



Ψηφιακή Επεξεργασία Σημάτων

3^η Εργαστηριακή Άσκηση

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

ΠΑΝΑΓΙΩΤΑΡΑΣ ΗΛΙΑΣ ΑΜ: 03115746

Ακαδ. Έτος 2017-18 | Ημ. Παράδοσης 13/07/18 | 6^ο Εξάμηνο

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. Αφού πρώτα κατανοήσουμε βασικές έννοιες, όπως το Beamforming, Beam pattern εισαγόμαστε στον delay-and-sum beamformer (conventional beamformer), τον οποίο και θα χρησιμοποιήσουμε στην παρούσα άσκηση πάνω σε ομοιόμορφες γραμμικές συστοιχίες μικροφώνων (uniform linear arrays).

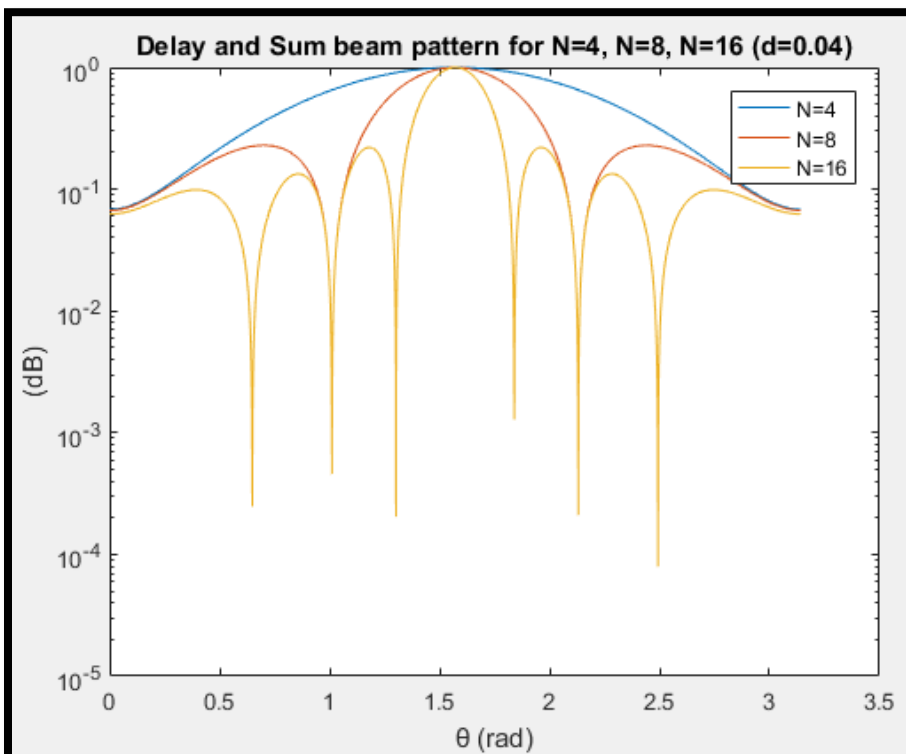
Μέρος 1) Συστοιχίες Μικροφώνων και Χωρικό Φιλτράρισμα (Spatial Filtering)

Σε αυτό το μέρος της άσκησης θεωρούμε ότι έχουμε μία γραμμική συστοιχία μικροφώνων και ένα σήμα που προέρχεται από μία πηγή που έχει γωνία $\theta_s = 90^\circ$ με την συστοιχία.

Υλοποιήσαμε τον τύπο $B(\omega, \theta) = \frac{1}{N} \frac{\sin[\frac{N\omega}{2c} d(\cos\theta - \cos\theta_s)]}{\sin[\frac{\omega}{2c} d(\cos\theta - \cos\theta_s)]}$ που μας δίνεται από την θεωρία της

άσκησης για το delay-and-sum beam pattern και ζητείται να σχεδιαστεί το μέτρο του που είναι συνάρτηση της συχνότητας ω και της γωνίας θ σε λογαριθμική κλίμακα (dB), μεταβάλλοντας κάποιες παραμέτρους (N: αριθμός μικροφώνων, d: απόσταση μεταξύ δύο διαδοχικών μικροφώνων, θ_s : γωνία συστοιχίας με την πηγή). Θεωρήσαμε ότι η γωνία θ μεταβάλλεται από 0 έως 2π rad, η σταθερά της ταχύτητας του ήχου στον αέρα είναι $c=340\text{m/s}$, και η συχνότητα είναι $f=2\text{kHz}$.

1.4.1)



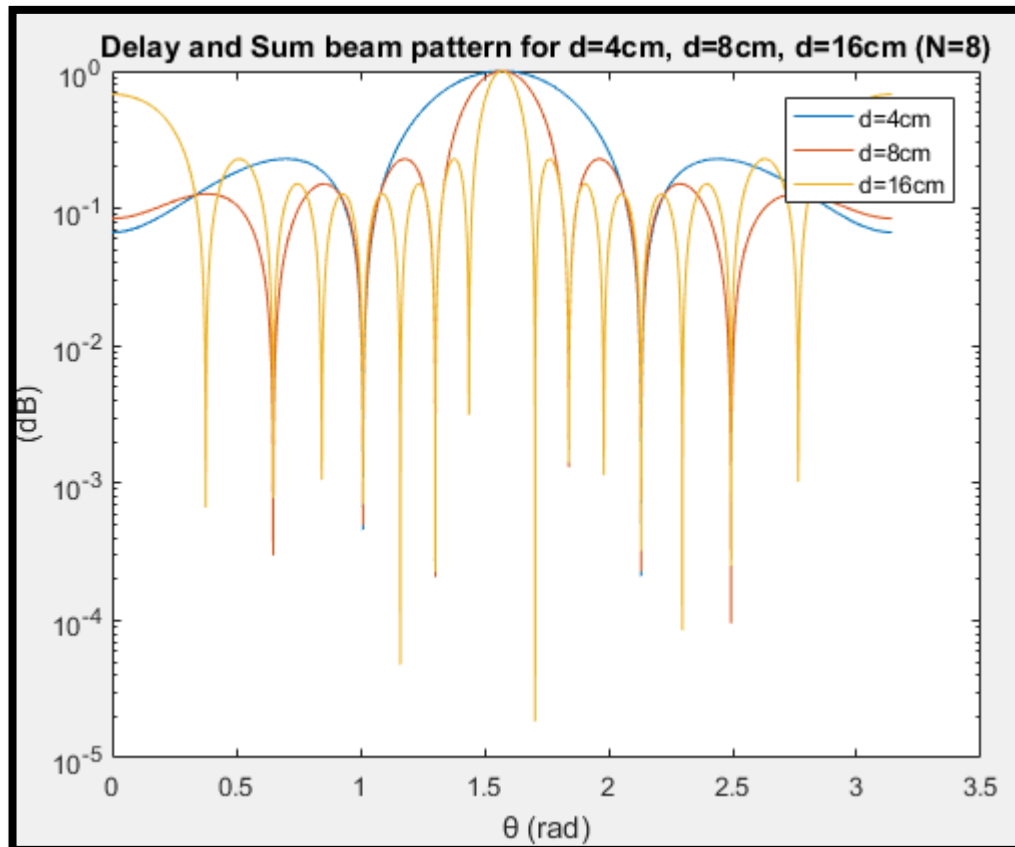
Απόσταση
μικροφώνων
 $d=4\text{cm}$ και
αριθμός
μικροφώνων
 $N=4,8,16$:

Ιδανικά θα θέλαμε να ενισχύσουμε τον ήχο που προέρχεται από μία συγκεκριμένη γωνία ($\theta_s = 90^\circ$). Σύμφωνα με το Παράρτημα Α) αν υπάρχει $\theta \neq \theta_s$ για το οποίο $|B(\omega, \theta)| = 1$, ο beamformer δεν μπορεί να διαχωρίσει το σήμα που προέρχεται από την κατεύθυνση θ_s από σήματα που προέρχονται από άλλες κατευθύνσεις για τις οποίες ισχύει η παραπάνω σχέση, αφού και αυτά τα σήματα θα παραμένουν αναλλοίωτα μετά το beamforming. Για αποφυγή του φαινομένου αυτού απαιτείται να ισχύει: $d < \frac{\lambda}{2}$. Αν δεν ικανοποιείται τότε δημιουργείται spatial aliasing, δηλαδή εμφανίζονται στο beam pattern grating lobes, πλευρικοί λοβοί μοναδιαίου πλάτους και συνεπώς ο beam former αδυνατεί να διαχωρίσει το σήμα που προέρχεται από το steering direction από τα σήματα που προέρχονται από τις κατευθύνσεις των grating lobes. Για $d=4\text{cm}$ και $f=2\text{kHz}$ η συνθήκη ισχύει.

Όπως βλέπουμε από το παραπάνω διάγραμμα δεν δημιουργούνται grating lobes. Παράλληλα, όσο αυξάνουμε το N (αριθμός μικροφώνων) παρατηρούμε πως ο κύριος λοβός στενεύει και μικραίνει το πλάτος των παράπλευρων λοβών με αποτέλεσμα να έχουμε καλύτερη ποιότητα σήματος στην έξοδο. Αυτό θα συμβαίνει καθώς ο θόρυβος μπορεί να είναι ένα σήμα που φτάνει στην γραμμική συστοιχία με μία διαφορετική γωνία από αυτής της πηγής θ_s . Στην γωνία πηγής το σήμα περνάει από το beamformer σε dB κλίμακα είναι 10 και άρα σε πλάτος είναι 1.

1.4.2)

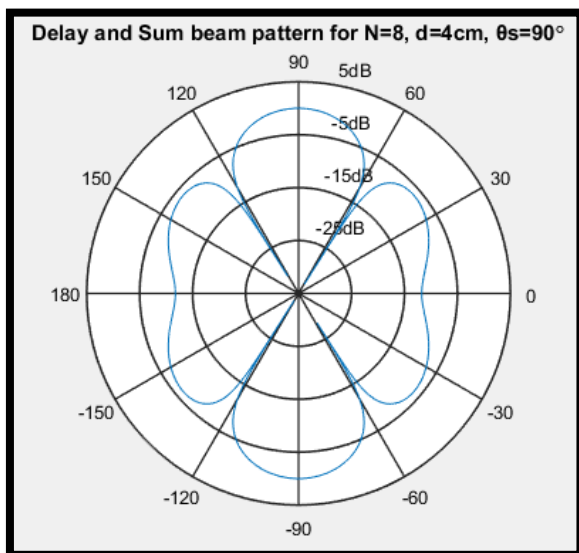
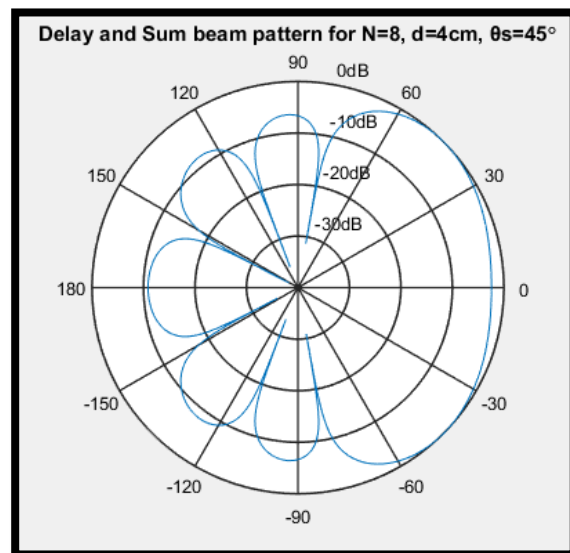
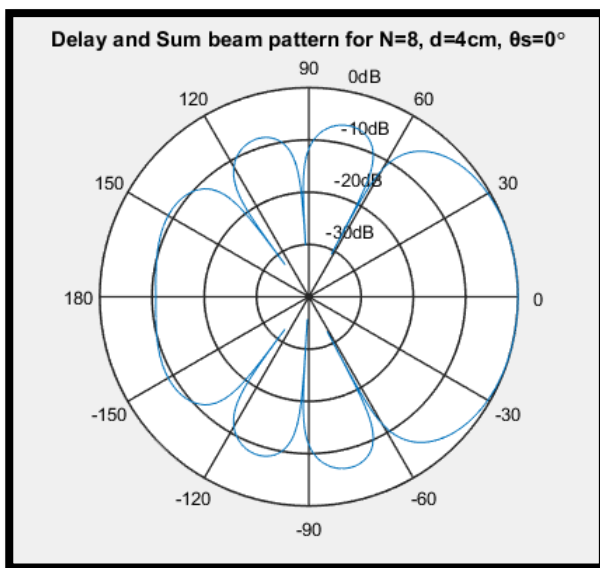
Αριθμός μικροφώνων $N=8$ και απόσταση μικροφώνων $d=4\text{cm}, 8\text{cm}, 16\text{cm}$:



Με βάση τις προηγούμενες παρατηρήσεις μας, βλέπουμε πως για σταθερό αριθμό μικροφώνων και αυξανόμενη απόσταση μεταξύ τους επιτυγχάνουμε «στενότερο» κύριο λοβό, κάτι το οποίο σημαίνει πως ενισχύουμε περισσότερο την επιθυμητή γωνία περισσότερο από τις υπόλοιπες. Ωστόσο, καθώς αυξάνουμε την απόσταση d οι παράπλευροι λοβοί τείνουν να πληθαίνουν και να αυξάνουν το μέτρο τους, κάτι το οποίο μπορεί να ενισχύει τους θορύβους από τις αντίστοιχες γωνίες. Συγκεκριμένα, για $d=16\text{cm}$, παρατηρούμε πως δεν ισχύει η σχέση : $d < \frac{\lambda}{2}$. Σε αυτή την περίπτωση έχουμε το ανεπιθύμητο φαινόμενο των grating lobes που περιγράφηκε παραπάνω.

1.4.3)

Θεωρούμε σταθερό το $N=8$, την απόσταση των μικροφώνων $d=4\text{cm}$ και την συχνότητα $f=2\text{kHz}$. Σχεδιάζουμε το μέτρο του delay-and-sum beam pattern σε λογαριθμική κλίμακα (dB) συναρτήσει της γωνίας $\theta \in [-180^\circ, 180^\circ]$ σε πολικό διάγραμμα, χρησιμοποιώντας την συνάρτηση semilogr_polar που δίνεται. Έχουμε για $\theta_s = 0^\circ, 45^\circ, 90^\circ$ αντίστοιχα:



Αρχικά, βλέπουμε πως η συνθήκη $d < \frac{\lambda}{2}$ ικανοποιείται και άρα δεν έχουμε το φαινόμενο των grating lobes. Για $\theta_s = 0^\circ$ και 45° παρατηρούμε πως το μέτρο του $B(\omega, \theta)$ γίνεται 1, όπως αναμένεται στο steering direction, ενισχύοντας τα σήματα που έρχονται από αυτές τις γωνίες (οι κύριοι λοβοί είναι μεγάλοι με αποτέλεσμα να ενισχύονται και γειτονικές γωνίες, μειώνοντας την ποιότητα του συστήματος). Για

$\theta_s = 90^\circ$ οι δύο κύριοι λοβοί έχουν ενωθεί σε έναν, με αποτέλεσμα το μέτρο του $B(\omega, \theta)$ στην γωνία θ_s να μην είναι ακριβώς 1 και να μην ενισχύει την ζητούμενη γωνία πλήρως.

Μέρος 2) Εφαρμογή Beamforming για Speech Enhancement

2.1) Beamforming σε προσομοιωμένα σήματα

Σε αυτό το μέρος της άσκησης θα δουλέψουμε με μία γραμμική συστοιχία $N=7$ μικροφώνων με απόσταση $d=4\text{cm}$ μεταξύ τους και καταγράφει σήματα που παράγονται από μία πηγή φωνής σε γωνία $\theta = 45^\circ$ σε σχέση με τη συστοιχία και μία πηγή θορύβου σε γωνία $\theta = 135^\circ$ σε σχέση με τη συστοιχία. Υποθέτουμε επως τα σήματα πηγής και θορύβου είναι ασυσχέτιστα.

A) Delay-and-sum beamforming

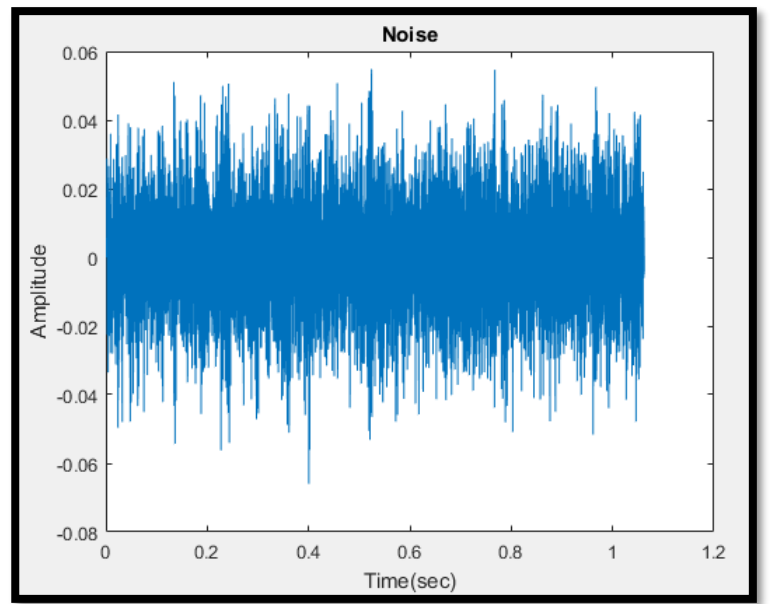
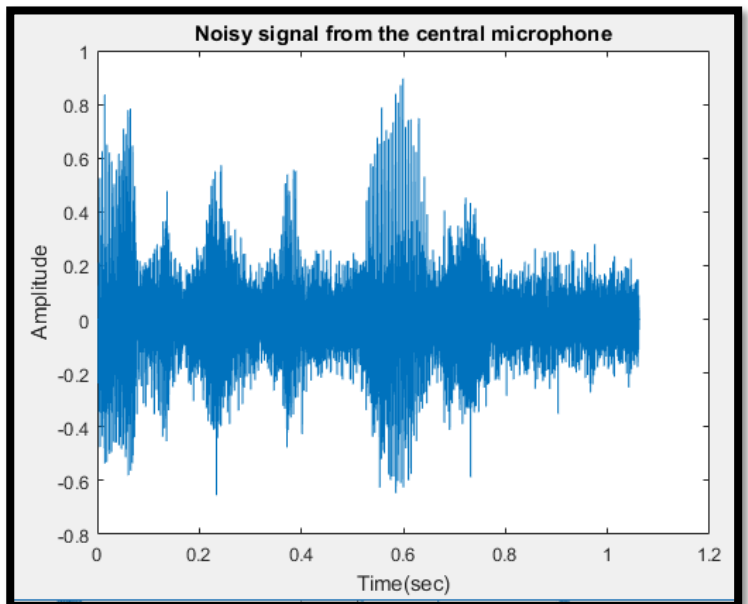
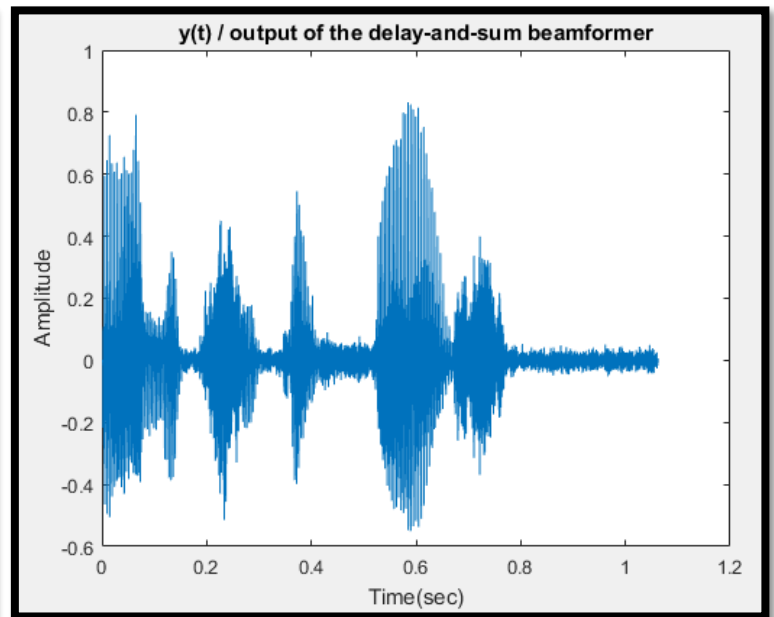
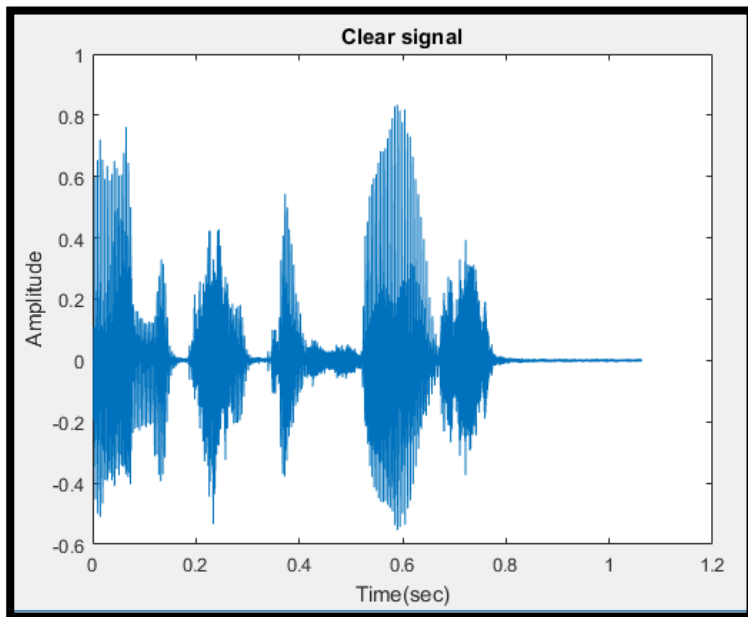
Προσπαθούμε να υλοποιήσουμε σύστημα delay-and-sum beamformer για τα προσομοιωμένα σήματα που μας δίνονται. Προκειμένου, λοιπόν, να αποθρομβοποιήσουμε το δοσμένο σήμα ακολουθούμε την παρακάτω διαδικασία:

- Γνωρίζουμε πως αν τα μικρόφωνα της γραμμικής συστοιχίας λαμβάνουν το καθαρό σήμα φωνή υπό γωνία $\theta_s = 45^\circ$, το σήμα σε κάθε ένα μικρόφωνο θα είναι το θορυβώδες μετατοπισμένο χρονικά κατά $\tau_n = \frac{n-\frac{N-1}{2}}{c} d \cos \theta$. Επομένως, πρώτα ολισθαίνουμε χρονικά τα σήματα αυτά ώστε να μπορέσουν να ευθυγραμμιστούν χρονικά. Αυτή η μετατόπιση χρονικά μπορεί να γίνει κλασματική. Έτσι, υλοποιούμε το διάνυσμα $d(k)$ (φίλτρο κλασματικής καθυστέρησης) το οποίο για κάθε μικρόφωνο ($n=0,1,2,3,4,5,6$) θα πολλαπλασιάζεται με τον Fourier του αντίστοιχου σήματος μικροφώνου και θα το ευθυγραμμίζει με το καθαρό σήμα. Στο πεδίο της συχνότητας το Fractional Delay Filter μπορεί να γραφεί ως $H_D(e^{j\omega}) = e^{-j\omega D}$. Το διάνυσμα αυτό θα έχει ίσο μήκος δειγμάτων με το αρχικό μας σήμα.
- Για κάθε μικρόφωνο βρίσκουμε τον μετασχηματισμό Fourier του αντίστοιχου σήματος και το πολλαπλασιάζουμε με τον αντίστοιχο πίνακα γραμμή του $d(k)$ στοιχείο προς στοιχείο προκειμένου να πάρουμε την απαραίτητη ευθυγράμμιση. Έπειτα, για κάθε ένα μικρόφωνο πάλι παίρνουμε το πραγματικό μέρος του αντίστροφου μετασχηματισμού Fourier. Σε κάθε έξοδο από τα 7 φίλτρα χρονικής μετατόπισης έχουμε ουσιαστικά το σήμα $f(t)$ που αποτελεί μια προσέγγιση του καθαρού σήματος φωνής που μας δίνεται στο αρχείο **source.wav**.
- Τελικά, προσθέτουμε όλα τα σήματα που προέκυψαν στο προηγούμενο βήμα και παίρνουμε τον μέσο όρο τους. Τώρα έχουμε στην έξοδο του delay-and-sum beamformer μία αρκετά καλή προσέγγιση του σήματος και χρονικά ευθυγραμμισμένη με αυτό.

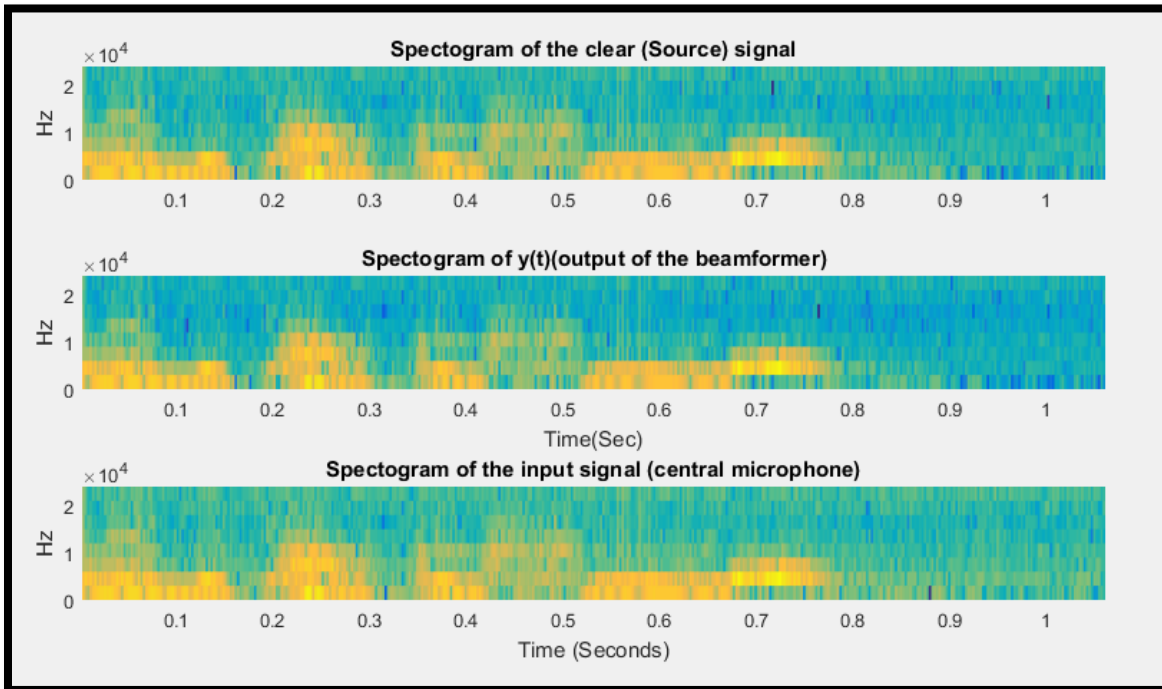
Για επαλήθευση της ορθότητας της παραπάνω διαδικασίας παίρνουμε την διαφορά της εξόδου του συστήματος delay-and-sum beamformer και του καθαρού σήματος φωνής και

παρατηρούμε πως πρόκειται για θόρυβο, αφού ακούγοντάς το φαίνεται πως ακούγεται μόνο θόρυβος χωρίς καθόλου σήμα φωνής.

Παρακάτω συγκρίνουμε τις κυματομορφές και τα σπεκτρογραφήματα για το καθαρό σήμα φωνής, το θορυβώδες σήμα στο κεντρικό μικρόφωνο και την έξοδο $y(t)$ του delay-and-sum beamformer. Παραθέτουμε τον θόρυβο μαζί με τις υπόλοιπες γραφικές παραστάσεις.



Είναι εύκολο να παρατηρήσει κανείς πως ενώ το θορυβώδες σήμα από το κεντρικό μικρόφωνο έχει κακή ποιότητα σχετικά με το καθαρό σήμα της πηγής, το σήμα εξόδου του συστήματος delay-and-sum beamformer το προσεγγίζει πολύ καλά. Ταυτόχρονα, παρατηρούμε πως η διαφορά της εξόδου του συστήματος και του καθαρού σήματος είναι ασυσχετιστος με το σήμα θόρυβος.



Οι παραπάνω παρατηρήσεις μας επαληθεύονται και με τα σπεκτρογραφήματα των παραπάνω σήματων, καθώς βλέπουμε πως η έξοδος του συστήματος προσεγγίζει πολύ καλύτερα το χρονοσυχνοτικό περιεχόμενο του καθαρού σήματος από την είσοδο του κεντρικού μικροφώνου.

Τέλος, ζητείται το SNR του θορυβώδους σήματος στο κεντρικό μικρόφωνο το οποίο βρίσκεται **4.8491** και το SNR της εξόδου του συστήματος το οποίο υπολογίζεται **17.5836**. Παρατηρείται, λοιπόν, αυτό που ήταν αναμενόμενο και από την θεωρία. Η σηματοθορυβική σχέση μας δείχνει την σχέση των ενεργειών του καθαρού σήματος φωνής και του θορύβου που περιέχεται σε κάθε ένα από τα δύο σήματα που συγκρίνουμε. Η έξοδος του beamformer έχει πολύ λιγότερο θόρυβο από το κεντρικό μικρόφωνο της συστοιχίας και αυτό οφείλεται στο γεγονός ότι το προσομοιωμένο σήμα στέλνεται υπό συγκεκριμένη γωνία και ο beamformer μπορεί και ενισχύει την κύρια γωνία που φτάνει το σήμα μειώνοντας τις επιδράσεις του θορύβου που έρχεται και από διαφορετικές γωνίες. Εν αντιθέση το κεντρικό μικρόφωνο λαμβάνει όλο τον θόρυβο με αποτέλεσμα να έχει χαμηλό SNR.

B) Μονοκαναλικό Wiener φιλτράρισμα

Στόχος μας είναι να συγκρίνουμε την πολυκαναλική μέθοδο αποθορυβοποίησης με τη μονοκαναλική μέθοδο Wiener filtering. Θεωρούμε το πλαίσιο (frame) $f(t, p_3)$, $t \in [0.36s, 0.39s]$ διάρκειας 30ms το οποίο περιέχει έναν έμφωνο ήχο καταγεγραμμένο από το κεντρικό μικρόφωνο τη συστοιχίας. Θα υλοποιήσουμε μονοκαναλική αποθορυβοποίηση του πλαισίου αυτού με Wiener φιλτράρισμα.

Για την δημιουργία του φίλτρου πρώτα υπολογίζουμε το φάσμα ισχύος του σήματος από το κεντρικό μικρόφωνο και το φάσμα ισχύος του καθαρού σήματος με τη μέθοδο Welch (χρήση συνάρτησης `pwelch`) και έπειτα σύμφωνα με τον ακόλουθο τύπο παίρνουμε την απόκριση συχνότητας του IIR Wiener φίλτρου:

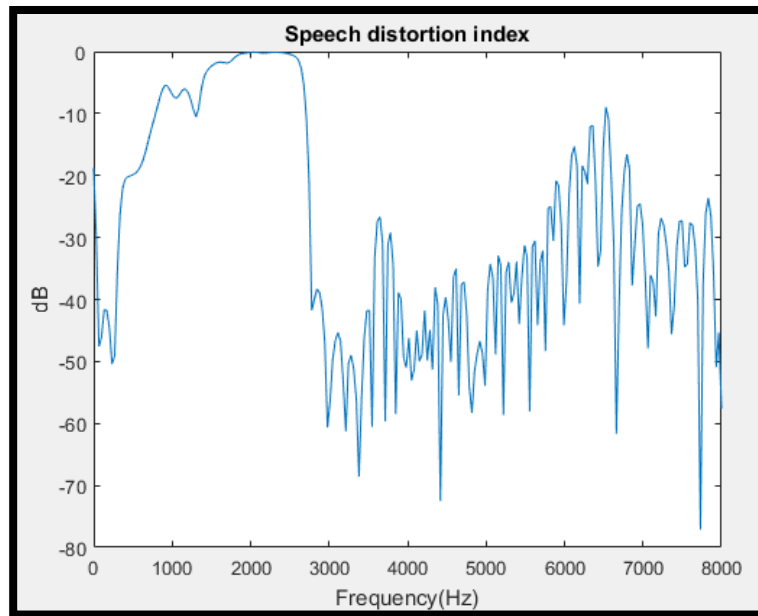
$$H_w(\omega) = \frac{P_s(\omega)}{P_x(\omega)}$$



Έπειτα, υπολογίζουμε την παραμόρφωση (distortion) που προκαλεί το Wiener φίλτρο στο σήμα φωνής χρησιμοποιώντας το speech distortion index το οποίο ορίζεται ως εξής:

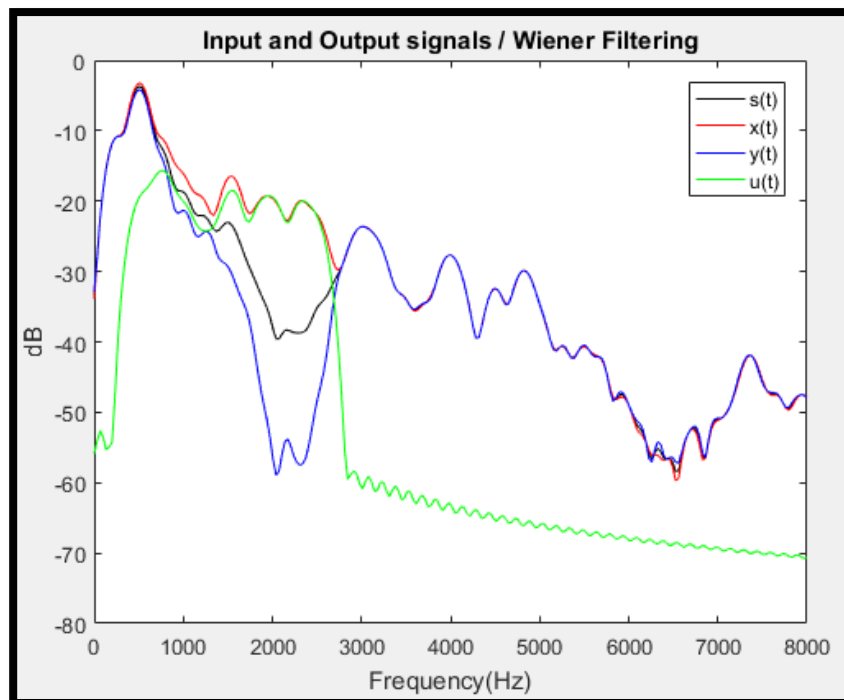
$$n_{sd}(\omega) = |1 - H_w(\omega)|^2$$

Όμοια με την απόκριση συχνότητας του φίλτρου σχεδιάζουμε το $n_{sd}(\omega)$ σε λογαριθμική κλίμακα (dB) για συχνότητες $f \in [0, 8kHz]$.



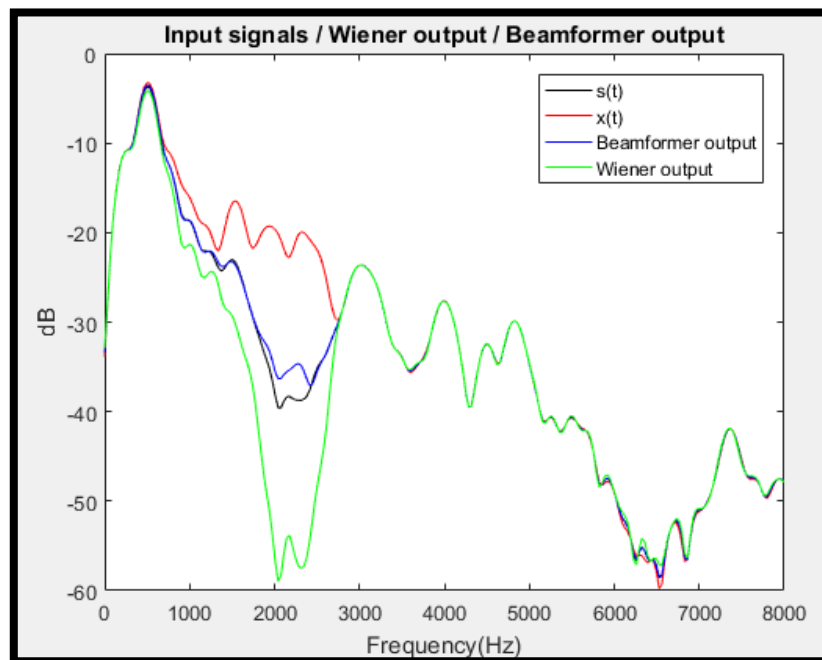
Παρατηρούμε πως το speech distortion index είναι πιο υψηλό στις συχνότητες $f \in [500Hz, 2.5kHz]$. Αυτό δεν είναι τυχαίο καθώς γνωρίζουμε η πηγή θορύβου παράγει θόρυβο ζωνοπερατό με ενέργεια συγκεντρωμένη σε αυτό ακριβώς το εύρος συχνοτήτων.

Παρακάτω φαίνονται στην ίδια γραφική παράσταση τα φάσματα ισχύος για το καθαρό σήμα φωνής $s(t)$, τη θορυβώδη είσοδο του φίλτρου $x(t)$, την έξοδο του φίλτρου Wiener $y(t)$ και το σήμα θορύβου στην είσοδο του φίλτρου $u(t)$ σε λογαριθμική κλίμακα για συχνότητες 0-8kHz:



Από το παραπάνω διαγράμμα γίνεται εμφανές πάλι το γεγονός πως ο θόρυβος είναι συγκεντρωμένος στις συχνότητες 500-2500 Hz, αφού εκεί η PSD του θορύβου είναι πιο θψηλή και σε αυτό το εύρος συχνοτήτων είναι που και το θορυβώδες σήμα στο κεντρικό μικρόφωνο και η έξοδος του φίλτρου αποκλίνουν περισσότερο από το καθαρό σήμα φωνής. Σε όλες τις άλλες συχνότητες και τα δύο σήματα προσεγγίζουν το καθαρό σήμα, με την καλύτερη προσέγγιση να την επιτυγχάνει φυσικά η έξοδος του φίλτρου.

Υπολογίζουμε το SNR εξόδου του φίλτρου **10.9104** και το SNR εισόδου του φίλτρου **7.9415**. Για να συγκρίνουμε την πολυκαναλική μέθοδο με την μονοκαναλική πραγματοποιούμε πάλι την μέθοδο delay-and-sum beamforming ακολουθώντας την διαδικασία που περιγράφεται στο προηγούμενο ερώτημα. Για το συγκεκριμένο πλαίσιο υπο επεξεργασία το SNR με την πολυκαναλική μέθοδο υπολογίζεται **21.1893** και διαφορά του με το αντίστοιχο της μονοκαναλικής μεθόδου είναι **10.2790**. Η βελτίωση είναι εμφανής και στο παρακάτω διάγραμμα όπου φαίνονται στην ίδια γραφική παράσταση τα φάσματα για το σήμα καθαρής φωνής για το πλαίσιο υπό μελέτη, το θορυβώδες πλαίσιο, την έξοδο του φίλτρου και την έξοδο του beamformer για το δεδομένο frame.



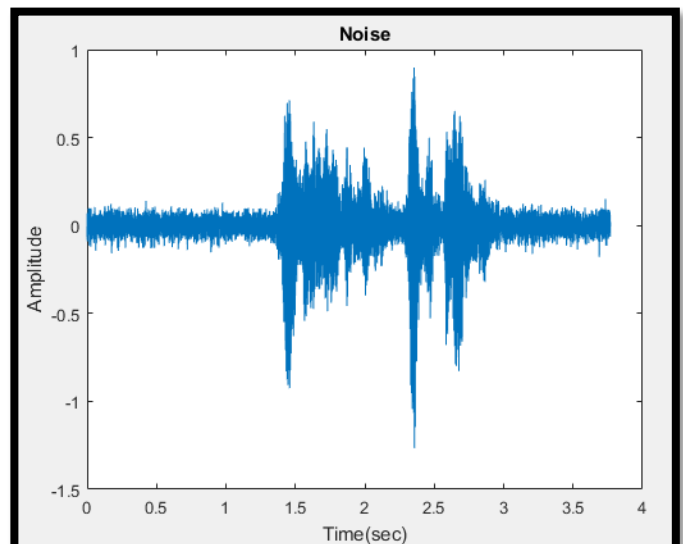
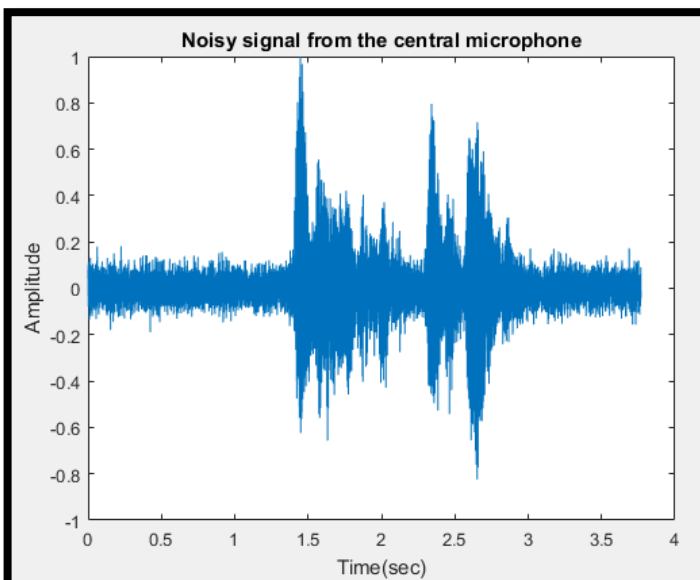
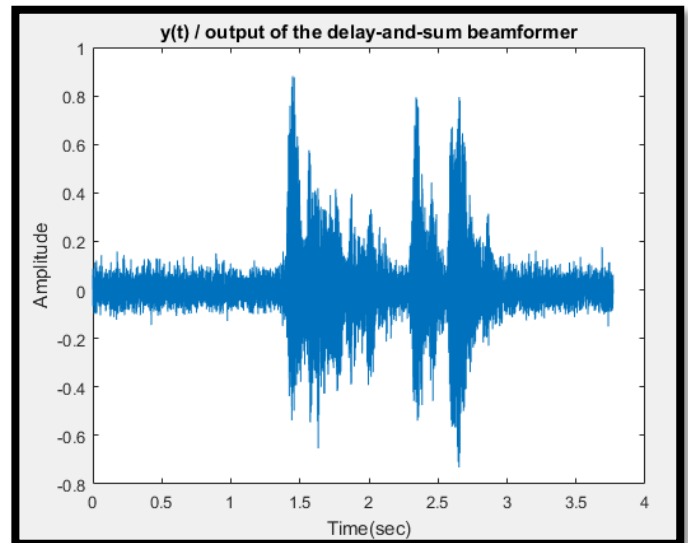
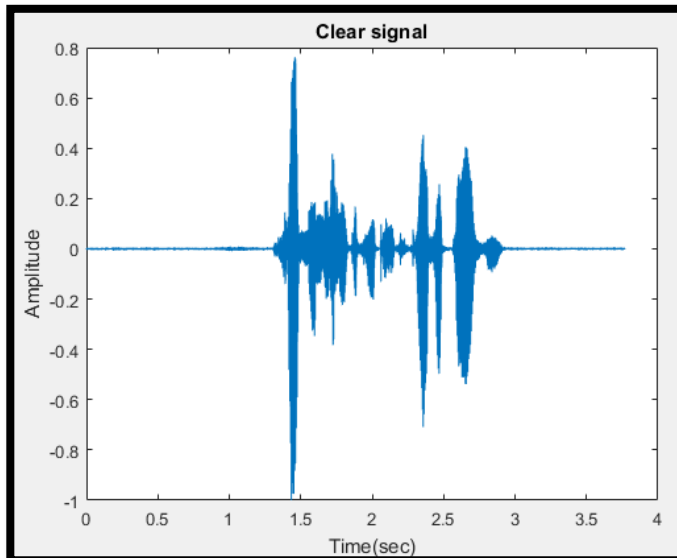
Είναι εμφανές πως η έξοδος του delay-and-sum beamformer προσεγγίζει πολύ καλύτερα το καθαρό σήμα φωνής από το φίλτρο Wiener κάτι το οποίο φαίνεται τόσο στο διάγραμμα με τα φάσματα ισχύος όσο και με τα SNR. Η πολυπλοκότητα του beamforming επιτρέπει την προσέγγιση του σήματος χρησιμοποιώντας όλα τα μικρόφωνα της συστοιχίας (πολυκαναλική μέθοδος) κάνοντάς τη πιο ακριβή σε αντίθεση με την απλή μονοκαναλική μέθοδο του φίλτρου όπου γίνεται χρήση του σήματος ενός μικροφώνου. Επιπλέον, το beamforming ενισχύει την γωνία από την οποία έρχεται το σήμα φωνής μειώνοντας την επίδραση του θορύβου.

2.2) Beamforming σε πραγματικά σήματα

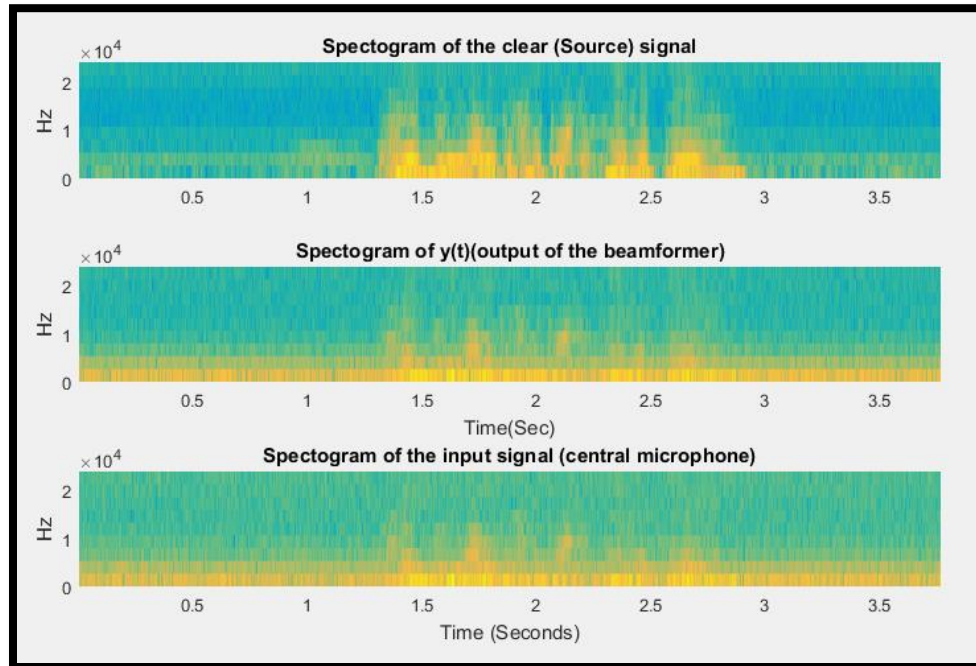
Σε αυτό το μέρος υλοποιούμε beamforming σε πραγματικά σήματα με τη χρήση συστοιχίας $N=7$ μικροφώνων με απόσταση $d=4\text{cm}$ μεταξύ τους σε θορυβώδες δωμάτιο με $\theta_s = 45^\circ$. Ο θόρυβος σε αντίθεση με το προηγούμενο μέρος δεν προέρχεται από σημειακή πηγή αλλά από διάφορες πηγές οι οποίες δημιουργούν ένα ιστροπικό και ομογενές πεδίο θορύβου (diffuse noise field). Ο θόρυβος θεωρείται στάσιμος (stationary random process).

A) Delay-and-sum beamforming

Ακολουθούμε την διαδικασία που περιγράφηκε στο 2.1.A) με σκοπό να βρούμε τα βάρη για τον delay-and-sum beamformer και να εφαρμόσουμε beamforming. Ακολούθως παρατίθενται οι κυματομορφές και τα σπεκτρογραφήματα για το καθαρό σήμα φωνής, το θορυβώδες σήμα στο κεντρικό μικρόφωνο της συστοιχίας και την έξοδο του beamformer:



Είναι εμφανές πως το σήμα από το κεντρικό μικρόφωνο έχει αρκετά κακή ποιότητα, όμοια με την έξοδο του beamformer. Αυτό αιτιολογείται από το γεγονός ότι τώρα δουλεύουμε με πραγματικά σήματα και με θόρυβο (διαφορά καθαρού σήματος με την έξοδο) ο οποίος όπως φαίνεται και από το παραπάνω διάγραμμα δεν είναι ασυσχέτιστος με το σήμα μας (ακούγοντάς τον μάλιστα παρατηρούμε πως ακούγεται σαν το σήμα φωνής!).



Οι παραπάνω παρατηρήσεις μας επαληθεύονται και από τα σπεκτρογραφήματα όπου βλέπουμε πως τόσο το σήμα από το κεντρικό μικρόφωνο όσο και η έξοδος του beamformer δεν προσεγγίζουν καλά το καθαρό σήμα φωνής.

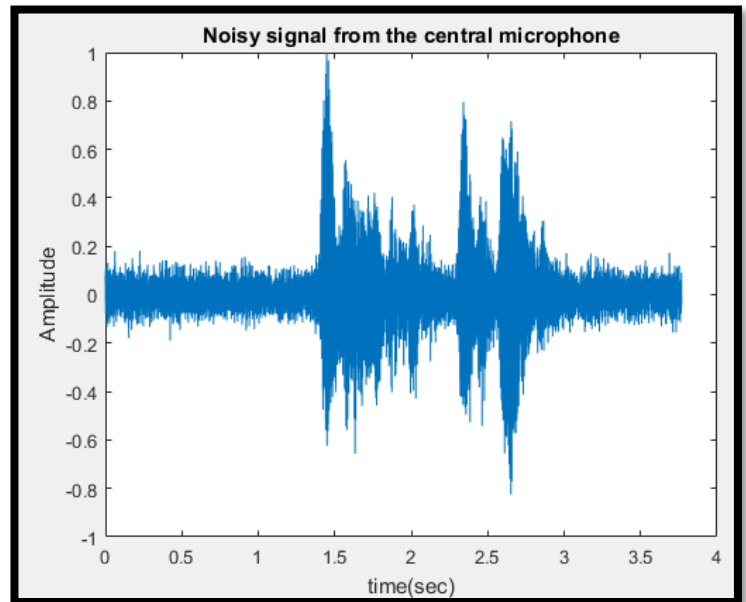
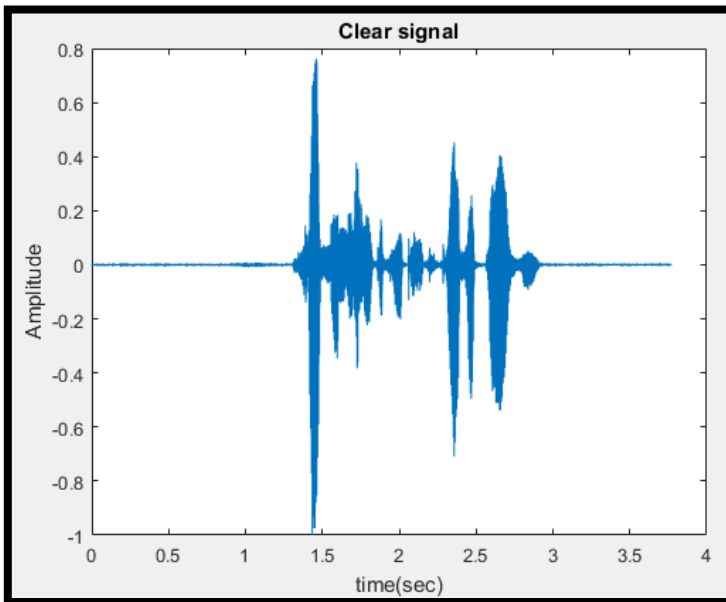
Έπειτα, ζητείται να βρεθεί το segmental SNR (SSNR) το οποίο ορίζεται ως το μέσο SNR των πλαίσιων βραχέος χρόνου του σήματος φωνής. Πλαίσια με SNR μεγαλύτερο των 35dB τίθενται στα 35dB, ενώ πλαίσια με τιμή μικρότερη των 0dB αγνοούνται κατά τον υπολογισμό του μέσου όρου. Σημειώνεται πως την ισχύ του θορύβου την υπολογίζουμε από το δεύτερο πλαίσιο του σήματος που περιέχει τον θόρυβο καθώς στην αρχή των σημάτων δεν υπάρχει σήμα φωνής. Υπολογίζουμε το SSNR στο κεντρικό μικρόφωνο της συστοιχίας **7.8981** και το SSNR στην έξοδο του beamformer **9.1327**. Η βελτίωση όπως φάνηκε τόσο από τις κυματομορφές και τα σπεκτρογραφήματα όσο και τον υπολογισμό των SSNR δεν είναι ικανοποιητική. Ακούγοντας το παραγόμενο σήμα στην έξοδο του beamformer καταλαβαίνουμε πως η διαφορά δεν είναι εύκολα αισθητή, ενώ παράλληλα η ακουστική εμπειρία παραμένει αρνητική.

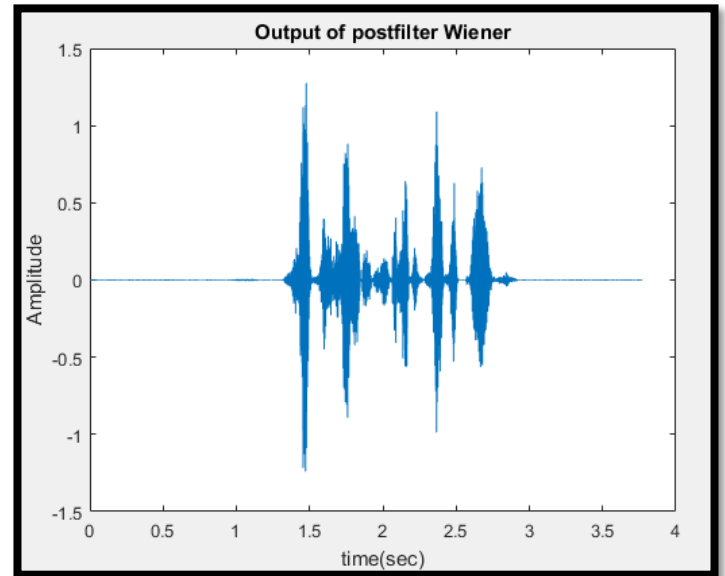
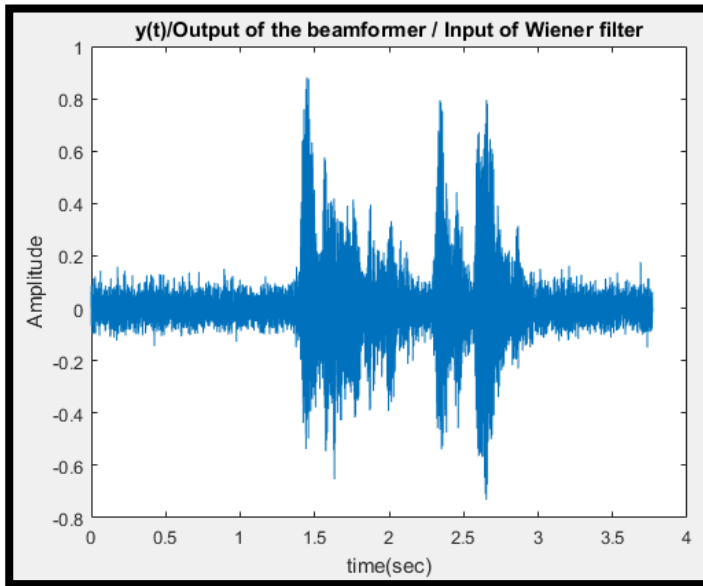
B) Post-filtering με Wiener φίλτρο

Προηγουμένως παρατηρήσαμε πως στην περίπτωση του diffuse noise field, ο delay-and-sum beamformer δεν έχει καλή απόδοση, διότι τα σήματα θορύβου εμφανίζουν μεγάλη συσχέτιση μεταξύ των μικροφώνων, ειδικά στις χαμηλές συχνότητες. Θα εφαρμόσουμε μονοκαναλικό φιλτράρισμα μετά το beamforming για περαιτέρω βελτίωση της ποιότητας του σήματος, την λεγόμενη post-filtering μέθοδο. Ακολουθήσαμε την ακόλουθη διαδικασία:

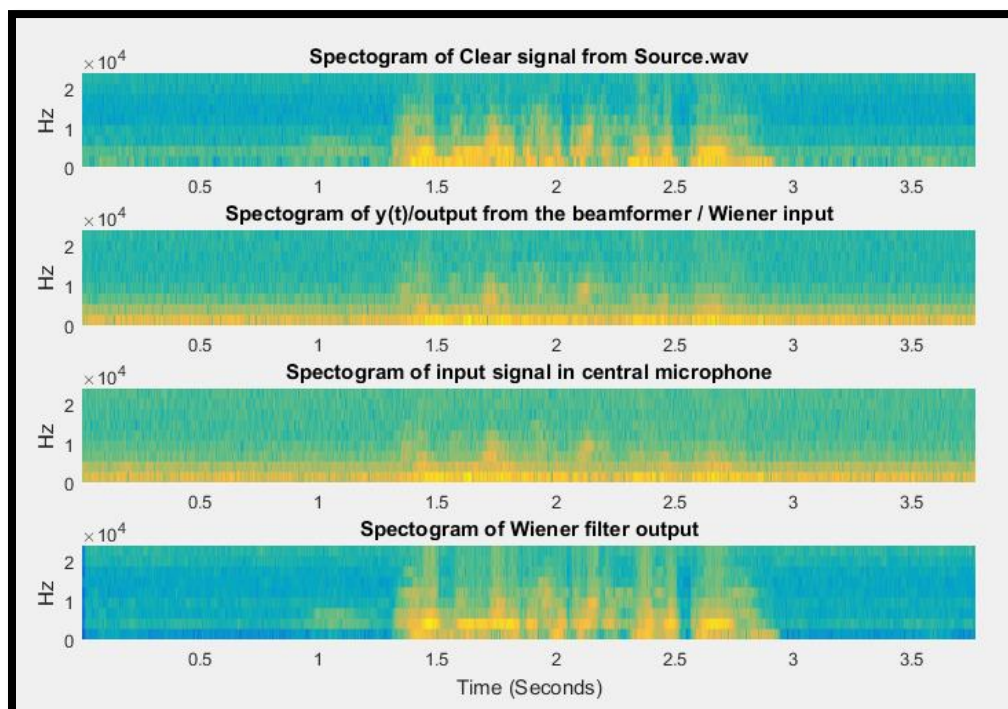
- Αρχικά, υλοποιήθηκε ο beamformer όπως και στο ερώτημα 2.1.A) και έπειτα στην έξοδο του beamformer διασπάσαμε το σήμα με παράθυρα Hamming (ανάλυση βραχέος χρόνου (short time analysis)) μήκους 30ms δηλαδή 1440 δειγμάτων. Επιλέξαμε overlap 50% του μήκους παραθύρου διότι πειραματικά φάνηκε πως αυτή η επικάλυψη δίνει καλύτερα αποτελέσματα.
- Σε κάθε ένα από αυτά τα παράθυρα εφαρμόστηκε μονοκαναλικό Wiener φιλτράρισμα ακριβώς όπως περιγράφηκε στο μέρος 2.1.B).
- Τέλος, υλοποιήθηκε η μέθοδος overlap-add ώστε να μπορέσουμε να ανακατασκευάσουμε ολόκληρο το σήμα στην έξοδο του συστήματος.

Πριν εξάγουμε τα συμπεράσματά μας παραθέτουμε τις απαραίτητες γραφικές παραστάσεις και τα σπεκτρογραφήματα για όλα τα επίπεδα του συνδιαστικού συστήματος που υλοποιήθηκε:





Όπως αναμενόταν το σήμα από την έξοδο delay-and-sum beamformer δεν έχει ικανοποιητική ποιότητα εξαιτίας του diffuse noise field. Συγκεκριμένα, σε αυτή την περίπτωση δεν έχουμε συγκεκριμένη γωνία από την οποία κατευθύνεται ο ήχος στα μικρόφωνα και αφού το beamforming ενισχύει μία συγκεκριμένη γωνία, υπάρχει θόρυβος και σε αυτή την γωνία που ενισχύει. Το ίδιο δεν συνέβαινε στο 2.1 μέρος της άσκησης με τα σήματα προσομοίωσης. Για αυτό τον λόγο τα σήματα στην είσοδο του Wiener και του κεντρικού μικροφώνου είναι τόσο θορυβώδη. Για κάθε πλαίσιο των 30ms κάνουμε μονοκαναλικό φιλτράρισμα Wiener και με αυτόν τον τρόπο καταφέρνουμε να μειώσουμε κατά έναν μεγάλο βαθμό τον θόρυβο στην έξοδο, κάτι το οποίο γίνεται εμφανές και από το παραπάνω διάγραμμα.



S

Τα παραπάνω συμπεράσματα ενισχύονται και από τα ανωτέρω σπεκτρογραφήματα, όπου φαίνεται πολύ καθαρά πως η έξοδος του συστήματος προσεγγίζει αρκετά καλά το καθαρό σήμα φωνής, σε αντίθεση με την έξοδο του beamformer και ειδικά το θορυβώδες σήμα του κεντρικού μικροφώνου. Άλλη μία απόδειξη πως το post-filtering κάνει σωστά την δουλειά του αποτελεί το wav. αρχείο στο οποίο εγγράφουμε την έξοδο του συστήματος με όνομα **real_mmse.wav** . Ακούγοντάς το παρατηρούμε αισθητή διαφορά τόσο στο χρονικό διάστημα που του σήματος φωνής όσο και στο αντίστοιχο της ηρεμίας.

Υπολογίζουμε τα SSNR στην έξοδο και στην είσοδο του Wiener φίλτρου για να διαπιστώσουμε πόσο έχουμε βελτιώσει το ακουστικό κομμάτι. Σημειώνεται πως για όλους τους υπολογισμούς του SSNR παίρνουμε ένα αρχικό frame του σήματος (2^ο) που μελετάμε, το οποίο προφανώς περιέχει μόνο θόρυβο, και το συγκρίνουμε με το καθαρό σήμα φωνής που δίνεται. Το threshold που χρησιμοποιήθηκε είναι στα -5dB και έτσι όταν ένα πλαίσιο έχει SNR μικρότερο από αυτό αγνοείται στον υπολογισμό του SSNR.

- SSNR Wiener Input = **6.5119**
- SSNR Wiener Output = **25.0946**

Η βελτίωση είναι κάτι παραπάνω από ικανοποιητική. Εφόσον υπάρχει τέτοια βελτίωση στην σηματοθορυβική σχέση πράγματι το σύστημα που μοντελοποιήσαμε βελτιώνει σημαντικά την ποιότητα του ήχου στο diffuse noise field μετά την έξοδο του beamformer.

Τέλος, υπολογίζουμε τον μέσο όρο των SSNRs των σημάτων εισόδου στο σύστημα delay-and-sum beamformer + Wiener post-filter **4.8792**. Παρατηρούμε πως έχουμε μία πολύ υψηλή βελτίωση κάτι το οποίο επαληθεύεται τόσο από τα διαγράμματα όσο και από την ακουστική εμπειρία του τελικού ακουστικού σήματος. Τονίζεται πως τα νούμερα για τα SSNR μπορεί να διαφέρουν αφού καθοριστικοί παράγοντες για τις τιμές τους είναι τα κατώφλια και των noisy frames.