

# Ψηφιακή Επεξεργασία Σήματος

## 3<sup>η</sup> Εργαστηριακή άσκηση

Σιφναίος Σάββας, AM: 03116080

Στούμπου Βασιλική, AM: 03116026

6<sup>ο</sup> εξάμηνο, HMMY

Στην άσκηση αυτή εξετάζουμε τη χρήση συστοιχιών μικροφώνων για καταγραφή και επεξεργασία ακουστικών σημάτων.

### Μέρος 1.4

Συγκεκριμένα, χρησιμοποιούμε μία ομοιόμορφη γραμμική συστοιχία με N μικρόφωνα που απέχουν απόσταση d. Τα μικρόφωνα είναι τοποθετημένα κατά μήκος του άξονα z σε θέσεις που δίνονται από τη σχέση:

$$p_{z_n} = \left( n - \frac{N-1}{2} \right) d, n = 0, 1, \dots, N-1$$

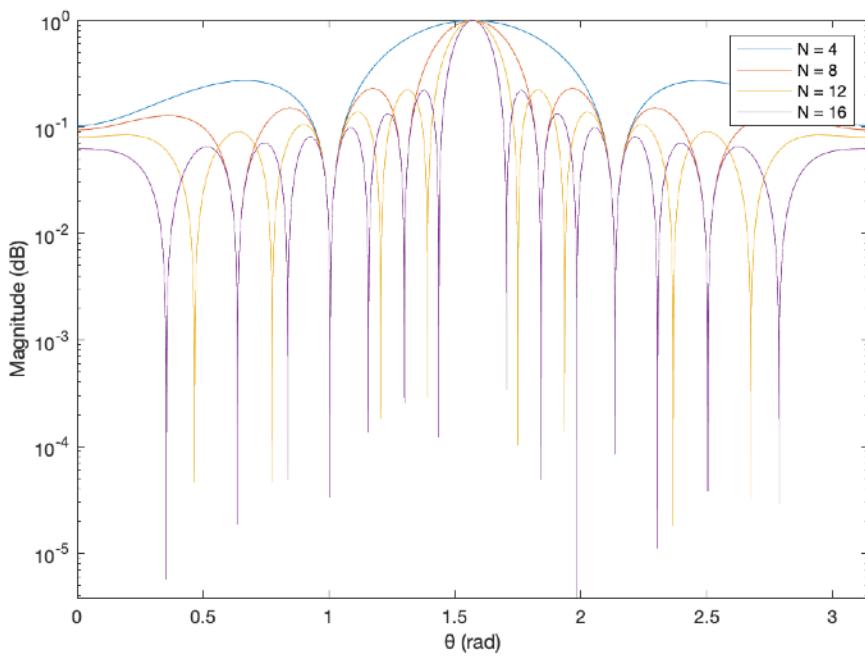
Τα σήματα που προέρχονται από τα διαφορετικά μικρόφωνα υφίστανται χωρικό φιλτράρισμα και στη συνέχεια συνδυάζονται ώστε να προκύψει η αποθρυβοποιημένη έξοδος. Η διαδικασία αυτή αποτελεί μέθοδο πολυκαναλικής επεξεργασίας σημάτων και ονομάζεται beamforming. Εδώ, χρησιμοποιούμε τον delay-and-sum beamformer, με τον οποίο γίνεται πρώτα χρονική ολίσθηση κάθε σήματος (array steering), ώστε να ευθυγραμμιστούν χρονικά και στη συνέχεια να αθροιιστούν. Η εξίσωση του delay-and-sum beam pattern είναι στην περίπτωση μας:

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

όπου  $\theta_s$  η γωνία από την οποία φτάνει το σήμα ( $\theta_s=90^\circ$ ). Για συχνότητα f = 2kHz σχεδιάζουμε το μέτρο του delay-and-sum beam pattern σε λογαριθμική κλίμακα (dB) συναρτήσει της γωνίας  $\theta \in [0, 180^\circ]$ .

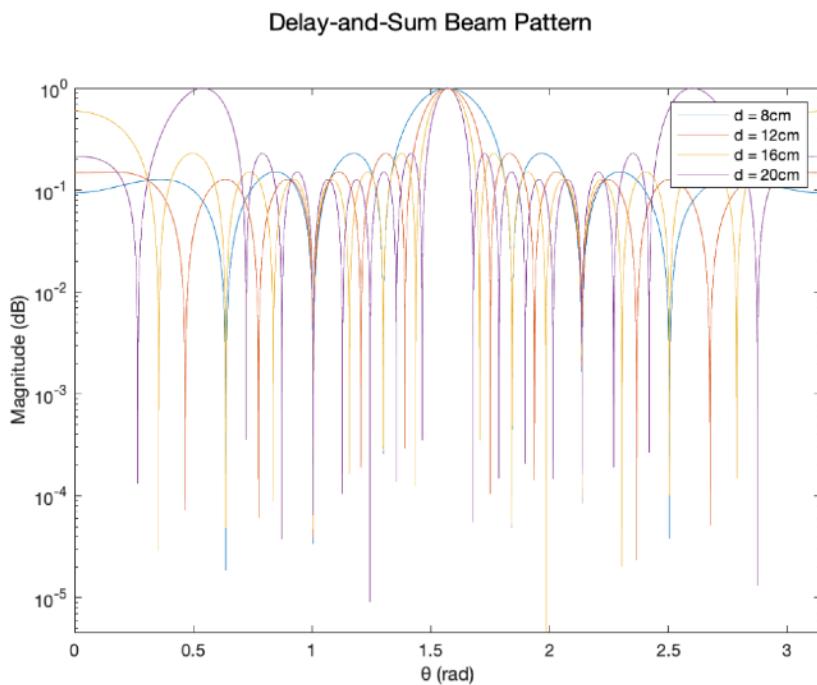
1) Αρχικά, η απόσταση μεταξύ των μικροφώνων είναι d=8cm και μεταβάλλουμε το πλήθος τους N.

Delay-and-Sum Beam Pattern



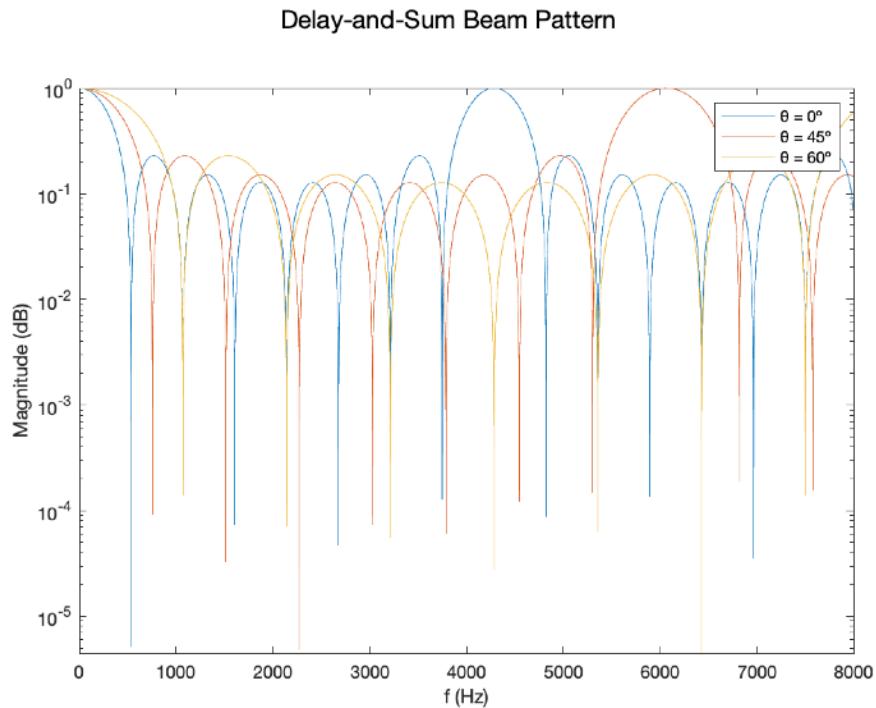
Παρατηρούμε ότι με αύξηση του αριθμού των μικροφώνων έχουμε περισσότερους και στενότερους λοβούς, οι οποίοι έχουν και μικρότερο πλάτος.

2) Στη συνέχεια, διατηρούμε τον αριθμό των μικροφώνων  $N=8$  και μεταβάλλουμε τη μεταξύ τους απόσταση  $d$ .



Παρατηρούμε ότι με αύξηση της απόστασης  $d$  εμφανίζονται περισσότεροι λοβοί οι οποίοι βρίσκονται περίπου στο ίδιο ύψος. Επίσης, αν το  $d$  αυξηθεί αρκετά ( $d=20$  cm), δημιουργούνται και άλλοι λοβοί με ύψος ίσο με αυτό του κεντρικού.

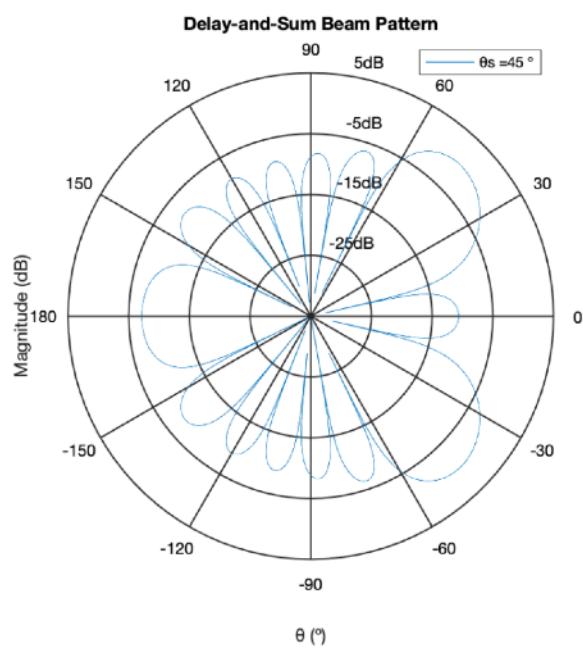
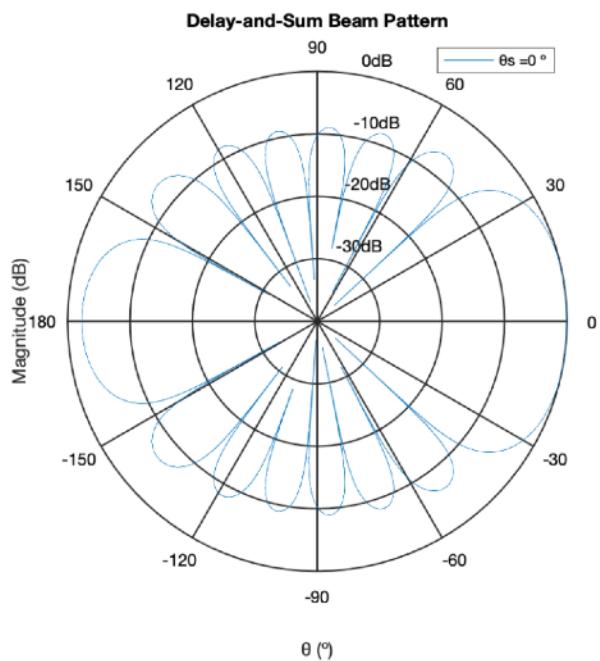
3) Στη συνέχεια, θεωρούμε ότι η συστοιχία αποτελείται από  $N=8$  μικρόφωνα με απόσταση  $d=8$  cm. Για γωνίες άφιξης των σημάτων  $\theta=0^\circ, 45^\circ, 60^\circ$  σχεδιάζουμε στην ίδια γραφική παράσταση το μέτρο του delay-and-sum beam pattern σε λογαριθμική κλίμακα (dB) συναρτήσει της συχνότητας του αφικνούμενου σήματος για συχνότητες  $f \in [0, 8\text{kHz}]$ .

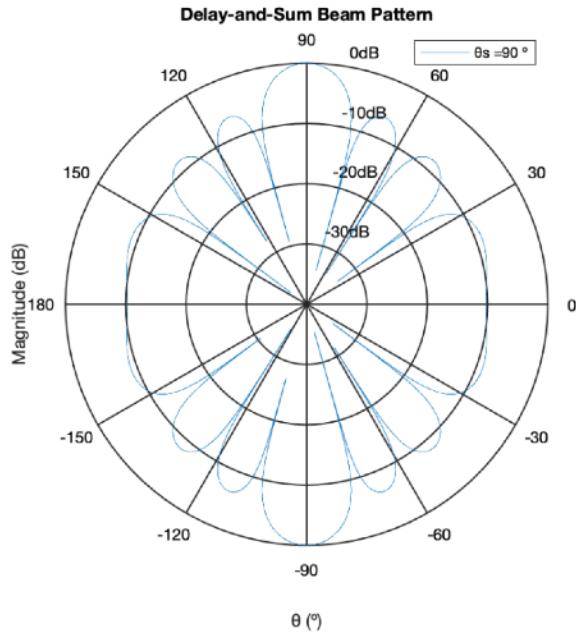


Παρατηρούμε ότι αύξηση της γωνίας άφιξης  $\theta$  ισοδυναμεί με μετατόπιση του κεντρικού λοβού προς μεγαλύτερες συχνότητες. Πλέον, δεν παρατηρείται συμμετρία ως προς τη γωνία  $\theta=180^\circ$ .

4) Θεωρούμε ότι η συστοιχία αποτελείται από  $N=8$  μικρόφωνα με απόσταση  $d=0.8$  cm. Η συχνότητα  $f$  ισούται με 2 kHz.

Σχεδιάζουμε το μέτρο του delay-and-sum beamformer σε λογαριθμική κλίμακα (dB) συναρτήσει της γωνίας  $\theta \in [-180^\circ, 180^\circ]$  για τιμές της γωνίας  $\theta_s = 0^\circ, 45^\circ, 90^\circ$ . Για τη σχεδίαση χρησιμοποιήσαμε τη συνάρτηση semilogr\_polar.





Παρατηρούμε ότι ο μέγιστος λοβός παρατηρείται κάθε φορά συμμετρικός γύρω από τη γωνία  $\theta_s$  και έχει πλάτος 1. Δηλαδή, σε κάθε περίπτωση ενισχύεται το σήμα που φτάνει από τη γωνία  $\theta_s$ , ενώ από τις υπόλοιπες κατευθύνσεις εξασθενεί. Αυτός άλλωστε είναι και ο σκοπός του array steering που έχουμε εφαρμόσει. Με αλλαγή της  $\theta_s$  μπορούμε να αλλάζουμε την κατεύθυνση που ενισχύει στο σύστημά μας.

## Μέρος 2. Εφαρμογή Beamforming για speech Enhancement

### **2.1 Beamforming σε προσομοιωμένα σήματα**

Η παρακάτω επεξεργασία πραγματοποιείται για σήματα που παράγονται από δύο πηγές, μία φωνής η οποία είναι σε γωνία  $45^\circ$  σε σχέση με την συστοιχία και μία θορύβου που βρίσκεται σε γωνία  $135^\circ$ . Τα σήματα αυτά έχουν καταγραφεί από ομοιόμορφη γραμμική συστοιχία  $N=7$  μικροφώνων τα οποία απέχουν μεταξύ τους απόσταση  $d=8\text{cm}$ . Επιπλέον, αξίζει να σημειωθεί ότι ο παραγόμενος θόρυβος είναι ζωνοπερατός με την ενέργειά του να είναι συγκεντρωμένη στις συχνότητες  $500\text{Hz}$  εώς και  $2.5\text{kHz}$ . Τέλος, τονίζουμε πώς τα σήματα φωνής και θορύβου είναι ασυσχέτιστα.

#### **A) Delay-and-Sum Beamforming**

Σε αυτό το σημείο της άσκησης καλούμαστε να εφαρμόσουμε την μέθοδο Beamforming, σε προσομοιωμένα σήματα, ώστε να απομακρύνουμε το θόρυβο από τελικό σήμα που καταγράφει η συστοιχία των μικροφώνων.

1. Αρχικά, υπολογίζουμε τα βάρη για το Delay-and-Sum Beamformer. Για να βρούμε τα βάρη, ώστοσο, χρείαζεται πρώτα να υπολογίσουμε τον πίνακα  $d(k_s)$  που δίνεται από την παρακάτω σχέση :

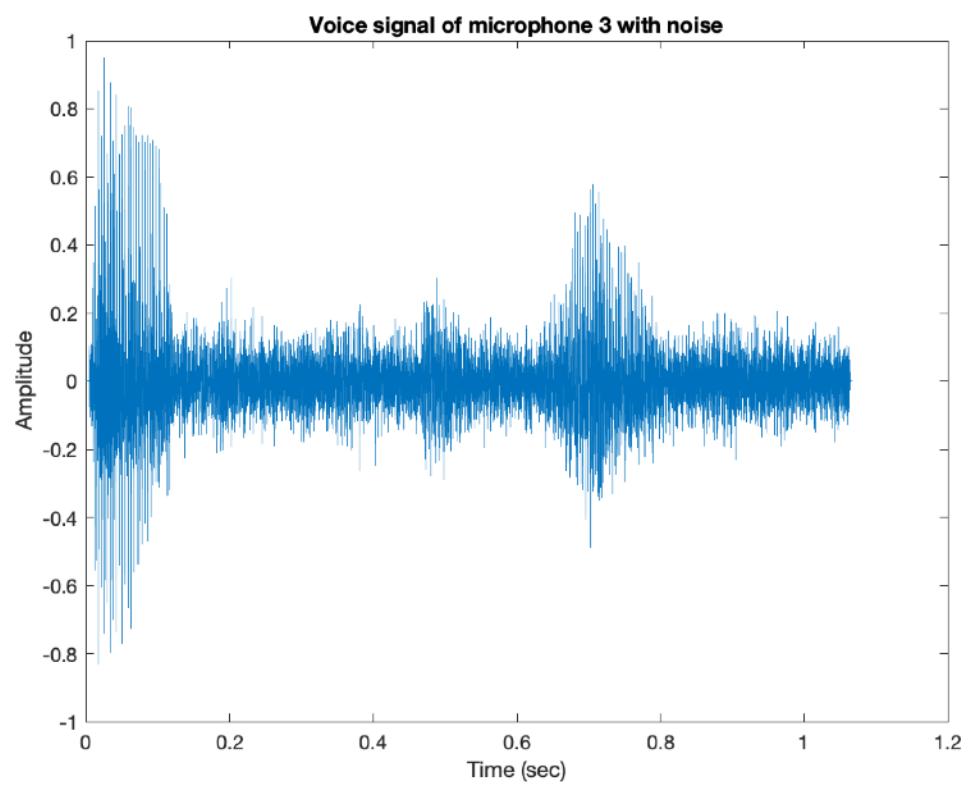
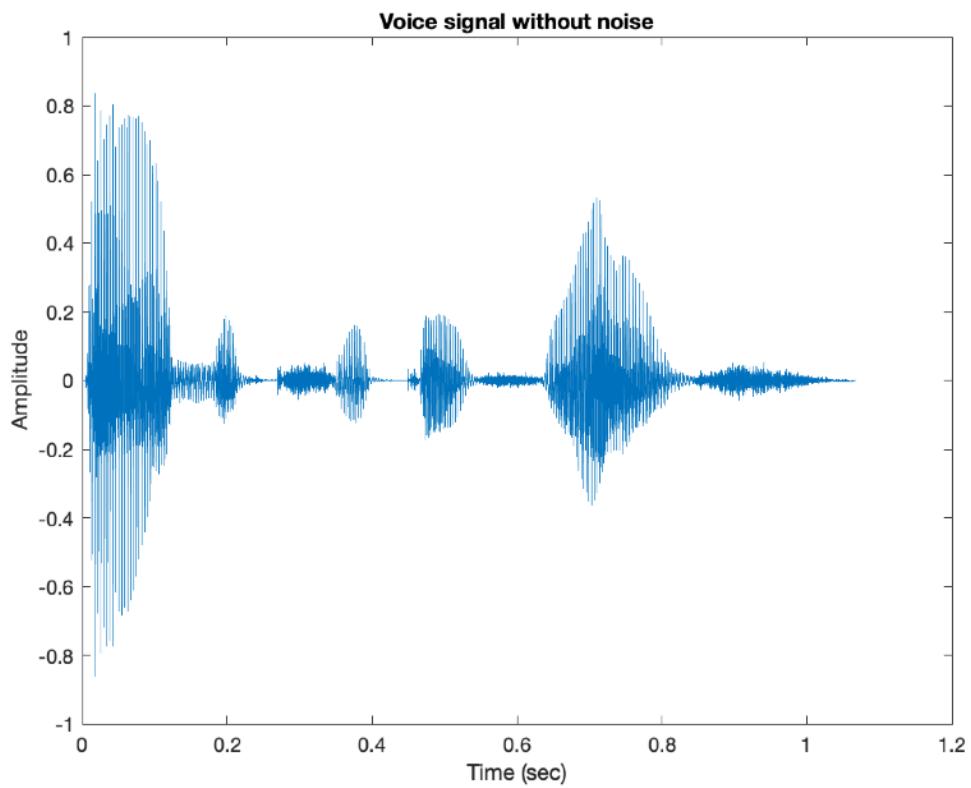
$$\mathbf{d}(\mathbf{k}) = e^{-j \frac{N-1}{2} \frac{\omega}{c} d \cos \theta} [1, e^{j \frac{\omega}{c} d \cos \theta}, \dots, e^{j(N-1) \frac{\omega}{c} d \cos \theta}]^T \quad \text{για } k = k_s$$

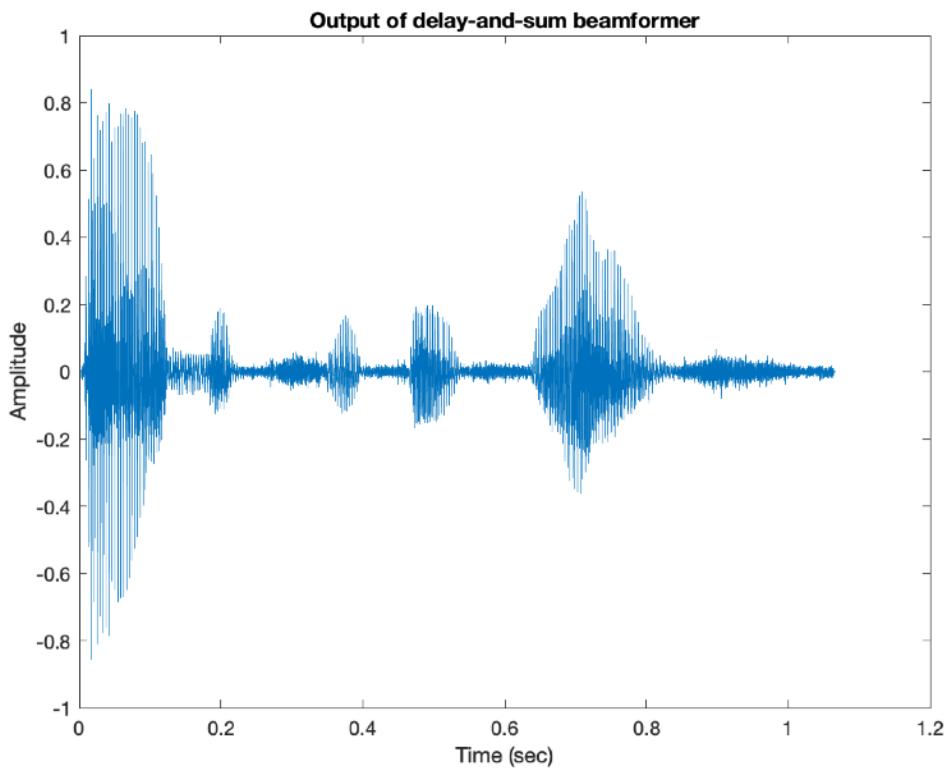
Επιπλέον, βρίσκουμε και τον ερμιτιανό ανάστροφο του  $d(k_s)$ . Στην συνέχεια, δημιουργούμε ένα time-shifting φίλτρο με απόκριση συχνότητας  $H$  που δίνεται από την παρακάτω σχέση :

$$H_{DS}^T(\omega) = \frac{1}{N} d^H(k_s)$$

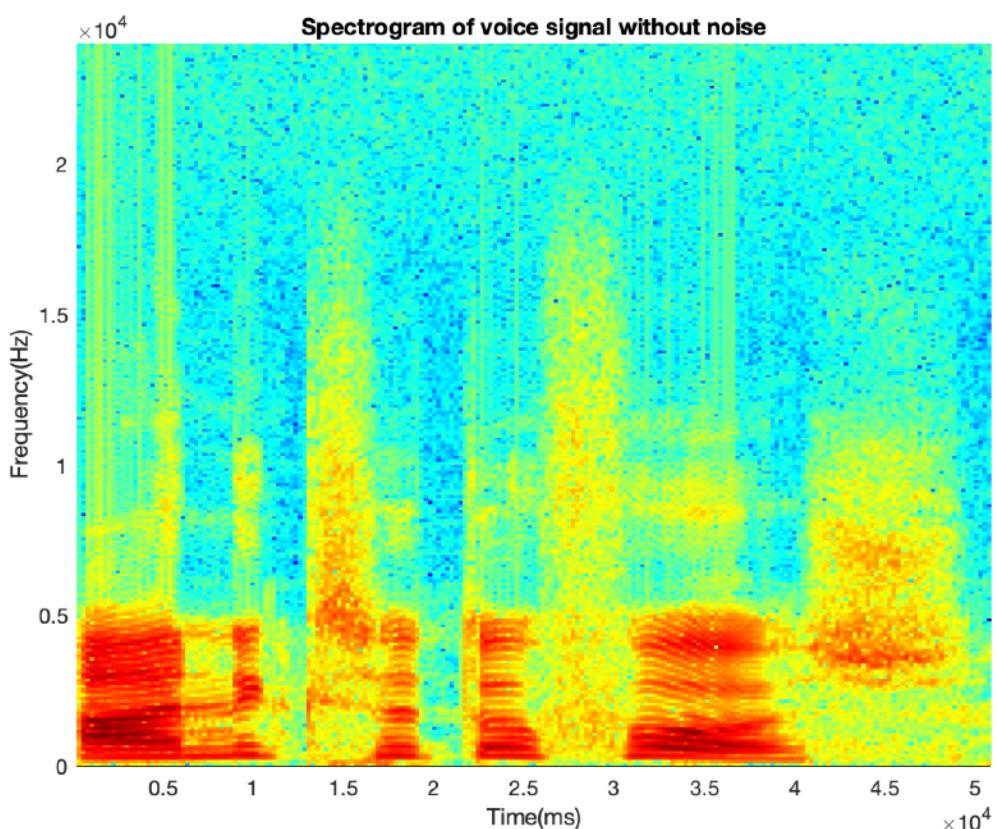
Τα καταγεγραμμένα από τα μικρόφωνα της συστοιχίας σήματα, στην συνέχεια, περνάνε από το παραπάνω φίλτρο, προκειμένου να ευθυγραμμιστούν κατάλληλα στο χρόνο, ώστε, τελικά, η άθροιση των μετατοπισμένων αύτων σημάτων να μας δώσει ένα συνολικό σήμα, απαλλαγμένο από θόρυβο. Το τελικά παραγόμενο σήμα το ονομάζουμε, *sim\_ds.wav*

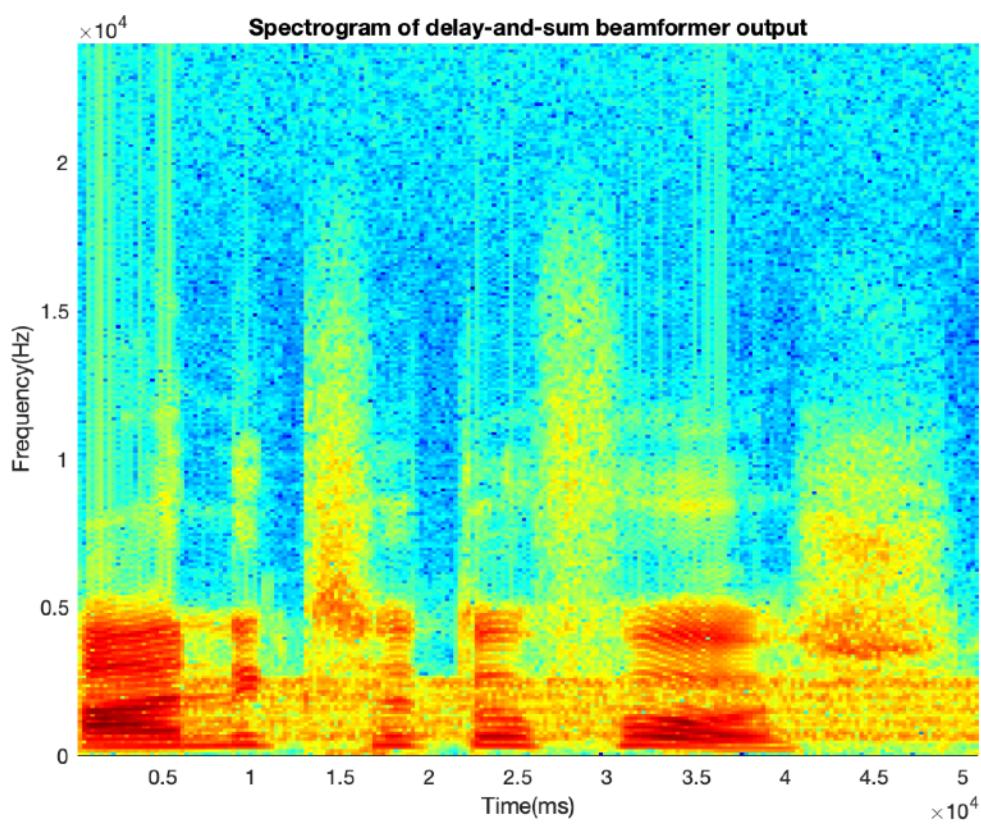
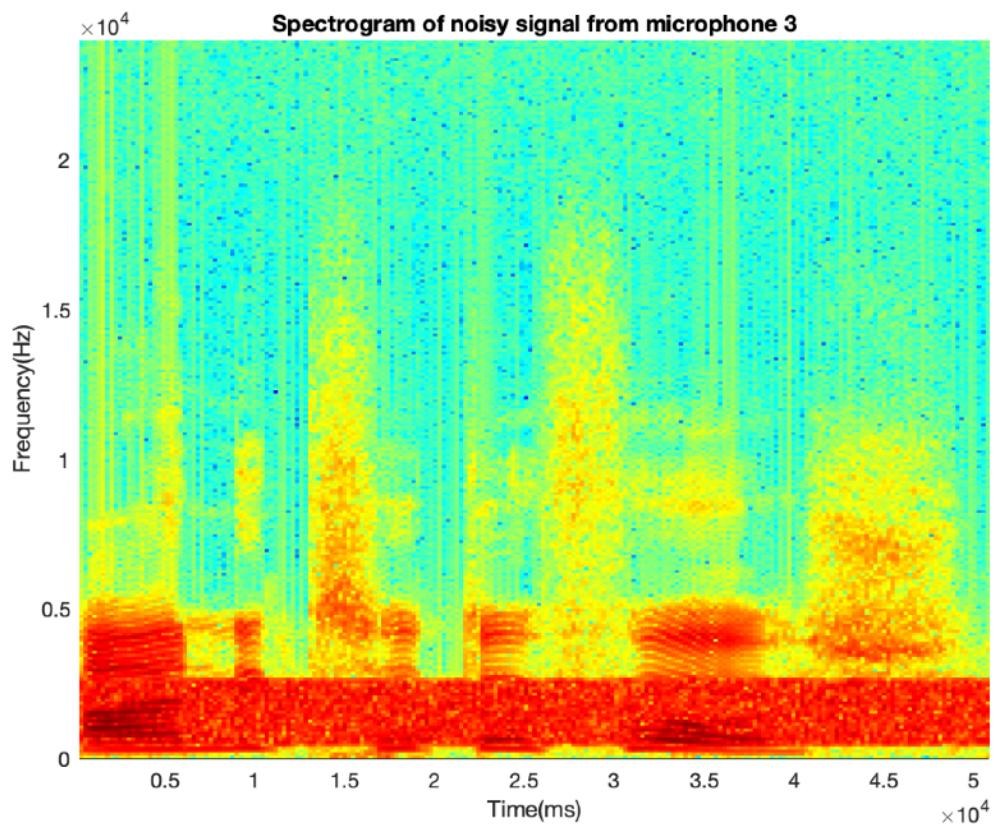
2. Παρακάτω ακολουθούν τα γραφήματα των κυματομορφών (α) του καθαρού σήματος φωνής, (β) του θορυβώδους σήματος στο κεντρικό μικρόφωνο της συστοιχίας ( $n=3$ ) και (γ) της εξόδου του *delay-and-sum beamformer*.





Εξετάζοντας τις κυματομορφές του καθαρού σήματος ( πρώτο γράφημα ) και της εξόδου του Beamformer ( τελευταίο γράφημα ), παρατηρούμε μεγάλη ομοιότητα μεταξύ των δύο σημάτων. Η οπτική αύτη ομοιότητα των κυματομορφών ισοδυναμεί με καλή αποθορυβοποίηση του σήματος που καταγράφεται από την συστοιχία των μικροφώνων. Το συμπέρασμα αυτό επιβεβαιώνεται από το παραγόμενο αρχείο ήχου, το οποίο πραγμάτι ακουστικά διαφέρει ελάχιστα από το καθαρό σήμα. Στην συνέχεια παρατίθενται τα αντίστοιχα σπεκτρογραφήματα :





Με τα παραπάνω σπεκτρογραφήματα επαληθεύμουμε ,αρχικά, πως ο θόρυβος είναι πράγματι ζωνοπερατός με την ενέργειά του να είναι συγκεντρωμένη σε μια συχνοτική περιοχή από 500 Hz έως και 2.5 kHz. Επιπλέον, συγκρίνοντας τα σπεκτρογραφήματα του τελικού σήματος και του καθαρού σήματος επιβεβαιώνουμε το παραπάνω συμπέρασμά μας περί ικανοποιητικής αποθορυβοποίησης του σήματος.

3. Στο σημείο αύτο, καλούμαστε να υπολογίσουμε το σηματοθορυβικό λόγο (SNR) για το θορυβώδες σήμα που καταγράφει το κεντρικό μικρόφωνο της συστοιχίας, καθώς και για το σήμα εξόδου του delay-and-sum beamformer. Σημείωνεται ότι ο θόρυβος στις δύο περιπτώσεις είναι διαφορετικός, ώστοσο σε κάθε περίπτωση ισούται με την διαφορά του καθαρού σήματος φωνής μείον το αντίστοιχο σήμα. Τελικά με την χρήση της συνάρτησης "snr()" βρίσκουμε τα εξής αποτελέσματα :

  - SNR στο κεντρικό μικρόφωνο = 3 dB
  - SNR στην έξοδο του Beamformer = 20.35 dB

Με σύγκριση των δύο παραπάνω σηματοθορυβικών λόγων φαίνεται ξεκάθαρα πως η ποιότητα του σήματος εξόδου είναι σημαντικά βελτιωμένη σε σχέση με αυτή του σήματος που καταγράφει το κεντρικό μικρόφωνο. Σε αυτό το συμπέρασμά καταλήγει κανείς , αναλογιζόμενος το γεγονός ότι ο σηματοθορυβικός λόγος ,εξ ορισμού, ισούται με το πηλικό της ισχύος του καθαρού σήματος προς την ισχύ του θορύβου. Επομένως, δεδομένου ότι το καθαρό σήμα και στις δύο περιπτώσεις είναι το ίδιο, η μεταβολή(αύξηση) του σηματοθορυβικού λόγου από την πρώτη στην δεύτερη περίπτωση μπορεί μόνο να ισοδυναμεί με σημαντική μείωση του θορύβου στο τελικά παραγόμενο σήμα.

#### **B) Μονοκαναλικό Wiener φίλτραρισμα**

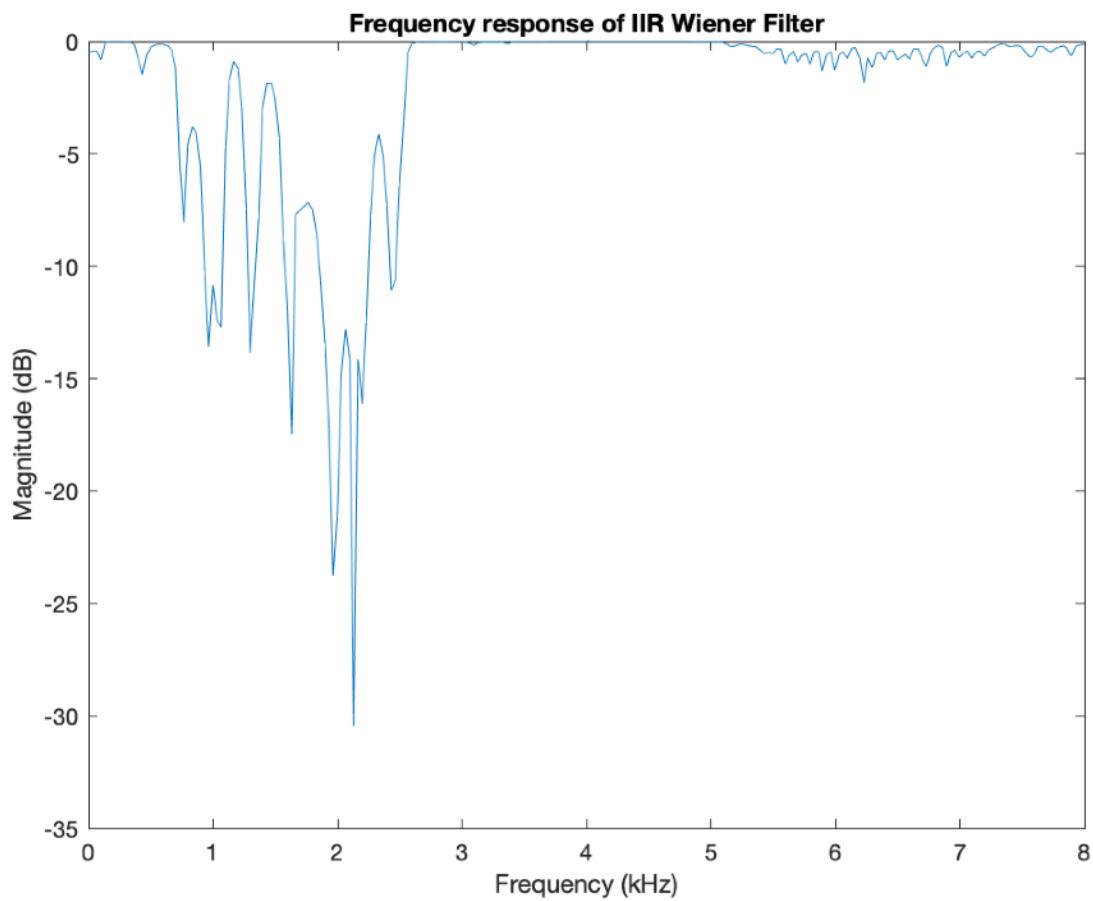
Για το ερώτημα αυτό, επιλέγουμε ένα μόνο πλαίσιο διάρκειας 30 msec , από 0.47 sec έως και 0.5 sec, από το σήμα που έχει καταγράψει το κεντρικό μικρόφωνο της συστοιχίας. Όπως και στο προηγούμενο ερώτημα, έτσι και τώρα θα επιχειρίσουμε αποθορυβοποίηση του συγκεκριμένου πλαισιού ,αυτή την φορά, με την μονοκαναλική μέθοδο του Wiener φίλτραρισματος.

1. Αρχικά, υπολογίζουμε την απόκριση συχνότητας του IIR Wiener φίλτρου, η οποία ,καθώς τα σήματα φωνής και θορύβου είναι ασυσχέτιστα, δίνεται από την σχέση :

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

όπου  $P_v(\omega)$ το φάσμα ισχύος (power spectrum) του θορύβου και  $P_x(\omega)$ το φάσμα ισχύος του συνολικού σήματος x(t).

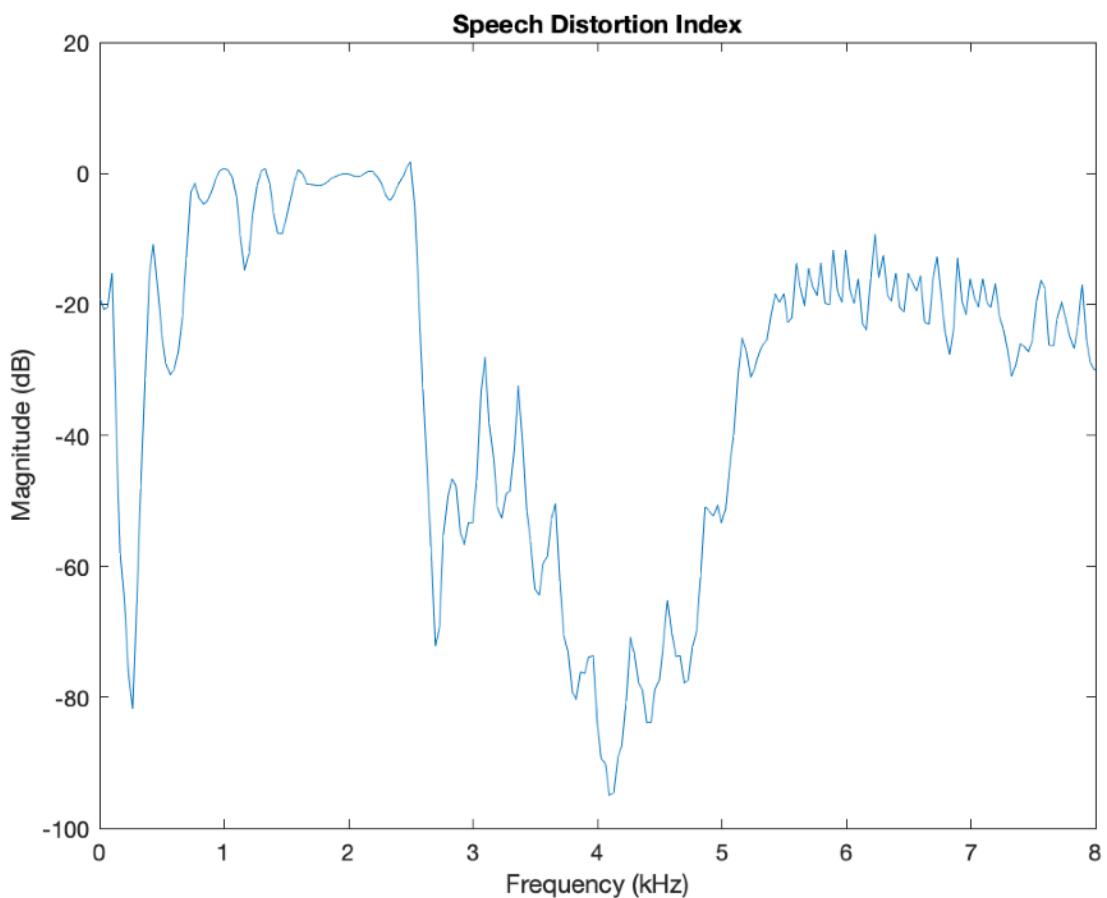
Παρακάτω φαίνεται το γράφημα της απόκρισης συχνότητας του IIR Wiener φίλτρου :



2. Η χρήση του Wiener φίλτρου προκαλεί παραμόρφωση (distortion) στο σήμα φωνής  $s(t)$ , η οποία ισούται με  $s(t) - h_W(t) * s(t)$ . Μία χρήσιμη μετρική για την παραμόρφωση αυτή είναι το speech distortion index που υπολογίζεται από την σχέση :

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

Παρακάτω φαίνεται σε λογαριθμική κλίμακα το  $n_{sd}$  για συχνότητες από 0 έως και 8 kHz.



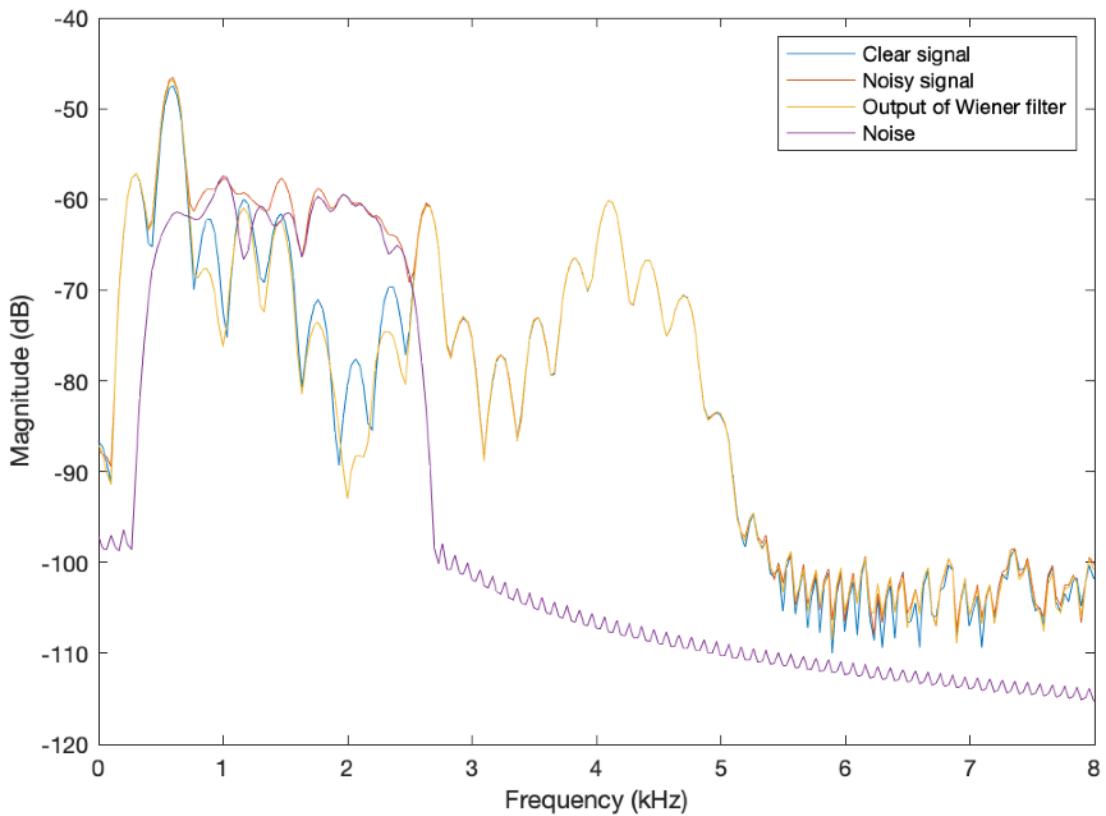
Στο σημείο αυτό, αξίζει να σημειωθεί πως στις χαμηλές συχνότητες έως και 800 Hz που αντιστοιχούν στο φάσμα συχνοτήτων της ανθρώπινης φωνής το φίλτρο εισάγει σχετικά μικρή παραμόρφωση, ενώ για συχνότητες από 1 kHz έως και περίπου 2.5 kHz το φίλτρο εισάγει μεγάλη παραμόρφωση.

### 3. Στην συνέχεια, εφαρμόζουμε το Wiener φιλτράρισμα για το παραπάνω πλαίσιο.

Παρακάτω παρατίθενται σε κοινό γράφημα τα φάσματα ισχύος ,σε λογαριθμηκή κλίμακα, των εξής σημάτων :

- Του καθαρού σήματος φωνής στην είσοδο του φίλτρου
- Του θορυβώδους σήματος στην είσοδο του φίλτρου
- Του σήματος εξόδου του φίλτρου
- Το σήμα θορύβου στην είσοδο του φίλτρου

## Power Spectra



Από το παραπάνω γράφημα, παρατηρούμε πως στην συχνοτική περιοχή 0 έως 2.5 kHz , όπου εντοπίζεται μεγάλη απόκλιση μεταξύ του καθαρού και του θορυβώδους σήματος , το Wiener φίλτρο επιτυγχάνει ,σε μεγάλο βαθμό, να φέρει την εξόδο κοντά στο επιθυμητό σήμα (καθαρό σήμα). Τέλος, για συχνότητες μεγαλύτερες των 3 kHz παρατηρούμε ότι η έξοδος του φίλτρου ταυτίζεται απόλυτα με το καθαρό σήμα φωνής.

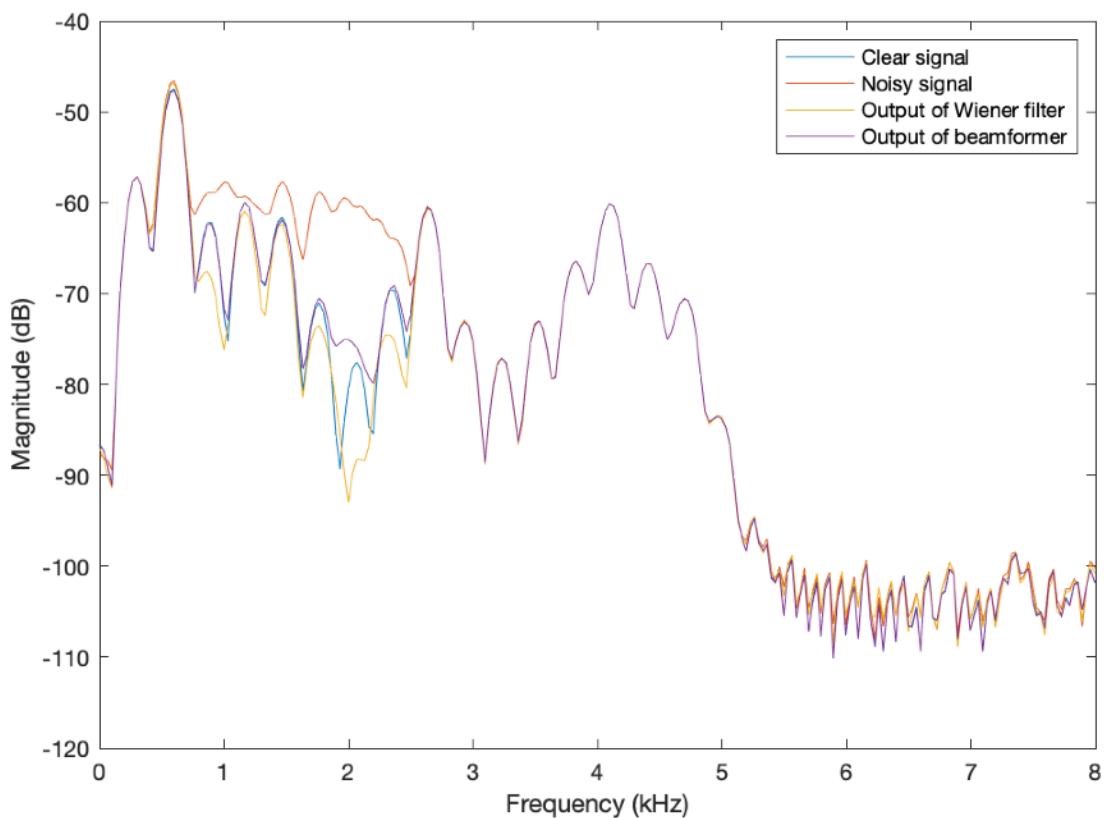
4. Τέλος , υπολογίζουμε το σηματοθορυβικό λόγο για (α) το σήμα που προκύπτει στην έξοδο του Wiener φίλτρου , (β) το θορυβώδες σήμα στην είσοδο του φίλτρου, (γ) την έξοδο του delay-and-sum beamformer αν ως είσοδός του εφαρμοζόταν το παραπάνω πλαίσιο. Με την χρήση της συνάρτησης "snr()" παίρνουμε τα παρακάτω αποτελέσματα :

- SNR στην έξοδο του φίλτρου = 10.83 dB
- SNR στην είσοδο του φίλτρου = 3.14 dB
- SNR στην έξοδο του delay-and-sum beamformer = 20.81 dB

Συνεπώς , η βελτίωση στο SNR του συγκεκριμένου πλαισίου που επιτεύχθηκε με την πολυκαναλική μέθοδο (delay-and-sum beamformer) είναι ίση με  $20.81 - 10.83 = 9.98$  dB.

Επομένως, στην συγκεκριμένη εφαρμογή , η πολυκαναλική μέθοδος αποθορυβοποίησης έχει καλύτερα αποτελέσματα συγκριτικά με την μονοκαναλική μέθοδο ( Wiener φιλτράρισμα ). Το παραπάνω συμπέρασμα επιβεβαιώνεται αν χαράξουμε τα φάσματα ισχύος , σε κοινό γράφημα και σε λογαριθμική κλίμακα , για το καθαρό σήμα φωνής για το δεδομένο πλαίσιο, το θορυβώδες πλαίσιο, την έξοδο του Wiener φίλτρου και την έξοδο του delay-and-sum beamformer για το δεδομένο πλαίσιο

Power Spectra



## Μέρος 2.2

Εδώ θεωρούμε γραμμική συστοιχία  $N = 7$  μικροφώνων με απόσταση  $d = 4$  cm που καταγράφει ένα σήμα φωνής που προέρχεται από γωνία  $\theta = 45^\circ$ . Σε αυτή την περίπτωση, έχουμε θόρυβο ο οποίος δεν προέρχεται από σημειακή πηγή αλλά από διάφορες πηγές, που δημιουργούν ένα diffuse noise field (ισοτροπικό και ομογενές πεδίο θορύβου, τον οποίο θεωρούμε στάσιμο).

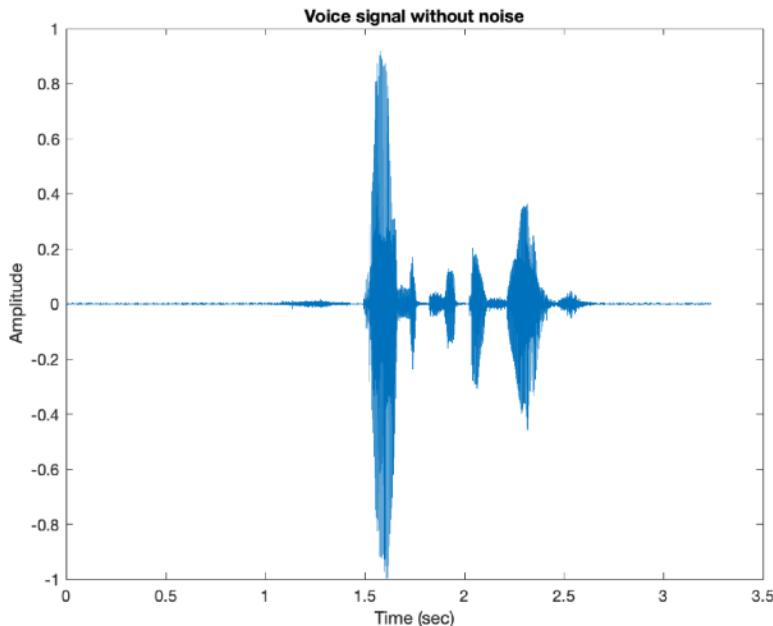
### A) Delay-and-sum Beamforming

1. Αρχικά, υπολογίζουμε τα βάρη για τον delay-and-sum beamformer (αφού πρώτα υπολογίσουμε, όπως και προηγουμένως, το διάνυσμα  $\mathbf{d}(\mathbf{k})$ ) και εφαρμόζουμε το beamforming. Είσοδος του beamformer είναι και εδώ το θορυβώδες σήμα που προέρχεται από το κεντρικό μικρόφωνο της συστοιχίας. Δηλαδή, μετά τη χρονική ευθυγράμμιση, αθροίζουμε τα σήματα που προκύπτουν από την έξοδο του beamformer και λαμβάνουμε το τελικό σήμα, το οποίο ηχογραφούμε στο αρχείο “real\_ds.wav”.

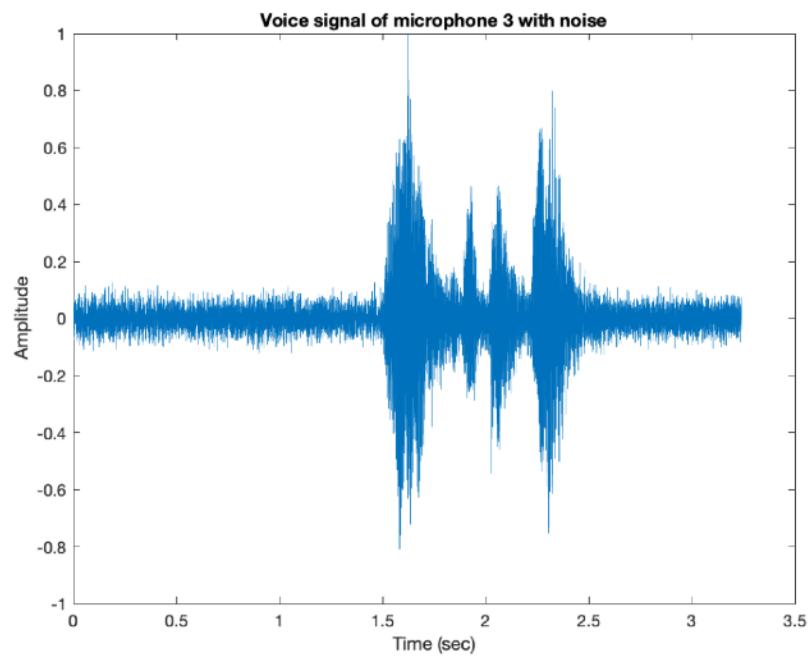
Παρατηρούμε ότι το σήμα που προκύπτει δεν παρουσιάζει, ακουστικά, ουσιαστική βελτίωση σε σχέση με το θορυβώδες σήμα του κεντρικού μικροφώνου.

2. Σχεδιάζουμε αρχικά τις εξής κυματομορφές:

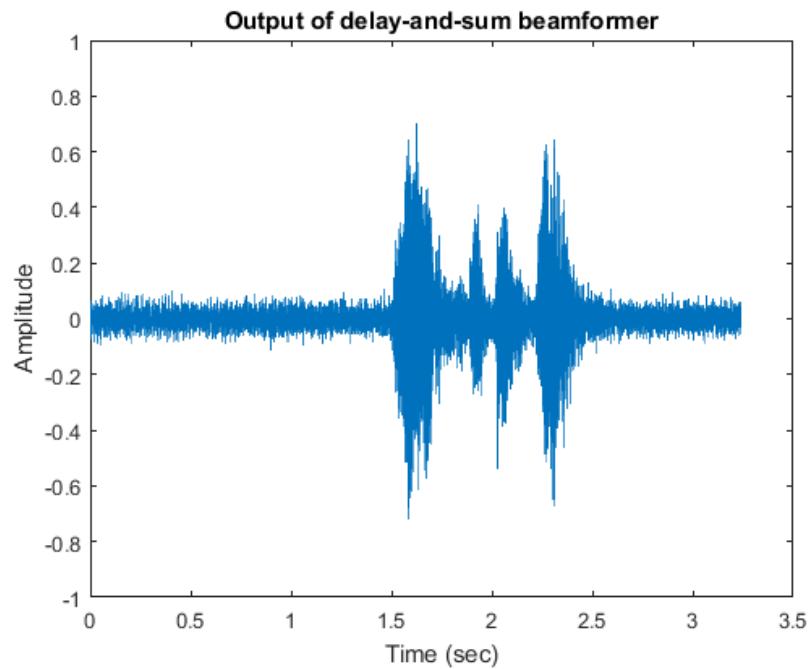
a) Καθαρό σήμα φωνής



β) Θορυβώδες σήμα στο κεντρικό μικρόφωνο της συστοιχίας ( $n=3$ )



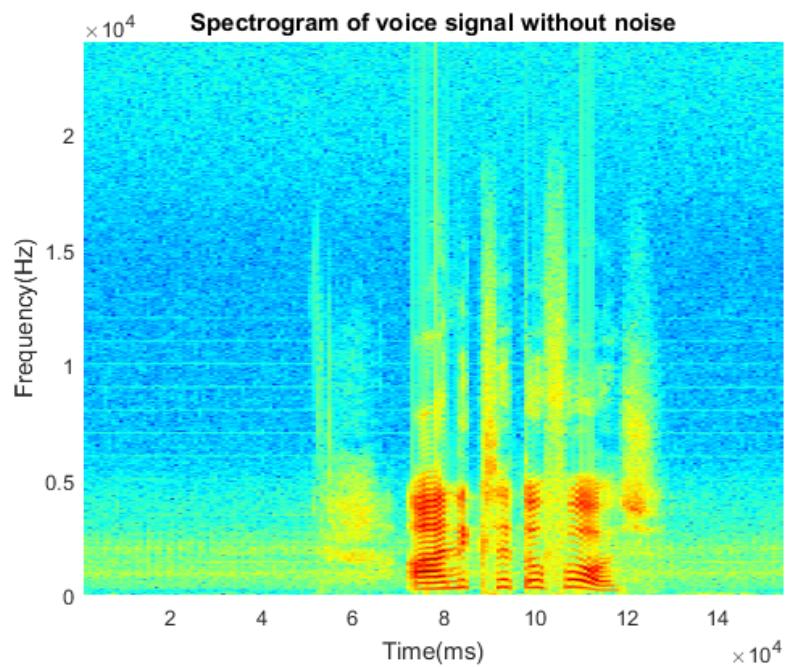
γ) Έξοδος του delay-and-sum beamformer



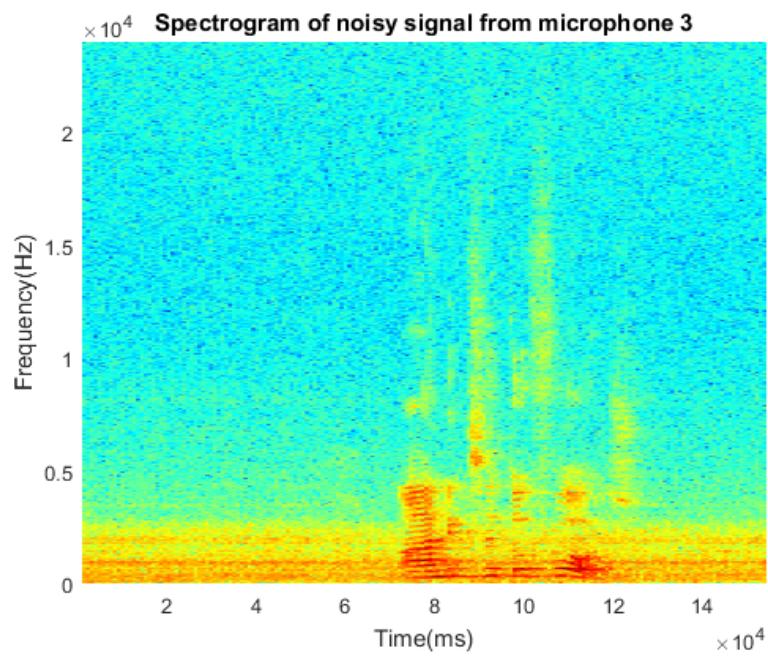
Όπως φαίνεται και από τις γραφικές παραστάσεις των σημάτων, το η έξοδος του beamformer παρουσιάζει ελάχιστες διαφορές σε σχέση με την είσοδό του, δηλαδή το θορυβώδες σήμα του κεντρικού μικροφώνου, ενώ και τα δύο σήματα φαίνεται να έχουν σημαντικό θόρυβο σε σχέση με το καθαρό σήμα φωνής.

Στη συνέχεια, σχεδιάζουμε τα σπεκτρογραφήματα για καθένα από τα τρία αυτά σήματα.

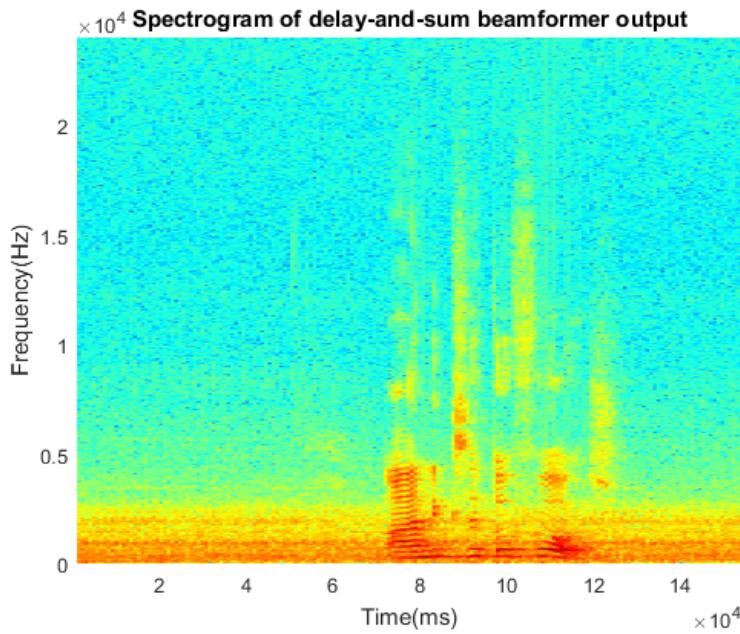
α) Καθαρό σήμα φωνής



β) Θορυβώδες σήμα στο κεντρικό μικρόφωνο της συστοιχίας ( $n=3$ )



γ) ) Έξοδος του delay-and-sum beamformer



Παρατηρούμε ότι το σπεκτρογράφημα της εξόδου του delay-and-sum beamformer εμφανίζει πολλές ομοιότητες με αυτό του σήματος του κεντρικού μικροφώνου, καθώς επίσης και ότι έχουν αξιοσημείωτο θόρυβο σε σχέση με το καθαρό σήμα.

3. Ένας ακόμη τρόπος αξιολόγησης της ποιότητας του σήματος στην έξοδο του beamformer σε σχέση με την είσοδο είναι ο υπολογισμός του segmental SNR, που μας δίνει πιο αξιόπιστη πληροφορία για την αντίληψη του ανθρώπου στα σήματα φωνής. Ουσιαστικά, πρόκειται για το μέσο SNR των πλαισίων βραχέος χρόνου, στα οποία χωρίζουμε το σήμα μας. Το μήκος του πλαισίου ορίζεται ίσο με 30 ms. Το SSNR δίνεται από τη σχέση:

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

όπου  $L$  το μήκος των πλαισίων,  $s(n)$  το καθαρό σήμα φωνής και  $v(n)$  ο θόρυβος και  $M$  το μήκος των πλαισίων στα οποία χωρίζεται το σήμα.

Αρχικά, υπολογίζουμε το SSNR για το σήμα από το κεντρικό μικρόφωνο.

Παρατηρούμε ότι το πρώτο 1 sec του σήματος αποτελείται αποκλειστικά από θόρυβο. Επομένως, για να υπολογίσουμε την ισχύ του θορύβου, απομονώνουμε, με τυχαία επιλογή από τα πρώτα παράθυρα του σήματος, το 15° παράθυρο. Θεωρώντας ότι το σήμα  $x(t)$  του μικροφώνου δίνεται από τον τύπο  $x(t) = s(t) + v(t)$ , καθώς επίσης και ότι τα  $s(t), v(t)$  είναι ασυσχέτιστα, μπορούμε να υπολογίσουμε την ισχύ του καθαρού σήματος από τον τύπο:  $\sigma_s^2 = \sigma_x^2 - \sigma_v^2$ .

Χρησιμοποιώντας αυτά τα δεδομένα, υπολογίζουμε το SNR για κάθε παράθυρο των 30 ms. Γενικά, αν σε κάποιο παράθυρο το SNR είναι μεγαλύτερο από 35 dB, δεν έχει σημαντική ακουστική διαφορά από ένα παράθυρο με SNR 35 dB. Γι' αυτό, το SNR ενός παραθύρου με  $SNR > 35dB$  τίθεται ίσο με 35 dB. Επίσης, αν το SNR σε κάποιο παράθυρο είναι μικρότερο από 0 dB, συμπεραίνουμε ότι το πλαίσιο αυτό είναι πλαίσιο σιωπής, οπότε το αγνοούμε στον υπολογισμό του SSNR.

Συνυπολογίζοντας όλες αυτές τις παραδοχές, το SSNR του κεντρικού μικροφώνου της συστοιχίας υπολογίστηκε ίσο με 11.59 dB.

Επίσης, ακολουθώντας την ίδια διαδικασία και επιλέγοντας εκ νέου το  $15^{\circ}$  πλαίσιο της παραθυρωμένης εξόδου του beamformer ως πλαίσιο θορύβου, υπολογίζουμε το SSNR για την έξοδο του beamformer, το οποίο προκύπτει ίσο με 12.23 dB.

Η βελτίωση που επιτυγχάνεται είναι 0.64 dB δεν είναι ικανοποιητική και επιβεβαιώνει την περιορισμένη αποθορυβοποίηση που επιτυγχάνεται με την εφαρμογή του beamforming.

### B) Post-filtering με Wiener φίλτρο

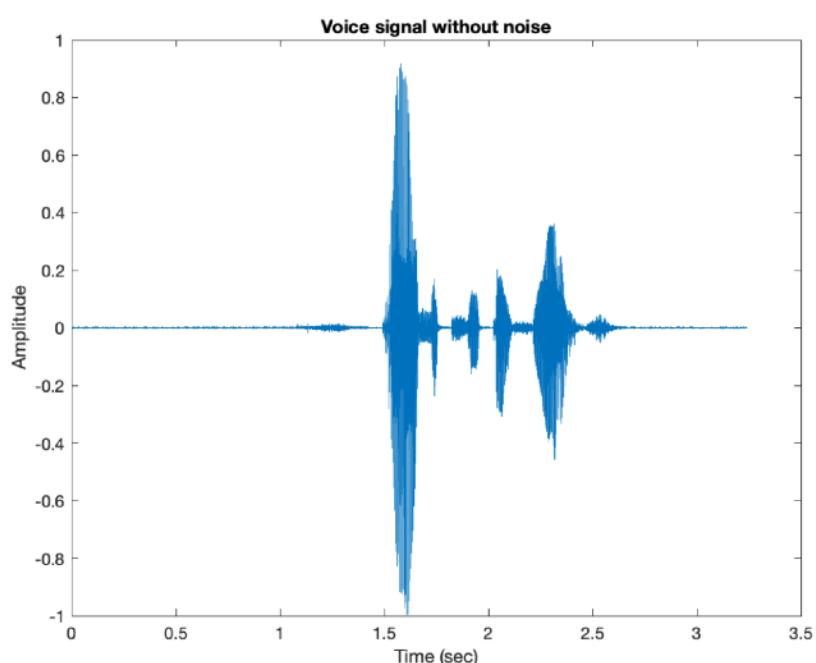
Δεδομένου ότι στην περίπτωσή μας έχουμε ένα diffuse noise field, τα σήματα θορύβου εμφανίζουν μεγάλη συσχέτιση μεταξύ των μικροφώνων, κυρίως στις χαμηλές συχνότητες.

1. Γι' αυτό, εφαρμόζουμε μονοκαναλικό IIR Wiener φιλτράρισμα στην έξοδο του beamformer.

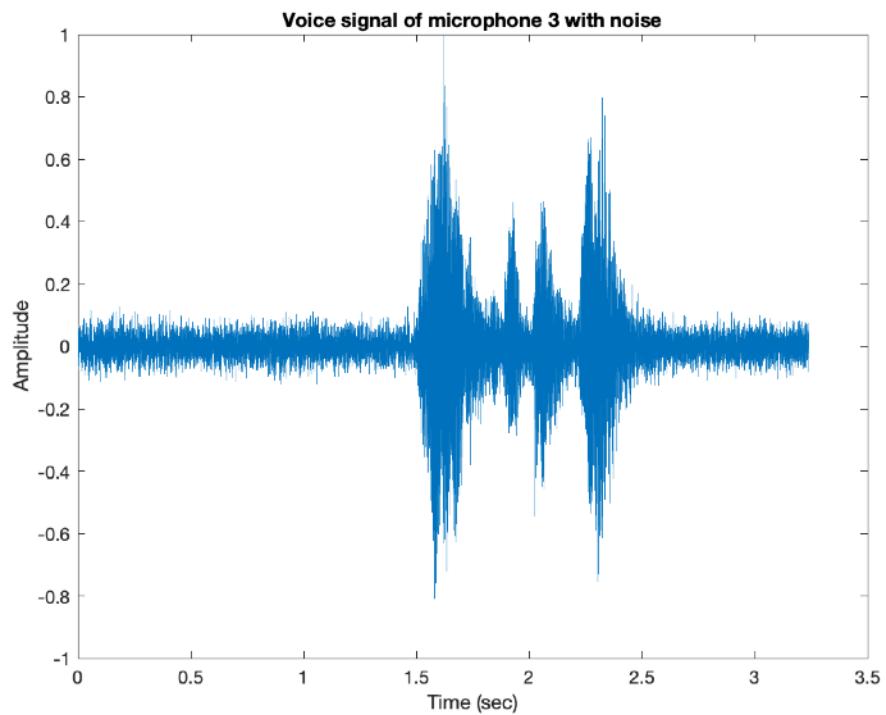
Αρχικά, παραθυροποιούμε την έξοδο του beamformer σε πλαίσια των 30 ms, με επικάλυψη 40%. Χρησιμοποιούμε παράθυρο Hamming. Για να υπολογίζουμε τη συνάρτηση μεταφοράς του φίλτρου, πρέπει να υπολογίζουμε το φάσμα ισχύος του θορύβου, καθώς επίσης και της εισόδου στο φίλτρο, δηλαδή της έξοδου του beamformer. Για το φάσμα ισχύος του θορύβου, που υπολογίζεται με τη χρήση της συνάρτησης pwelch, χρησιμοποιούμε το  $15^{\circ}$  πλαίσιο του παραθυροποιημένου σήματος. Στη συνέχεια, φιλτράρουμε κάθε πλαίσιο και το σήμα που προκύπτει στην έξοδο του φίλτρου, το οποίο είναι ακόμα παραθυροποιημένο, πρέπει να ανακατασκευαστεί με χρήση της μεθόδου overlap-add. Έτσι, λαμβάνουμε το τελικό σήμα, το οποίο ηχογραφούμε στο αρχείο “real\_mmse.wav”. Παρατηρούμε ότι ακουστικά υπάρχει σημαντική βελτίωση σε σχέση με το αρχείο που προέκυψε από την έξοδο του beamformer. Συγκεκριμένα, έχει επιτευχθεί καλύτερη αποθορυβοποίηση, ειδικά στο κομμάτι του σήματος φωνής.

2. Σχεδιάζουμε αρχικά τις εξής κυματομορφές:

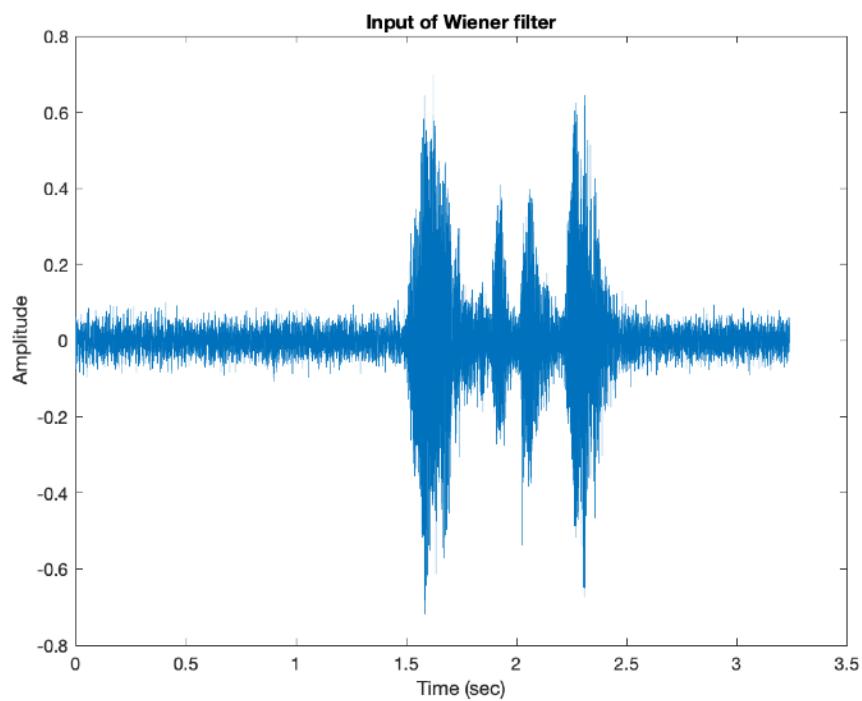
α) Καθαρό σήμα φωνής



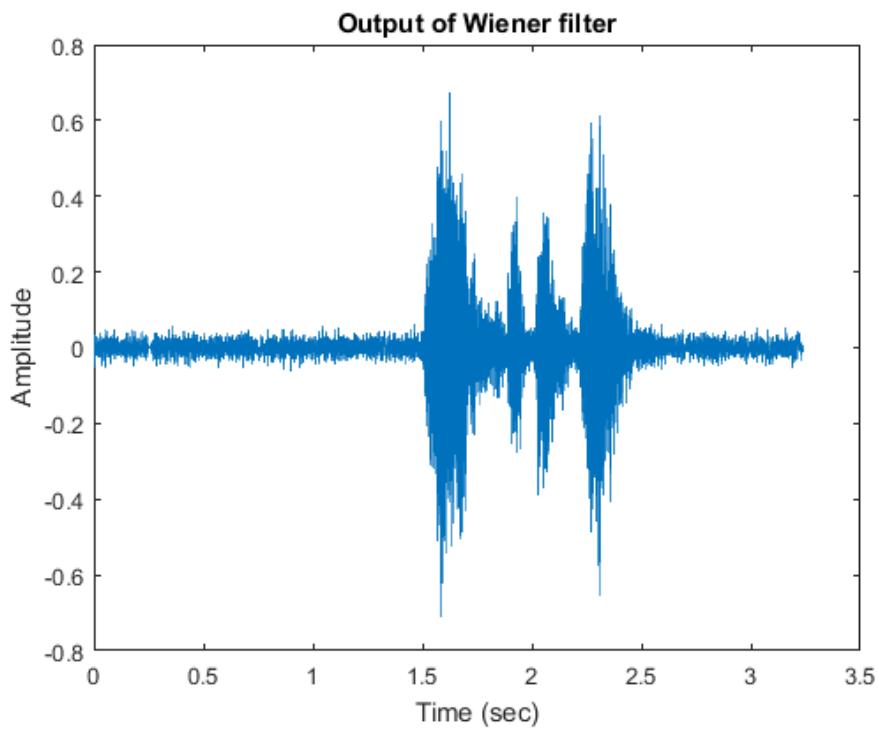
β) Θορυβώδες σήμα στο κεντρικό μικρόφωνο ( $n=3$ )



γ) Είσοδος του Wiener φίλτρου



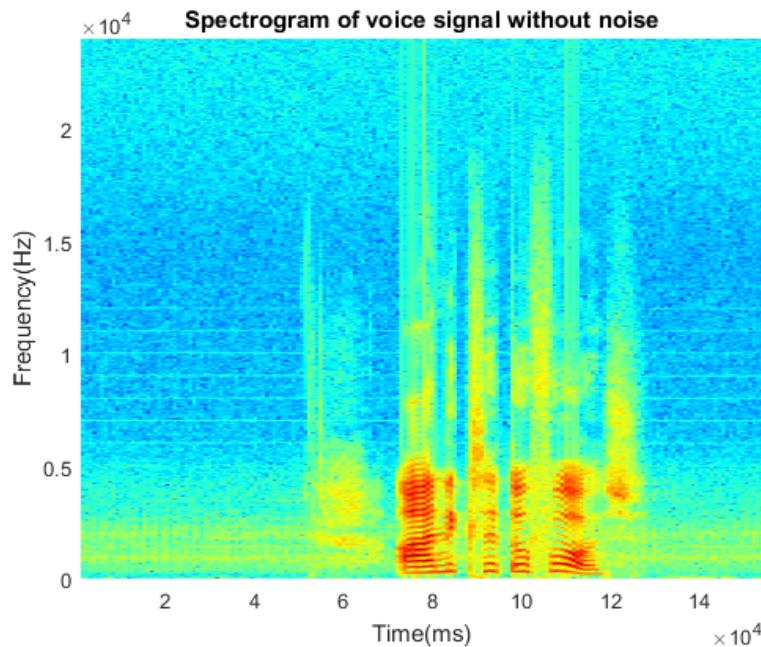
δ) Έξοδος του Wiener φίλτρου



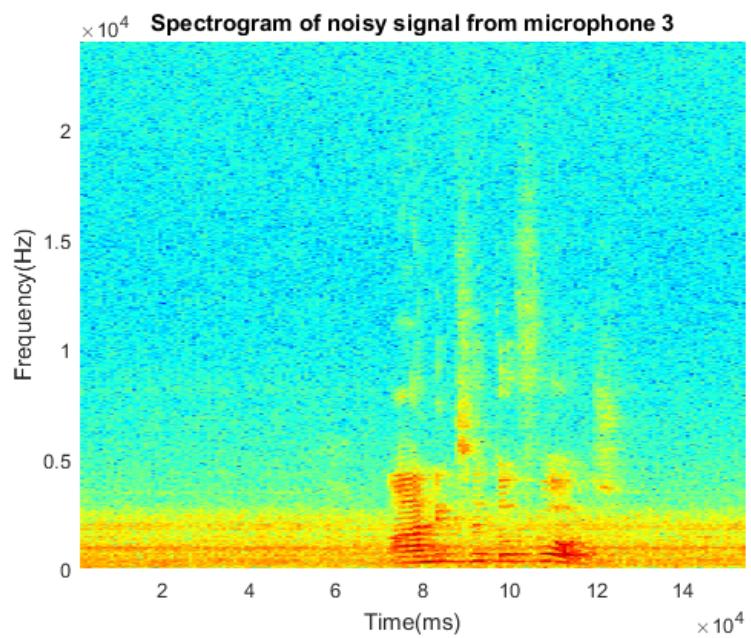
Παρατηρούμε από την κυματομορφή της εξόδου του Wiener φίλτρου ότι ο θόρυβος είναι σημαντικά μικρότερος σε σχέση με την είσοδο, οπότε το φίλτρο συμβάλλει στην περαιτέρω αποθορυβοποίηση του σήματος. Ωστόσο, αφού το σήμα είναι πραγματικό δεν μπορεί να εξαλειφθεί εντελώς.

Στη συνέχεια, σχεδιάζουμε τα σπεκτρογραφήματα αυτών των σημάτων:

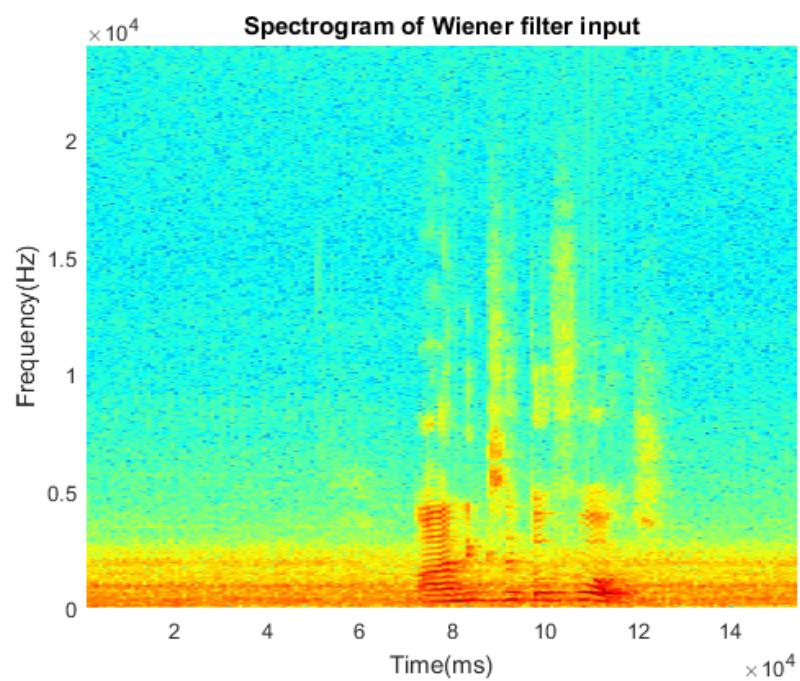
a) Καθαρό σήμα



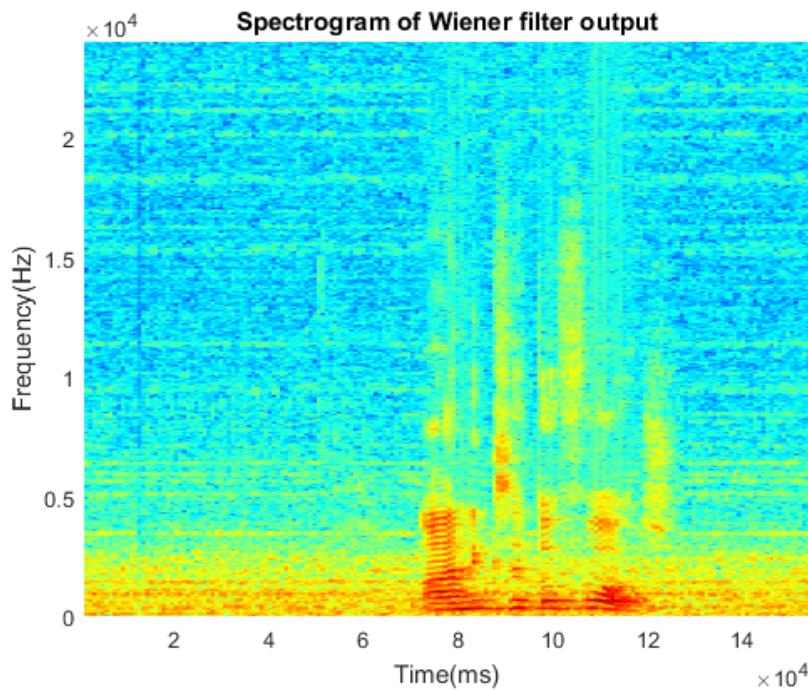
β) Θορυβώδες σήμα στο κεντρικό μικρόφωνο ( $n=3$ )



γ) Είσοδος του Wiener φίλτρου



δ) Έξοδος του Wiener φίλτρου



Παρατηρούμε τώρα ότι το σπεκτρογράφημα του σήματος εξόδου του Wiener φίλτρου εμφανίζει περισσότερες ομοιότητες με το σπεκτρογράφημα του καθαρού σήματος. Είναι φανερό ότι ο θόρυβος έχει περιοριστεί, τόσο στις χαμηλές όσο και στις υψηλές συχνότητες.

3. Το SSNR στην είσοδο του φίλτρου ταυτίζεται με το SSNR στην έξοδο του beamformer, το οποίο υπολογίστηκε προηγουμένως ίσο με 12.23 dB. Για να υπολογίσουμε το SSNR στην έξοδο του Wiener φίλτρου, την παραθυροποιούμε σε πλαίσια των 30 ms και επιλέγουμε το 15<sup>o</sup> ως πλαίσιο θορύβου. Από το πλαίσιο αυτό υπολογίζουμε το φάσμα ισχύος του θορύβου. Με την ίδια διαδικασία και παραδοχές με το ερώτημα 2.2.A.3 και με κατώτατο κατώφλι και εδώ τα 0 dB, το SSNR της εξόδου υπολογίζεται ίσο με 14.43 dB. Παρατηρούμε ότι το SSNR βελτιώθηκε κατά 2.2 dB, που αποτελεί ικανοποιητική βελτίωση σε σχέση με τη βελτίωση των 0.6 dB που παρουσίασαν τα σήματα εισόδου και εξόδου του beamformer. Το αποτέλεσμα εξακολουθεί να είναι θορυβώδες, όμως είναι αρκετά καλύτερο, τόσο από το αρχικό, όσο και από την έξοδο του beamformer.

4. Τέλος, υπολογίζουμε το SSNR για κάθε σήμα εισόδου του delay-and-sum beamformer και στη συνέχεια τον μέσο όρο τους. Ο μέσος όρος των SSNR είναι ίσος με 11.45 dB. Στην τελική έξοδο του συστήματος, το SSNR υπολογίστηκε ίσο με 14.43 dB από το προηγούμενο ερώτημα. Άρα, η συνολική βελτίωση που επιτεύχθηκε είναι 2.98 dB, η οποία μεταφράζεται σίγουρα σε αποθορυβοποίηση του αρχικού σήματος, χωρίς όμως να επιτυγχάνεται απομάκρυνση του θορύβου τόσο μεγάλη όσο στα προσομοιωμένα σήματα του ερωτήματος 2.1.