκαρδιοειδή	0.75
Καρδιοειδές	0.5
Υπερκαρδιοειδές	0.25
Δικατευθυντικό	0

Πίνακας 4: Πολικά μοτίβα για διαφορετικές τιμές του  $\alpha$



Εικόνα 17: Διαφορική συστοιχία 1ης τάξης : (a) 4-κάναλη κυκλική συστοιχία ακτίνας  $q$ , (b) διάγραμμα κατευθυντικότητας της μικροφωνικής διάταξης  $\alpha_{1,1} = 0$ .

Στην μελέτη που έγινε στην παρούσα εργασία, πραγματοποιήθηκε η υλοποίηση δικατευθυντικού μοτίβου για κάθε ζεύγος μικροφώνων (1 – 3 και 2 – 4 σχήμα 17a) με την χρήση τεσσάρων πανκατευθυντικών μικροφώνων. Για την μέθοδο που ακολουθήσαμε ο κάθε άξονας περιλαμβάνει τα κωδικοποιημένα B-format σήματα  $X, Y$  τα οποία προκύπτουν με την χρήση διαφορικής συστοιχίας 1ης τάξης. Το κανάλι κωδικοποίησης

W προκύπτει παίρνοντας το σήμα του 1ου μικροφώνου του παραπάνω σχήματος. Τέλος, βλέπουμε ότι με τη διαθέσιμη συστοιχία τεσσάρων μικροφώνων μπορούμε να κάνουμε ανάλυση του ηχητικού πεδίου κατά τους άξονες  $-x$  και  $-y$ , αλλά όχι κατά τον άξονα  $-z$ . Δηλαδή η κυκλική συστοιχία δε μπορεί να διακρίνει τη γωνία ανύψωσης της ηχητικής πηγής, και επομένως κατά την αναπαραγωγή το ηχητικό πεδίο περιορίζεται στις δύο διαστάσεις αντί για τις τρεις. Για να μπορεί να είναι δυνατή η σύγκριση μεταξύ του Shenheiser και της κυκλικής συστοιχίας, όλες οι ηχογραφήσεις έγιναν σε διδιάστατο ηχητικό πεδίο.

#### 4.2.2 Βαθμονόμηση Συστοιχίας Μικροφώνων

Για τη σωστή υλοποίηση της αμφιηχίας, κρίθηκε απαραίτητη η πραγματοποίηση βαθμονόμησης των μικροφώνων της συστοιχίας. Οι συντελεστές βαθμονόμησης  $\lambda(k)$  υπολογίστηκαν στο πεδίο της συχνότητας. Τα τελικά μικροφωνικά σήματα που προέκυψαν με την χρήση των παραπάνω συντελεστών είναι:

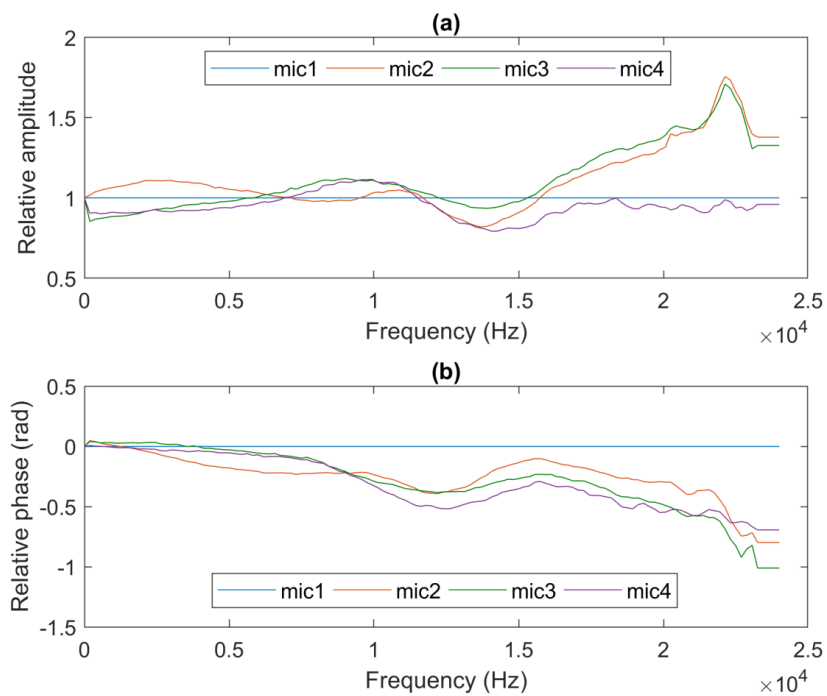
$$X_m(\tau, k) = \lambda_m(k)X_m(\tau, k) \quad (37)$$

όπου  $m = 2, \dots, 4$  είναι ο δείκτης μικροφώνου. Παρόμοια με την περίπτωση της κάρτας ήχου, η βαθμονόμηση έγινε κανονικοποιώντας τις αποκρίσεις των καναλιών της συστοιχίας σε σχέση με την απόκριση του πρώτου καναλιού, δηλαδή  $\lambda_1(k) = 1$  για κάθε δείκτη συχνότητας. Σε αντίθεση ωστόσο με την περίπτωση του Shenheiser, οι συντελεστές βαθμονόμησης ήταν μιγαδικοί αριθμοί, έτσι ώστε να υπεισέρχεται και διόρθωση της φάσης εκτός από διόρθωση του πλάτους.

Για την βαθμονόμηση της συστοιχίας μικροφώνων ακολουθήσαμε την εξής διαδικασία: Κάθε μικρόφωνο της συστοιχίας αφαιρέθηκε από την θήκη του και τοποθετήθηκε σε προκαθορισμένο σημείο, 1 μέτρο περίπου μπροστά από ένα ηχείο. Το ηχητικό σήμα που παράχθηκε από το ηχείο ήταν λευκός θόρυβος διάρκειας 5 δευτερολέπτων. Κύριος σκοπός αυτής της διαδικασίας τοποθέτησης είναι να εξασφαλιστεί ότι τα κανάλια των μικροφώνων έχουν τις ίδιες ακουστικές συνθήκες. Κατ' αυτόν τον τρόπο τα βάρη βαθμονόμησης που υπολογίστηκαν, κρίθηκαν ικανά να αντισταθμίσουν όχι μόνο αποκλίσεις στα ποτενσιόμετρα των προενισχυτών του κάθε καναλιού, αλλά και σε αποκλίσεις μεταξύ των μικροφώνων, τα οποία αν και είναι ίδιο μοντέλο, έχουν σημαντικές αποκλίσεις στην απόκρισή τους κυρίως λόγω γήρανσης.

Στο παρακάτω σχήμα φαίνονται τα διορθωτικά βάρη πλάτους και φάσης για το σύστημα κάρτα ήχου – συστοιχία μικροφώνων. Οι τιμές που προκύπτουν σε κάθε περίπτωση είναι αναφορικά με το πρώτο μικρόφωνο της διάταξης. Και στις δύο περιπτώσεις ακόμα και αν τα μικρόφωνα είναι κατασκευαστικά πανομοιότυπα, παρατηρούνται αποκλίσεις

μεταξύ των σχετικών τιμών των μικροφώνων όσο στο πλάτος αλλά και στην φάση.



Εικόνα 18: Βάρη διόρθωσης πλάτους (a) και φάσης (b) συναρτήσει της συχνότητας για το σύστημα κάρτα ήχου – συστοιχία μικροφώνων.

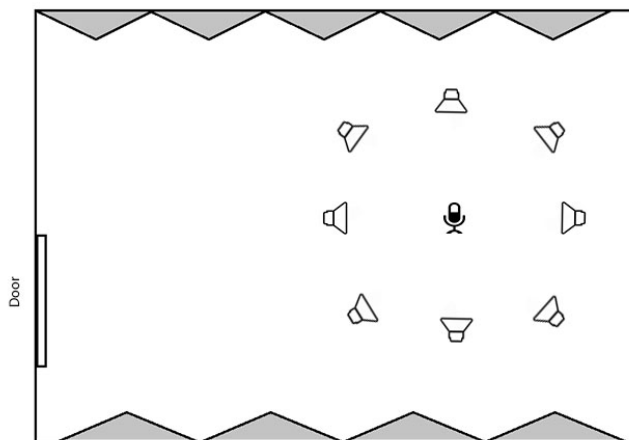


## 5 Αξιολόγηση

### 5.1 Πειραματική Διαδικασία

#### 5.1.1 Ηχογραφήσεις (Στατικές - Κινούμενες πηγές)

Η διεξαγωγή των πειραμάτων έγιναν στο Εργαστήριο Ακουστικής του Ινστιτούτου Πληροφορικής. Το εργαστήριο ακουστικής περιλαμβάνει μια αίθουσα control Room στην οποία γίνεται η αποθήκευση και επεξεργασία ηχητικών μετρήσεων καθώς και έναν ξεχωριστό χώρο για την διεξαγωγή πειραμάτων. Στο παρακάτω σχήμα φαίνεται η κάτοψη της αίθουσας.



Εικόνα 19: Κάτοψη αίθουσας διεξαγωγής πειραμάτων

Η αίθουσα διεξαγωγής πειραματικών διαδικασιών είναι ένας ειδικά διαμορφωμένος χώρος ο οποίος είναι επενδυμένος με ηχοαπορροφητικά υλικά. Ο χρόνος αντήχησης της αίθουσας είναι 0.18 s στα 500 Hz.

Για τις ανάγκες της αξιολόγησης έγιναν ηχογραφήσεις εντός του χώρου διεξαγωγής πειραμάτων. Οι ηχογραφήσεις χωρίστηκαν σε δυο μέρη, τις ηχογραφήσεις σταθερής πηγής και τις ηχογραφήσεις κινούμενης πηγής. Για τον σκοπό αυτό, αξιοποιήθηκαν οκτώ ηχεία σε κυκλική διάταξη και ίσες αποστάσεις και γωνίες μεταξύ τους. Τα οκτώ ηχεία ήταν τοποθετημένα ανά 45° πάνω σε κύκλο ακτίνας 1.65 m. Η θέση του μικροφώ-

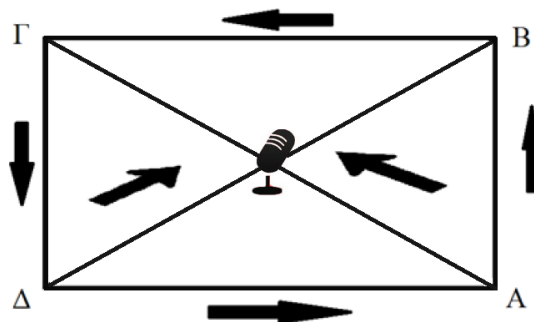


Εικόνα 20: Τοποθέτηση των δύο τύπων μικροφώνων κατά τη φάση της σύλληψης του ηχητικού πεδίου

νου **Shennheizer** καθώς και της συστοιχίας μικροφώνων ήταν ακριβώς στο κέντρο της κυκλικής συστοιχίας των ηχείων (εικόνα 20).

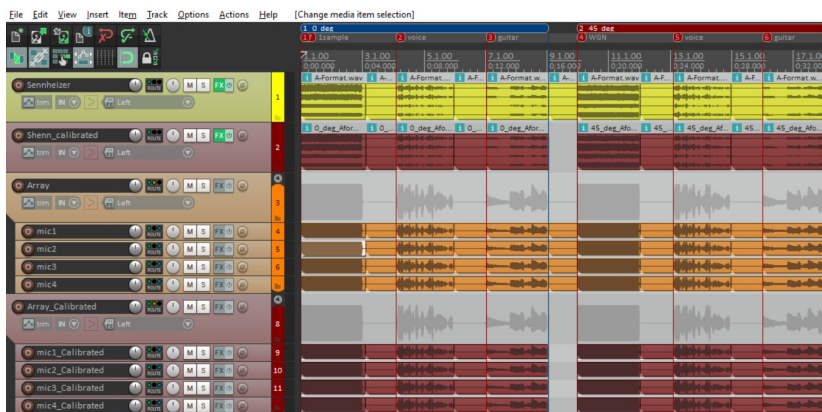
Για την ηχογράφηση με τις σταθερές πηγές χρησιμοποιήθηκαν τρία διαφορετικά σήματα: λευκός θόρυβος, απόσπασμα ηλεκτρικής κιθάρας και απόσπασμα γυναικείας φωνής. Για την ηχογράφηση με κινούμενες πηγές, τα σήματα ήταν λευκός θόρυβος, ήχος σφίγγας και απόσπασμα γυναικείας φωνής. Επιπλέον, οι ηχογραφήσεις για τις κινούμενες πηγές υλοποιήθηκαν με την χρήση φορητού ηχείου **bluetooth**. Το **bluetooth** ηχείο μετακινείτο πάνω σε 6 διαφορετικά ευθύγραμμα τμήματα τα οποία αντιστοιχούσαν στις πλευρές και τις δύο διαγώνιες ενός ορθογωνίου παραλληλόγραμμου με κέντρο που συνέπιπτε με το κέντρο της κυκλικής συστοιχίας των ηχείων, σύμφωνα με το (σχήμα 21).

Τα μικροφωνικά σήματα από τους δύο τύπους μικροφώνων ηχογραφήθηκαν ταυτόχρονα, δεσμεύοντας  $2 \times 4 = 8$  κανάλια εισόδου του συστήματος **ProTools**, με συχνότητα δειγμα-



Εικόνα 21: Διάγραμμα κινούμενων πηγών στον χώρο

τοληψίας  $f_s = 48000\text{Hz}$ . Η μετατροπή των A-format σημάτων από το Shenheizer σε B-format έγινε με χρήση plugin στο Reaper. Ωστόσο, για την μετατροπή των B-format σημάτων από το πεδίο του χρόνου στο πεδίο της συχνότητας, για την υλοποίηση των φίλτρων διόρθωσης και για την αποκωδικοποίηση των σημάτων για να προκύψουν τα σήματα των ηχείων χρησιμοποιήθηκε κώδικας στο Matlab. Όσον αφορά την κυκλική συστοιχία μικροφώνων, τόσο η κωδικοποίηση όσο και η αποκωδικοποίηση έγινε εξολοκλήρου στο Matlab.



Εικόνα 22: Περιβάλλον λογισμικού Reaper για την μετατροπή σημάτων από A-Format σε B-Format.

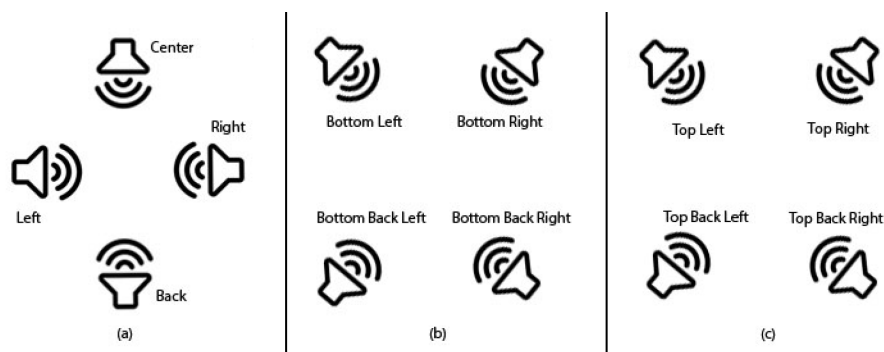
### 5.1.2 Επεξεργασία για Αναπαραγωγή Στο **Acoustic Lab**

Για την διεξαγωγή ακουστικών πειραμάτων χρησιμοποιήθηκαν συνολικά 12 ηχεία σε διάταξη σφαίρας των οποίων η χωροθέτηση τους έγινε κοντά στο κέντρο της αίθουσας. Η εικόνα 23 απεικονίζει την διάταξη των ηχείων στην αίθουσα πειραμάτων.

Τα ηχεία είναι τοποθετημένα πάνω σε σφαίρα ακτίνας 1.46 m από το κέντρο εκτός από τα κάτω ηχεία, των οποίων η ακτίνα είναι 1.28 m. Τα ηχεία που βρίσκονται στο μεσαίο επίπεδο (σχήμα 24(a)) είναι τοποθετημένα σε σχήμα σταυρού ανά  $90^\circ$  μεταξύ τους από το κέντρο της διάταξης. Για τα ηχεία που βρίσκονται στο πάνω και κάτω τμήμα (σχήμα 24(b)), οι θέσεις για το αζιμούθιο καλύπτουν τα ενδιάμεσα τμήματα του μεσαίου επιπέδου και η γωνία ανύψωσης παραμένει σχεδόν ίδια. Στον πίνακα 5 φαίνονται αναλυτικά οι τιμές των γωνιών των ηχείων της διάταξης.



Εικόνα 23: Διάταξη 12 ηχείων στο Acoustic Lab σε σχήμα σφαίρας.



Εικόνα 24: Διάγραμμα τοποθέτησης ηχείων: (a) μεσαίο επίπεδο, (b) κάτω επίπεδο, (c) πάνω επίπεδο.

Ηχείο	αζιμουθια	ανύψωσης	ακτίνα (m)
C	0°	0°	1.46
L	90°	0°	1.46
BC	180°	0°	1.46
R	-90°	0°	1.46
BottomL	45°	-40.5°	1.28
BottomBL	135°	-40.5°	1.28
BottomBR	-135°	-40.5°	1.28
BottomR	-45°	-40.5°	1.28
TopL	45°	39.5°	1.46
TopBL	135°	39.5°	1.46
TopBR	-135°	39.5°	1.46
BottomR	-45°	39.5°	1.46

Πίνακας 5: Τιμές των γωνιών και ακτίνας χωροθεσίας των ηχείων.

Τα ηχεία που χρησιμοποιήθηκαν ως πηγή ήχου είναι τα **Monitor Genelec 8050** (σχήμα 25), τα οποία είναι ενεργά ηχεία δύο δρόμων. Ο ρυθμιστή έντασης τοποθετήθηκε στη χαμηλότερη στάθμη για όλα τα ηχεία. Ο ενσωματωμένος ισοσταθμιστής (**equalizer**) επιλέχτηκε ως **flat** για όλα τα ηχεία εκτός από αυτά που ήταν στο πάτωμα, όπου επιλέχτηκε η ρύθμιση **-3 dB**. Αυτό έγινε για την αντιστάθμιση φαινομένων προσθετικής συμβολής λόγω των ισχυρών ανακλάσεων από το πάτωμα (τα άλλα ηχεία απέχουν αρκετή απόσταση από ανακλαστικές επιφάνειες σε σχέση με αυτά του πατώματος).

Στο σχήμα 26 φαίνονται οι αποστάσεις των ηχείων και για τα 3 επίπεδα. Τα ηχεία του κάτω επιπέδου έχουν μικρότερη ακτίνα συγκριτικά με τα υπόλοιπα ηχεία της διά-



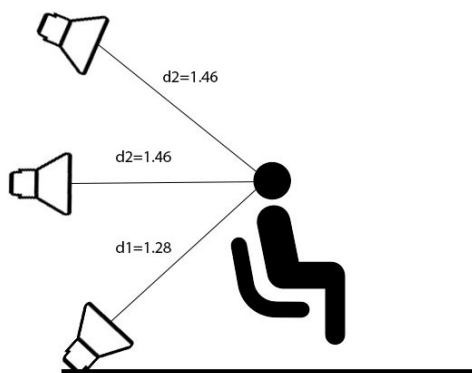
Εικόνα 25: Ηχείο Monitor Genelec 8040b

τάξης  $d_1 < d_2$ . Για τον λόγο αυτό, θα πρέπει να προστεθεί μια καθυστέρηση ώστε να μη δημιουργείται διαφωνία φάσης με τα υπόλοιπα ηχεία. Ο υπολογισμός αυτός προκύπτει από τα παρακάτω:

$$D_{samples} = \frac{dr}{c} f_s \quad (38)$$

με  $dr = 0.18$  η διαφορά της απόστασης των ηχείων του κάτω επιπέδου από τα υπόλοιπα,  $c$  η ταχύτητα διάδοσης του ήχου και  $f_s$  η συχνότητα δειγματοληψίας.

Τέλος όσο αφορά την ρύθμιση των εντάσεων των ηχείων της διάταξης, κρίθηκε ότι δεν χρειάστηκε κάποιου είδους βαθμονόμηση σε αυτά.



Εικόνα 26: Αποστάσεις ηχείων της διάταξης

## 5.2 Υποκειμενική Αξιολόγηση

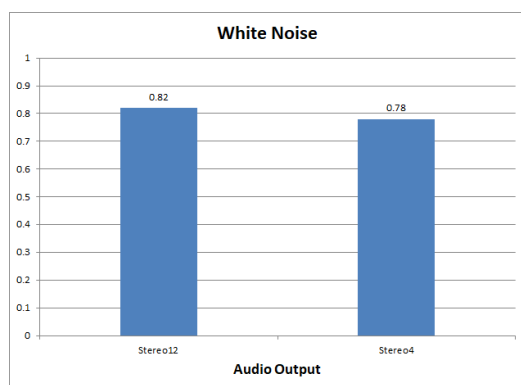
Για την επιβεβαίωση της ορθής λειτουργίας του συστήματος κρίθηκε απαραίτητη η δημιουργία τεστ ακρόασης με κριτήριο τη σωστή αντίληψη της θέσης στατικής πηγής και τη σωστή αντίληψη της φοράς κίνησης κινούμενης πηγής. Η υλοποίηση της Αμφιηχίας έλαβε χώρα με διαφοροποίηση όσον αφορά 1) τον τύπο του μικροφώνου (δηλαδή Shenheizer ή κυκλική συστοιχία) και 2) τη διάταξη των ηχείων αναπαραγωγή (4 ή 12 ενεργά ηχεία). Όλα αυτά αναλύονται με περισσότερη ακρίβεια στις παρακάτω ενότητες.

### 5.2.1 Ερωτηματολόγιο και Αποτελέσματα

Η διεξαγωγή πειραματικών διαδικασιών για την αξιολόγηση του συστήματος έγινε με την τοποθέτηση ακροατών στο κέντρο της ηχητικής διάταξης. Ο αριθμός των ακροατών που χρησιμοποιήθηκε για την αξιολόγηση ήταν συνολικά 19. Κατά την διάρκεια του πειράματος αναπαράχθηκαν συνολικά 52 ηχητικά αποσπάσματα μικρής διάρκειας για κάθε υποκείμενο. Το πλήθος των ηχητικών αποσπασμάτων χωρίστηκε σε μια συλλογή των 16, και δυο ζεύγη των 18 ηχογραφήσεων.

Το ερωτηματολόγιο που δημιουργήθηκε για το πλήθος των ηχητικών αποσπασμάτων χωρίστηκε σε 3 μέρη. Στο πρώτο μέρος έγινε αναπαραγωγή σήματος λευκού θορύβου στατικής πηγής στις παρακάτω γωνίες  $0, 45^\circ, 90^\circ, \dots, 315^\circ$  από 12 και 4 ηχεία. Στο δεύτερο μέρος πραγματοποιήθηκε αναπαραγωγή ηχητικών αποσπασμάτων στατικής πηγής κατά τον ίδιο τρόπο. Και στις δύο περιπτώσεις, το υποκείμενο καλέστηκε να προσδιορίσει την κατεύθυνση της ηχητικής πηγής έχοντας ως δεδομένο ότι είναι μία από τις γωνίες  $0, 45, 90, 135, 180, 225, 270$  και  $315$  μοίρες. Τέλος, στο τρίτο μέρος του πειράματος έγινε αναπαραγωγή κινούμενων πηγών όπου το δείγμα καλέστηκε να εντοπίσει την φορά κίνησης της πηγής έχοντας ως δεδομένο ότι είναι μία από τις 6 πιθανές διδρομές που απεικονίζονται στο σχήμα 21.

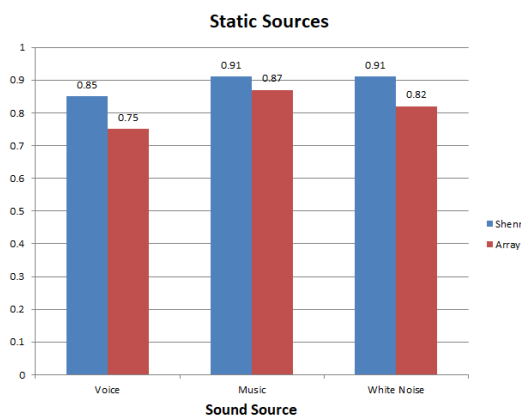
Σε αυτό το σημείο παρουσιάζονται τα αποτελέσματα του πρώτου μέρους του ερωτηματολογίου. Η σύγκριση αφορά την χωρική απεικόνιση μιας ηχητικής πηγής σε μία διάταξη 12 και 4 ηχείων με την εφαρμογή της μεθόδου της αμφιηχίας και την χρήση της μικροφωνικής συστοιχίας Shennheizer.



Εικόνα 27: Ποσοστά σωστών απαντήσεων για στατικές πηγές με χρήση του Shenhizer για 12 και 4 ηχεία.

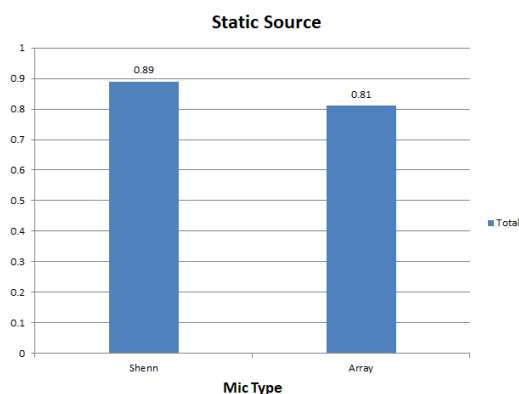
Όπως φαίνεται στο παραπάνω ιστόγραμμα, και στις δύο περιπτώσεις η μέθοδος της αμφιηχίας λειτουργεί ικανοποιητικά για διατάξεις 12 και 4 ηχείων. Στην περίπτωση των 12 ηχείων φαίνεται μάλιστα ένα μικρό πλεονέκτημα συγκριτικά με τα 4 ηχεία.

Τα αποτελέσματα του δεύτερου μέρους του ερωτηματολογίου αποδίδονται παρακάτω. Τα ιστογράμματα που προκύπτουν περιγράφουν την αντίληψη της κατεύθυνσης μιας ηχητικής πηγής σε ακουστικό περιβάλλον 12 ηχείων. Τα αποτελέσματα επιτρέπουν τη σύγκριση μεταξύ Sennheizer και κυκλικής συστοιχίας, δείχνοντας ότι η σύλληψη και επεξεργασία με Sennheizer υπερτερεί έναντι της κυκλικής συστοιχίας.



Εικόνα 28: Ποσοστά σωστών απαντήσεων για στατικές πηγές αναφορικά με το ηχητικό σήμα και τον τύπο μικροφώνου που έγινε η σύλληψη και επεξεργασία.

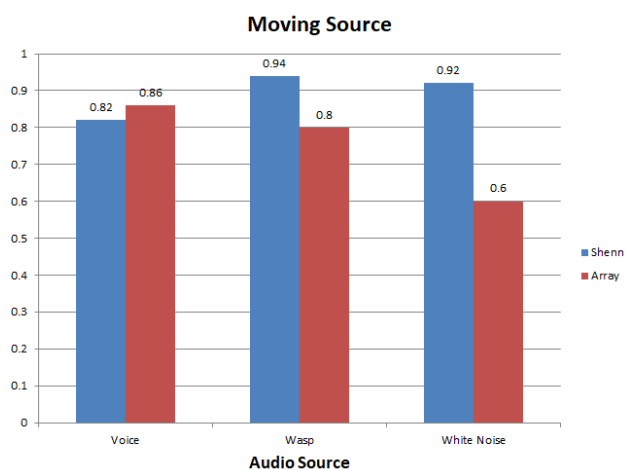




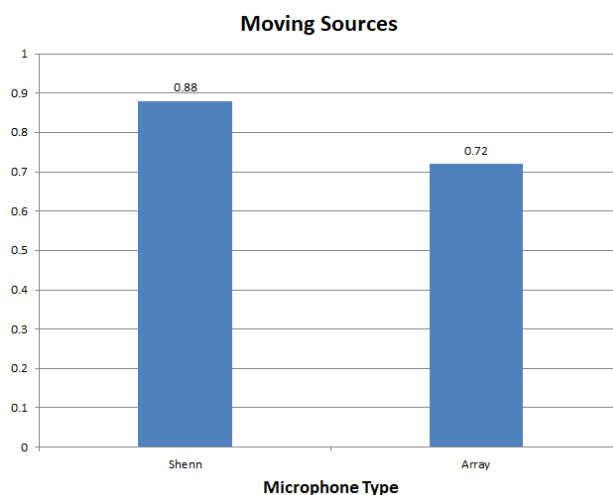
Εικόνα 29: Συνολικά ποσοστά σωστών απαντήσεων για στατικές πηγές αναφορικά με τον τύπο μικροφώνου που έγινε η σύλληψη και επεξεργασία.

Στα παραπάνω διαγράμματα 28 και 29, με την χρήση του Sennheizer φαίνεται να υπ-  
άρχει μεγαλύτερο ποσοστό σωστών απαντήσεων έναντι της κυκλικής συστοιχίας. Και  
στις δύο περιπτώσεις τα αποτελέσματα δείχνουν ικανοποιητικά όσο αφορά την αντίληψη  
της κατεύθυνσης μιας πηγής .

Τέλος, εξετάζοντας τα αποτελέσματα του τρίτου μέρους του ερωτηματολογίου πραγ-  
ματοποιείται σύγκριση και των δύο μικροφωνικών συστοιχιών όσον αφορά την αντί-  
ληψη φοράς κίνησης των πηγών.



Εικόνα 30: Ποσοστά σωστών απαντήσεων για κινούμενες πηγές αναφορικά με το ηχητικό σήμα και τον τύπο μικροφώνου που έγινε η σύλληψη και επεξεργασία.



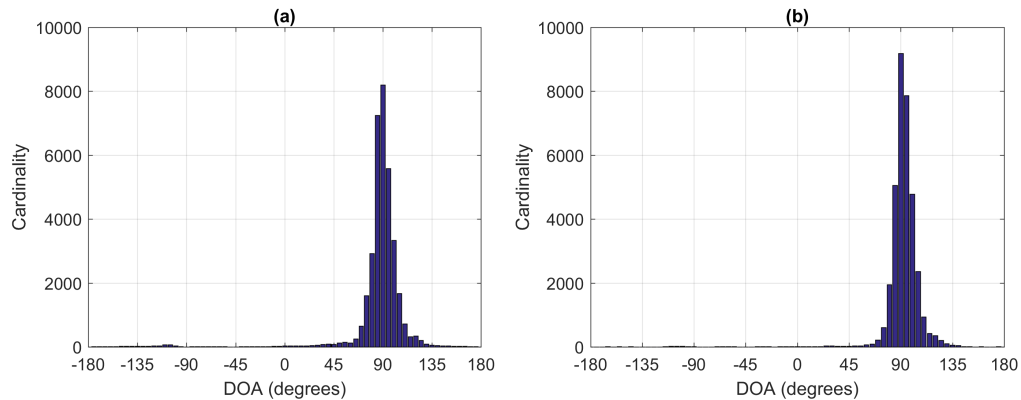
Εικόνα 31: Ποσοστά σωστών απαντήσεων για κινούμενες πηγές αναφορικά με το ηχητικό σήμα και τον τύπο μικροφώνου που έγινε η σύλληψη και επεξεργασία.

Στα παραπάνω διαγράμματα 30 και 31, τα αποτελέσματα που προκύπτουν για την αντίληψη της φοράς κίνησης μιας κινούμενης πηγής διαφοροποιούνται συγκριτικά με τις στατικές πηγές. Τελικά, και στις δύο περιπτώσεις (στατικές και κινούμενες πηγές) το **Shennheizer** φαίνεται να δίνει και εδώ καλύτερα αποτελέσματα συγκριτικά με την κυκλική συστοιχία όσο αφορά την κατεύθυνση μιας κινούμενης πηγής.

### 5.2.2 Αντικειμενική Αξιολόγηση

Εκτός από το υποκειμενικό τεστ, θα είχε ενδιαφέρον να εξάγει κανείς και παραμέτρους που προκύπτουν απευθείας από το ηχητικό σήμα για να ελέγξει τη σωστή λειτουργία του συστήματος. Συγκεκριμένα, χρησιμοποιώντας τις τεχνικές εκτίμησης της κατεύθυνσης της πηγής που αναλύθηκαν στην ενότητα 3.1.1, εξετάζουμε κατά πόσο ένα σύνολο από τέτοιες εκτιμήσεις συμπίπτουν με τις πραγματικές θέσεις των πηγών.

Στο σχήμα 32 απεικονίζεται η κατανομή των εκτιμώμενων αζιμουθιακών γωνιών κυκλικής συστοιχίας κατά την αναπαραγωγή λευκού θορύβου στις 90°. Όπως φαίνεται και στα δύο ιστογράμματα, παρατηρείται καλύτερη εκτίμηση της γωνίας άφιξης της ηχητικής πηγής στην περίπτωση υλοποίησης βαθμονόμησης του μικροφώνου (σχήμα 32(b)).



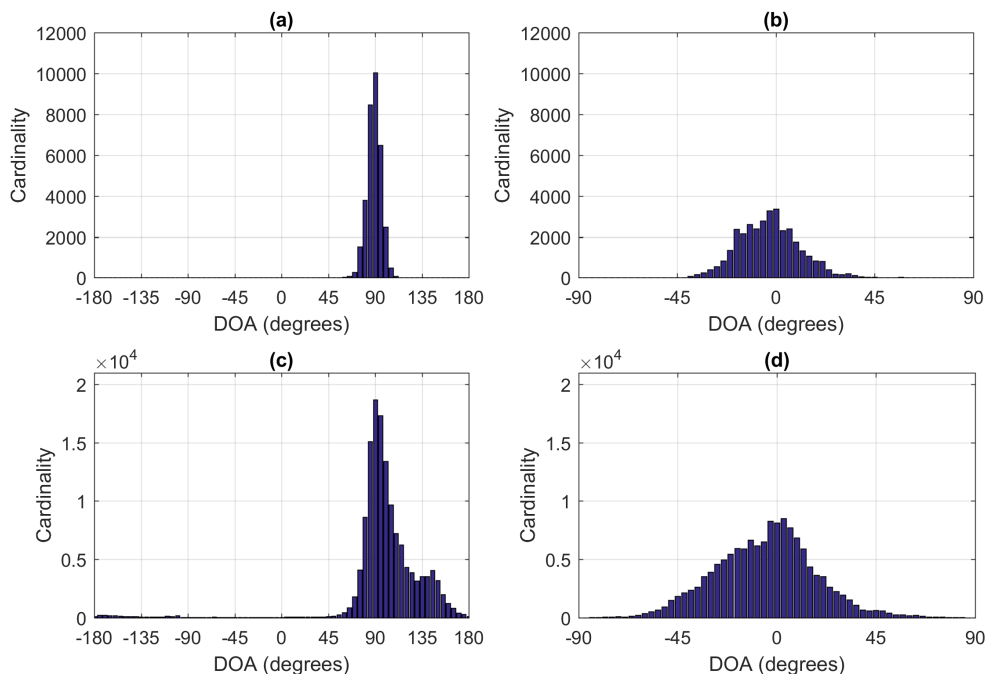
Εικόνα 32: Κατανομή των εκτιμώμενων γωνιών άφιξης με χρήση της κυκλικής συστοιχίας για πηγή λευκού θορύβου στις 90°: (a) χωρίς βαθμονόμηση, (b) με βαθμονόμηση

Στα παρακάτω διαγράμματα (σχήμα 33) παρουσιάζεται η κατανομή των εκτιμώμενων γωνιών άφιξης στατικής πηγής για το μικρόφωνο **Sennheizer**. Στα εικονιζόμενα διαγράμματα για το αζιμούθιο η γωνία κατεύθυνσης άφιξης είναι στις 90° και για την γωνία ανύψωσης η τιμή της βρίσκεται στις 0°. Στην περίπτωση των αποτελεσμάτων έχει προηγηθεί διαδικασία βαθμονόμησης του σήματος του μικροφώνου.

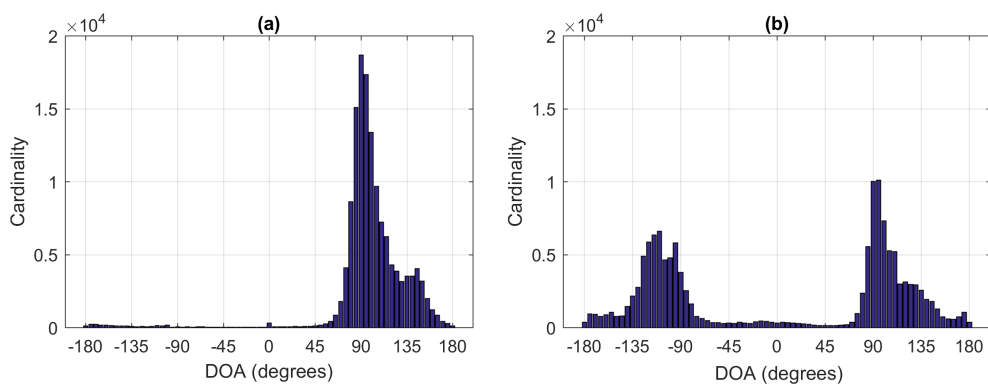
Από τα διαγράμματα 33(a) και (c) εμφανίζεται μια διαφορά στην ακρίβεια του εντοπισμού της γωνίας της ηχητικής πηγής. Στο διάγραμμα 33(c) φαίνεται να παρουσιάζονται και άλλες γωνίες εκτός της πραγματικής γωνίας. Αυτό οφείλεται κατά κύριο λόγο σε φαινόμενα σκίασης. Αυτά τα φαινόμενα ξεκινάνε από μια συχνότητα και πάνω, γεγονός που αποδεικνύεται από το διάγραμμα (c), όπου οι εκτιμώμενες γωνίες άφιξης υπολογίζονται μόνο μέχρι τη γωνία των 6 kHz.

Τέλος στο διάγραμμα του σχήματος 34 γίνεται παρουσίαση της κατανομής εκτιμώμενων γωνιών κατεύθυνσης άφιξης στατικής πηγής λευκού θορύβου μεταξύ των δύο μικροφωνικών διατάξεων **Shennheizer** και κυκλικής συστοιχίας. Το διάγραμμα 34(a) αναπαριστά το ιστόγραμμα γωνιών άφιξης του **Shennheizer** και το 34(b) της κυκλικής συστοιχίας όταν λαμβάνονται υπόψιν όλες οι συχνότητες μέχρι και τη συχνότητα Nyquist.

Όπως φαίνεται από τα αποτελέσματα των παραπάνω σχημάτων η πηγή βρίσκεται στις 90°. Συγκρίνοντας τα δύο γραφήματα, στο **Sennheizer** το πλήθος των εκτιμήσεων που αποκλίνουν από την πραγματική γωνία δείχνει να είναι αρκετά μικρότερο από αυτό για την περίπτωση της κυκλικής συστοιχίας. Επιπλέον παρατηρείται η εμφάνιση μιας εκ-



Εικόνα 33: Κατανομή των εκτιμώμενων γωνιών άφιξης με χρήση του Shenheizer για πηγή λευκού θορύβου στις 90°. Οι εκτιμήσεις φαίνονται για ανάλυση μέχρι τα 6kHz στα (a) και (b) και για ανάλυση μέχρι τη συχνότητα Nyquist στα (c) και (d). Οι εκτιμήσεις κατά το αζιμούθιο απεικονίζονται στα (a) και (c) και η γωνία ανύψωσης στα (b) και (d).



Εικόνα 34: Κατανομή εκτιμώμενων γωνιών άφιξης με την χρήση: (a) Shenheizer, (b) κυκλική συστοιχία

τιμήσεων κοντά στις  $135^\circ$  κάτι το οποίο οφείλεται σε φαινόμενα ηχητικής σκίασης. Ωστόσο, για την περίπτωση της κυκλικής συστοιχίας παρατηρείται ένα μεγάλο πλήθος γωνιών περίπου στις  $-120^\circ$ . Αυτό οφείλεται σε φαινόμενα ηχητικής αναδίπλωσης της μικροφωνικής διάταξης. Και στις δύο περιπτώσεις έχει προηγηθεί βαθμονόμηση των σημάτων των μικροφωνικών διατάξεων. Ως συμπέρασμα φαίνεται ότι η κυκλική διάταξη αισθητήρων πάσχει από φαινόμενα χωρικής αναδίπλωσης, με αποτέλεσμα οι σφαιρικές αρμονικές πρώτης τάξης να εμφανίζονται αλλοιωμένες από ένα συχνοτικό όριο και πάνω. Από την άλλη, το επαγγελματικό μικρόφωνο φαίνεται να μπορεί να αποδώσει σωστά τις αρμονικές πρώτης τάξης για ένα μεγαλύτερο συχνοτικό εύρος, πάσχοντας ωστόσο και αυτό στις υψηλές συχνότητες κυρίως από φαινόμενα ηχητικής σκίασης.

## 6 Συμπεράσματα

Η Αμφιηχία αποτελεί μια από τις πιο παλιές θεωρητικές προσεγγίσεις για σύλληψη και αναπαραγωγή μιας ηχητικής σκηνής. Ωστόσο η ορθή εφαρμογή της στην πράξη θέτει προκλήσεις, ανάλογα και με τον εξοπλισμό που είναι διαθέσιμος. Στην εργασία αυτή, χρησιμοποιήθηκαν δύο διαφορετικοί αισθητήρες για την σύλληψη του ηχητικού πεδίου: το επαγγελματικό μικρόφωνο **Shennheizer AMBEO VR** και μια αυτοσχέδια διάταξη πανκατευθυντικών μικροφώνων κυκλικής γεωμετρίας. Για την αναπαραγωγή στήθηκε μια τρισδιάστατη διάταξη 12 ηχείων, προσαρμοσμένη στις ήδη υπάρχουσες υποδομές του Εργαστηρίου Ακουστικής.

Η αξιολόγηση της υλοποίησης βασίστηκε τόσο σε αντικειμενικά κριτήρια όσο και σε υποκειμενικά τεστ. Τα τελευταία πραγματοποιήθηκαν με τη δημιουργία ηχογραφήσεων σε ελεγχόμενο ακουστικό περιβάλλον και με τη συλλογή απαντήσεων από ακροατές σχετικά με την αντίληψη της κατεύθυνσης του ήχου. Τα υποκειμενικά τεστ φανέρωσαν το ορθό της λειτουργίας όσον αφορά τη διάταξη των 12 ηχείων, δείχνοντας μάλιστα και ένα σχετικό πλεονέκτημα συγκριτικά με τη χρήση 4 μόνο ηχείων, κατά την αναπαραγωγή μιας δυσδιάστατης ηχητικής σκηνής στο επίπεδο που ορίζεται από τα 4 ηχεία.

Η τεχνική υλοποιήθηκε επιτυχώς και με τους δύο τύπους αισθητήρων, ωστόσο τα τεστ υποκειμενικής αξιολόγησης φανέρωσαν ότι το επαγγελματικό μικρόφωνο μπορεί να επιτύχει κάπως καλύτερη αντίληψη της κατεύθυνσης του ήχου από ότι το αυτοσχέδιο. Περαιτέρω ανάλυση που πραγματοποιήθηκε χρησιμοποιώντας τεχνικές εκτίμησης κατεύθυνσης άφιξης φανέρωσε ότι η κυκλική διάταξη αισθητήρων πάσχει από φαινόμενα χωρικής αναδίπλωσης, με αποτέλεσμα οι σφαιρικές αρμονικές πρώτης τάξης να εμφανίζονται αλλοιωμένες από ένα συχνοτικό όριο και πάνω. Από την άλλη, το επαγγελματικό μικρόφωνο φαίνεται να αποδίδει σωστά τις αρμονικές πρώτης τάξης για ένα μεγαλύτερο συχνοτικό εύρος, πάσχοντας ωστόσο και αυτό στις υψηλές συχνότητες κυρίως από φαινόμενα ηχητικής σκίασης. Επιπλέον αποτελέσματα που παρουσιάστηκαν ανέδειξαν τη σημασία της διαδικασίας βαθμονόμησης που προτάθηκε, κυρίως όσον αφορά την αυτοσχέδια συστοιχία μικροφώνων.

Μελλοντικά θα είχε ενδιαφέρον η αξιολόγηση να επικεντρωθεί στην αναπαραγωγή τρισδιάστατου ηχητικού πεδίου. Επιπλέον, θα είχε νόημα η υποκειμενική αξιολόγηση να μην περιοριστεί μόνο στην αντίληψη της κατεύθυνσης του ήχου, αλλά και στην αντίληψη της διάχυτης συνιστώσας του ήχου που είναι έντονο χαρακτηριστικό σε χώρους με μεγάλο χρόνο αντήχησης. Τέλος, θα είχε ενδιαφέρον να ερευνηθεί η δυνατότητα υλοποίησης των παραμετρικών τεχνικών χωρικού ήχου που έχουν αναπτυχθεί στο εργαστήριο με χρήση του **Sennheizer** μικροφώνου αντί για τις κυκλικές συστοιχίες που χρησιμοποιούνταν έως τώρα.

## 7 Παράρτημα

### 7.1 VBAP

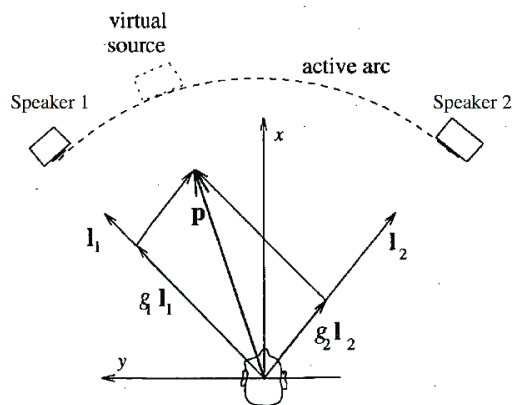
Η τεχνική VBAP, όπως και η στεροφωνία, βασίζεται στη διαφοροποίηση της έντασης των ηχείων για την δημιουργία του ακουστικού ειδώλου. Ωστόσο, η VBAP γενικεύεται για περισσότερα από δύο ηχεία, τα οποία μπορεί να είναι κατανομημένα και στις τρεις διαστάσεις. Ως εκ τούτου, μπορεί η VBAP να τοποθετήσει ακουστικά είδωλα και στις τρεις διαστάσεις.

Εστω δύο ηχεία τα οποία παράγουν ηχητικό σήμα τα με κοινό συνεκτικό περιεχόμενο. Τα σήματα αυτά μπορεί να έχουν διαφορετικό πλάτος το καθένα. Ο ακροατής τοποθετημένος σε μία θέση, αντιλαμβάνεται την φανταστική εικόνα (Phantom image) του ηχητικού σήματος, είτε μια εικονική πηγή (Virtual source). Έτσι για ένα ηχητικό σήμα  $s(t)$  κάθε ηχείο με διαφορετικό πλάτος, μπορεί να οριστεί ως :

$$x_i(t) = g_i s(t), \quad i = 1, \dots, N, \quad (39)$$

όπου  $N$  ο αριθμός των ηχείων,  $x_i(t)$  το σήμα κάθε καναλιού και  $g_i$  το αντίστοιχο κέρδος.

Με αυτόν τον τρόπο, ο ακροατής αντιλαμβάνεται την κατεύθυνση της εικονικής πηγής σε διαφορετική θέση ανάλογα με τις διαφορετικές τιμές των κερδών. Μια τυπική αναπαράσταση της μεθόδου φαίνεται στην παρακάτω εικόνα (Σχήμα 35).



Εικόνα 35: Στεροφωνική αναπαράσταση εικονικής πηγής με χρήση 2 ηχείων

Στο παραπάνω σχήμα φαίνεται η κατεύθυνση της πηγής ανάμεσα στα 2 ηχεία. Στην περιπτωσή μας, αν  $g_1$  και  $g_2$  τα κέρδη των αντίστοιχων σημάτων που παράγονται από κάθε ηχείο, τότε ισχύει :

$$g_1^2 + g_2^2 = 1, \quad (40)$$

Η παραπάνω εξίσωση μπορεί να γενικευτεί για  $N$  ηχεία ως :

$$\sum_{n=1}^{n=N} g_n^2 = 1. \quad (41)$$

Στην μέθοδο **VBAP** η μεθοδολογία τοποθέτησης ακουστικών ειδώλων διατυπώνεται εύγλωττα με χρήση διανυσμάτων. Στην περίπτωση του **2-D amplitude panning**, ορίζουμε ένα μοναδιαίου μήκους διάνυσμα βάσης  $\mathbf{l}_n$ , με  $n = 1, 2, \dots, N$  ο δείκτης του ηχείου. Το διάνυσμα  $\mathbf{l}_n$  ορίζει την κατεύθυνση του εκάστοτε ηχείου σε σχέση με την ακουστική θέση του ακροατή. Για παράδειγμα, για δύο ηχεία θα έχουμε διανύσματα βάσης  $\mathbf{l}_1, \mathbf{l}_2$  με  $\mathbf{l}_1 = [l_{11} \ l_{12}]^T$  και  $\mathbf{l}_2 = [l_{21} \ l_{22}]^T$  τα οποία δείχνουν απευθείας πάνω στην κατεύθυνση των δύο ηχείων (βλ. σχήμα 35). Επιπλέον ορίζουμε ένα μοναδιαίου μήκους διάνυσμα  $\mathbf{p}$ , το οποίο δείχνει την κατεύθυνση της εικονικής πηγής. Με την χρήση δύο ηχείων σε μια διάταξη το διάνυσμα  $\mathbf{p}$  αποτελεί τον γραμμικό συνδυασμό του διανύσματος βάσης  $\mathbf{L}$  με τα αντίστοιχα κέρδη  $g_1, g_2$ . Έτσι το διάνυσμα  $\mathbf{p} = [p_1 \ p_2]^T$ , μπορεί να κατασκευαστεί ως :

$$\mathbf{p} = g_1 \mathbf{l}_1 + g_2 \mathbf{l}_2. \quad (42)$$

Οι τιμές των μεταβλητών  $g_1, g_2$  της παραπάνω εξίσωσης θα πρέπει να είναι θετικοί αριθμοί. Επιπλέον μπορούμε να γράψουμε την (42) σε μορφή πίνακα ως :

$$\mathbf{p}^T = \mathbf{g} \mathbf{L}_{12} \quad (43)$$

με  $\mathbf{g} = [g_1 \ g_2]$  και  $\mathbf{L}_{12} = [\mathbf{l}_1 \ \mathbf{l}_2]^T$ . Το διάνυσμα  $\mathbf{g}$  μπορούμε να το υπολογίσουμε και ως εξής :

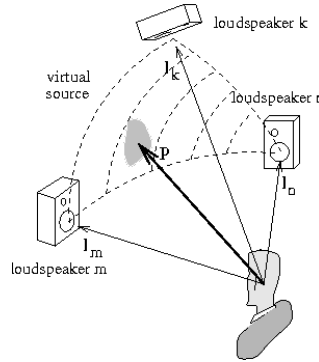
$$\mathbf{g} = \mathbf{p}^T \mathbf{L}_{12}^{-1} = [p_1 \ p_2] \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}^{-1}. \quad (44)$$

Αν υπάρχει η αντίστροφη μορφή του πίνακα  $\mathbf{L}_{12}^{-1}$  τότε το διάνυσμα  $\mathbf{L}_{12}$  ορίζεται σε ένα δισδιάστατο χώρο. Η (44) μας δίνει τις τιμές των κερδών για το παραπάνω σύστημα.

Αντίστοιχα σε μια διάταξη ηχείων στον τρισδιάστατο χώρο μπορούμε να εφαρμόσουμε την τεχνική **VBAP** επίσης με την χρήση διανυσμάτων όπως και προηγουμένως. Έστω λοιπόν μοναδιαίο διάνυσμα  $\mathbf{l}_n = [l_{n1} \ l_{n2} \ l_{n3}]^T$  το οποίο δείχνει πάνω στην κατεύθυνση των ηχείων  $n$ , από την θέση του ακροατή. Πάνω στη διάταξη τριπλέτας ηχείων τα μοναδιαία διανύσματα  $\mathbf{l}_n, \mathbf{l}_m$  και  $\mathbf{l}_k$  ορίζουν τις κατευθύνσεις  $n, m$  και  $k$  στον χώρο αντίστοιχα. Η κατεύθυνση της εικονικής πηγής αποτελεί ένα 3διάστατο διάνυσμα



$\mathbf{p} = [p_n \ p_m \ p_k]^T$ . Η διάταξη αυτή φαίνεται στο παρακάτω σχήμα.



Εικόνα 36: 3D αναπαράσταση εικονικής πηγής με χρήση 3 ηχείων

Το διάνυσμα κατεύθυνσης της εικονικής πηγής εκφράζεται ως γραμμικός συνδυασμός των διανυσμάτων βάσης  $\mathbf{l}_n$ ,  $\mathbf{l}_m$  και  $\mathbf{l}_k$  στην παρακάτω μορφή πίνακα:

$$\mathbf{p} = g_n \mathbf{l}_n + g_m \mathbf{l}_m + g_k \mathbf{l}_k, \quad (45)$$

$$\mathbf{p}^T = \mathbf{g} \mathbf{L}_{nmk}. \quad (46)$$

Οι τιμές των  $g_n$ ,  $g_m$  και  $g_k$  αποτελούν τα κέρδη,  $\mathbf{g} = [g_n \ g_m \ g_k]$  και  $\mathbf{L} = [l_n \ l_m \ l_k]^T$ . Το διάνυσμα  $\mathbf{g}$  λύνεται απο:

$$\mathbf{g} = \mathbf{p}^T \mathbf{L}_{nmk}^{-1} = [p_n \ p_m \ p_k] \begin{bmatrix} l_{n1} & l_{n2} & l_{n3} \\ l_{m1} & l_{m2} & l_{m3} \\ l_{k1} & l_{k2} & l_{k3} \end{bmatrix}^{-1}. \quad (47)$$

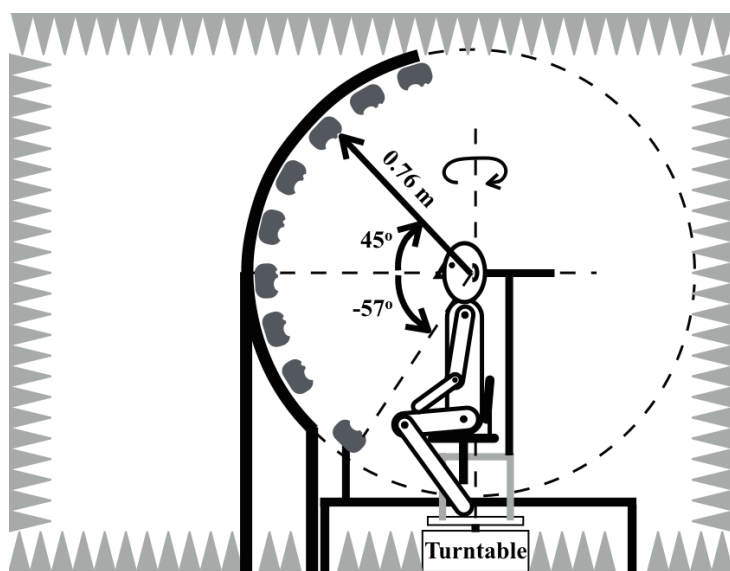
Ο  $\mathbf{L}_{nmk}^{-1}$  υπάρχει αν το διάνυσμα βάσης  $\mathbf{L}_{nmk}$  ορίζει τρισδιάστατο διανυσματικό χώρο. Η εξίσωση 47 υπολογίζει τα τις τιμές των βαρών για το διάνυσμα  $\mathbf{p}$  το οποίο ορίζεται απο το  $\mathbf{L}_{nmk}$  καθώς το περιεχόμενο του  $\mathbf{g}$ , αποτελεί τις διαφορετικές τιμές των κερδών. Τέλος είναι σημαντικό να σημειωθεί ότι η μέθοδος VBAP είναι σχεδιασμένη να ενεργοποιεί ένα μόνο υποσύνολο μιας διάταξης ηχείων συναρτήσει του διανύσματος κατεύθυνσης  $\mathbf{p}$ . Ο αριθμός των ηχείων που ενεργοποιούνται είναι το πολύ δύο για διατάξεις στο επίπεδο (2D) και το πολύ τρεις για διατάξεις στο τρισδιάστατο χώρο (3D).

## 7.2 Binaural Reproduction με HRTFs

Μια ειδική τεχνική που μπορεί να συγκριθεί με τη στερεοφωνική αναπαραγωγή είναι η διπλωτική αναπαραγωγή (binaural reproduction). Κατά την ηχογράφηση, δύο

μικρόφωνα τοποθετούνται στην είσοδο του ακουστικού πόρου της κεφαλής μιας κούκλας, και το ηχογραφημένο σήμα μετά αναπαράγεται σε έναν ακροατή με ακουστικά. Κατά αυτόν τον τρόπο, είναι δυνατόν να δημιουργηθούν πολύ πιστά ηχητικά είδωλα, γεγονός που στηρίζεται στο ότι το ηχογραφημένο σήμα είναι συσχετισμένο με τα φαινόμενα ανάκλασης και διάθλασης που δημιουργούνται από την κεφαλή και το πτερύγιο του αυτιού της κούκλας, που ιδανικά προσεγγίζει τα ανατομικά χαρακτηριστικά του αυτιού του ίδιου του ακροατή. Αν και χρησιμοποιούνται μόνο δύο κανάλια, μπορεί να δημιουργηθεί η εντύπωση ότι ο ήχος προέρχεται από όλες σχεδόν τις πιθανές κατευθύνσεις, και υπό κάποιες ιδανικές συνθήκες, ο ακροατής θα ταυτιστεί με τη θέση της κούκλας στο χώρο και το χρόνο της ηχογράφησης.

Η συνηθέστερη μέθοδος για την επίτευξη της διπλωματικής αναπαραγωγής είναι η χρήση HRTFs (Head-related transfer functions). Η μέθοδος αυτή αποτελεί ένα σύνολο φίλτρων προερχόμενα από οποιαδήποτε πιθανή κατεύθυνση. Τα φίλτρα αυτά σε ένα ακουστικό περιβάλλον περιγράφονται ως η χρονική απόκριση του συστήματος HRIR (head-related impulse response) και ο μετασχηματισμός τους στο πεδίο της συχνότητας (HRTF). Η σύλληψη του ηχητικού σήματος της πηγής γίνεται πάνω στην είσοδο των ακουστικών πόρων της κεφαλής καθώς είτε ο ακροατής, είτε η ηχητική πηγή περιστρέφεται στο χώρο σχήμα 37. Η συνέλιξη οποιαδήποτε ηχητικού σήματος με το παραγόμενο HRIR, δίνει στο σήμα την κατεύθυνση της θέσης που παράχθηκε από την θέση που το έλαβε ο ακροατής.



Εικόνα 37: Διάταξη χρήστη και ηχητικών πηγών σε ακουστικό περιβάλλον για τον υπολογισμό HRTFs

Ωστόσο ένα μειονέκτημα που προκύπτει με την χρήση των HRTFs είναι η λανθασμένη περισυλλογή δεδομένων από τον χρήστη για την σωστή αναπαράσταση της διπλωτικής αναπαραγωγής. Έχει παρατηρηθεί ότι αν τα σχεδιαστικά χαρακτηριστικά της κούκλας απέχουν πολύ από αυτά του ακροατή, είναι πολύ πιθανόν το αντιληπτό αποτέλεσμα να μην είναι ιδιαίτερα ρεαλιστικό. Σήμερα ωστόσο υπάρχει εξοπλισμός με τον οποίο η ηχογράφηση μπορεί να γίνει στον ακουστικό πόρο του ίδιου του ακροατή. Ένα μειονέκτημα της διπλωτικής αναπαραγωγής είναι ότι όταν ο ακροατής γυρίζει το κεφάλι του τότε το ακουστικό γεγονός δεν ακολουθεί την ανάλογη μεταβολή της ηχητικής εικόνας, κάτι που προφανώς οφείλεται στη χρήση των ακουστικών.

Η δυνατότητα διπλωτικής αναπαραγωγής είναι επίσης εφικτή με χρήση ηχείων (transaural reproduction), αλλά είναι πολύ πιο περίπλοκη διαδικασία από τη στερεοφωνία γιατί πρέπει να λαμβάνονται μέτρα για την ακύρωση της διαφωνίας μεταξύ των ηχείων (crosstalk cancellation). Αυτή η ακύρωση διαφωνίας μπορεί να εξαιρεθεί ικανοποιώντας την απαίτηση κάθε ένα ηχείο να είναι υπεύθυνο για το σήμα σε ένα και μόνο αυτί και συγκεκριμένα, το αριστερό ηχείο για το αριστερό αυτί και αντίστοιχα το δεξί. Η επίλυση του προβλήματος επιτυγχάνεται με τεχνικές οι οποίες αφορούν στην πρόσθεση, στο σήμα του δεξιού καναλιού, μιας εκδοχής εκτός φάσης του σήματος του αριστερού καναλιού που αναμένεται να φτάσει στο δεξί αυτί μέσω διαφωνίας, και στην πρόσθεση, στο σήμα του αριστερού καναλιού, μιας εκδοχής εκτός φάσης του σήματος του δεξιού καναλιού που αναμένεται να φτάσει στο αριστερό αυτί μέσω διαφωνίας. Η ιδέα είναι ότι αυτά τα επιπλέον εκτός φάσης σήματα ακυρώνουν τα ανεπιθύμητα σήματα διαφωνίας, με αποτέλεσμα το ισοδύναμο των αρχικών διπλωτικών σημάτων να φτάνει στα κατάλληλα αυτιά. Ένα μεγάλο μέρος της έρευνας που αφορά την διπλωτική αναπαραγωγή ασχολείται με την αύξηση της ευρωστίας αυτού του τύπου αναπαραγωγής ο οποίος χαρακτηρίζεται από μεγάλες αποκλίσεις για μικρές αλλαγές της θέσης του ακροατή [14]. Η χρήση δύο πηγών τοποθετημένων πολύ κοντά η μία στην άλλη που βρίσκονται μπροστά από τον ακροατή παρουσιάζει ιδιαίτερη αύξηση της ευρωστίας. Το αναπαραγόμενο ηχητικό πεδίο έχει όχι τυχαίως μεγάλες ομοιότητες με αυτό ενός ακουστικού δίπολου και για αυτό το λόγο έχει αποδοθεί με τον όρο στερεοφωνικό δίπολο (stereo dipole).

## Βιβλιογραφία

- [1] “Ambisonics-wikipedia,the free encyclopedia,” <https://en.wikipedia.org/w/index.php?title=Ambisonics&oldid=656474391>, 2015.
- [2] Jerome Daniel, “Ambisonics encoding of other audio formats for multiple listening conditions,” 01 1998.
- [3] Daniel Arteaga, “Introduction to ambisonics,” 06 2015.
- [4] Daniel Arteaga, “An ambisonics decoder for irregular 3d loudspeaker arrays,” 01 2013.
- [5] Ortolani Francesca, “Introduction to ambisonics a tutorial for beginners in 3d audio,” 2015.
- [6] Ville Pulkki, “Spatial sound reproduction with directional audio coding,” *JAES*, vol. 55, pp. 503–516, 06 2007.
- [7] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, 1997.
- [8] Christof Faller and Ville Pulkki, “Directional audio coding: Filterbank and stft-based design,” in *Audio Engineering Society Convention 120*, May 2006.
- [9] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, “Real-time multiple sound source localization and counting using a circular microphone array,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2193–2206, Oct 2013.
- [10] Anastasios Alexandridis, Anthony Griffin, and Athanasios Mouchtaris, “Capturing and reproducing spatial audio based on a circular microphone array,” *Journal of Electrical and Computer Engineering*, vol. 2013, 03 2013.
- [11] Farina Angelo, “A-format to b-format conversion,” <http://pcfarina.eng.unipr.it/Public/B-format/A2B-conversion/A2B.htm>, 2006.
- [12] “Shennheiser ambeo vr mic,” online: <https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic>.
- [13] Messner Elmar, “Differential microphone arrays,” 12 2013.
- [14] Ole Kirkeby, Philip Nelson, F Orduna-bustamente, and Hareo Hamada, “Local sound field reproduction using digital signal processing,” *Journal of The Acoustical Society of America - J ACOUST SOC AMER*, vol. 100, 09 1996.