

# ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ

## **Αναφορά 2<sup>ης</sup> Εργαστηριακής Άσκησης**

Μιχαλίτσης Αλκιβιάδης – Παναγιώτης: 03118868

Κάπρος Παναγιώτης : 03118926



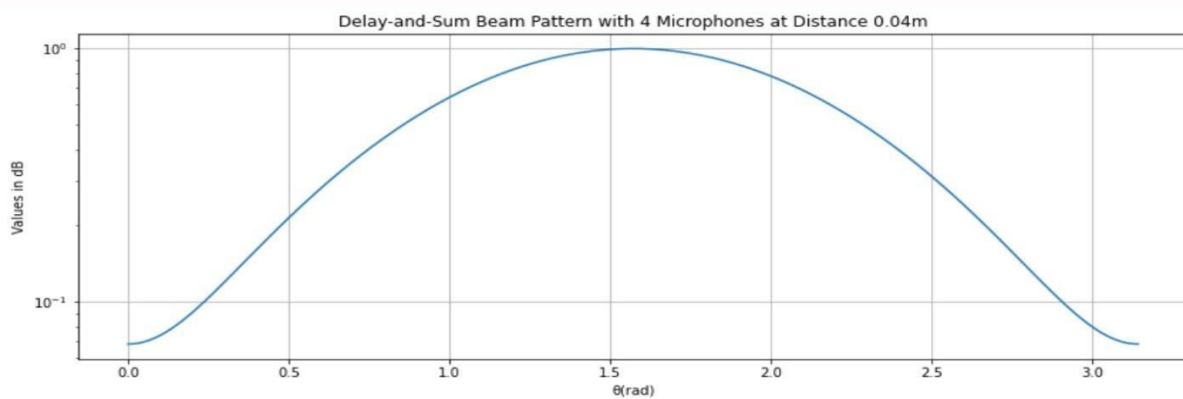
## 1ο Μέρος

### 1.5

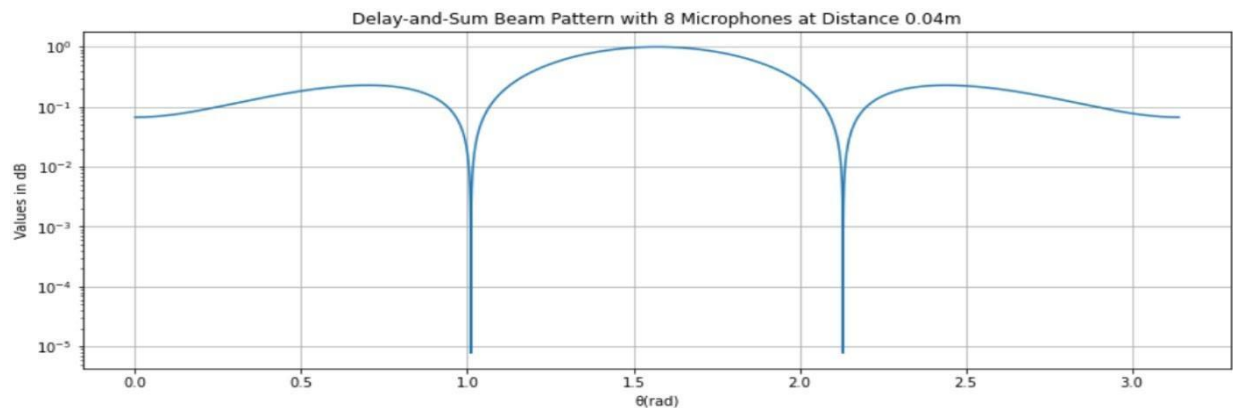
Θεωρούμε ότι έχουμε μία γραμμική συστοιχία μικροφώνων και ένα σήμα που προέρχεται από μία πηγή τοποθετημένη σε γωνία  $\theta_s$  ως προς την συστοιχία. Για την υλοποίηση του ερωτήματος, χρησιμοποιήθηκε ο τύπος 19 από την εκφώνηση - θεωρία της άσκησης εκτυπώνοντας το μέτρο του όπου είναι συναρτήσει της συχνότητας  $\omega$  και της γωνίας  $\theta$ , μεταβάλλοντας τις παραμέτρους  $N$ ,  $d$  και  $\theta_s$  (όπου  $N$  ο αριθμός των μικροφώνων,  $d$  η απόσταση μεταξύ δύο διαδοχικών μικροφώνων και  $\theta_s$  η γωνία της συστοιχίας με την πηγή). Η γωνία  $\theta$  μεταβάλλεται από 0 έως  $2\pi$  rad, η σταθερά της ταχύτητας του ήχου στον αέρα είναι  $c=340\text{m/s}$ , και η συχνότητα είναι  $f=2\text{kHz}$ .

1) Αρχικά θεωρούμε  $d=4\text{cm}$  και μεταβάλλουμε τον αριθμό των μικροφώνων  $N$ :

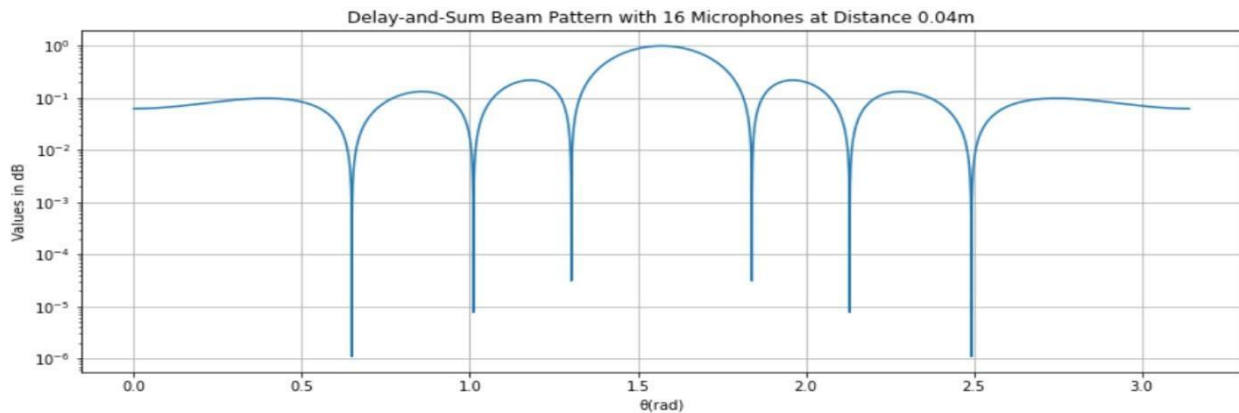
**Για  $N=4$ .**



**Για  $N=8$ .**



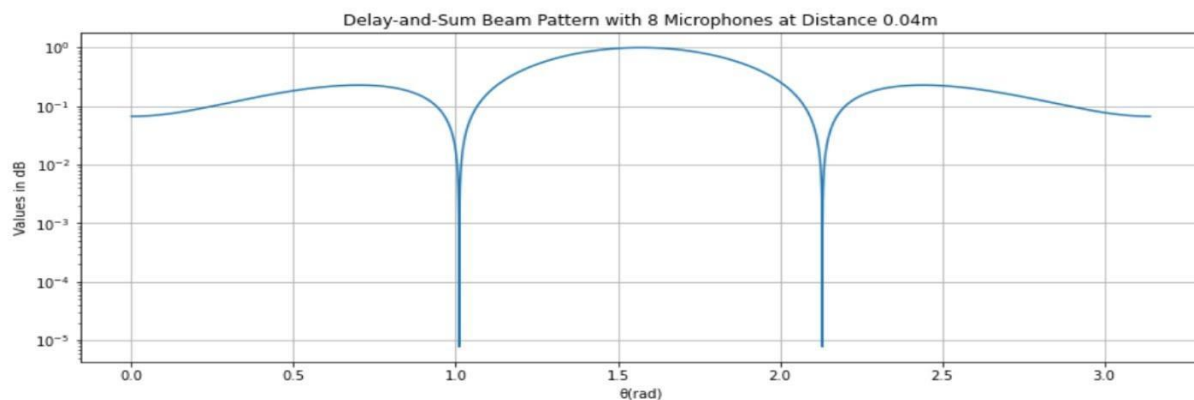
Για  $N=16$ .



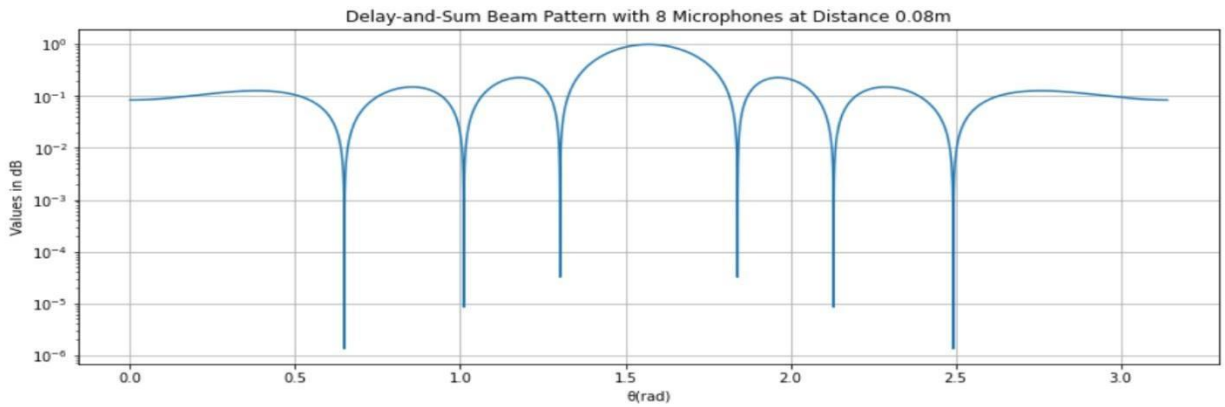
Προσδοκούμε στην ενίσχυση του ήχου που προέρχεται από μία συγκεκριμένη γωνία. Έτσι στα πλάτη του Beamformer θα θέλαμε πολύ μικρό κύριο λοβό για να ενισχύουμε μόνο τον ήχο που έχουμε εντοπίσει από την γωνία που προέρχεται. Το γεγονός ότι οι υπόλοιποι λοβοί (ή ο κύριος λοβός) έχουν φάσμα, αποτελεί ένδειξη μίας μη επιθυμητής κατάστασης όπου ενισχύονται εκτός της επιθυμητής γωνίας τα ηχητικά σήματα που φτάνουν στην συστοιχία από τις άλλες γωνίες. Αυτό δεν είναι επιθυμητό καθώς παρατηρούμε ότι όσο αυξάνουμε τον αριθμό των μικροφώνων στενεύουμε όλο και περισσότερο τον κύριο λοβό και μικραίνουμε κατά πλάτος τους παράπλευρους λοβούς με αποτέλεσμα να έχουμε καλύτερη ποιότητα σήματος στην έξοδο μας. Έτσι, θα προκύψει ένα σήμα με λιγότερο θόρυβο μιας και ο θόρυβος μπορεί να είναι ένα σήμα που απλά φτάνει στην γραμμική συστοιχία με μία διαφορετική γωνία από αυτής της πηγής θs. Στην συχνότητα που επικεντρωνόμαστε έχουμε ότι το σήμα περνάει αυτούσιο αφού το μέτρο του Beamformer έχει πλάτος 1.

2) Κρατώντας σταθερό τον αριθμό των μικροφώνων με  $N=8$  και μεταβάλλοντας την απόσταση  $d$  μεταξύ τους έχουμε:

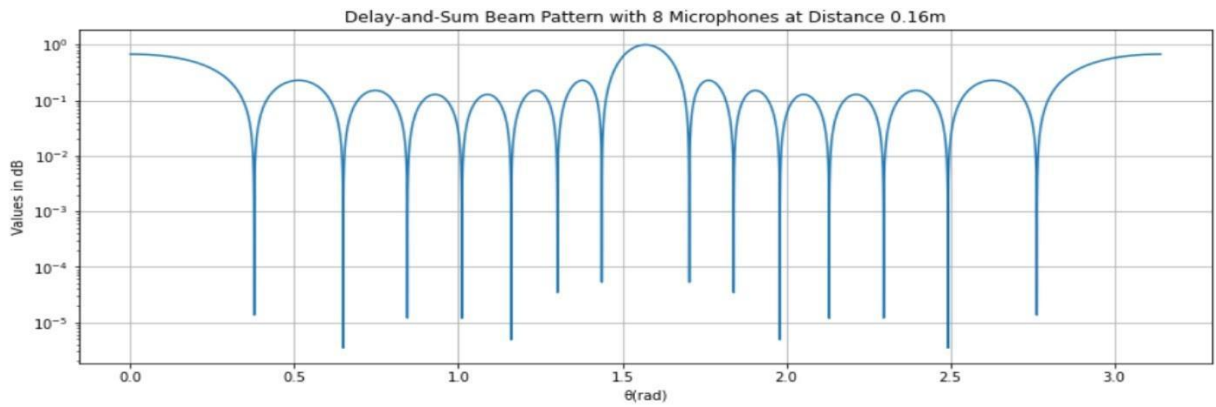
Για  $d=4\text{cm}$ .



Για  $d=8\text{cm}$ .



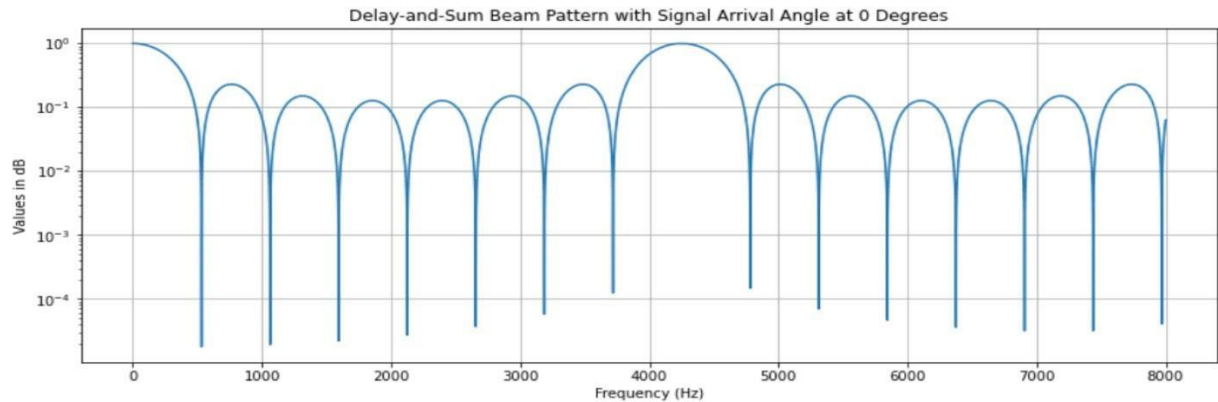
Για  $d=16\text{cm}$ .



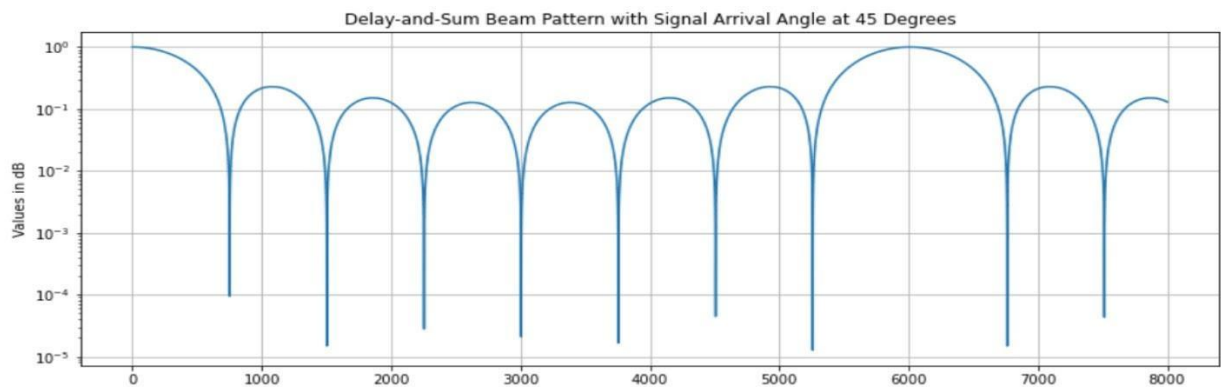
Σύμφωνα με τα συμπεράσματα που προέκυψαν παραπάνω, παρατηρούμε ότι αυξάνοντας την απόσταση μεταξύ των μικροφώνων προκύπτει μικρότερος κύριος λοβός και έτσι έχουμε καλύτερη ποιότητα του συστήματος που υλοποιούμε καθώς ενισχύουμε την επιθυμητή γωνία συγκριτικά περισσότερο από τις άλλες γωνίες. Αυξάνοντας την απόσταση οι παράπλευροι λοβοί είναι γενικά μικρότεροι αλλά κοντά στα άκρα του διαστήματος  $[0,180]$  μεγαλώνουν αρκετά με την αύξηση του  $d$  γεγονός που μπορεί να ενισχύσει θορύβους από τις αντίστοιχες γωνίες.

3) Κρατώντας σταθερό τον αριθμό των μικροφώνων στα  $N=8$  μικρόφωνα με απόσταση μικροφώνων  $d=4\text{cm}$  και  $f=2\text{kHz}$  και βλέποντας την μεταβολή του μέτρου του Beamformer σε λογαριθμική κλίμακα καθώς μεταβάλλουμε την γωνία με την οποία φτάνει στην συστοιχία το σήμα της πηγής  $\theta_s$  έχουμε:

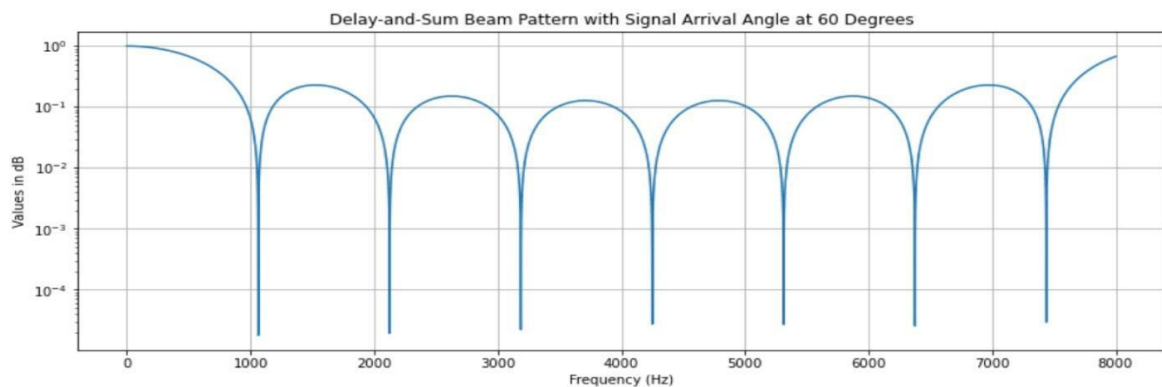
**Για  $\theta_s=0$**



**Για  $\theta_s=45$**



**Για  $\theta_s=60$**



Παρατηρούμε ότι οι αναπαραστάσεις μας έχουν την αναμενόμενη μορφή. Αυτό συμβαίνει διότι από τον τύπο που έχουμε χρησιμοποιήσει για το Beamformer  $B(\omega, \theta)$ , το μέτρο του  $B$  γίνεται '1' στο steering direction. Εξαιτίας αυτής της σχέσης, όταν το μέτρο του Beamformer γίνεται μονάδα, έχουμε ότι στην περίπτωση που  $|\cos\theta \pm \lambda d| \leq 1$  τότε θα υπάρχει γωνία  $\neq \theta_s$  για την οποία θα ισχύει ότι το μέτρο του Beamformer θα γίνεται μονάδα. Αυτό μας οδηγεί στο φαινόμενο του Spatial Aliasing όπου στο μέτρο του Beam pattern εμφανίζονται grating lobes. Για να αποφύγουμε το συγκεκριμένο φαινόμενο αρκεί να επιλέξουμε κατάλληλο συνδυασμό των  $\lambda, d$  έτσι ώστε να πληρείται η σχέση  $d < \lambda/2$ . Εφόσον έχουμε σταθερό  $d=0.04m$  και μήκος κύματος  $\lambda=c*f=340ms*2000Hz=0.17m$  η παραπάνω συνθήκη ικανοποιείται πάντα και γι αυτό δεν παρατηρείται το φαινόμενο που αναφέρθηκε παραπάνω. Ωστόσο καθώς μεταβάλλουμε την γωνία  $\theta_s$  μπορούμε να μετακινήσουμε τους λοβούς σε διαφορετικές γωνίες αλλάζοντας έτσι το ποιες γωνίες ενισχύουν το σύστημά μας. Παρατηρείται επίσης ότι για κάποιες τιμές γωνίας άφιξης σήματος (πχ  $\theta_s = 45^\circ$ ) έχουμε πολύ μεγάλο κύριο λοβό και άρα θα ενισχύουμε σήματα από πολλές παραπλήσιες γωνίες, πράγμα που δεν είναι επιθυμητό διότι μειώνεται η ποιότητα του συστήματός μας.

## **2ο Μέρος**

### **2.1 - A**

Υλοποιούμε τη μέθοδο του delay and sum beamformer για το simulated audio signal που μας δίνεται. Υποθέτουμε ότι έχουμε  $N=7$  μικρόφωνα τα οποία δέχονται όλα στην είσοδό τους ένα θορυβώδες σήμα που με γωνία  $\theta_s = 45^\circ$  φτάνει το καθαρό κομμάτι του σήματος και με  $\theta = 135^\circ$  φτάνει το θορυβώδες κομμάτι του και μία πηγή. Προκειμένου λοιπόν να αποθορυβοποιήσουμε το σήμα μας θα την μέθοδο delay and sum beamformer ως εξής:

Τα μικρόφωνα της γραμμικής συστοιχίας λαμβάνουν το καθαρό σήμα φωνής υπό γωνία  $\theta_s = 45^\circ$  και τα σήματα που τα μικρόφωνα της συστοιχίας λαμβάνουν, περιέχουν θόρυβο και είναι μετατοπισμένα χρονικά κατά μία ποσότητα:  $t_n=(n - N+1)c*d*\cos\theta$ . Επομένως αρχικά ολισθαίνουμε χρονικά τα σήματα αυτά για να μπορέσουν να ευθυγραμμιστούν μεταξύ τους θέτοντας ένα από τα μικρόφωνα ως μικρόφωνο αναφοράς (στην συγκεκριμένη άσκηση το Μικρόφωνο 4). Γι αυτό τον λόγο επειδή αυτή η μετατόπιση χρονικά μπορεί να γίνει κλασματική θα υλοποιούμε το διάνυσμα  $d(k)$  (βάρος) όπου για καθένα από τα μικρόφωνα ( $N$ ) θα πολλαπλασιάζεται με τον Fourier του αντίστοιχου σήματος μικροφώνου και θα το ευθυγραμμίζεται με το καθαρό σήμα. Το  $d(k)$  ορίζεται ως εξής:

για κάθε ένα από τα  $N$  μικρόφωνα στο πεδίο της συχνότητας:

$d(k)= \exp(-j\omega t_n f_s) \quad \forall n \in \{1,2,...,7\}$ . Σημειώνεται ότι το  $\omega \in [-\pi, \pi]$  και πρέπει να πολλαπλασιάσουμε με την συχνότητα δειγματοληψίας του σήματος. Προφανώς το διάνυσμα έχει ίσο μήκος δειγμάτων με το αρχικό μας σήμα.

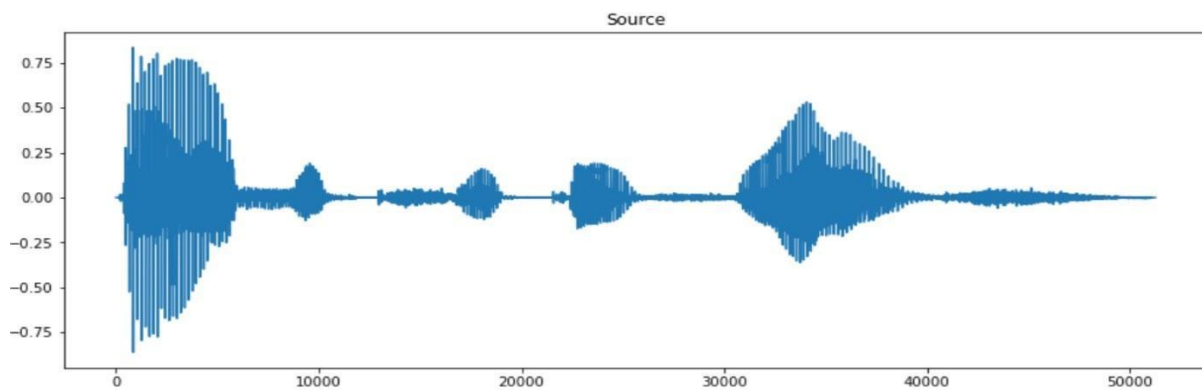
Για κάθε μικρόφωνο υπολογίζουμε τον μετασχηματισμό Fourier του σήματός του και το πολλαπλασιάζουμε με τον αντίστοιχο πίνακα γραμμή του  $d(k)$  στοιχείο με στοιχείο προκειμένου να πάρουμε την απαραίτητη ευθυγράμμιση των σημάτων μας.

Για κάθε ένα μικρόφωνο παίρνουμε τον αντίστροφο μετασχηματισμό Fourier και κρατάμε το πραγματικό μέρος του αντίστροφου μετασχηματισμού. Σε κάθε έξοδο από τα 7 φίλτρα χρονικής μετατόπισης έχουμε ουσιαστικά το σήμα  $f(t)$  που περιγράφεται και στην εκφώνηση της άσκησης και αποτελεί μία προσέγγιση του καθαρού σήματος φωνής που μας δίνεται.

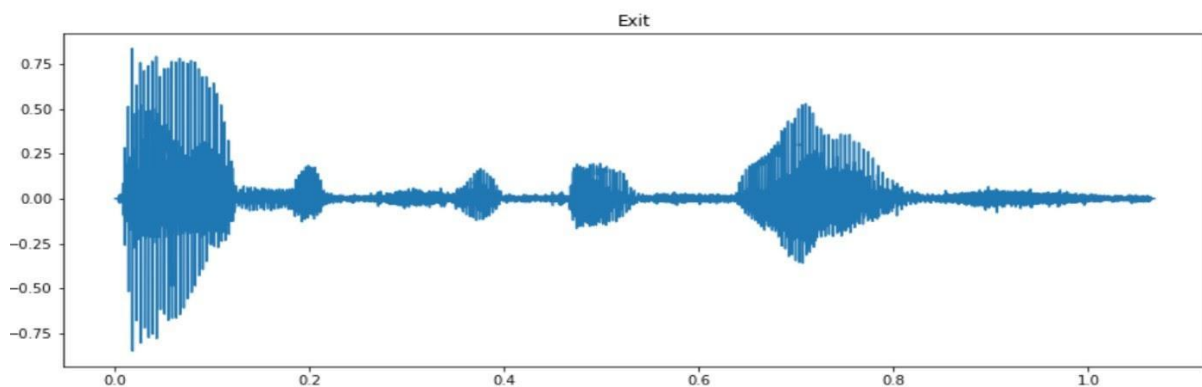
Τέλος, προσθέτουμε όλα αυτά τα σήματα που προέκυψαν από την παραπάνω διαδικασία και βρίσκουμε τον μέσο όρο τους. Αποτέλεσμα όλων αυτών των βημάτων είναι μία πολύ καλή προσέγγιση του σήματος στην έξοδο του delay and sum beamformer με το καθαρό σήμα.

Παίρνοντας την διαφορά της εξόδου και του καθαρού σήματος φωνής source.wav μας προκύπτει αποκλειστικά θόρυβος. Ο θόρυβος φαίνεται μαζί με την έξοδο του συστήματός μας.

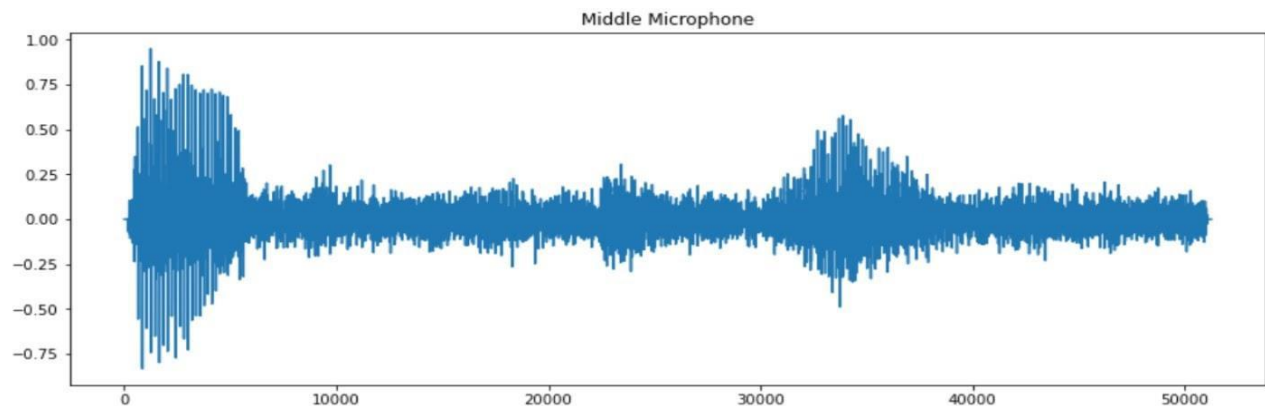
**Το καθαρό σήμα φωνής:**



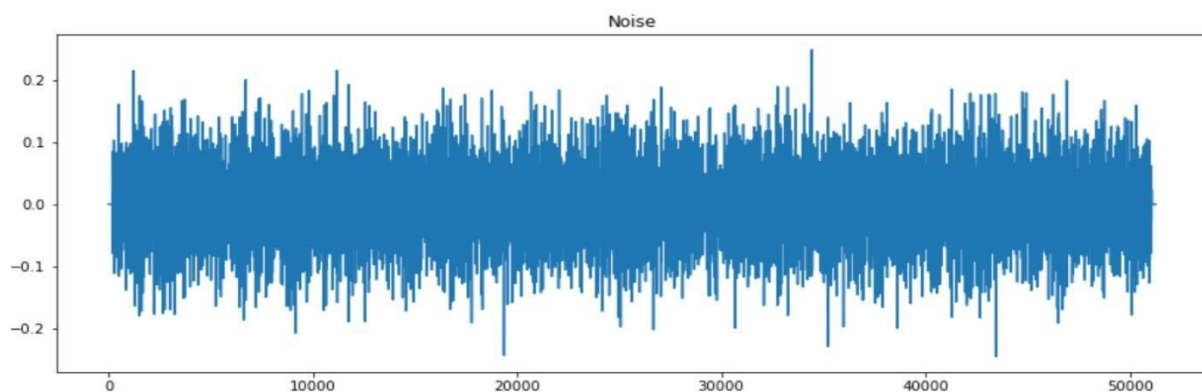
**Η έξοδος του συστήματος Delay and Sum Beamformer:**



**Το θορυβώδες σήμα που προέρχεται από το κεντρικό μικρόφωνο της συστοιχίας:**



**Ο θόρυβος που προκύπτει αφαιρώντας από την έξοδο του συστήματός μας το καθαρό σήμα φωνής πού μας δίνεται:**



Υπολογίζουμε και συγκρίνουμε τα SNR του θορυβώδους σήματος από το κεντρικό μικρόφωνο της συστοιχίας και του σήματος εξόδου από το σύστημα delay and sum Beamformer:

```
SNR of middle sensor signal: 2.9999999759490685
SNR of exit y(t) signal: 20.382190062342502
General SNR of mean and middle: 6.794063408582018
```

**SNR κεντρικού μικροφώνου = 2.999**

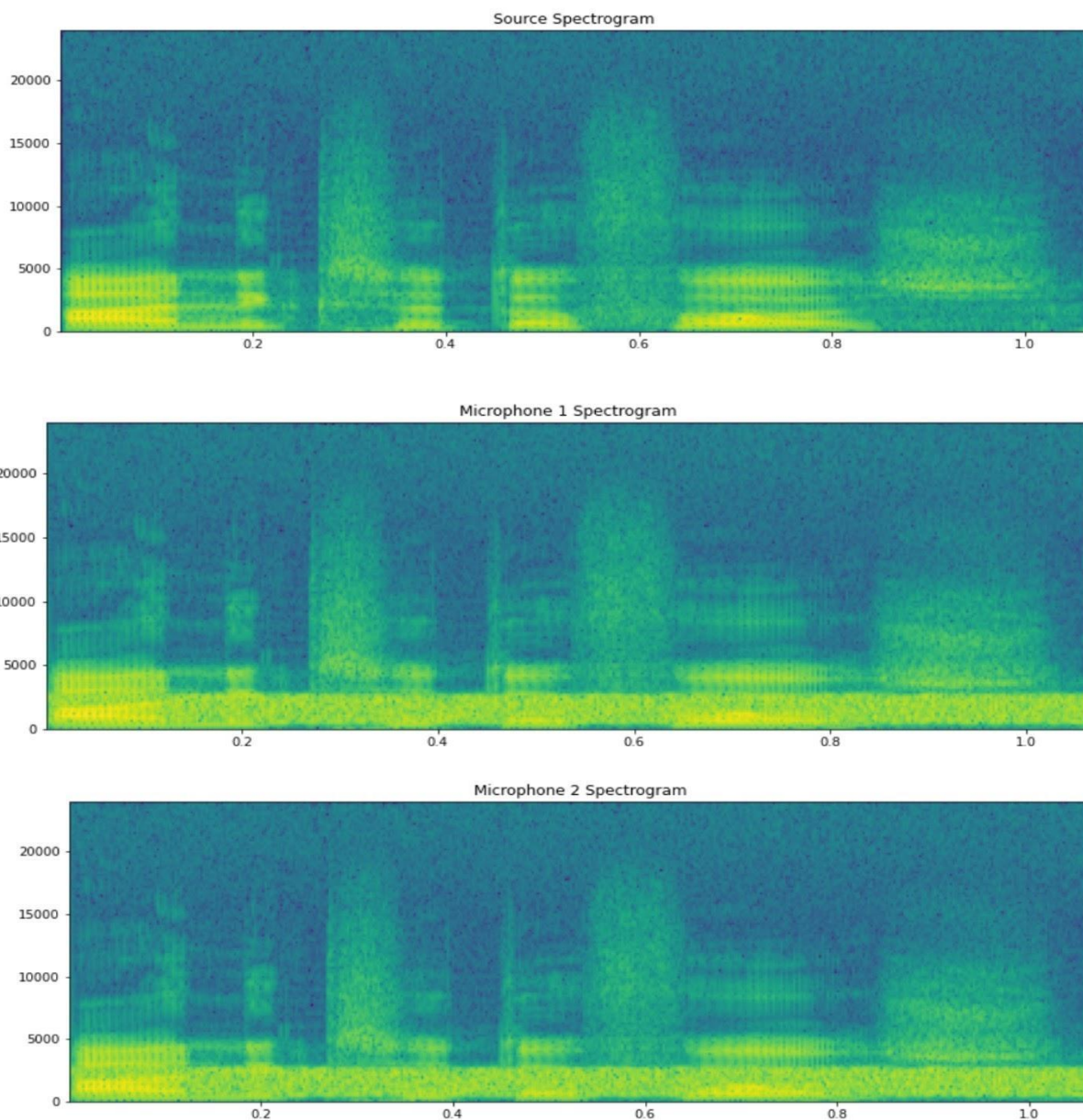
**SNR εξόδου Beamformer = 20.38210**

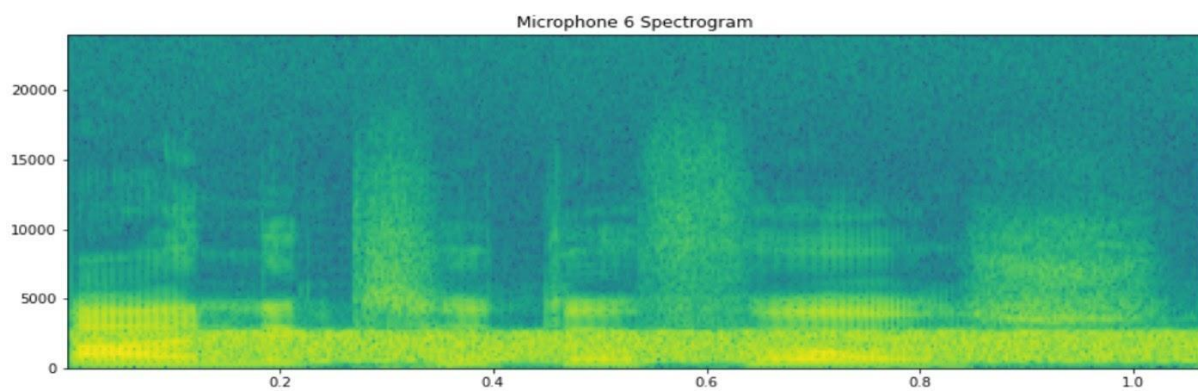
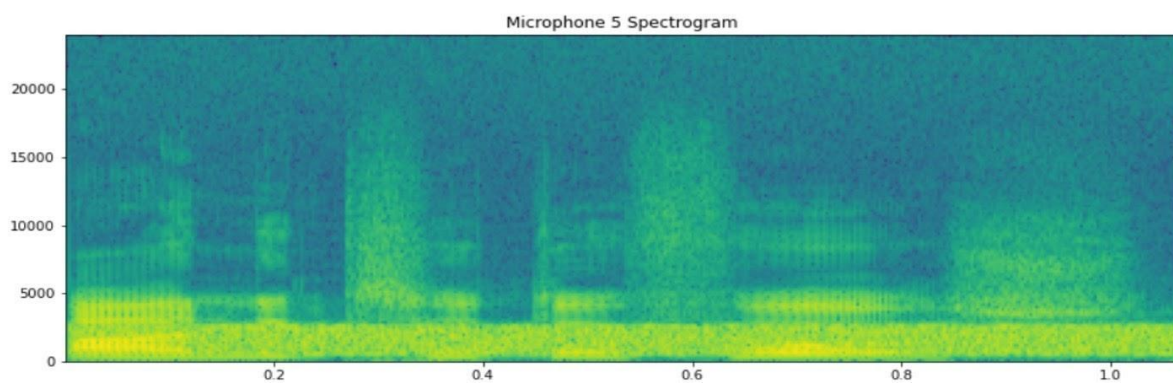
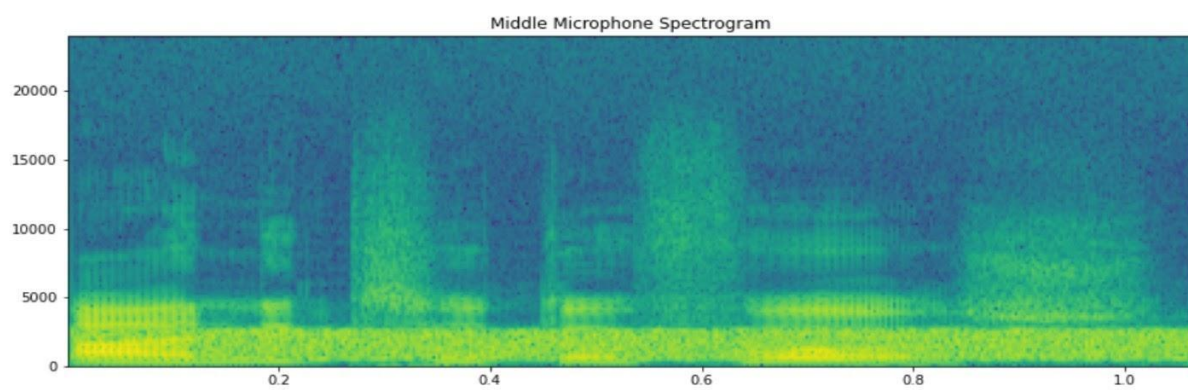
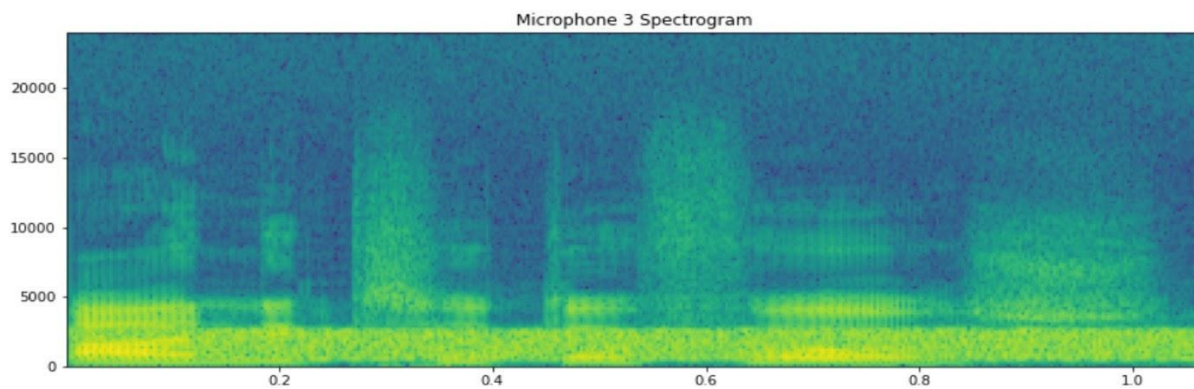
**Λόγος SNR = Beamformer / central mic = 6.794**



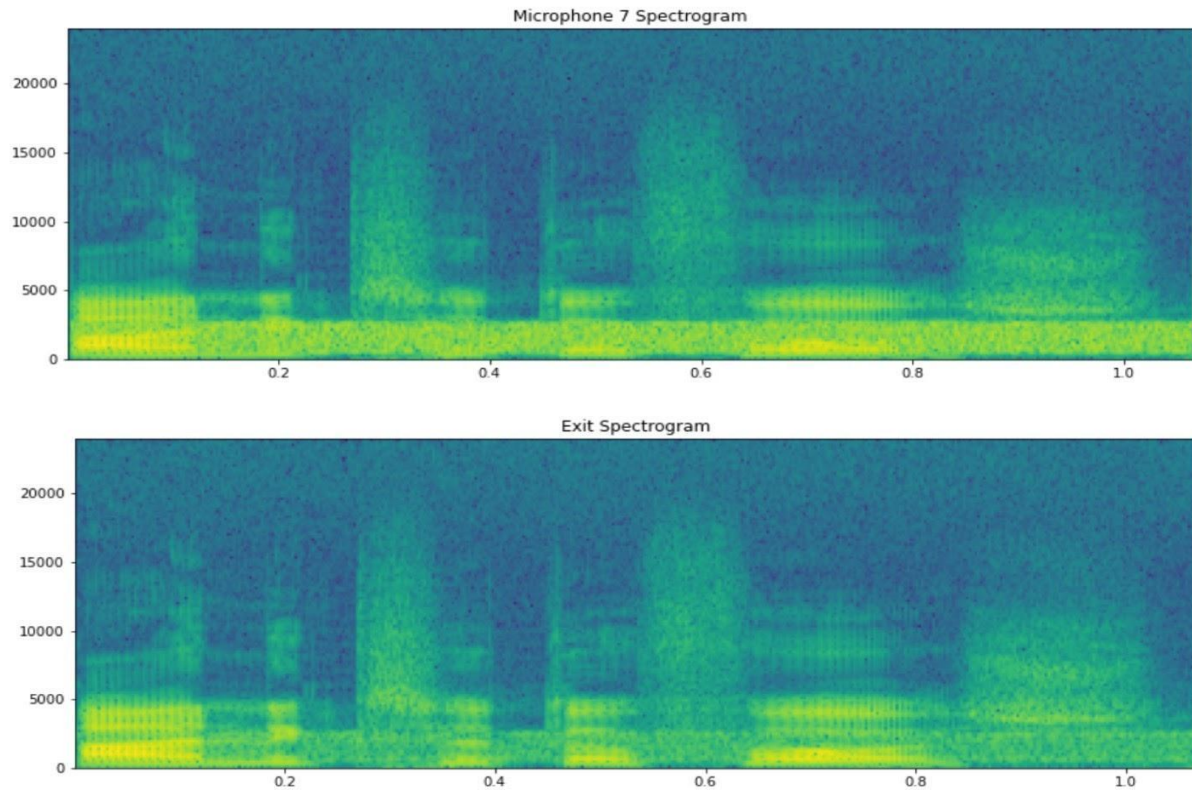
Παρατηρούμε ότι η σηματοθορυβική σχέση, μας δείχνει την σχέση ενέργειας του καθαρού σήματος φωνής και του θορύβου που περιέχεται σε κάθε ένα από τα δύο σήματα που συγκρίνουμε. Έτσι παρατηρούμε ότι η έξοδος του Beamformer έχει πολύ μικρότερο θόρυβο από το κεντρικό μικρόφωνο της συστοιχίας και αυτό οφείλεται στο γεγονός ότι το προσομοιωμένο σήμα, στέλνεται υπό συγκεκριμένη γωνία και έτσι ο beamformer μπορεί να ενισχύσει την κύρια γωνία που φτάνει το σήμα μας μειώνοντας τις επιδράσεις του θορύβου που έρχεται από διαφορετική γωνία. Αντίθετα το κεντρικό μικρόφωνο της συστοιχίας λαμβάνει όλο τον θόρυβο χωρίς να τον μειώνει και έτσι δικαιολογείται η μικρή τιμή του SNR του.

Παρακάτω ακολουθούν τα σπεκτρογραφήματα των σημάτων προς ανάλυση αλλά και των τελικών σημάτων:









Το καθαρό σήμα μοιάζει με το σήμα εξόδου του beamformer, κάτι που αποδεικνύει την αποτελεσματικότητα της μεθόδου για αποθορυβοποίηση του σήματος καθώς ο θόρυβος που ανήκει στις συχνότητες  $[500, 2500]$  Hz έχει σχεδόν εξαφανιστεί και γι' αυτό το σήμα εξόδου ακούγεται πολύ πιο καθαρό.

## 2.1 - B

1) Παίρνουμε το πλαίσιο **[0.75,0.78]** από το κεντρικό μικρόφωνο της συστοιχίας και το περνάμε από το φίλτρο Wiener που κατασκευάζουμε στη συνέχεια.

Για την υλοποίηση του φίλτρου Wiener ακολουθούμε την εξής διαδικασία :

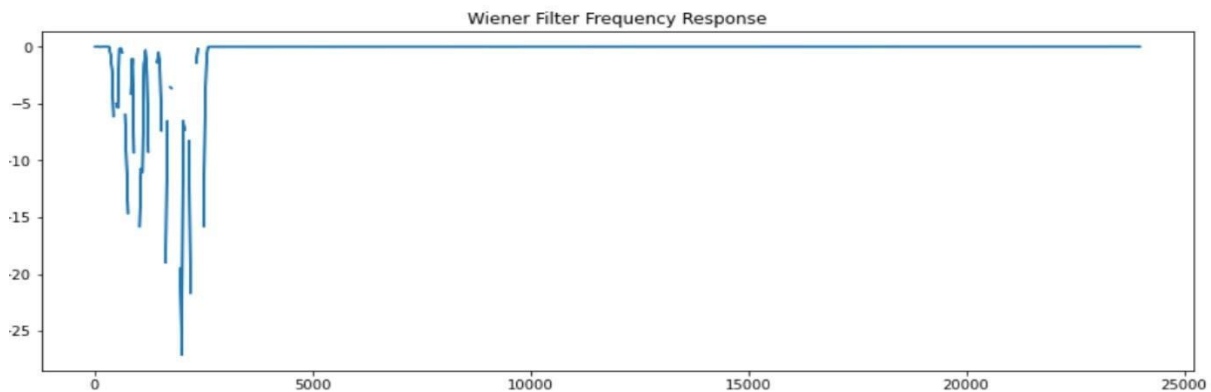
-Υπολογίζουμε το φάσμα ισχύος του σήματος από το κεντρικό μικρόφωνο ( $x(t)$ )

-Υπολογίζουμε το φάσμα ισχύος του καθαρού σήματος ( $s(t)$ )

-Παίρνουμε το πηλίκο των δυο φασμάτων ισχύος.

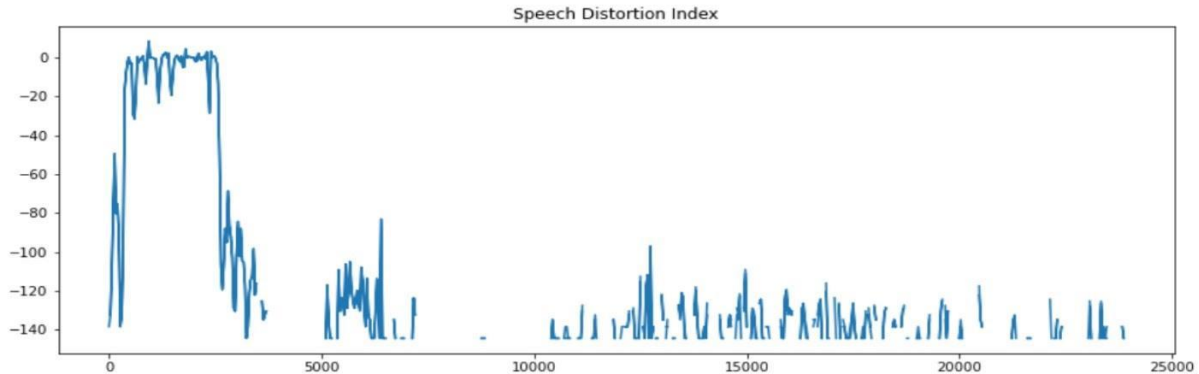
Με τη χρήση της συνάρτησης `scipy.signal.welch` παίρνουμε την απόκριση συχνότητας του φίλτρου Wiener:

```
[<matplotlib.lines.Line2D at 0x7f5d98284220>,  
 <matplotlib.lines.Line2D at 0x7f5d982840d0>,  
 <matplotlib.lines.Line2D at 0x7f5d98284250>,  
 <matplotlib.lines.Line2D at 0x7f5d982843a0>]
```



2) Για τον υπολογισμό του speech distortion index εφαρμόζουμε την παρακάτω σχέση όπου προκύπτει:

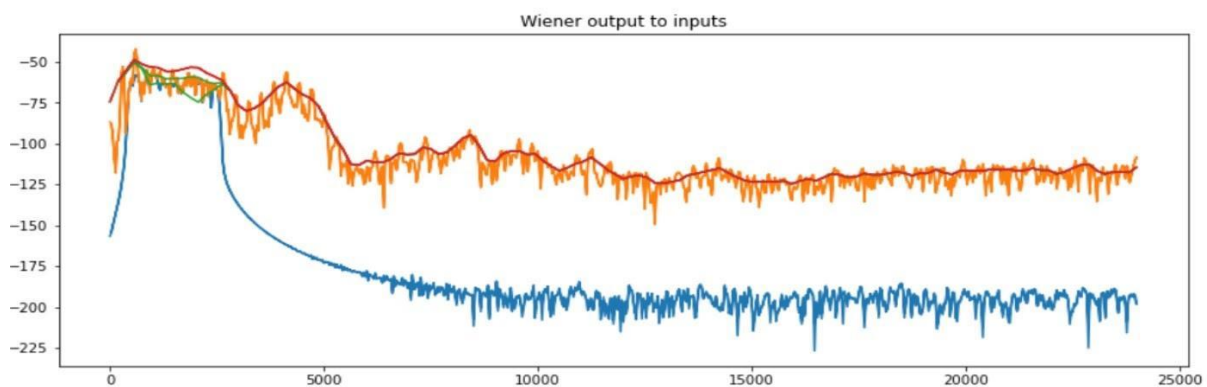
$$\eta(w) = |1 - Hw|^2$$



Στις συχνότητες ομιλίας του ανθρώπου έχουμε καλή απόκριση στο speech distortion (20-250 Hz) και στις συχνότητες θορύβου έχουμε αρκετά μεγάλο distortion και συνεπώς αρκετά μεγάλη παραμόρφωση.

Να αναφέρουμε πως για να πάρουμε τις συχνότητες [0-8kHz] , παίρνουμε το 1/6 των δειγμάτων καθώς η συχνότητα δειγματοληψίας είναι 48.000 , κάτι που αντιστοιχεί στο 0-2π κανονικά, αλλά εφόσον θέλουμε 0-8kHz θα πάρουμε  $48.000/6 = 8.000$ .

3) Υπολογίζουμε και σχεδιάζουμε το καθαρό σήμα φωνής, το θορυβώδες σήμα στην είσοδο του Wiener, την έξοδο του Wiener καθώς και το σήμα θορύβου σε ένα γράφημα.



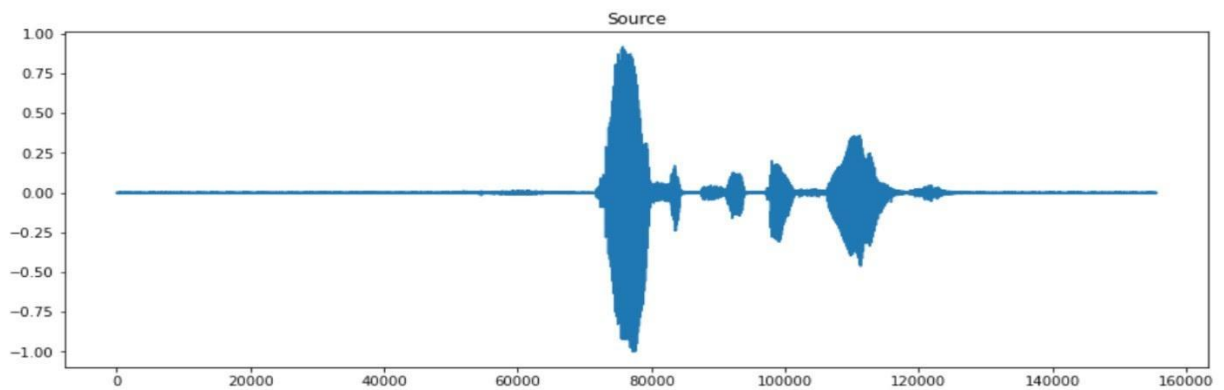
Παρατηρούμε ότι η έξοδος του φίλτρου , έχει κάποια απόκλιση από την ιδανική έξοδο σήματος χωρίς θόρυβο, και κοντά στις συχνότητες 6000-8000 αυτή η απόκλιση ελαχιστοποιείται.

## 2.2 - A

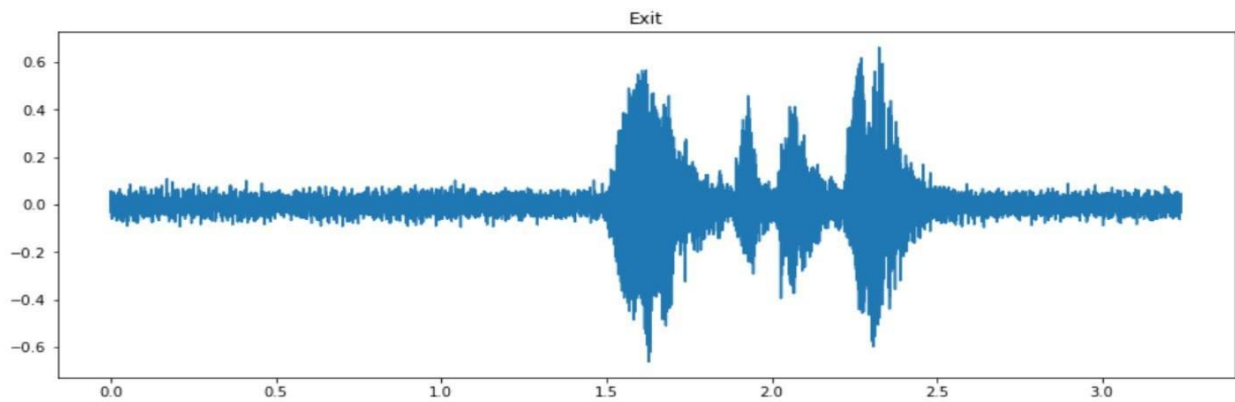
1) Υπολογίζουμε τα βάρη του delay and sum beamformer, ακολουθώντας την ίδια διαδικασία με αυτή του ερωτήματος 2.1.A. Έτσι έχουμε:

2) Παρακάτω ακολουθούν τα εξής γραφήματα:

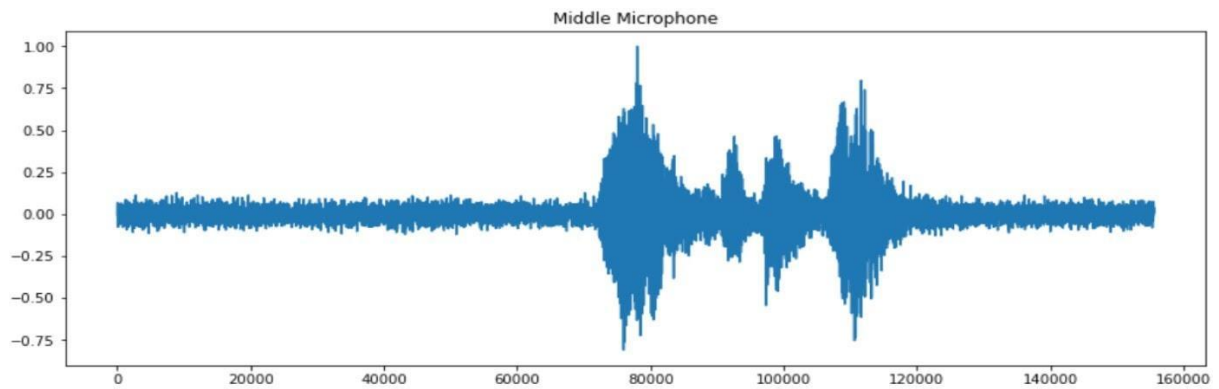
**Καθαρό σήμα φωνής :**



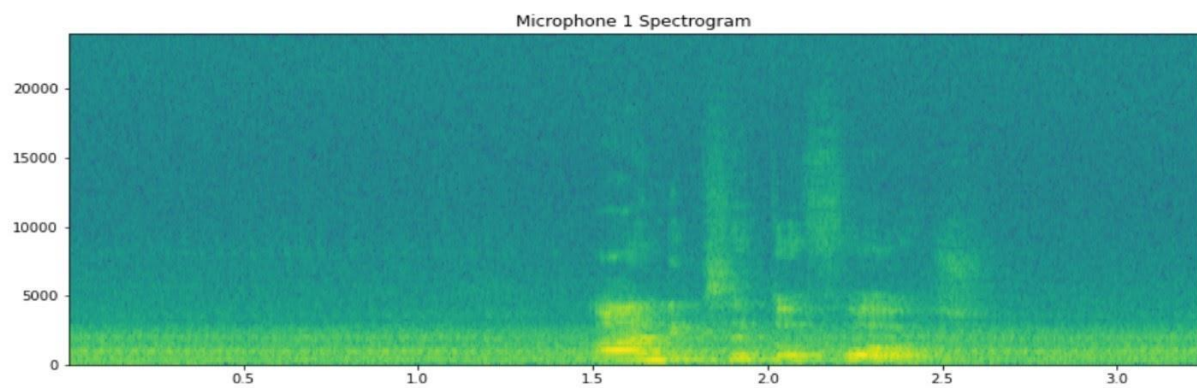
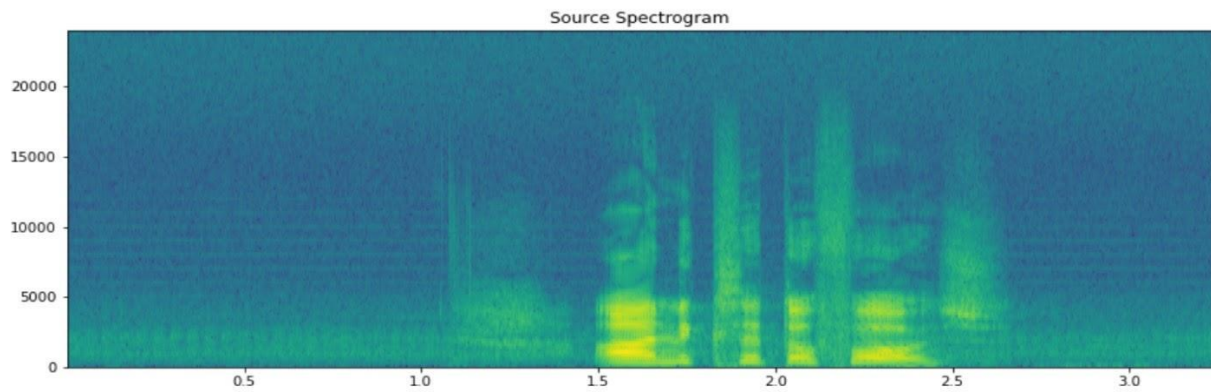
**Έξοδος του beamformer :**



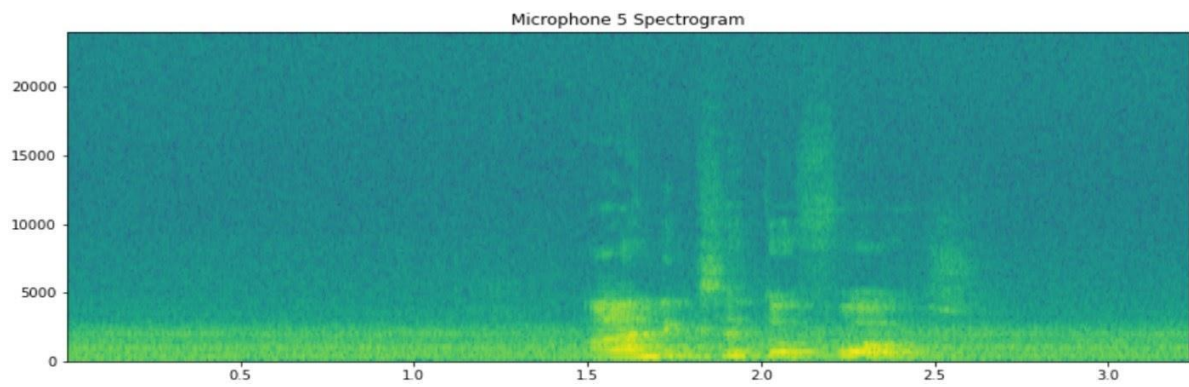
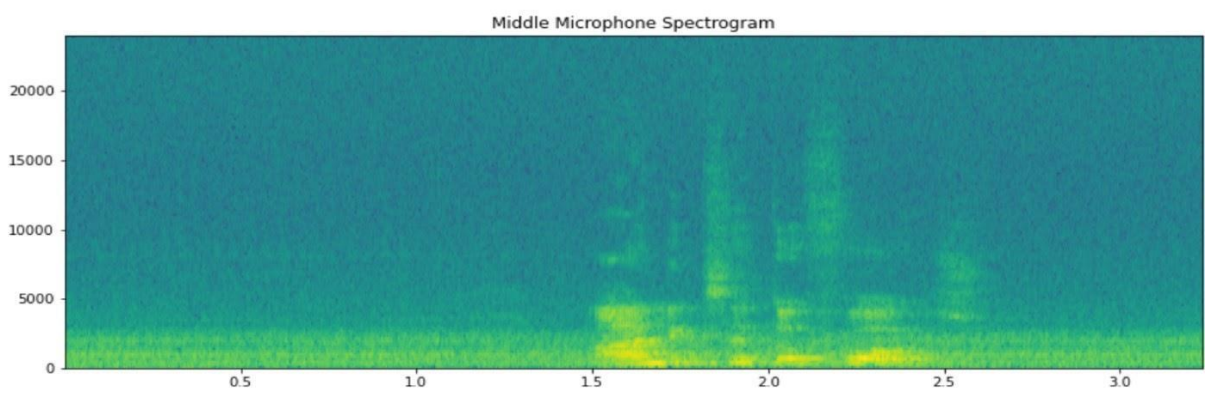
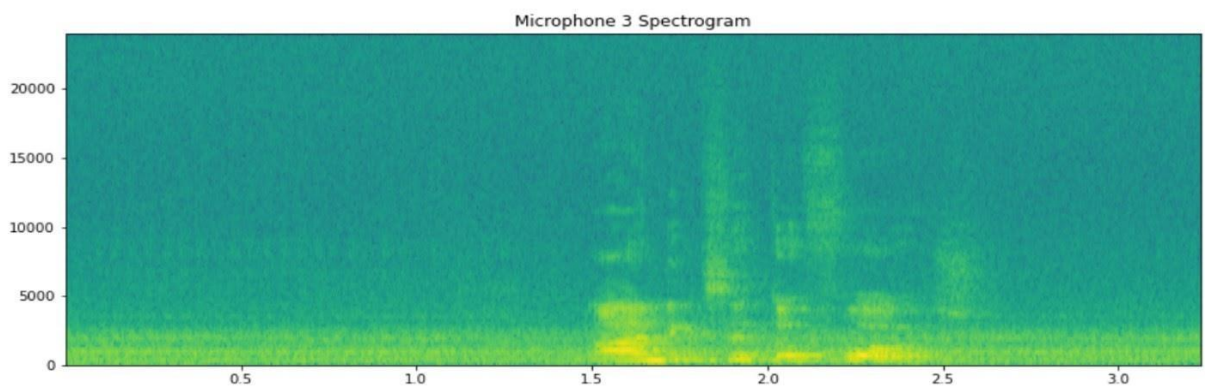
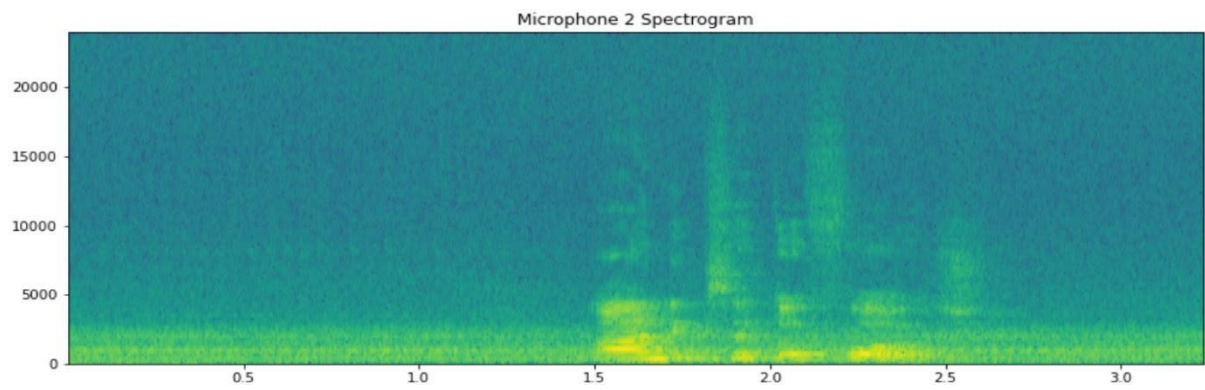
### Σήμα κεντρικού μικροφώνου:



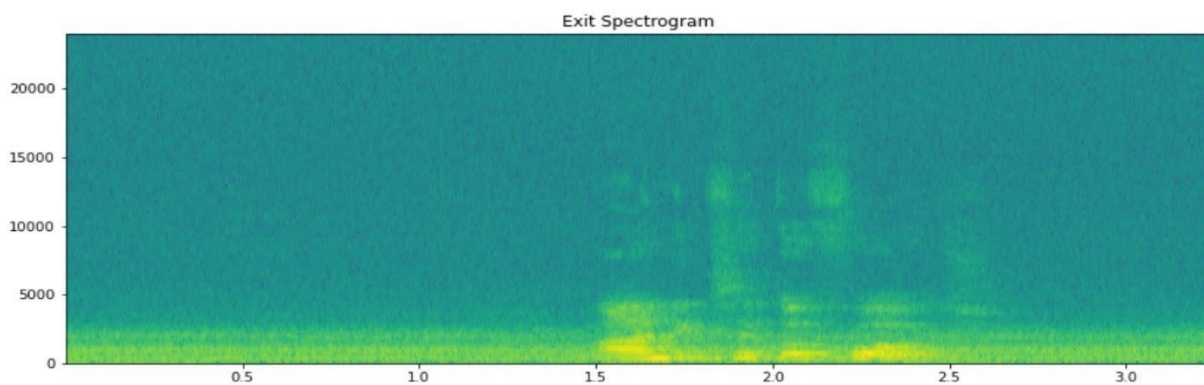
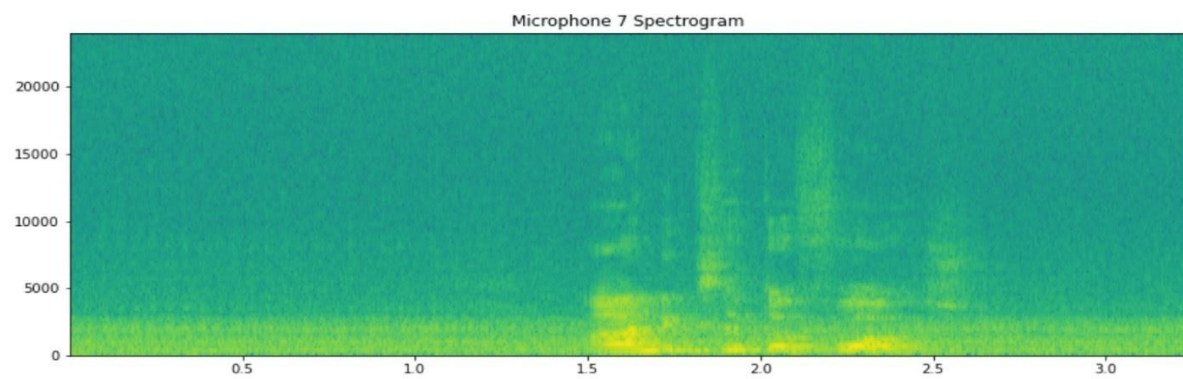
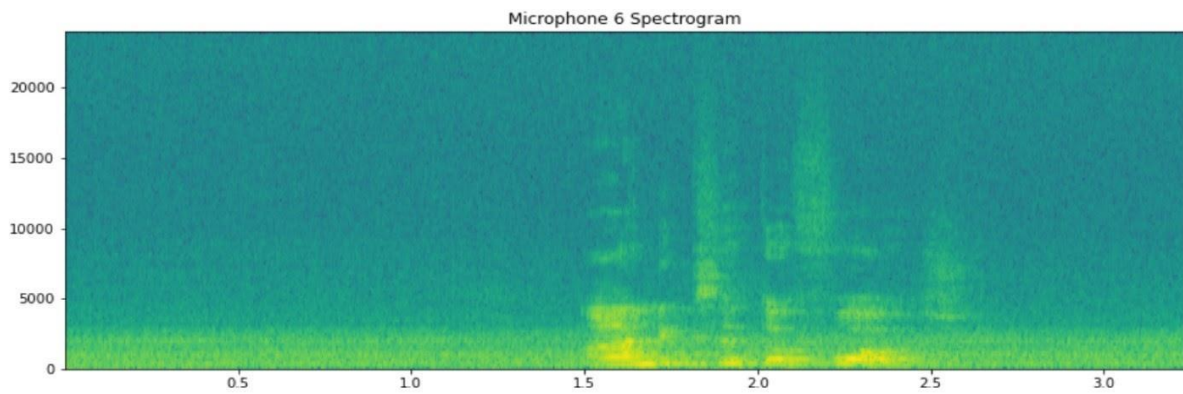
Το σήμα μετά το beamforming, προσεγγίζει σε μεγάλο βαθμό το καθαρό σήμα , κάτι που φαίνεται και από τα παρακάτω σπεκτρογραφήματα :











3) Δίνουμε ως είσοδο το πλαίσιο του καθαρού σήματος, και το πλαίσιο του θορύβου , και υπολογίζουμε το  $snr$  τους.

Κάνουμε το ίδιο για κάθε ένα από τα πλαίσια και στη συνέχεια τα προσθέσουμε και διαιρούμε με το πλήθος.

Αν ο  $snr$ , είναι κάτω από το κατώφλι που έχουμε ορίσει, τότε το συγκεκριμένο πλαίσιο δεν λαμβάνει μέρος στον υπολογισμό του  $snr$ .

Μετά από αυτή τη διαδικασία έχουμε :

---

```
SSNR of Central Microphone is: 8.146842586237858  
SSNR of Beamformer Result is: 9.094192855028705
```

**SSNR(κεντρικού μικροφώνου) = 8.14684**

**SSNR(έξοδος beamformer) = 9.0941**

Παρατηρείται αισθητή βελτίωση ακουστικά όμως η μέθοδος του delay-and-sum beamforming δεν αποτελεί την βέλτιστη καθώς υπάρχει διάχυση του ήχου στον χώρο.