# Ψηφιακή Επεξεργασία Σήματος

## 3η Εργαστηριακή Άσκηση

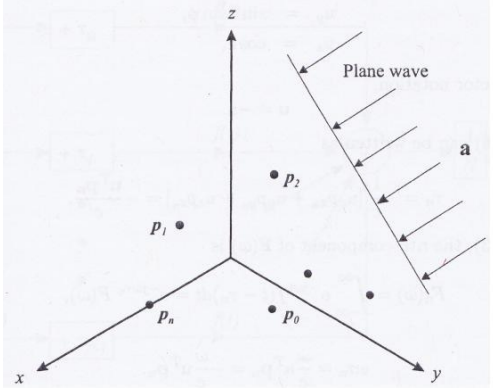
**Θέμα: Συστοιχίες Μικροφώνων (Microphone Arrays) και Πολυκαναλική Επεξεργασία Σημάτων (Multichannel Signal Processing)**

Διαμάντη Ιωάννα

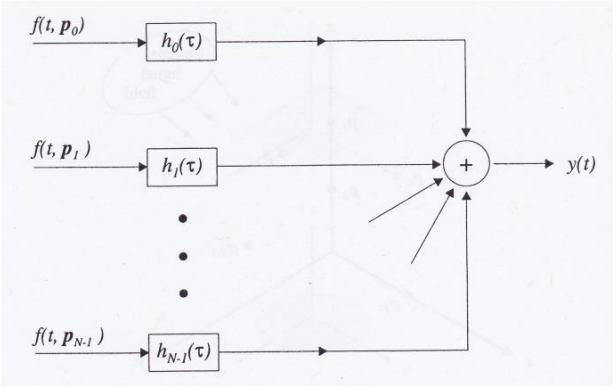
ΑΜ: 03115035

## Εισαγωγή

Η χρήση συστοιχιών μικροφώνων για καταγραφή και επεξεργασία ακουστικών σημάτων γίνεται όλο και πιο διαδεδομένη. Το πλεονέκτημα που παρουσιάζει η χρήση μικροφώνων κατανεμημένων στο χώρο είναι η δυνατότητα καταγραφής και αξιοποίησης όχι μόνο των χρονικών (temporal), αλλά και των χωρικών χαρακτηριστικών (spatial characteristics) των ακουστικών σημάτων. Τα χωρικά χαρακτηριστικά αυτά μπορούν να αξιοποιηθούν σε εφαρμογές όπως εύρεση κατεύθυνσης και εντοπισμός θέσης ακουστικής πηγής (direction-finding και source localization), αποθορυβοποίηση σημάτων ομιλίας (speech enhancement) κ.ά. Η εφαρμογή που θα μελετηθεί στην παρούσα άσκηση είναι το speech enhancement. Με τη χρήση συστοιχιών μπορεί να γίνει χωρικό φιλτράρισμα των ακουστικών σημάτων προκειμένου να ενισχυθούν ή να απορριφθούν σήματα που καταφθάνουν στη συστοιχία από συγκεκριμένη κατεύθυνση. Αυτό επιτυγχάνεται με κατάλληλο συνδυασμό των σημάτων που καταγράφονται από τα διάφορα μικρόφωνα ώστε το επιθυμητό σήμα που καταφθάνει από συγκεκριμένη κατεύθυνση να ενισχυθεί με ενισχυτική συμβολή, ενώ θόρυβος από τις υπόλοιπες κατευθύνσεις να εξασθενηθεί με αποσβεστική συμβολή.




Σχήμα 1: Συστοιχία μικροφώνων




Σχήμα 2: Πολυκαναλική επεξεργασία σημάτων

## Μέρος 1. Συστοιχίες Μικροφώνων και Χωρικό Φιλτράρισμα (Spatial Filtering)

### Beamforming

Το χωρικό φιλτράρισμα είναι μία τεχνική επεξεργασίας σημάτων, η οποία χρησιμοποιείται σε συστοιχίες κεραιών για κατευθυντική μετάδοση ή λήψη σημάτων. Αυτό επιτυγχάνεται συνδυάζοντας στοιχεία σε μία συστοιχία κεραίας έτσι ώστε σήματα προερχόμενα από συγκεκριμένες γωνίες να υπόκεινται σε εποικοδομητική, ενώ σήματα από άλλες γωνίες να υπόκεινται σε καταστροφική παρεμβολή. Το beamforming μπορεί να χρησιμοποιηθεί και στα δύο άκρα (πομπός και δέκτης) προκειμένου να επιτευχθεί χωρική επιλεκτικότητα. Μία συστοιχία N μικροφώνων, τα οποία βρίσκονται στα σημεία του χώρου $p_n$ , n = 0,1,…,N-1 δειγματολειπτεί το ακουστικό πεδίο, καταγράφοντας ένα σύνολο σημάτων:

$$\mathbf{f}(t,\mathbf{p}) = [f(t,p_0), f(t,p_1) , \dots , f(t,p_{N-1})]^T.$$

Κάθε καταγεγραμμένο σήμα φιλτράρεται από ένα γραμμικό, χρονικά αναλλοίωτο φίλτρο με κρουστική απόκριση $h_n(t)$ και στη συνέχεια τα σήματα αθροίζονται δίνοντας την τελική έξοδο:

$$y(t) = \sum_{n=0}^{N-1} \int_{-\infty}^{\infty} h_n(t-\tau) f(\tau, p_n) d\tau = \int_{-\infty}^{\infty} \boldsymbol{h}^T(t-\tau)\, f(\tau, p_n) \mathrm{d}\tau$$

όπου

$$\mathbf{h}(\tau) = [h_0(\tau), h_1(\tau), \dots , h_{N-1}(\tau)]^T.$$

Στο πεδίο της συχνότητας επομένως είναι:

$$Y(\omega) = \mathbf{H}^T(\omega)\mathbf{F}(\omega), \quad \text{με} \quad \mathbf{H}(\omega) = [H_0(\omega), \dots, H_{N-1}(\omega)]^T , \mathbf{F}(\omega) = [F_0(\omega) , \dots , F_{N-1}(\omega)]^T, \quad (1)$$

όπου $H_n(\omega) = \mathcal{F}\{h_n(t)\}$ και $F_n(\omega) = \mathcal{F}\{f(t,p_n)\}$.

Για δεδομένη γεωμετρία της συστοιχίας μικροφώνων, η επιλογή των φίλτρων $H_n(\omega)$ καθορίζει τα χαρακτηριστικά του χωρικού φιλτραρίσματος που επιτυγχάνεται. Οπότε το πρόβλημα σχεδιασμού ενός beamformer , ώστε το χωρικό φίλτρο που προκύπτει να έχει συγκεκριμένα χαρακτηριστικά, έγκειται στην επιλογή των φίλτρων $H_n(\omega)$.

Λόγω της καθυστέρησης διάδοσης του ηχητικού σήματος, κάθε μικρόφωνο καταγράφει το ηχητικό σήμα με μία χρονική μετατόπιση σε σχέση με τα υπόλοιπα μικρόφωνα. Έτσι θεωρούμε το διάνυσμα d, με $\mathrm{d}(k) = [e^{-jk^T p_0} , e^{-jk^T p_1} , \dots , e^{-jk^T p_{N-1}}]^T$ , όπου $k = \frac{\omega}{c}$ a (a : κατεύθυνση άφιξης του σήματος, c = 340m/s η ταχύτητα του ήχου στον αέρα). Το διάνυσμα αυτό περιέχει όλη την πληροφορία των χωρικών χαρακτηριστικών της συστοιχίας και ονομάζεται array manifold vector. Όποτε η παραπάνω εξίσωση (1) γίνεται:

$$Y(\omega,\alpha) = \mathbf{H}^T(\omega)\mathrm{d}(k)\mathbf{F}(\omega)$$

Το beam pattern του beamformer, το οποίο είναι το χωρικό ανάλογο της απόκρισης συχνότητας ενός χρονικού φίλτρου είναι :

$$B(\omega,a) = \mathbf{H}^T(\omega)\mathrm{d}(k)|_{k=\frac{\omega}{c}a}$$

Το B(ω,a) περιγράφει πλήρως τη χωρο-χρονική επεξεργασία που γίνεται από τον beamformer.

Το delay-and-sum beam pattern για μία ομοιόμορφη γραμμική συστοιχία μικροφώνων (μικρόφωνα πάνω σε μία ευθεία που ισαπέχουν) είναι:

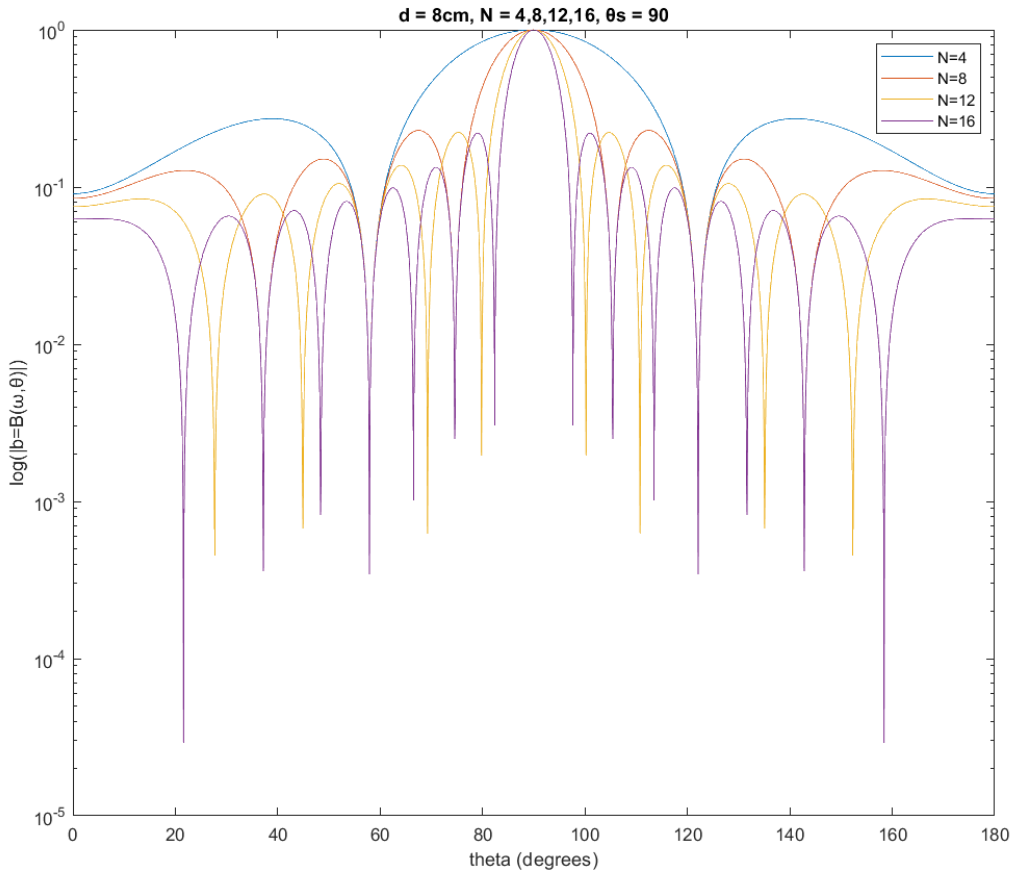
$$B(\omega,\theta) = \frac{1}{N} \frac{\sin[\frac{N\omega}{2\,c} d(cos\theta - cos\theta_s)]}{\sin[\frac{1\omega}{2\,c} d(cos\theta - cos\theta_s)]}$$

,όπου d η απόσταση των μικροφώνων, θs η γωνία του επιθυμητού σήματος (κατεύθυνση $a_s$).

**1.4 Μελέτη χαρακτηριστικών του delay-and-sum beam pattern για ομοιόμορφη γραμμική συστοιχία**

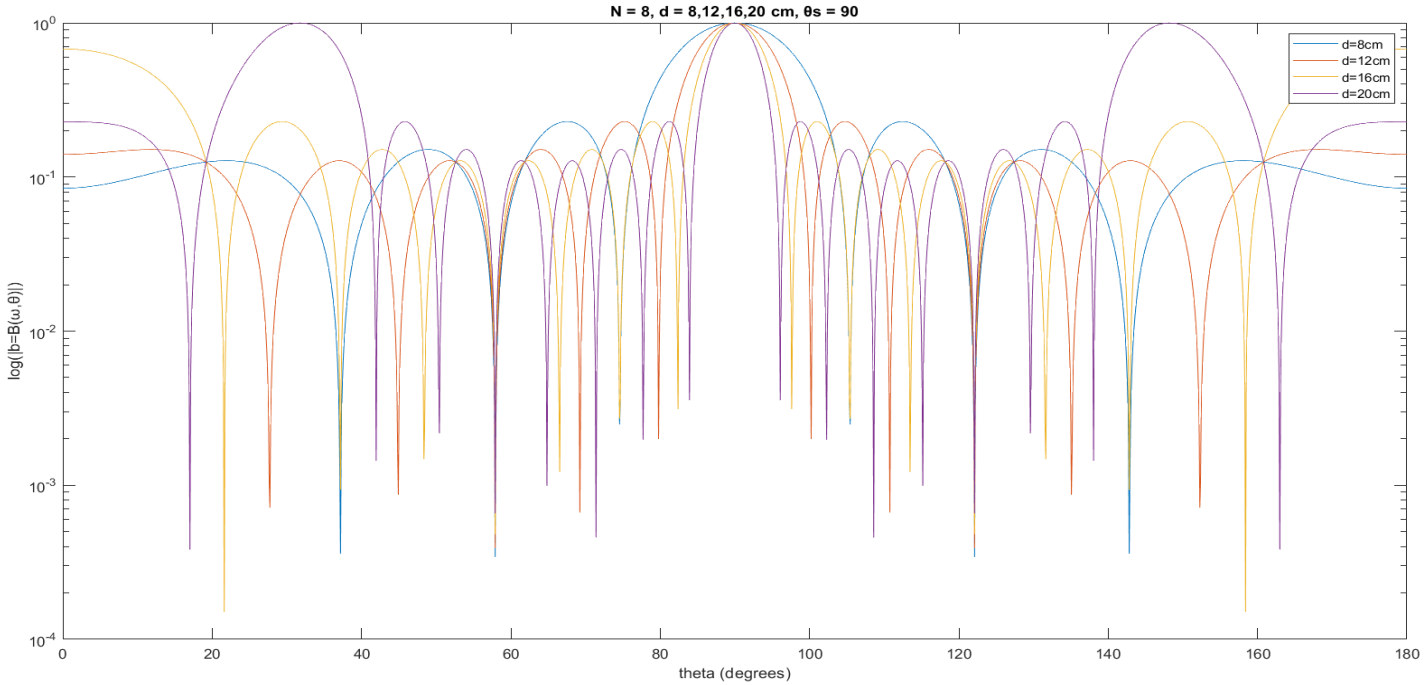
Θεωρούμε αρχικά ότι το επιθυμητό σήμα καταφθάνει στη συστοιχία από γωνία $\theta_s = 90^o$ και έχει συχνότητα f = 2kHz.

<u>Μεταβολή αριθμού μικροφώνων N = 4, 8, 12, 16 με σταθερή απόσταση d = 8cm</u>




Παρατηρούμε ότι καθώς ο αριθμός των μικροφώνων αυξάνεται, οι λοβοί του beam pattern στενεύουν όλο και περισσότερο γύρω από τις 90$^o$, δηλαδή το πλάτος του αποσβένει όλο και πιο γρήγορα και τα σήματα που προέρχονται από διαφορετική κατεύθυνση από αυτή του επιθυμητού αποκόβονται όλο και περισσότερο.

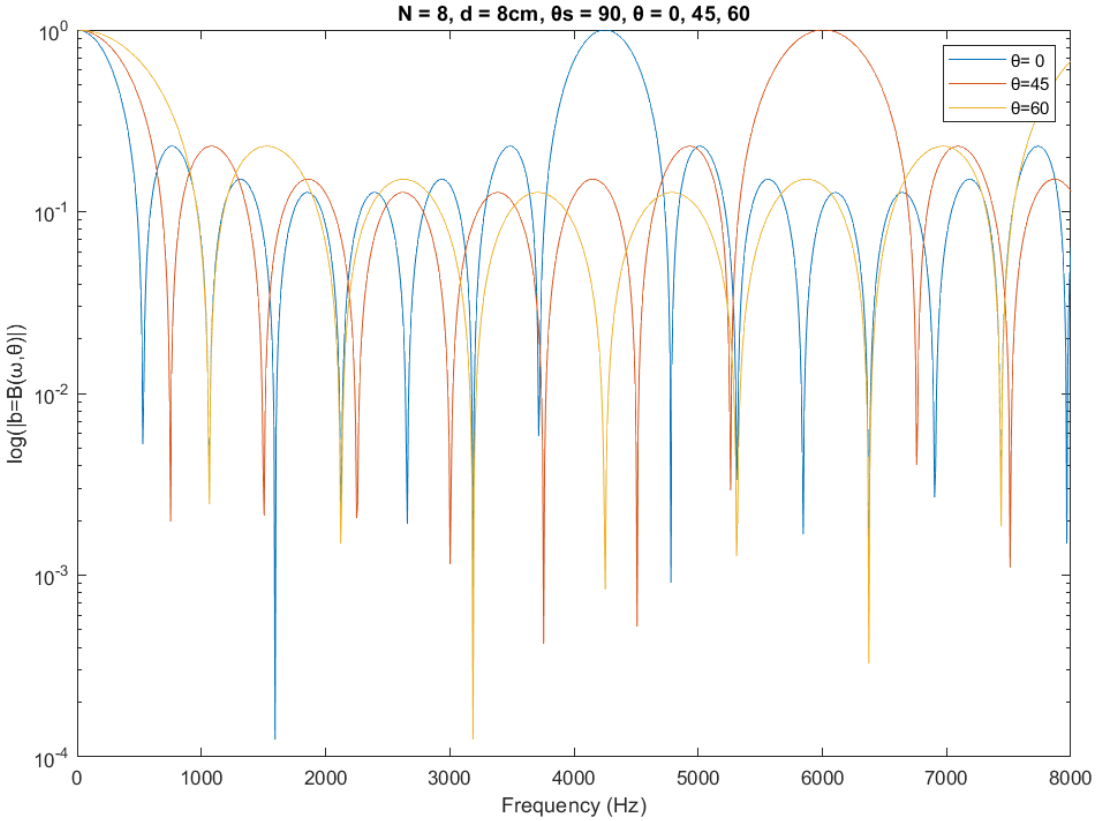
Μεταβολή απόστασης μικροφώνων d = 8, 12, 16, 20cm με σταθερό πλήθος N = 8




Παρατηρούμε πως για d = 8, 12 cm έχουμε πιο γρήγορη απόσβεση του πλάτους όσο οι γωνίες απομακρύνονται από τις 90°. Παρόλαυτα, για d = 16, 20 cm παρατηρούμε ότι το πλάτος αυξάνεται απότομα στις 0, 30, 150 και 180 μοίρες, δηλαδή ενισχύονται σήματα από μη επιθυμητές κατευθύνσεις.

Θεωρούμε τώρα ότι η συστοιχία αποτελείται από N = 8 μικρόφωνα με απόσταση d = 8cm και ότι όπως και προηγουμένως $\theta_s = 90^o$.

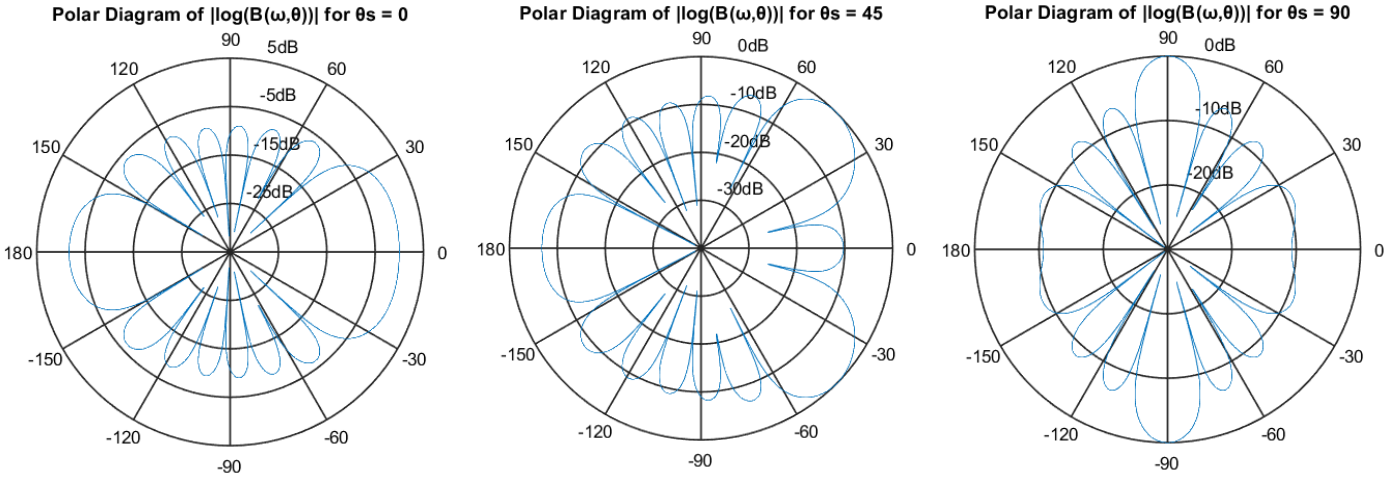
Μεταβολή γωνίας άφιξης σημάτων θ = 0°, 45°, 60° , f ∈ [0, 8kHz]



Παρατηρούμε ότι όσο αυξάνεται η γωνία άφιξης του σήματος θ, τόσο μεγαλώνει το πλάτος των λοβών. Αυτό είναι λογικό, καθώς όσο η γωνία θ προσεγγίζει την επιθυμητή γωνία $\theta_s$ ($90^o$), τόσο η κατεύθυνση του σήματος προσεγγίζει αυτήν του επιθυμητού και συνεπώς λαμβάνεται όλο και περισσότερο υπόψιν από τη συστοιχία.

Θεωρούμε ότι η συστοιχία αποτελείται από N = 8 μικρόφωνα με απόσταση d = 8cm. Συχνότητα f = 2kHz

Μεταβολή γωνίας άφιξης επιθυμητού σήματος $\theta_s = 0^o, 45^o, 90^o$ , $\theta \in [-180^o , 180^o]$




Παρατηρούμε ότι, όπως ήταν αναμενόμενο, σε κάθε πολικό διάγραμμα ο μεγαλύτερος λοβός εμφανίζεται στις γωνίες που αντιστοιχούν στην επιθυμητή γωνία άφιξης $\theta_s$. Επίσης παρατηρούμε ότι καθώς η επιθυμητή γωνία άφιξης προσεγγίζει τις $90^o$, οι λοβοί στενεύουν, έχουμε δηλαδή καλύτερη απόσβεση.

## Μέρος 2. Εφαρμογή Beamforming για Speech Enhancement

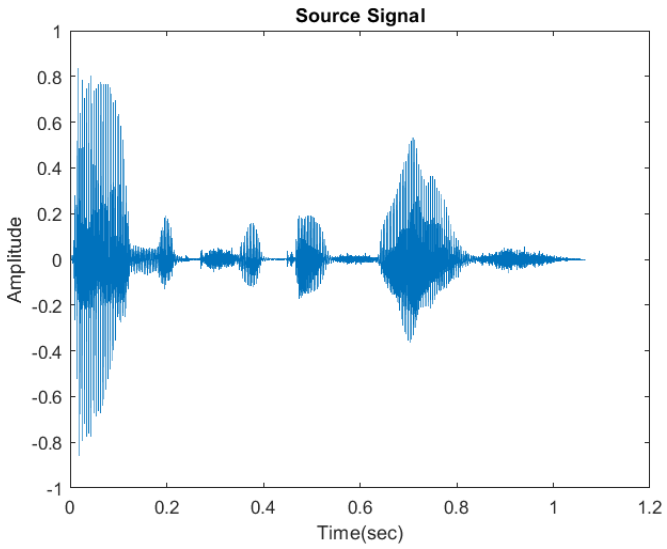
### 2.1 Beamforming σε προσομοιωμένα σήματα

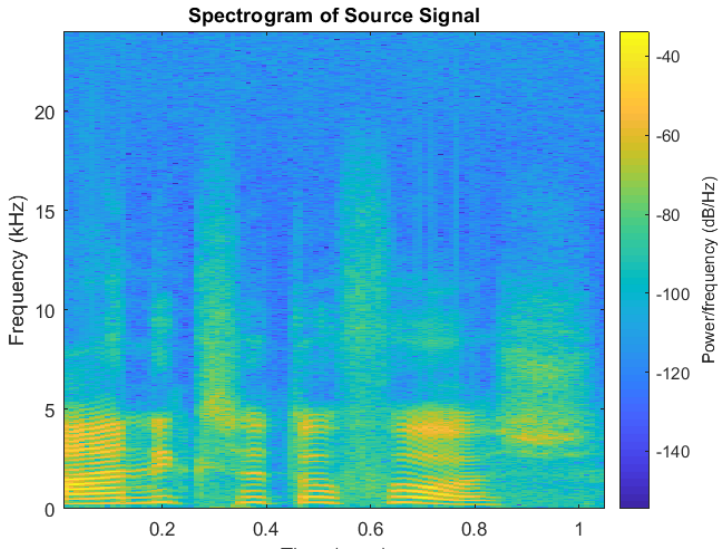
Θεωρούμε ότι μία γραμμική συστοιχία N = 7 μικροφώνων στοιχείων με απόσταση d = 8cm καταγράφει σήματα που παράγονται από δύο σημειακές πηγές. Η μία πηγή παράγει ένα σήμα φωνής που βρίσκεται σε γωνία $\theta_s = 45^o$ σε σχέση με τη συστοιχία, ενώ η δεύτερη πηγή παράγει ένα σήμα θορύβου και βρίσκεται σε γωνία $\theta = 135^o$ σε σχέση με τη συστοιχία. Ο θόρυβος είναι ζωνοπερατός και η ενέργεια του είναι συγκεντρωμένη στις συχνότητες f ∈ [500Hz, 2.5kHz]. Τα σήματα πηγής και θορύβου είναι ασυσχέτιστα. Η συχνότητα δειγματοληψίας είναι Fs = 48kHz.

**A)** Delay-and-sum beamforming
Επιχειρούμε να αποθορυβοποιήσουμε το σήμα φωνής με delay-and-sum beamforming.

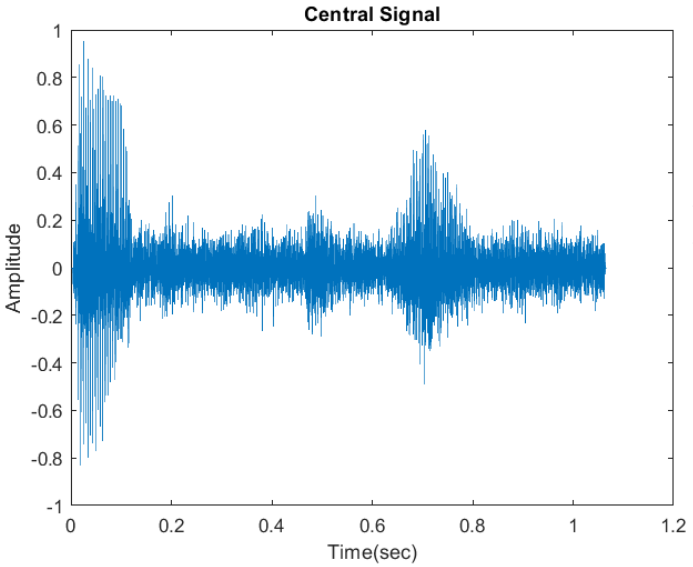
Υπολογίζουμε αρχικά τα βάρη για τον delay-and-sum beamformer, όπως προκύπτουν από την εξίσωση $\mathbf{H}^{T}_{DS}(\omega) = \frac{1}{N} d^H(k_s)$ , $k_s = \frac{\omega}{c} a_s$ , $d(k_s) = [e^{-jk_s{}^T p_0}, e^{-jk_s{}^T p_1}, \ldots, e^{-jk_s{}^T p_{N-1}}]^T$ και εφαρμόζουμε το beamforming, βρίσκοντας τον DFT των σημάτων εισόδου, πολλαπλασιάζοντας τα με την $\mathbf{H}^{T}_{DS}(\omega)$ και αθροίζοντας τα αποτελέσματα. Η επιθυμητή έξοδος y(t) προκύπτει με αντίστροφο DFT.

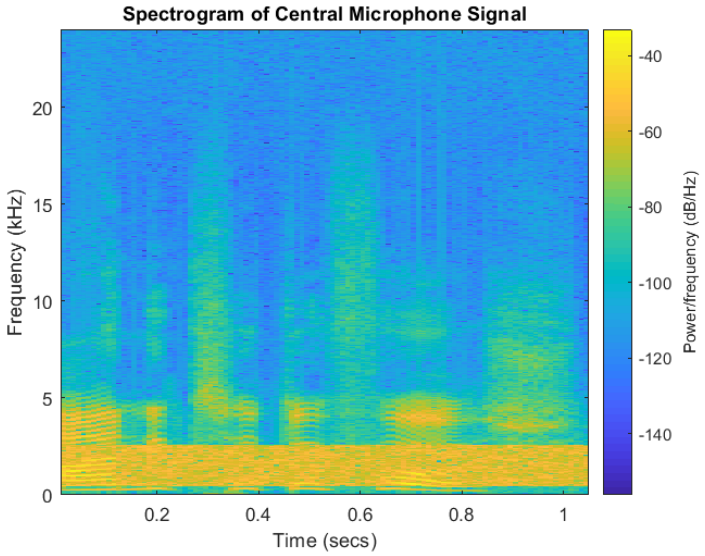
Κυματομορφή και Σπεκτρογράφημα καθαρού σήματος φωνής



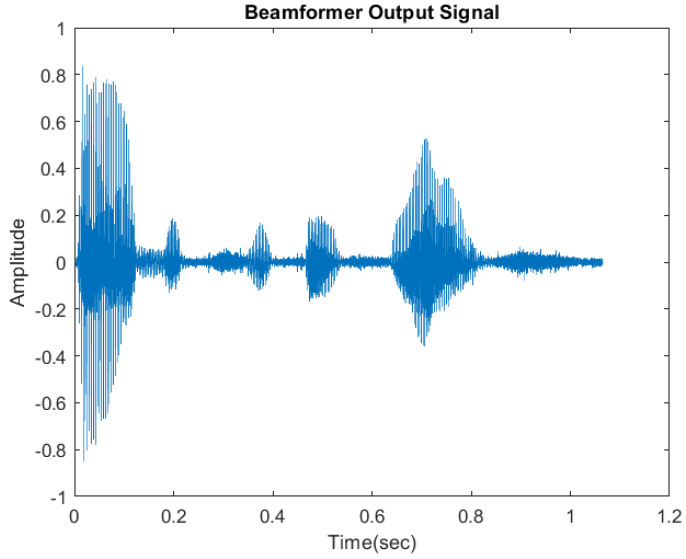



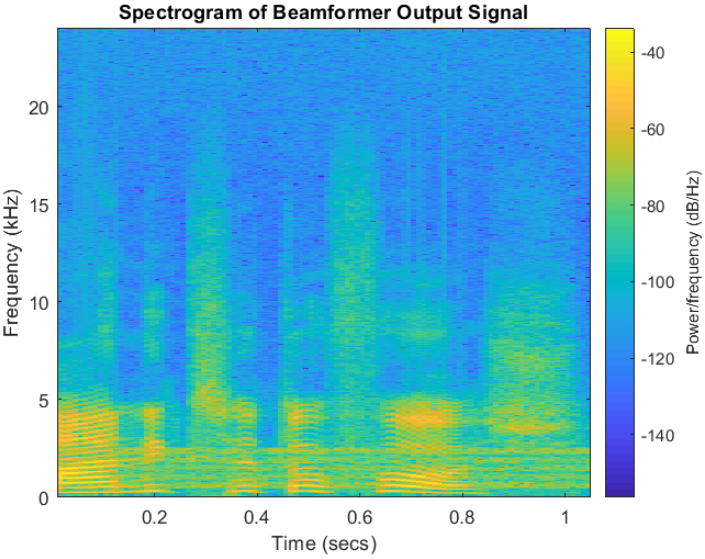

Κυματομορφή και Σπεκτρογράφημα θορυβώδους σήματος κεντρικού μικροφώνου (n=3)




Κυματομορφή και Σπεκτρογράφημα σήματος εξόδου y(t) του delay-and-sum beamformer



Για την γραφική παράσταση του σπεκτρογραφήματος έχουμε διαλέξει Hamming παράθυρο μήκους L = 30ms = 0.03·$F_s$ = 1440 δειγμάτων και επικάλυψης P = 2L/3 δειγμάτων. Παρατηρούμε τόσο στο πεδίο του χρόνου όσο και της συχνότητας (μέσω του σπεκτρογραφήματος) πως το καθαρό σήμα φωνής και το σήμα εξόδου y(t) έχουν αρκετά παρόμοια συμπεριφορά και στα δύο πεδία. Αντίθετα βλέπουμε πως το σήμα του κεντρικού μικροφώνου της συστοιχίας έχει έντονη παρουσία θορύβου. Συνεπώς ο beamformer αποθορυβοποίησε το σήμα αρκετά καλά. Ακουστικά τα συμπεράσματα αυτά επιβεβαιώνονται. Υπολογίστηκε επίσης το SNR του θορυβώδους σήματος του κεντρικού μικροφώνου, καθώς και το SNR της εξόδου y(t) και βρέθηκαν: $SNR_{central\_microphone}$ = 3dB, $SNR_y$ = 20.3822. Η διαφορά των δύο σηματοθορυβικών λόγων είναι πολύ μεγάλη, πράγμα που επιβεβαιώνει άλλη μία φορά πως ο beamformer έδωσε ένα πολύ καθαρότερο σήμα. Η έξοδος y(t) του beamformer αποθηκεύτηκε στο αρχείο sim_ds.wav, που περιέχεται στον φάκελο της εργασίας.

**B)** Μονοκαναλικό Wiener φιλτράρισμα

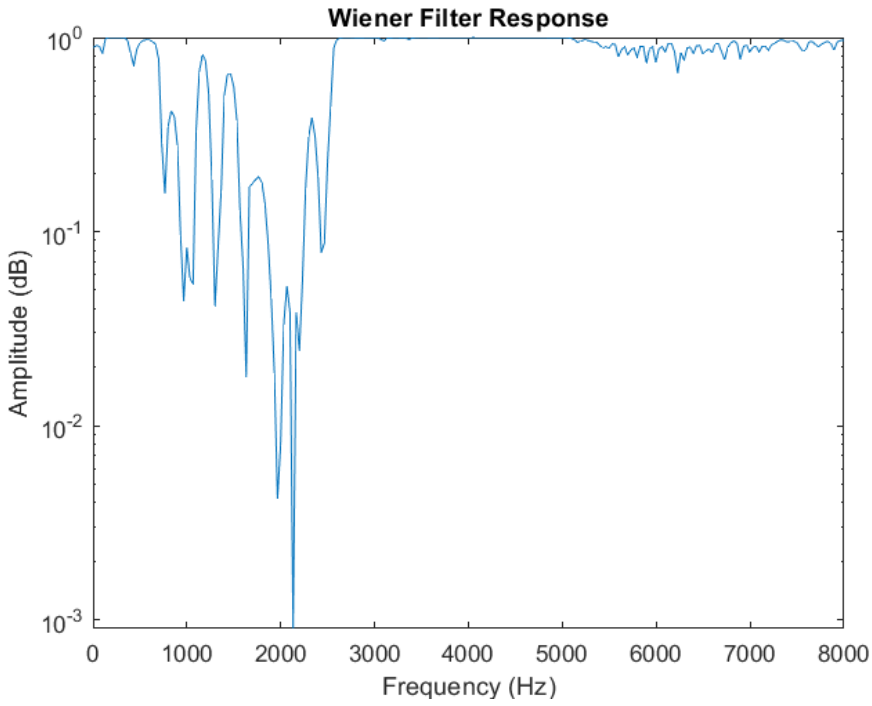
Στο ερώτημα αυτό θα επιχειρήσουμε να αποθορυβοποιήσουμε ένα πλαίσιο του σήματος με τη μονοκαναλική μέθοδο Wiener Filtering και να συγκρίνουμε το τελικό αποτέλεσμα με την εφαρμογή της πολυκαναλικής μεθόδου στο πλαίσιο, όπως την υλοποιήσαμε και στο **A)** ερώτημα.

Θεωρούμε το πλαίσιο $f(t,p_3)$ (πλαίσιο του σήματος του κεντρικού μικροφώνου) , με $t \in [0.47s, 0.5s]$ (διάρκειας 30ms), το οποίο περιέχει τον έμφωνο ήχο (/oh/). Το πλαίσιο αυτό μπορεί να μοντελοποιηθεί ως x(t) = s(t) + v(t), όπου s(t) το επιθυμητό σήμα φωνής, ενώ v(t) ο θόρυβος.

Υπολογίζουμε αρχικά την απόκριση συχνότητας του IIR Wiener φίλτρου, η οποία εφόσον τα σήματα φωνής και θορύβου είναι ασυσχέτιστα είναι:

$$H_w(\omega) = 1 - \frac{P_v(\omega)}{P_x(\omega)}$$

όπου $P_v(\omega)$ το φάσμα ισχύος (power spectrum) του θορύβου και $P_x(\omega)$ το φάσμα ισχύος του συνολικού σήματος x(t). Τα φάσματα ισχύος εκτιμήθηκαν με τη μέθοδο Welch, χρησιμοποιώντας τις εξής παραμέτρους: Μήκος παραθύρου = 10ms, επικάλυψη = 5ms. Η συνάρτηση μεταφοράς $H_w(\omega)$ σε λογαριθμική κλίμακα (dB) στο πεδίο της συχνότητας για $f \in [0, 8\ kHz]$ είναι:

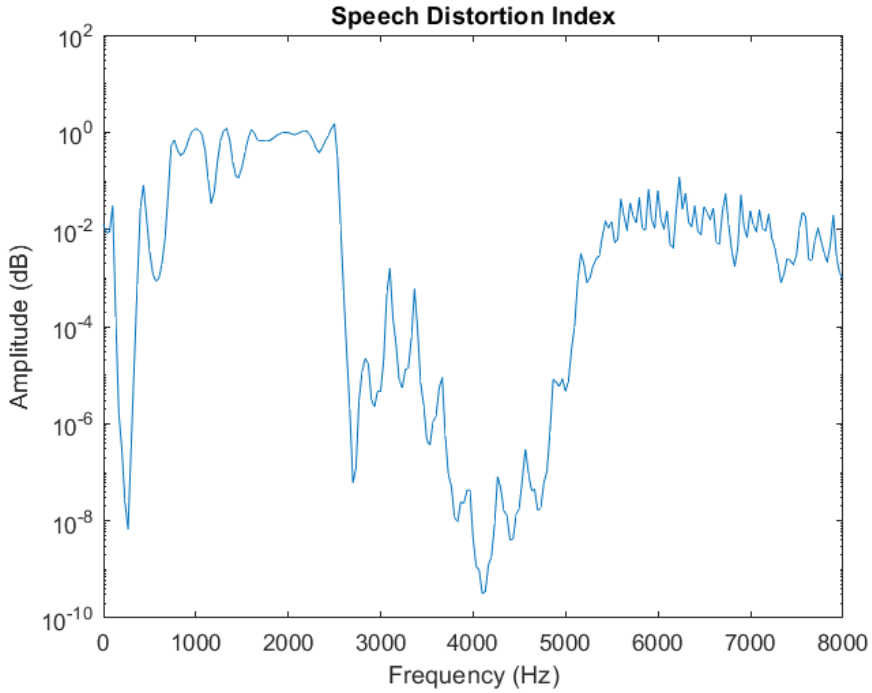


Παρατηρούμε ότι το Wiener filter έχει πολύ χαμηλή απόκριση συχνότητας για f ∈ [500Hz, 2.5kHz], δηλαδή αποσβένει το ζωνοπερατό θόρυβο που βρίσκεται σε αυτή την μπάντα συχνοτήτων, πράγμα επιθυμητό και αναμενόμενο.

Το Wiener φίλτρο προκαλεί παραμόρφωση (distortion) στο σήμα φωνής s(t), η οποία ισούται με $s(t) - h_w(t) * s(t)$. Ένας τρόπος να μετρηθεί η παραμόρφωση αυτή είναι το speech distortion index, το οποίο ορίζεται ως ο λόγος του φάσματος ισχύος της παραμόρφωσης προς το φάσμα ισχύος του σήματος φωνής:

$$n_{sd}(\omega) = \frac{E\left[|S(\omega) - H_w(\omega)S(\omega)|^2\right]}{P_x(\omega)} = |1 - H_w(\omega)|^2$$

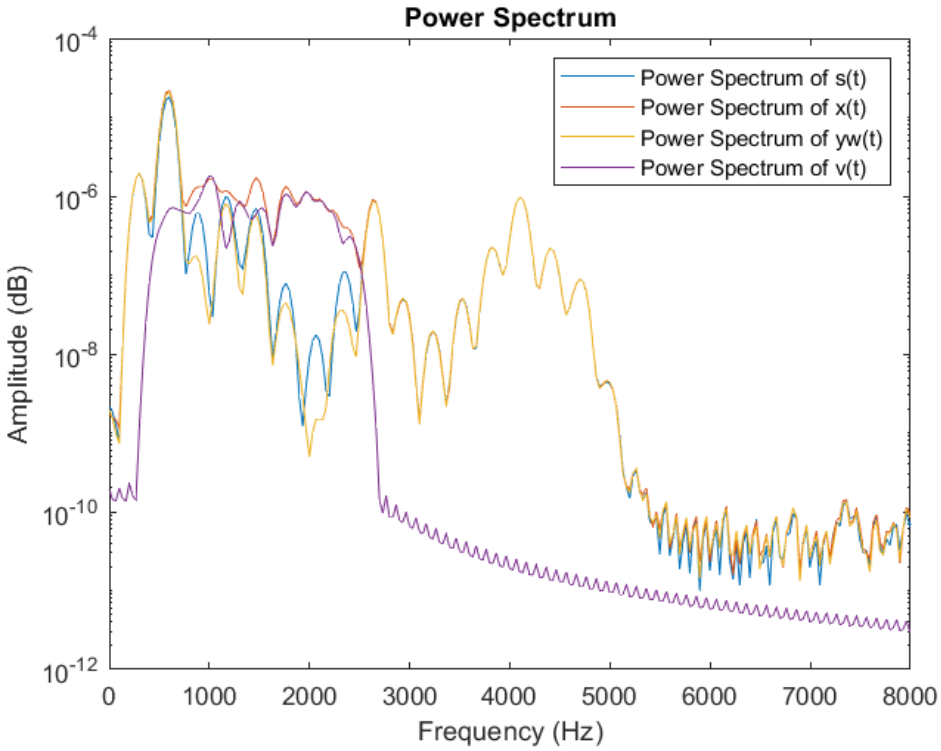
Η παραμόρφωση αυτή φαίνεται παρακάτω σε λογαριθμική κλίμακα για f ∈ [0, 8kHz]:




Παρατηρούμε ότι η παραμόρφωση του σήματος φωνής είναι υψηλότερη στις συχνότητες f ∈ [500Hz, 2.5kHz]. Αυτό συμβαίνει διότι ο θόρυβος είναι ζωνοπερατός και βρίσκεται σε αυτές τις συχνότητες με αποτέλεσμα η αποθορυβοποίηση να γίνεται κυρίως σε αυτές τις συχνότητες και συνεπώς το σήμα να παραμορφώνεται περισσότερο.

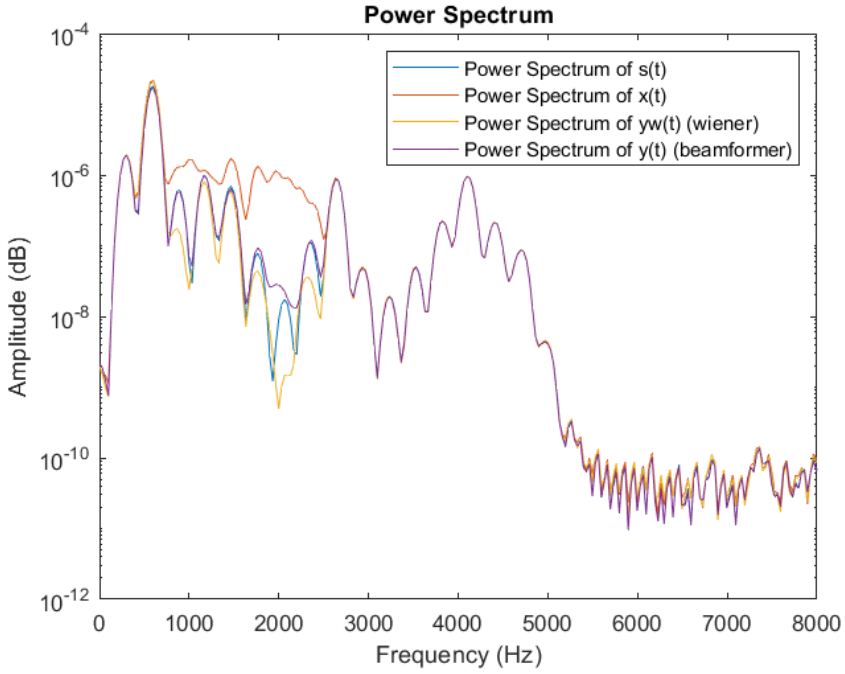
Στη συνέχεια υπολογίζουμε στον DFT μετασχηματισμό του θορυβώδους σήματος x(t) και εφαρμόζουμε το Wiener φιλτράρισμα, απ' όπου με αντίστροφο DFT προκύπτει η έξοδος yw(t).

Σχεδιάσαμε για f ∈ [0, 8kHz] τα φάσματα ισχύος του καθαρού σήματος φωνής, του σήματος εισόδου x(t), του σήματος εξόδου yw(t) καθώς και του σήματος θορύβου v(t) στην είσοδο του φίλτρου το οποίο ορίζεται ως v(t) = x(t) – s(t):




Για τα φάσματα ισχύος παρατηρούμε ότι, τα φάσματα των s(t), yw(t) ταυτίζονται σε πολύ μεγάλο βαθμό, με εξαίρεση τις συχνότητες f ∈ [500Hz, 2.5kHz], όπως ήταν αναμενόμενο, στις οποίες παρουσιάζεται μία μικρή διαφοροποίηση, η οποία οφείλεται στην παρουσία θορύβου στην έξοδο yw(t), η οποία φαίνεται ότι αποθορυβοποιήθηκε αρκετά καλά, αλλά όχι πλήρως. Το φάσμα ισχύος της εισόδου x(t) ταυτίζεται με αυτό των s(t), yw(t) σε όλες τις περιοχές συχνοτήτων εκτός από την ζώνη f ∈ [500Hz, 2.5kHz], στην οποία υπάρχει ο θόρυβος, κάτι το οποίο προφανώς περιμέναμε και γι' αυτό χρησιμοποιήθηκε και το φιλτράρισμα. Τέλος για το φάσμα ισχύος του θορύβου v(t) παρατηρούμε ότι όπως ήταν λογικό, αυτό συγκεντρώνεται στην περιοχή f ∈ [500Hz, 2.5kHz], ενώ στις υπόλοιπες συχνότητες προσεγγίζει το μηδέν.

Στη συνέχεια σχεδιάζουμε σε λογαριθμική κλίμακα και για συχνότητες f ∈ [0, 8kHz], τα φάσματα ισχύος του καθαρού σήματος φωνής για το υπό μελέτη πλαίσιο s(t), του θορυβώδους σήματος εισόδου x(t), του σήματος εξόδου yw(t) και του σήματος εξόδου y(t) του delay-and-sum beamformer για το υπό μελέτη πλαίσιο και προκύπτει :



Τέλος υπολογίσαμε τα SNR της εισόδου x(t), εξόδου yw(t), καθώς και το SNR για το συγκεκριμένο πλαίσιο της πολυκαναλικής μεθόδου που υλοποιήθηκε στο **Α)** ερώτημα και βρήκαμε: $SNR_x$ = 5.2696dB , $SNR_{yw}$ = 10.8257dB , $SNR_{beamformer}$ = 20.666dB. Παρατηρούμε ότι $SNR_{beamformer} > SNR_{yw} > SNR_x$ . Επίσης από τα φάσματα ισχύος παρατηρούμε ότι παρά το γεγονός ότι τα φάσματα των s(t), y(t), yw(t) είναι πολύ κοντά, το φάσμα ισχύος του y(t) του beamformer προσεγγίζει σε πολλά σημεία καλύτερα το φάσμα ισχύος του καθαρού σήματος φωνής s(t). Αυτό σημαίνει πως και οι δύο μέθοδοι λειτουργούν αρκετά καλά, επιτυγχάνοντας σημαντική βελτίωση στην ποιότητα του ήχου. Στην συγκεκριμένη εφαρμογή είναι φανερό πως *το beamforming υπερέχει σημαντικά του μονοκαναλικού Wiener filtering,* με την αποθορυβοποίηση του σήματος να είναι αρκετά καλύτερη σε αυτή την μέθοδο.

### 2.2 Beamforming σε πραγματικά σήματα

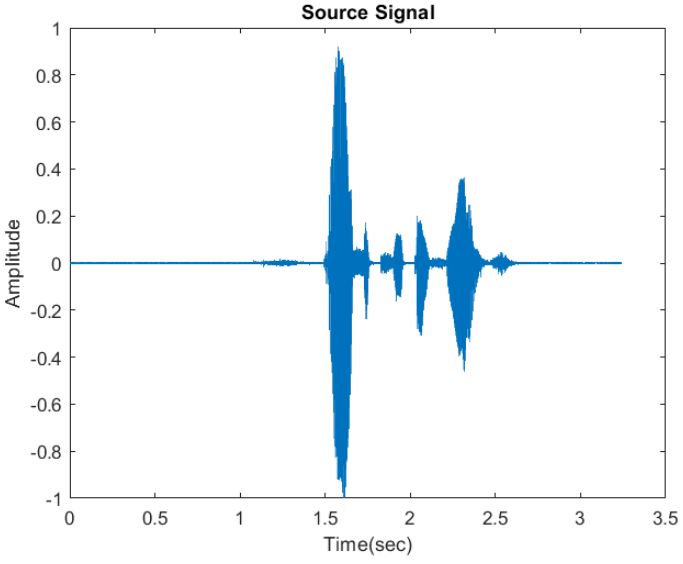
Θεωρούμε μία γραμμική συστοιχία N = 7 μικροφώνων με απόσταση d = 4cm, η οποία βρίσκεται σε ένα θορυβώδες δωμάτιο και καταγράφει το σήμα φωνής εκφωνεί ένας άνθρωπος σε γωνία θ = 45°. Ο θόρυβος τώρα δεν προέρχεται από σημειακή πηγή αλλά από διάφορες πηγές, οι οποίες δημιουργούν ένα ισοτροπικό και ομογενές πεδίο θορύβου που ονομάζεται diffuse noise field. Ο θόρυβος θεωρείται στάσιμος (stationary random process - WSS). Η συχνότητα δειγματοληψίας είναι 48kHz.

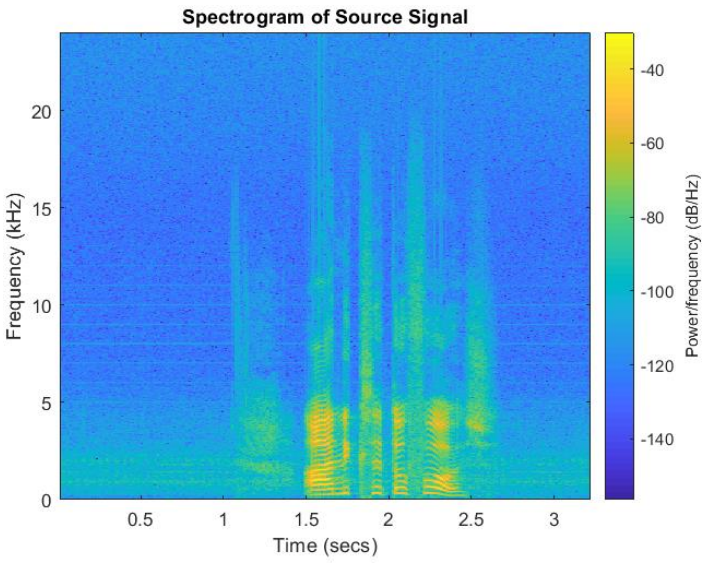
**Α)** Delay-and-sum beamforming

Επιχειρούμε να αποθορυβοποιήσουμε το σήμα φωνής με delay-and-sum beamforming.

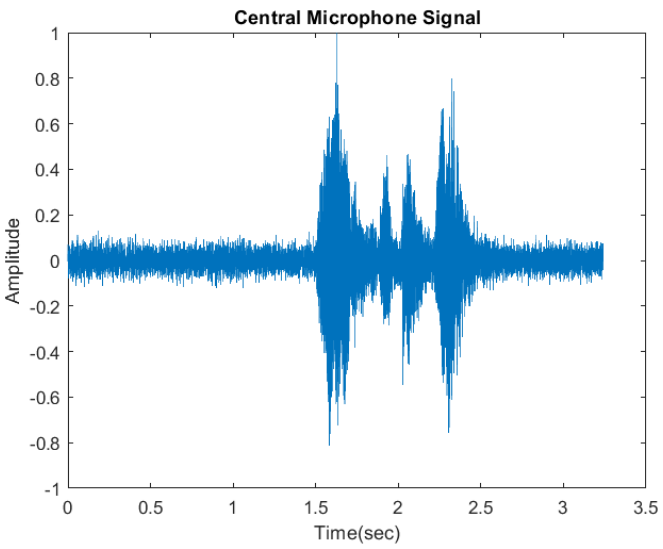
Υπολογίζουμε αρχικά τα βάρη για τον delay-and-sum beamformer, όπως προκύπτουν από την εξίσωση $\mathbf{H}^{T}_{DS}(\omega) = \frac{1}{N} d^{H}(k_s)$ , $k_s = \frac{\omega}{c} a_s$ , $d(k_s) = [e^{-jk_s{}^T p_0}, e^{-jk_s{}^T p_1}, \ldots, e^{-jk_s{}^T p_{N-1}}]^T$ και εφαρμόζουμε το beamforming, βρίσκοντας τον DFT των σημάτων εισόδου, πολλαπλασιάζοντας τα με την $\mathbf{H}^{T}_{DS}(\omega)$ και αθροίζοντας τα αποτελέσματα. Η επιθυμητή έξοδος y2(t) προκύπτει με αντίστροφο DFT.

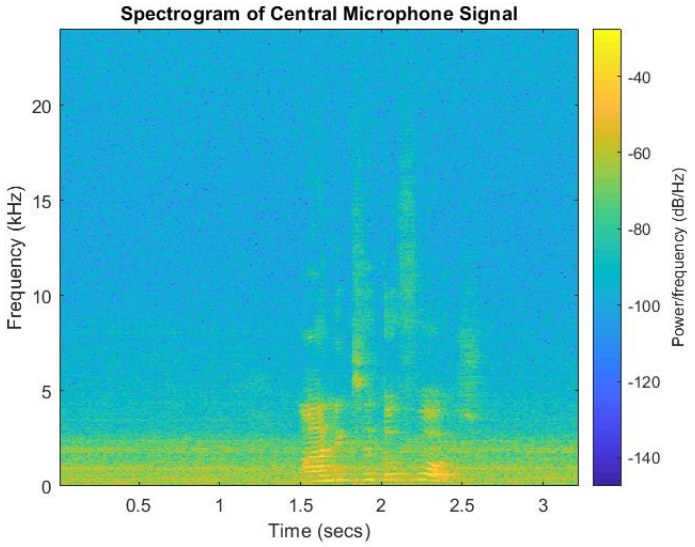
<u>Κυματομορφή και Σπεκτρογράφημα καθαρού σήματος φωνής</u>



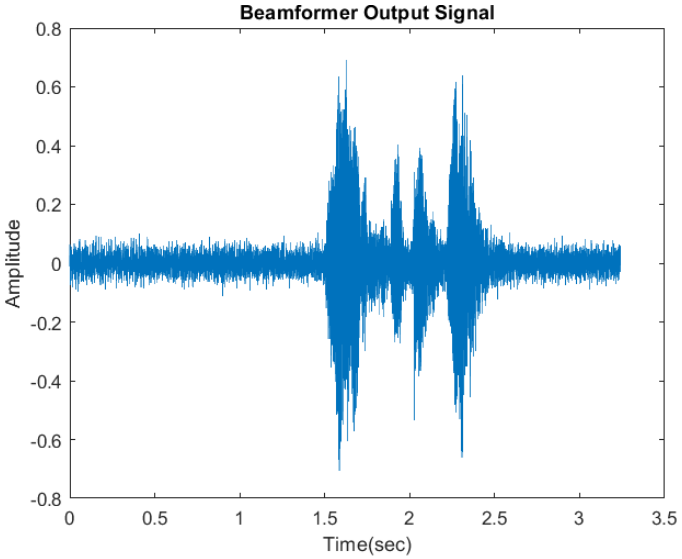



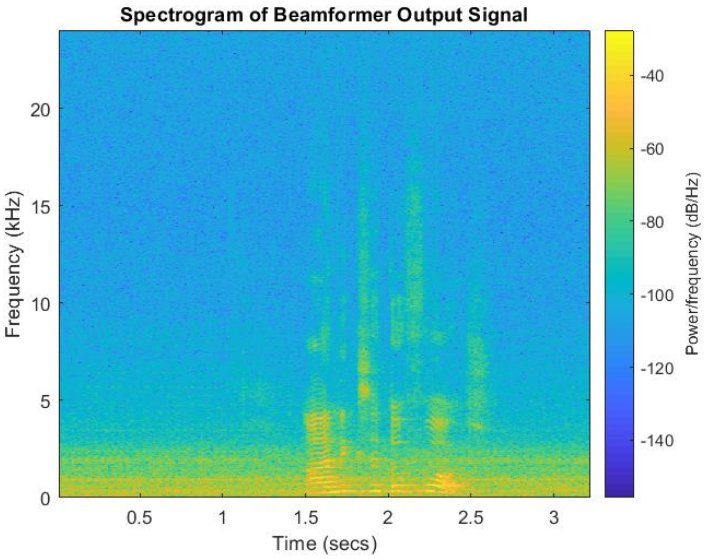
Κυματομορφή και Σπεκτρογράφημα θορυβώδους σήματος κεντρικού μικροφώνου συστοιχίας



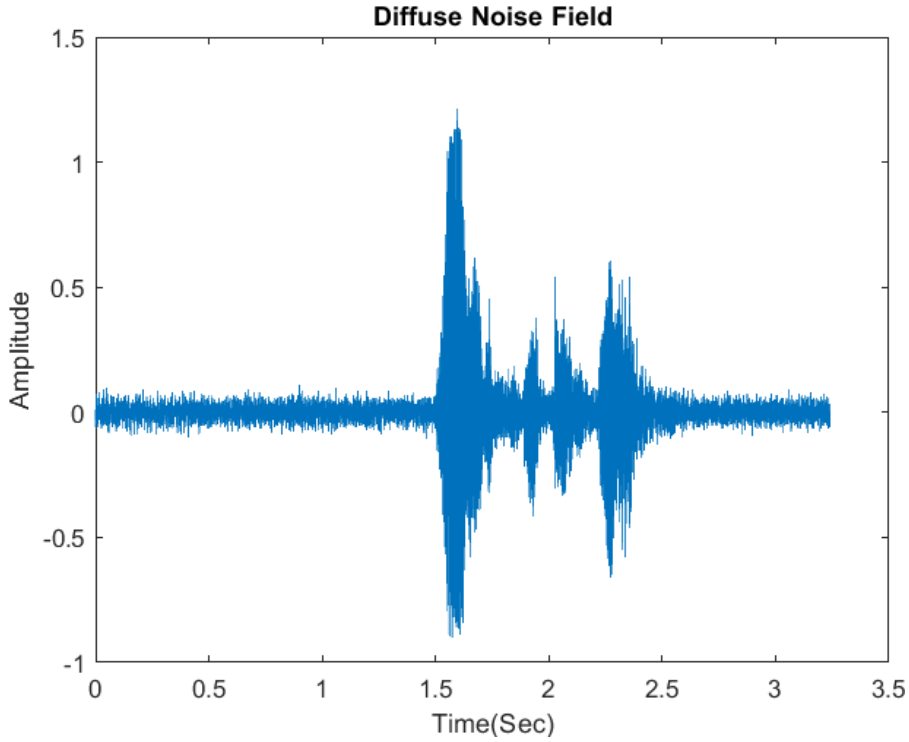




Κυματομορφή και Σπεκτρογράφημα σήματος εξόδου y2(t) beamformer




Συγκρίνοντας τις παραπάνω κυματομορφές και σπεκτρογραφήματα, παρατηρούμε ότι στο καθαρό σήμα φωνής δεν υπάρχει παρουσία θορύβου. Αντίθετα το θορυβώδες σήμα που καταγράφηκε από το κεντρικό μικρόφωνο της συστοιχίας φαίνεται να υπάρχει αρκετός θόρυβος στις συχνότητες $f \in [0, 2.5kHz]$ , κάτι που επιβεβαιώνεται και από το αντίστοιχο σπεκτρογράφημα. Βλέπουμε επίσης ότι η παρουσία θορύβου στην έξοδο είναι ακόμα πολύ έντονη, κάτι το οποίο επιβεβαιώνεται και ακουστικά, καθώς η υψηλή ισχύς του τον κάνει ιδιαίτερα αισθητό. Συνεπώς η εφαρμογή του beamforming για την αποθορυβοποίηση του σήματος απέτυχε, καθώς η είσοδος και η έξοδος του beamformer είναι σε πολύ μεγάλο βαθμό παρόμοιες. Η αποτυχία του beamforming ήταν αναμενόμενη, καθώς για την μέθοδο αυτή απαραίτητη προϋπόθεση είναι να ισχύει: $y(t) = s(t)+v(t)$. Στην περίπτωση του diffuse noise field τα σήματα θορύβου εμφανίζουν μεγάλη συσχέτιση μεταξύ μικροφώνων. Η παραπάνω προϋπόθεση δεν ισχύει, καθώς βλέπουμε από την οπτικοποίηση του σήματος θορύβου, πως σε αυτό περιέχεται πληροφορία και όχι μόνο θόρυβος:




Για τα σήματα φωνής, μία καταλληλότερη μετρική ποιότητας από το SNR είναι το segmental SNR (SSNR), το οποίο ορίζεται ως το μέσο SNR των πλαισίων βραχέος χρόνου του σήματος φωνής:

$$\text{SSNR} = \frac{10}{M}\sum_{m=0}^{M-1} log_{10}\frac{\sum_{n=Lm}^{Lm+L-1} s^2(n)}{\sum_{n=Lm}^{Lm+L-1} v^2(n)}$$

,όπου L το μήκος των πλαισίων, M ο αριθμός των πλαισίων, s(n) το σήμα φωνής και v(n) ο θόρυβος. Πλαίσια με SNR μεγαλύτερο των 35dB δεν έχουν σημαντικές διαφορές στην ποιότητα του σήματος και τίθενται στα 35dB για την εξαγωγή του μέσου όρου. Σε πλαίσια σιωπής το SNR είναι έντονα αρνητικό, οπότε πλαίσια με SNR μικρότερο ενός κατωφλιού με τιμή ∈ [-20dB, 0dB] αγνοούνται. Επιλέχθηκε τιμή κατωφλιού ίση με -10dB.

Υπολογίσαμε το SSNR στο κεντρικό μικρόφωνο της συστοιχίας (n = 3) και στην έξοδο του beamformer μέσω της συνάρτησης ssnr.m που υλοποιήσαμε και για μήκος πλαισίων ίσο με 100 δείγματα βρέθηκαν:

$\text{SSNR}_{\text{central\_microphone}}$ = 1.8386dB

και

$\text{SSNR}_{\text{beamformer}}$ = 1.2241dB

Πράγματι λοιπόν ισχύει ότι $\text{SSNR}_{\text{central\_microphone}} > \text{SSNR}_{\text{beamformer}}$, δηλαδή το beamforming απέτυχε.

Η έξοδος του beamformer περιέχεται στον φάκελο της παρούσας άσκησης και είναι καταγεγραμμένη στο αρχείο real_ds.wav.

**B)** Post-filtering με Wiener φίλτρο

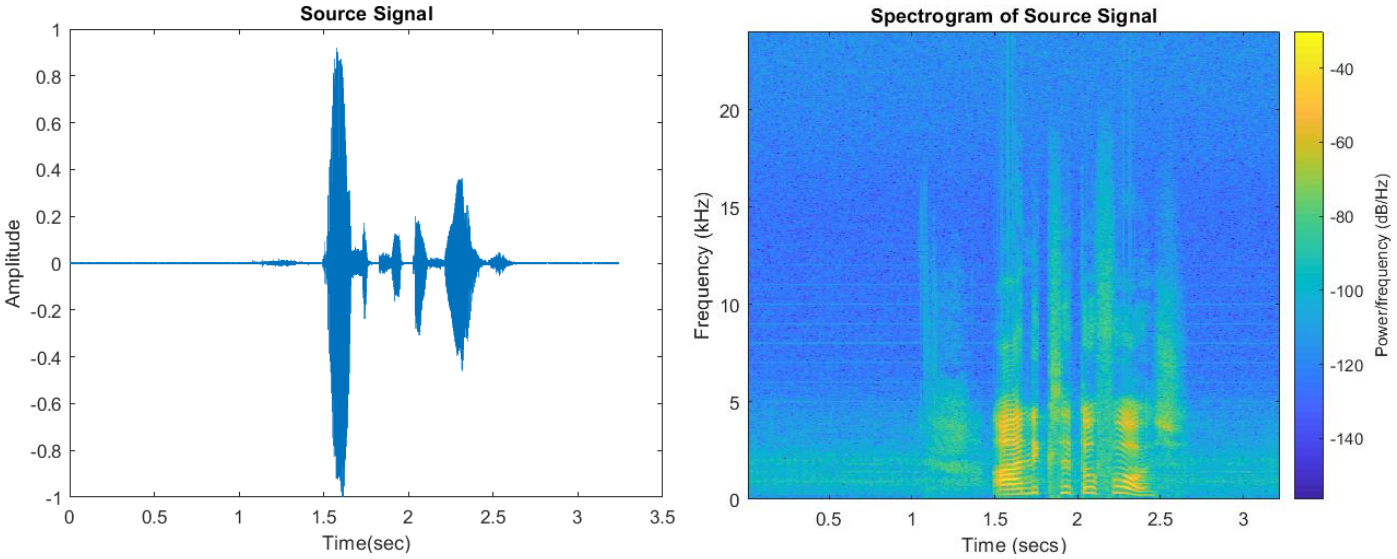
Όπως διαπιστώθηκε, στην περίπτωση του diffuse noise field, ο delay-and-sum beamformer δεν έχει καλή απόδοση, διότι τα σήματα θορύβου εμφανίζουν μεγάλη συσχέτιση μεταξύ μικροφώνων, ειδικά στις χαμηλές συχνότητες. Είναι συνήθης πρακτική να εφαρμόζεται και μονικαναλικό φιλτράρισμα μετά το beamforming για περαιτέρω βελτίωση της ποιότητας του σήματος. Η διαδικασία αυτή ονομάζεται post-filtering και υλοποιήθηκε ως εξής:

Χωρίζουμε την έξοδο του beamformer σε επικαλυπτόμενα πλαίσια, μήκους 30ms και επικάλυψης 15ms, τα οποία παραθυμώνουμε και με Hamming window. Στη συνέχεια εκτιμούμε το φάσμα ισχύος κάθε παραθυμωμένου πλαισίου με την μέθοδο Welch, διαιρώντας κάθε πλαίσιο σε υποπλαίσια μήκους 10ms και επικάλυψης 5ms. Έτσι βρίσκουμε για κάθε πλαίσιο i την απόκριση συχνότητας Hw_i(ω) του IIR Wiener φίλτρου από τον τύπο:

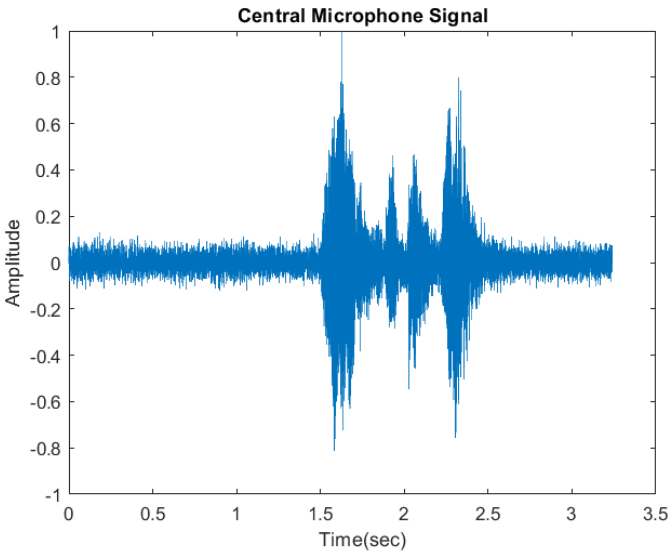
$$H_{w\_i}(\omega) = 1 - \frac{P_v(\omega)}{P_{frame\_i}(\omega)}$$

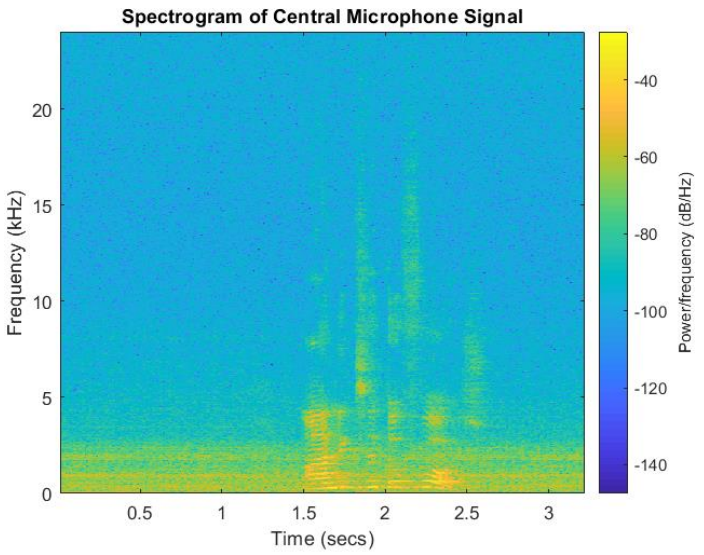
όπου $P_v(\omega)$ το φάσμα ισχύος (power spectrum) του θορύβου και $P_{frame\_i}(\omega)$ το φάσμα ισχύος του κάθε πλαισίου. Ο θόρυβος έχει θεωρηθεί στάσιμος, δηλαδή δεν αλλάζει από πλαίσιο σε πλαίσιο, συνεπώς το φάσμα ισχύος $P_v(\omega)$ του θορύβου είναι αμετάβλητο. Στη συνέχεια υπολογίζουμε τον DFT κάθε πλαισίου και πολλαπλασιάζοντας με την αντίστοιχη απόκριση συχνότητας προκύπτει η έξοδος Yw_i(ω), απ' όπου με αντίστροφο DFT βρίσκουμε την έξοδο yw_i(t) στο πεδίο του χρόνου. Τέλος με overlap-add σύνθεση προκύπτει η συνολική έξοδος του Wiener φίλτρου.

<u>Κυματομορφή και Σπεκτρογράφημα καθαρού σήματος φωνής</u>

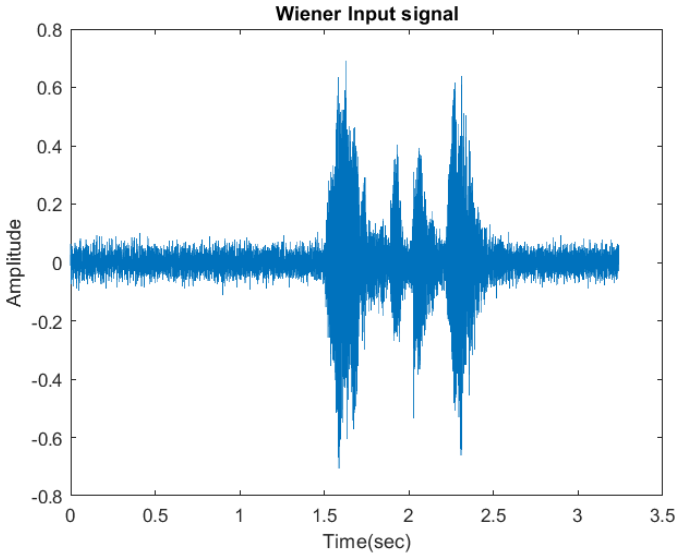


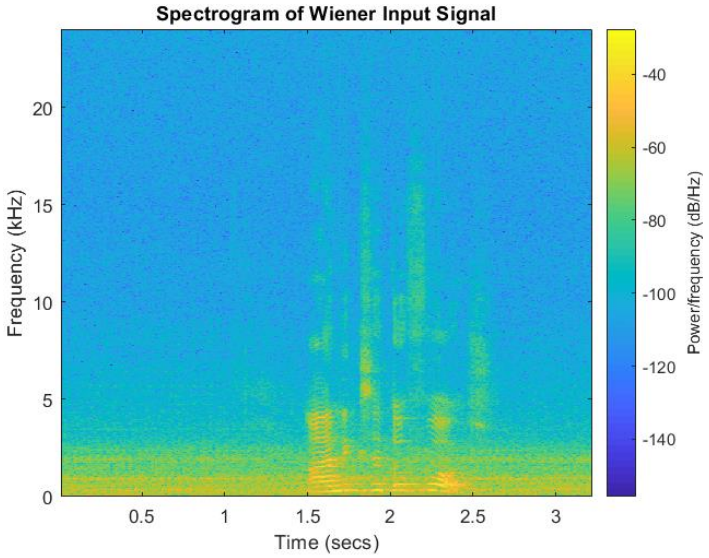
Κυματομορφή και Σπεκτρογράφημα θορυβώδους σήματος κεντρικού μικροφώνου συστοιχίας



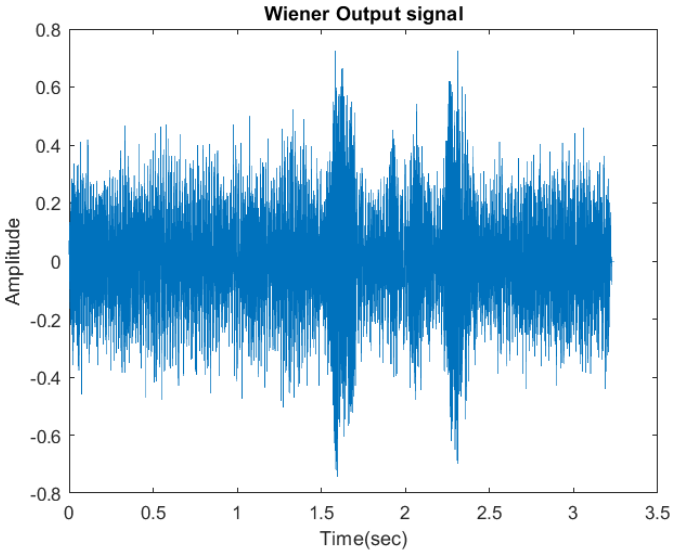



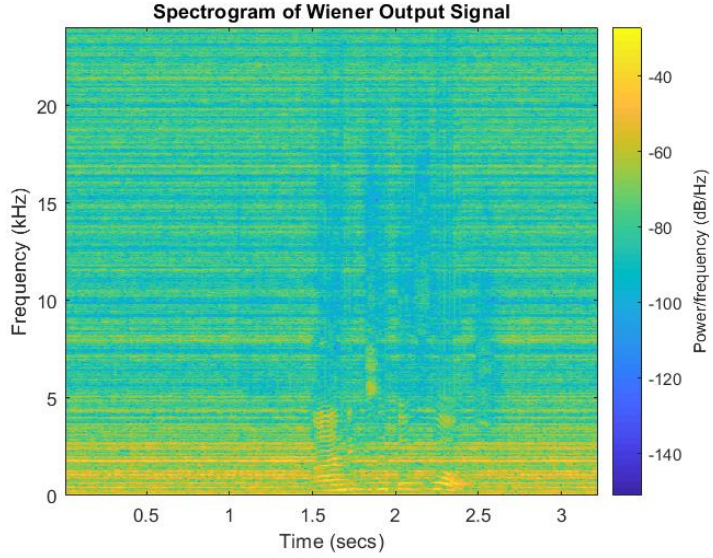

Κυματομορφή και Σπεκτρογράφημα σήματος εισόδου Wiener φίλτρου ( ≡ έξοδος beamformer)




Κυματομορφή και Σπεκτρογράφημα σήματος εξόδου Wiener φίλτρου



Παρατηρούμε ότι στην έξοδο του Wiener φίλτρου υπάρχει αισθητή παρουσία θορύβου, ο οποίος μάλιστα αντί να μειωθεί φαίνεται ότι πλέον υπερισχύει του αρχικού σήματος. Η αποθορυβοποίηση λοιπόν του σήματος έχει αποτύχει, κάτι το οποίο επιβεβαιώνεται και ακουστικά. Η έξοδος του Wiener φίλτρου περιέχεται στον φάκελο της εργασίας και είναι καταγεγραμμένη στο αρχείο real_mmse.wav.

Παρόλαυτα, υπολογίζοντας όπως εξηγήθηκε παραπάνω τα SSNR στην είσοδο και την έξοδο του Wiener φίλτρου βρήκαμε ότι:

$SSNR_{input\_wiener}$ = 1.2241dB

$SSNR_{output\_wiener}$ = 11.7102dB

Αυτό σημαίνει ότι, παρά την αποτυχία του φιλτραρίσματος στην ελάττωση του θορύβου, ο σηματοθορυβικός λόγος αυξήθηκε κατά 10.4861dB, δηλαδή η ισχύς του σήματος εξόδου συγκρινόμενη με την ισχύ του θορύβου έχει αυξηθεί σημαντικά. Έτσι, παρά τον έντονο θόρυβο, έχουμε τώρα ισχυρότερη παρουσία του επιθυμητού σήματος.

Τέλος, υπολογίσαμε τον μέσο όρο των SSNRs των σημάτων εισόδου στο σύστημα delay-and-sum beamformer + Wiener post-filter και βρήκαμε:

$MEAN\_SSNR_{input\_signals}$ = 2.2047dB

Βλέπουμε λοιπόν και πάλι πως το αρχικό σήμα έχει βελτιωθεί αρκετά μετά το φιλτράρισμα από το σύστημα delay-and-sum beamformer + Wiener post-filter, κάτι το οποίο και περιμέναμε.

## ***Συμπεράσματα***

1. Το beamforming λειτουργεί αρκετά καλά για παρουσία θορύβου σημειακής πηγής
2. Το beamforming δεν είναι αποτελεσματικό όταν το πεδίο θορύβου είναι διάχυτο
3. Το Wiener φιλτράρισμα μειώνει τον θόρυβο και στις δύο περιπτώσεις