

ΤΕΧΝΟΛΟΓΙΚΟ  
ΕΚΠΑΙΔΕΥΤΙΚΟ  
Ι Δ Ρ Υ Μ Α



ΠΕΛΟΠΟΝΝΗΣΟΥ

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ  
ΙΔΡΥΜΑ ΠΕΛΟΠΟΝΝΗΣΟΥ  
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ Τ.Ε.

## ΑΠΟΘΡΥΒΟΠΟΙΗΣΗ ΣΗΜΑΤΩΝ ΟΜΙΛΙΑΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

των

ΚΑΤΣΙΩΝΗ ΝΙΚΟΛΑΟΥ

ΕΙΛΑΡΧΟΥ ΝΙΚΟΛΑΟΥ

Επιβλέπων Καθηγητής: Νόκας Γεώργιος

Σπάρτη, Δεκέμβριος 2018

**Copyright © Σπάρτη –Δεκέμβριος, 2018**

*Με επιφύλαξη παντός δικαιώματος. All rights reserved.*

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Τ.Ε.Ι. Πελοποννήσου.

## ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ

### ΕΥΘΥΝΗΣ

"Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάση επιστημονικής παράφρασης.

Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δε μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας."

Όνομα και Επώνυμο Συγγραφέων: ΚΑΤΣΙΩΝΗΣ ΝΙΚΟΛΑΟΣ, ΞΙΑΡΧΟΣ ΝΙΚΟΛΑΟΣ

Υπογραφές :

Ημερομηνία : 20 Δεκεμβρίου 2018

## Ευχαριστίες

Έχοντας φτάσει στο τέλος της πτυχιακής μας εργασίας, αισθανόμαστε υποχρεωμένοι να μιλήσουμε για κάποιους ανθρώπους, που ο καθένας με τον δικό του τρόπο σηματοδότησε την πορεία των χρόνων μας στις προπτυχιακές μας σπουδές. Πρώτα απ' όλα, θα θέλαμε να ευχαριστήσουμε τον επιβλέποντά μας, κύριο Νόκα Γεώργιο, Εργαστηριακό Συνεργάτη του Τμήματος Μηχανικών Πληροφορικής ΤΕ του Τ.Ε.Ι. Πελοποννήσου, για την καθοδήγησή του και την πολύτιμη βοήθειά του κατά τη διάρκεια εκπόνησης της εργασίας διότι η συνεργασία μαζί του ήταν ένας καταλύτης για την ολοκλήρωση των προπτυχιακών μας σπουδών. Τέλος, θα θέλαμε να ευχαριστήσουμε τους γονείς μας για την αμέριστη υποστήριξή τους όλα αυτά τα χρόνια, των προπτυχιακών σπουδών μας. Τους αφιερώνουμε αυτή την εργασία, ως ελάχιστη ευγνωμοσύνη για την κατανόηση και την υπομονή τους όλα αυτά τα χρόνια.

Κατσιώνης Νικόλαος

Ξιάρχος Νικόλαος

Σπάρτη, 20/12/2018

## Περίληψη

Το θέμα της παρούσας εργασίας είναι η αποθορυβοποίηση των σημάτων ομιλίας. Στο πρώτο κεφάλαιο γίνεται αναφορά στην φύση του ήχου καθώς και στις μεθόδους καταγραφής και ψηφιοποίησης του. Στο δεύτερο κεφάλαιο αναφέρονται οι κύριες ιδιότητες και χαρακτηριστικά του θορύβου, γίνεται μία εισαγωγή στην ανάλυση Fourier, βασικό εργαλείο των αλγορίθμων αφαίρεσης θορύβου, και αναφέρονται, γενικά, οι διαθέσιμες λύσεις για την αντιμετώπιση του προβλήματος του θορύβου. Στο τρίτο κεφάλαιο περιγράφονται οι μέθοδοι που χρησιμοποιούν οι αλγόριθμοι αποθορυβοποίησης και για την καλύτερη εποπτεία αυτών καθώς και του πως εφαρμόζονται περισσότερες της μία μεθόδους στους αλγόριθμους αποθορυβοποίησης, παρουσιάζεται αρκετά αναλυτικά ένας από αυτούς μαζί με τα αποτελέσματα που αυτός δίνει. Τέλος, στο τέταρτο κεφάλαιο παρουσιάζεται ένα πρόγραμμα (script) που υλοποιήθηκε με χρήση του MatLab όπου εφαρμόζονται τα βασικά που πρέπει να εφαρμοστούν από αυτά που αναφέρθηκαν στην εργασία αυτή, για την αποθορυβοποίηση ενός σήματος ομιλίας.

## Λέξεις-κλειδιά

Ήχος, θόρυβος, σήμα, αποθορυβοποίηση σημάτων, μετασχηματισμός Fourier, φασματική αφαίρεση, αλγόριθμος Favrot-Faller, χαμηλοπερατό φίλτρο, πρόσθεση θορύβου, ενθόρυβο σήμα, αφαίρεση θορύβου, εξομάλυνση

## ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη .....	Σελ. 5
Περιεχόμενα .....	Σελ. 6
Κατάλογος Εικόνων .....	Σελ. 8
Κατάλογος Πινάκων .....	Σελ. 9
<b>ΚΕΦΑΛΑΙΟ 1 - <i>Ηχητικά κύματα &amp; Ήχος</i></b>	Σελ. 10
1.1- Ηχητικά κύματα .....	Σελ. 10
1.2- Καταγραφή Ήχου .....	Σελ. 15
1.3- Ψηφιακός Ήχος .....	Σελ. 19
<b>ΚΕΦΑΛΑΙΟ 2 – <i>Θόρυβος και αντιμετώπισή του</i></b>	Σελ. 23
2.1- Θόρυβος .....	Σελ. 23
2.2- Ανάλυση Fourier .....	Σελ. 28
2.3- Αντιμετώπιση του προβλήματος του Θορύβου .....	Σελ. 38
<b>ΚΕΦΑΛΑΙΟ 3 - <i>Αποθορυβοποίηση</i></b>	Σελ. 41
3.1- Γραμμικό Φιλτράρισμα Ψηφιακών Σημάτων .....	Σελ. 42
3.2- Αποθορυβοποίηση στο πεδίο του Χρόνου .....	Σελ. 48
3.3- Αποθορυβοποίηση στο πεδίο των Συχνοτήτων .....	Σελ. 50
3.4- Καταστολή θορύβου σε σήματα ομιλίας με χρήση φίλτρου εξομάλυνσης, παραμετροποιημένο με αντιληπτικά κριτήρια (Favrot-Faller) .....	Σελ. 52
<b>ΚΕΦΑΛΑΙΟ 4 – <i>Αλγόριθμοι Αποθορυβοποίησης</i></b>	Σελ. 65

4.1-Πρόσθεση θορύβου .....	Σελ. 65
4.1.1-Ανάλυση του αλγορίθμου του προγράμματος προσθήκης θορύβου .....	Σελ. 66
4.2-Αφαίρεση θορύβου με χρήση χαμηλοπερατού φίλτρου .....	Σελ. 68
4.2.1-Εφαρμογή του αλγορίθμου του προγράμματος αφαίρεσης θορύβου με χρήση χαμηλοπερατού φίλτρου .....	Σελ. 71
4.3-Αφαίρεση θορύβου με χρήση φασματικής αφαίρεσης .....	Σελ. 75
4.3.1-Εφαρμογή του αλγορίθμου του προγράμματος αφαίρεσης θορύβου με χρήση φασματικής αφαίρεσης .....	Σελ. 79
ΚΕΦΑΛΑΙΟ 5 .....	Σελ. 86
5.1-Συμπεράσματα .....	Σελ. 86
ΠΑΡΑΡΤΗΜΑ Α: Κώδικας Προγράμματος Προσθήκης Θορύβου .....	Σελ. 87
ΠΑΡΑΡΤΗΜΑ Β: Κώδικας Προγράμματος Αφαίρεσης Θορύβου με χρήση χαμηλοπερατού φίλτρου .....	Σελ.88
ΠΑΡΑΡΤΗΜΑ Γ: Κώδικας Προγράμματος Αφαίρεσης Θορύβου με χρήση φασματικής αφαίρεσης .....	Σελ.90
ΠΗΓΕΣ .....	Σελ. 95

## ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1.1: Ηχητικά κύματα	Σελ. 10
Εικόνα 1.2: Σύστημα Παλμοκωδικής Διαμόρφωσης (PCM)	Σελ. 19
Εικόνα 3.1: Μπλοκ διάγραμμα αρχής λειτουργίας προσαρμοστικού φίλτρου	Σελ. 46
Εικόνα 3.2: Μπλοκ διάγραμμα αρχής λειτουργίας ALE συστήματος	Σελ. 47
Εικόνα 3.3: Διάγραμμα βαθμίδων που υλοποιούν τον μετασχηματισμό φάσματος	Σελ. 55
Εικόνα 3.4: Ζώνες διαιρέσεων του φάσματος	Σελ. 58
Εικόνα 3.5: a) φίλτρα παρεμβολής Hann b) τιμές του κέρδους φίλτρου πριν (έντονες κουκκίδες) και μετά (συνεχής γραμμή) την παρεμβολή.	Σελ. 58
Εικόνα 3.6: Το κατώφλι ακοής (Συνεχής γραμμή), το αντίστοιχο κατώφλι μάσκας (διακεκομένη) για έναν τόνο με $f=500\text{Hz}$ και ισχύ $60\text{dB}$	Σελ. 60
Εικόνα 3.7: a) Το αρχικό, μη εξομαλυσμένο κέρδος φίλτρου στο διαιρεμένο πεδίο συχνοτήτων, b) Σταθερά εξομάλυνσης $\mu$ , c) Το εξομαλυσμένο κέρδος φίλτρου	Σελ. 64
Εικόνα 4.1: Κυματομορφές των δεδομένων του καθαρού ήχου, του θορύβου και του ενθόρυβου σήματος	Σελ. 72
Εικόνα 4.2: Φάσμα του καθαρού ήχου, του θορύβου και του ενθόρυβου σήματος	Σελ. 72
Εικόνα 4.3: Κυματομορφές του ενθόρυβου και του καθαρισμένου σήματος	Σελ. 73



Εικόνα 4.4: Φάσμα του ενθόρυβου και του καθαρισμένου σήματος	Σελ. 74
Εικόνα 4.5: Διάγραμμα ροής του αλγόριθμου	Σελ. 78
Εικόνα 4.6: Αρχικός θορυβώδης ήχος	Σελ. 79
Εικόνα 4.7: “Καθαρός” ήχος (με χρήση των εξισώσεων των σχέσεων 4.12 και 4.13)	Σελ. 80
Εικόνα 4.8: “Καθαρός” ήχος (με χρήση της εξίσωσης της σχέσης 4.12 αλλά όχι της 4.13)	Σελ. 80
Εικόνα 4.9: “Καθαρός” ήχος (με χρήση της σχέσης 4.11, με εξομάλυνση 10 σημείων)	Σελ. 82
Εικόνα 4.10: “Καθαρός” ήχος (με χρήση της σχέσης 4.11 χωρίς εξομάλυνση)	Σελ. 83
Εικόνα 4.11: “Καθαρός” ήχος (με χρήση της σχέσης 4.11 με εξομάλυνση 5 σημείων)	Σελ. 83
Εικόνα 4.12: “Καθαρός” ήχος (με χρήση της σχέσης 4.11 με εξομάλυνση 20 σημείων)	Σελ. 84
Εικόνα 4.13: “Καθαρός” ήχος (με χρήση της σχέσης 4.11 με εξομάλυνση 40 σημείων)	Σελ. 84

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

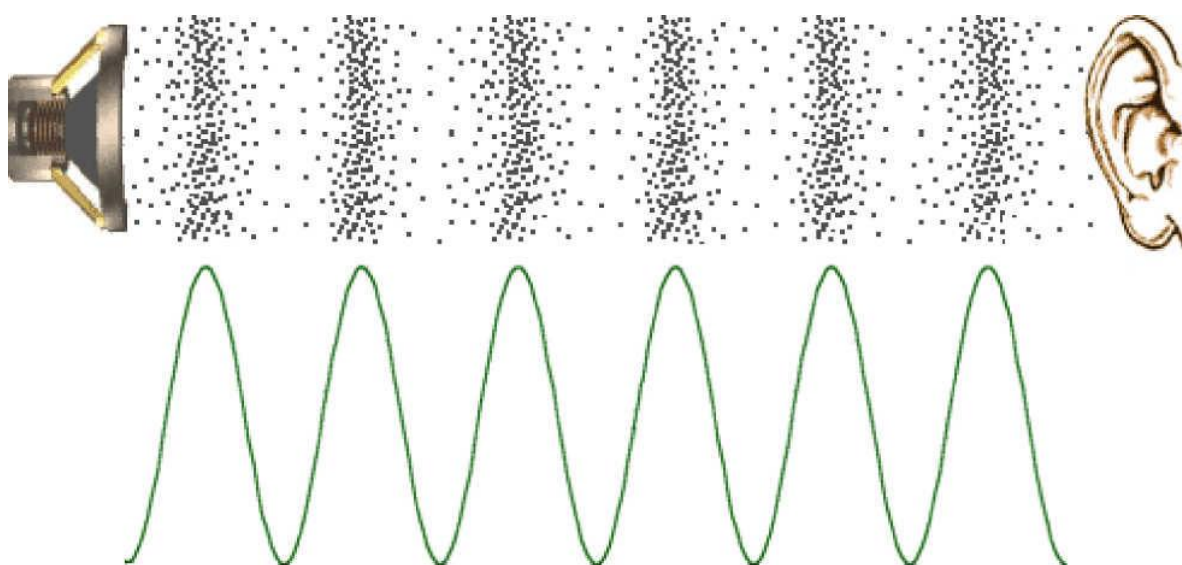
Πίνακας 1: Χρονική σταθερά $T_m$ για κάθε είδος σήματος που ορίζεται στο αντίστοιχο διάστημα του μέτρου διακύμανσης $\Phi[k, b]$ .	Σελ. 63
--	---------

# ΚΕΦΑΛΑΙΟ 1

## *ΗΧΗΤΙΚΑ ΚΥΜΑΤΑ & ΉΧΟΣ*

### **1.1- Ηχητικά κύματα**

Η ικανότητα αντίληψης του “ήχου” είναι ο σημαντικότερος τρόπος με τον οποίο επικοινωνούν οι ζωντανοί οργανισμοί με το περιβάλλον τους αλλά και το αντιλαμβάνονται. Η παραγωγή και αντίληψη του ήχου περιλαμβάνονται στα σημαντικότερα μέσα με τα οποία τα ζώα επικοινωνούν με τα υπόλοιπα, εντοπίζουν την τροφή τους και αντιλαμβάνονται τον κίνδυνο. Τι είναι όμως ο “ήχος”;



Εικόνα 1.1: Ηχητικά κύματα

Ο “ήχος” είναι ένα κύμα, μία ταλάντωση. Είναι η μετάδοση μίας διέγερσης που προκαλείται από την πηγή του κύματος, και την οποία την ονομάζουμε “ηχητικό κύμα”. Για να παραχθεί ένα τέτοιο κύμα από μία πηγή, θα πρέπει η πηγή αυτή να προκαλέσει κάποια αυξομείωση στην πίεση του μέσου στο οποίο βρίσκεται, η οποία θα μεταδοθεί μέσω του μέσου αυτού με την μορφή του ηχητικού κύματος, όπως φαίνεται στην παραπάνω εικόνα (Εικόνα 1.1), στην οποία το πάνω σχήμα αναπαριστά το ηχητικό κύμα και το κάτω την

μεταβολή της πίεσης του μέσου μέσω του οποίου διαδίδεται το κύμα. Τα κύματα αυτά χαρακτηρίζονται ως “μηχανικά κύματα”, γιατί το μέγεθος που μεταβάλλεται είναι η πίεση και γιατί με τα κύματα αυτά η ενέργεια που μεταφέρεται είναι μηχανική, με ένα μέρος να είναι κινητική και ένα άλλο δυναμική. Βασικό χαρακτηριστικό των μηχανικών κυμάτων είναι το ότι είναι απαραίτητο να υπάρχει κάποιο υλικό μέσον μέσω του οποίου θα διαδίδεται το ηχητικό κύμα, και μάλιστα, όσο πυκνότερο είναι το μέσον αυτό, τόσο καλύτερα διαδίδεται ο ήχος. Οπότε, το ηχητικό κύμα θα διαδίδεται, γενικά<sup>1</sup>, καλύτερα στα στερεά, λιγότερο “καλά” στα υγρά και ακόμα λιγότερο στα αέρια.

Το πόσο “καλά” διαδίδεται ένα ηχητικό κύμα σε ένα μέσον αφορά την απόσβεση που παρατηρείται στο διαδιδόμενο κύμα και την απόσταση μέχρι την οποία μπορεί αυτή να γίνει αντιληπτή. Οι αποσβέσεις που υφίσταται ένα μηχανικό κύμα οφείλονται στο γεγονός ότι είναι μηχανικό κύμα και για την διάδοση του, η οποία γίνεται με την ταλάντωση των μορίων του υλικού μέσου στο οποίο αυτό διαδίδεται. Για την ταλάντωση λοιπόν των μορίων αυτών και την διάδοση του κύματος απαιτείται κάποια ενέργεια, η οποία είναι η ενέργεια που χάνεται από το κύμα, δηλαδή οι αποσβέσεις που αυτό υφίσταται. Έτσι, αν ένα ηχητικό κύμα διαδίδεται σε ένα υλικό μέσο με μικρή πυκνότητα, για να διαδοθεί το κύμα στο μέσον αυτό θα πρέπει τα μόρια του μέσου, αφού λόγω μικρής πυκνότητας θα είναι αρκετά απομακρυσμένα μεταξύ τους, να καλύψουν μεγάλη απόσταση, προκαλώντας μεγαλύτερη απώλεια ενέργειας στο κύμα, με την ενέργεια που χάνεται από το κύμα να μετατρέπεται σε θερμότητα. Αν όμως το κύμα διαδίδεται σε κάποιο πυκνότερο μέσον, τότε η απόσταση που θα έχουν να διανύσουν τα μόρια του μέσου αυτού θα είναι μικρότερη, και αντίστοιχα μικρότερες θα είναι και οι απώλειες (ή αποσβέσεις) που θα υφίσταται το κύμα αυτό. Αντίστοιχη ανάλυση μπορεί να γίνει και για την θερμοκρασία. Και η πυκνότητα και η

---

<sup>1</sup>Το κατά πόσον “καλά” θα διαδίδεται ένα ηχητικό κύμα, εκτός της πυκνότητας του υλικού, θα εξαρτηθεί και από το πόσο ισχυροί είναι οι δεσμοί μεταξύ των μορίων του υλικού, γιατί από αυτό θα εξαρτηθεί το κατά πόσον εύκολα ή δύσκολα, δηλαδή αν θα έχει πολλές ή λίγες απώλειες ενέργειας, μπορούν να κινηθούν τα μόρια του υλικού μέσω των οποίων θα διαδοθεί το ηχητικό κύμα. Στην γενικότερη περίπτωση όμως, θεωρούμε ότι διαδίδεται καλύτερα στα στερεά, μετά στα υγρά και τέλος στα αέρια.

θερμοκρασία του μέσου καθορίζουν το πόσο ισχυρές θα είναι οι αποσβέσεις τις οποίες θα υποστεί το κύμα καθώς και την ταχύτητα με την οποία θα διαδίδεται το κύμα στο μέσον αυτό.

Ως μηχανικό, αρμονικό κύμα, ο ήχος έχει όλα τα χαρακτηριστικά και τις ιδιότητες των αρμονικών κυμάτων, όπως είναι η συχνότητα “f”, η περίοδος “T”, το μήκος κύματος “λ”, το πλάτος της ταλάντωσης και η θέση ή η απομάκρυνση από την πηγή. Η συχνότητα ενός κύματος ή μίας ταλάντωσης εκφράζει το πόσες φορές επαναλαμβάνεται πλήρως το κύμα ή η ταλάντωση στη μονάδα του χρόνου ενώ η περίοδος είναι το αντίστροφο της συχνότητας ( $f=1/T$ ) και εκφράζει το πόσο χρόνο χρειάζεται για να ολοκληρωθεί μία πλήρης ταλάντωση. Όπως αναφέρθηκε παραπάνω, το μέγεθος που μεταβάλλεται στα ηχητικά κύματα είναι η πίεση και η μαθηματική έκφραση που περιγράφει την μεταβολή αυτή είναι η παρακάτω εξίσωση (Σχέση 1.1):

$$P = P_0 \cdot \sin 2\pi \left( ft - \frac{x}{\lambda} \right) = P_0 \cdot \sin 2\pi \left( \frac{t}{T} - \frac{x}{\lambda} \right) \text{ (Σχέση 1.1)}$$

όπου το μέγεθος “P<sub>0</sub>” είναι το πλάτος του κύματος, δηλαδή το πλάτος της πίεσης (N/m<sup>2</sup>), το “x” είναι η θέση ή η απομάκρυνση από την πηγή του κύματος (meter) και τα μεγέθη f, T και λ είναι η συχνότητα, η περίοδος και το μήκος κύματος (μεγέθη που αναφέρθηκαν παραπάνω), αντίστοιχα.

Η παραπάνω σχέση (Σχέση 1.1) εκφράζει, όπως προαναφέρθηκε, την μεταβολή της πίεσης που διαδίδεται σε έναν χώρο, με τις μεταβλητές να είναι ο χρόνος και η απόσταση x. Περιγράφει, δηλαδή, ένα γραμμικό κύμα. Α θεωρηθεί ότι το ηχητικό κύμα είναι σφαιρικό, τότε η σχέση που θα το ορίζει θα είναι η παρακάτω (Σχέση 1.2):

$$u(r, t) = \frac{A_0}{r} \cdot e^{j(\omega t - kr)} \text{ (Σχέση 1.2)}$$

που είναι η εξίσωση περιγραφής ενός σφαιρικού αρμονικού κύματος, με τον όρο  $k$  να ονομάζεται “κυματάριθμος” και να είναι ίσος με  $k=1/\lambda$ , εκφράζοντας τον αριθμό κύκλων του κύματος ανά μονάδα μήκους.

Χρησιμοποιώντας τις παραπάνω σχέσεις (Σχέσεις 1.1 και 1.2), μπορεί κάποιος να περιγράψει την περιοδικότητα της πίεσης στον χώρο ( $x$ ) και τον χρόνο ( $t$ ). Εκτός όμως αυτών, θα ισχύει και η θεμελιώδης εξίσωση της κυματικής, η οποία συνδέει την ταχύτητα διάδοσης, το μήκος κύματος και την συχνότητα ενός κύματος μεταξύ τους και είναι η εξής:

$$u = \lambda \cdot f \text{ (Σχέση 1.3)}$$

όπου  $u$  είναι η ταχύτητα διάδοσης του κύματος,  $\lambda$  το μήκος κύματος και  $f$  η συχνότητα του.

Εκτός των παραπάνω χαρακτηριστικών που έχει ο ήχος ως αρμονικό κύμα, θα ακολουθεί και τους γενικότερους κανόνες στους οποίους υπακούν τα κύματα και θα παρατηρούνται τα φαινόμενα την διάθλασης, της περίθλασης, της ανάκλασης και της συμβολής στα ηχητικά κύματα. Η περαιτέρω αναφορά στα φαινόμενα αυτά ξεφεύγει από τα πλαίσια της παρούσας εργασίας και δεν θα πραγματοποιηθεί, εκτός της συμβολής στην οποία οφείλεται ο θόρυβος και στην οποία θα γίνει αναφορά σε επόμενη παράγραφο.

Για την αντίληψη των ηχητικών κυμάτων θα πρέπει κάποιος να έχει έναν κατάλληλο δέκτη που να μπορεί να αναγνωρίζει και να λαμβάνει τα ηχητικά κύματα. Ο δέκτης αυτός στα ζώα είναι το “αυτί”, το οποίο μπορεί να λαμβάνει τα ηχητικά κύματα, δηλαδή να καταλαβαίνει την αυξομείωση της πίεσης του μέσου. Λαμβάνοντας τα μηχανικά αυτά κύματα, το αυτί τα μετατρέπει σε ηλεκτρικά σήματα, τα οποία διοχετεύει στον εγκέφαλο, που τα ερμηνεύει κατάλληλα και εξάγεται έτσι η “πληροφορία” που περιείχαν τα κύματα

αυτά. Οπότε, το αυτί ως όργανο, είναι ένας μετασχηματιστής μηχανικής ενέργειας σε ηλεκτρική.

Τα ηχητικά κύματα, όπως αναφέρθηκε παραπάνω, έχουν κάποια βασικά χαρακτηριστικά, με την συχνότητα του κύματος να είναι το μέγεθος σύμφωνα με το οποίο διαιρείται το συχνοτικό φάσμα των ακουστικών φαινομένων. Έτσι, οι ήχοι με συχνότητα μικρότερη των 20Hz ανήκουν στην κατηγορία των “υποήχων” (infrasonics). Οι υποήχοι παράγονται σε σημαντική ποσότητα στους χώρους δοκιμής των κινητήρων αεροπλάνων και στους χώρους όπου γίνονται εργασίες με κομπρεσέρ. Λόγω του μεγάλου μήκους κύματος που έχουν<sup>2</sup> είναι πολύ δύσκολο να απορροφηθούν από τα συνηθισμένα, δεν γίνονται αντιληπτοί από τον άνθρωπο και έχουν πολύ σοβαρές βιολογικές επιδράσεις.

Οι ήχοι με συχνότητα μεγαλύτερη των 20Hz και μικρότερη των 20.000Hz ανήκουν στην κατηγορία των ακουστών συχνοτήτων από το ανθρώπινο αυτί και στην κατηγορία αυτή ανήκουν οι ήχοι που μπορεί να αντιλαμβάνεται ο άνθρωπος. Τέλος, οι ήχοι που έχουν συχνότητα μεγαλύτερη των 20.000Hz ανήκουν στην κατηγορία των “υπερήχων” (ultrasonics). Δεν γίνονται αντιληπτοί από τον άνθρωπο και δημιουργούνται κυρίως γύρω από αντιδραστήρες. Όταν έχουν χαμηλό επίπεδο έντασης, δεν θεωρείται ότι έχουν σημαντικές βιολογικές επιπτώσεις γιατί το δέρμα ανακλά τα κύματα αυτά ενώ η ελεγχόμενη πρόσπτωση τους πάνω στους ιστούς μπορεί να έχουν σημαντικά οφέλη για τους ιστούς, με χαρακτηριστικότερο παράδειγμα να είναι οι εφαρμογές που βρίσκουν οι υπέρηχοι στην ιατρική.

---

<sup>2</sup>Το μήκος κύματος είναι αντιστρόφως ανάλογο με την συχνότητα του κύματος και για τον λόγο αυτό, τα ηχητικά κύματα που ανήκουν στην κατηγορία των “υποήχων” θα έχουν την μικρότερη συχνότητα σε σχέση με τα ηχητικά κύματα των άλλων κατηγοριών και άρα το μεγαλύτερο μήκος κύματος από όλες τις κατηγορίες.

## 1.2- Καταγραφή Ήχου

Όπως αναφέρθηκε παραπάνω, η παραγωγή και ανίχνευση του ήχου είναι ένας από τους σημαντικότερους τρόπους που έχουν οι ζωντανοί οργανισμοί για να επικοινωνούν και να αντιλαμβάνονται το περιβάλλον τους. Όμως, ο άνθρωπος χρησιμοποιεί τον ήχο και για την ψυχαγωγία του ήδη για πολλές χιλιετίες. Για τον λόγο αυτό, η καταγραφή του ήχου έχει σημαντικό ρόλο στην ανθρώπινη ιστορία, με πρώτη τέτοια επιτυχημένη προσπάθεια καταγραφής ήχου σε κάποιο μέσο να γίνεται το 1857 από τον Γάλλο εφευρέτη Εντουάρ-Λεόν Σκοτ Ντε Μαρτενβίλ (Édouard-Léon Scott de Martinville), ο οποίος εφηύρε τον “Φωνοαυτογράφο”. Η συσκευή αποτελείτο από ένα χωνί που λάμβανε τα ηχητικά κύματα και τα μετέδιδε μέσω μεμβράνης σε μια προσαρμοσμένη γουρουνότριχα η οποία χάραζε την κυματομορφή του ήχου πάνω σε έναν περιστρεφόμενο γυάλινο κύλινδρο, ο οποίος ήταν καλυμμένος με φούμο. Άλλες εκδοχές της συσκευής χρησιμοποιούσαν για καταγραφή ένα ρολό χαρτί, έτσι ώστε το φωνοαυτογράφημα να είναι μόνιμο. Όλες όμως είχαν ένα μεγάλο μειονέκτημα. Μπορούσαν να καταγράψουν, είτε στον γυάλινο κύλινδρο είτε στον χάρτινο, την κυματομορφή του ήχου αλλά δεν μπορούσαν να τον αναπαράγουν.

Λίγα χρόνια αργότερα, η ιδέα του φωνοαυτόγραφου εξελίχθηκε με την εφεύρεση του “Φωνόγραφου” (Phonograph) για την οποία απόκτησε, το 1878 την πατέντα ο Αμερικάνος Τόμας Άλβα Έντισον (Thomas A. Edison). Η συσκευή αυτή βασιζόταν στις ίδιες αρχές με τον Φωνοαυτόγραφο και χρησιμοποιούσε μια βελόνα για να καταγράφει τον ήχο σε έναν κύλινδρο με αυλάκια, επικαλυμμένο με κάποιο μεταλλικό υλικό, ο οποίος περιστρεφόταν με σταθερή ταχύτητα. Τον κύλινδρο στον οποίο είχε καταγραφεί ο ήχος, η “ηχογράφηση”, μπορούσε να “διαβάσει” η συσκευή και να αναπαράγει τον καταγεγραμμένο ήχο. Αργότερα, αντικατέστησε την επικάλυψη από μέταλλο με επικάλυψη από σκληρό κερί στον κύλινδρο στον οποίο καταγραφόταν ο ήχος. Αυτό έδινε την δυνατότητα του “σβησίματος” ενός

κυλίνδρου με το ξύσιμο της στρώσης που είχε χρησιμοποιηθεί και την καταγραφή νέου ήχου στον ίδιο κύλινδρο, καθιστώντας έτσι το σύστημα αυτό κατάλληλο για οικιακή χρήση, και το προϊόν αυτό γνώρισε μεγάλη εμπορική επιτυχία.

Η εξέλιξη της παραπάνω εφεύρεσης έγινε με την εφεύρεση που το 1886 πρότεινε ο Εμίλ Μπερλίνερ (Emile Berliner), την οποία ονόμασε “Γραμμόφωνο” (Grammophone). Η ειδοποιός διαφορά από τον φωνόγραφο ήταν ότι ο ήχος καταγραφόταν σε έναν δίσκο, σχηματίζοντας μία ελικοειδής διαδρομή, ένα συνολικό “αυλάκι” σε όλη την επιφάνεια της πλάκας. Την ελικοειδή αυτή διαδρομή ακολουθούσε η βελόνα, η οποία έκανε εγκάρσιες, στην διαδρομή της, ταλαντώσεις, οι οποίες περιείχαν τον καταγεγραμμένο ήχο, και τις οποίες μπορούσε να μετατρέπει σε ήχο, αναπαράγοντας την ηχογράφιση. Η τεχνολογία αυτή είναι άμεσος πρόγονος των δίσκων βινυλίου που χρησιμοποιούνται μέχρι και σήμερα.

Σε όλες τις παραπάνω εφευρέσεις, ένα από τα πολύ βασικά μέρη τους ήταν ο μετατροπέας του ήχου σε κάποια μορφή η οποία να μπορεί να καταγραφεί. Στις παραπάνω εφευρέσεις, η μορφή αυτή η κυματομορφή του ήχου που καθόριζε η διαδρομή της βελόνας, κατά την καταγραφή. Από τις αρχές του 20<sup>ου</sup> αιώνα όμως, άρχισε να υλοποιείται και η “ηλεκτρική” καταγραφή του ήχου, μετατρέποντας τα ηχητικά κύματα σε ηλεκτρικά, που με τον τρόπο αυτό γίνονται οι ηχογραφήσεις στην σύγχρονη εποχή. Η εφεύρεση που χρησιμοποιείται για την μετατροπή του ήχου σε ηλεκτρικά σήματα είναι το “Μικρόφωνο”. Πήρε την ονομασία του από τον Ντέιβιντ Χιουζ, στον οποίο αποδίδεται η εφεύρεση αυτή και ο οποίος επινόησε μια διάταξη μεταφοράς ήχου που ήταν τόσο ευαίσθητη, που τη θεωρούσε κάτι σαν «μικροσκόπιο ήχου» και την ονόμασε «μικρόφωνο» (microphone). Το μικρόφωνο, λοιπόν, είναι μία συσκευή που μετατρέπει τα ηχητικά κύματα σε κατάλληλης μορφής ηλεκτρικά σήματα, όπως και το ανθρώπινο αυτί. Τα κατάλληλα διαμορφωμένα αυτά ηλεκτρικά σήματα μπορούν να μεταδοθούν ή να αποθηκευθούν κάπου και να μετατραπούν στον αρχικό ήχο. Οι δυνατότητες αυτές οδήγησαν σε (ή βοήθησαν στην εξέλιξη τους) πολλές



άλλες πολύ σημαντικές εφευρέσεις, όπως είναι το ραδιόφωνο και το τηλέφωνο, δύο εφευρέσεις που δημιούργησαν τον κλάδο των Τηλεπικοινωνιών.

Υπάρχουν διάφορα είδη μικροφώνων, κάποια από τα οποία είναι τα εξής:

- **Μικρόφωνο άνθρακα:** το μικρόφωνο άνθρακα αποτελείται από ένα μεταλλικό σώμα μέσα στο οποίο είναι τοποθετημένο ένα μικρό δοχείο που έχει μονωτικές ιδιότητες. Η λειτουργία του μικροφώνου αυτού, δηλαδή η μετατροπή των ηχητικών κυμάτων σε ηλεκτρικά, βασίζεται στην μεταβολή της ηλεκτρικής αντίστασης των κόκκων του άνθρακα, εξαιτίας της μεταβολής της πίεσης που ασκείται στο διάφραγμα από τα ηχητικά κύματα. Ήταν το πρώτο μικρόφωνο που εφευρέθηκε και χρησιμοποιήθηκε.
- **Μικρόφωνο ταινίας:** το μικρόφωνο ταινίας αποτελείται από μια λεπτή πτυχωτή ταινία, συνήθως από αλουμίνιο, η οποία μπορεί και πάλλεται ελεύθερα μέσα στο ηλεκτρικό πεδίο που δημιουργεί ένας ισχυρός μαγνήτης. Το μικρόφωνο αυτό έχει κατευθυνόμενη λήψη από δύο αντίθετες κατευθύνσεις και χρησιμοποιείται για ταυτόχρονη εξυπηρέτηση δύο ομιλητών, λόγω της καλής του απόκρισης.
- **Δυναμικό μικρόφωνο:** το δυναμικό μικρόφωνο αποτελείται από έναν ισχυρό μόνιμο μαγνήτη κ' ένα πηνίο τοποθετημένο ανάμεσα στους πόλους του, ώστε να κινείται ελεύθερα. Η λειτουργία του βασίζεται στο φαινόμενο της επαγωγής: όταν ηχητικά κύματα πέφτουν στο διάφραγμα, το πηνίο πάλλεται στο πεδίο του μαγνήτη, τέμνονται οι μαγνητικές γραμμές κι εμφανίζεται στα άκρα του πηνίου επαγωγική τάση. Χρησιμοποιείται κυρίως σε μικροφωνικές εγκαταστάσεις που απαιτούν ιδιαίτερη πιστότητα.
- **Πυκνωτικό μικρόφωνο:** Η λειτουργία του στηρίζεται στις μεταβολές χωρητικότητας ενός ενσωματωμένου πυκνωτή, σύμφωνα με τις μεταβολές της πίεσης που προκαλούνται από τα ηχητικά κύματα. Είναι ο πλέον σύγχρονος και αποδοτικός - από άποψη ποιότητας - τύπος μικροφώνου. Τα πυκνωτικά μικρόφωνα συναντιούνται στις

θεατρικές παραστάσεις, στη μουσική, λόγω της αυξημένης ευαισθησίας του μικροφώνου.

Ανάλογα με τον τρόπο διαμόρφωσης των ηλεκτρικών σημάτων στα οποία “ηχογραφείται” ο ήχος, αυτά μπορεί να είναι αναλογικά ή ψηφιακά. Όπως ισχύει για όλα τα σήματα, ανάλογα με τις τιμές που μπορεί να πάρει ο χρόνος και το φυσικό μέγεθος που μεταβάλλεται, τα σήματα χωρίζονται σε σήματα συνεχούς και διακριτού χρόνου και σε σήματα συνεχών και διακριτών τιμών (κβαντισμένα). Τα σήματα που είναι διακριτού χρόνου και διακριτών τιμών, των οποίων οι τιμές έχουν κωδικοποιηθεί, ονομάζονται ψηφιακά.

Όλα τα σήματα που παράγονται από την φύση είναι αναλογικά, δηλαδή είναι συνεχούς χρόνου και συνεχών τιμών. Οπότε, αφού το μικρόφωνο μετατρέπει τα ηχητικά κύματα σε ηλεκτρικά, θα είναι και αυτά αναλογικά. Παραδείγματα τέτοιων αναλογικών μέσων αποθήκευσης του ήχου είναι η μαγνητική ταινία (κασέτες) και οι δίσκοι βινυλίου<sup>3</sup>, ενώ οι οπτικοί δίσκοι (CD, DVD, BlueRay Disk) καθώς και όλα τα μέσα αποθήκευσης που χρησιμοποιούνται από έναν ηλεκτρονικό υπολογιστή είναι μέσα ψηφιακής αποθήκευσης.

Με όποια μορφή και αν διαμορφωθεί και κωδικοποιηθεί ο ήχος, θα έχει πάντα το πρόβλημα του θορύβου κατά την ηχογράφηση του. Λόγω της προσθετικής ιδιότητας που έχει η πίεση, δηλαδή του φαινομένου της συμβολής διαφορετικών κυμάτων που διαδίδονται στον ίδιο χώρο, σε ένα ηχητικό σήμα το οποίο θα ηχογραφηθεί θα προστεθούν και τα υπόλοιπα ηχητικά σήματα που διαδίδονται στον ίδιο χώρο. Έτσι, το σήμα που θα ηχογραφηθεί θα είναι το άθροισμα όλων αυτών των σημάτων, δηλαδή του επιθυμητού σήματος και των υπόλοιπων σημάτων που διαδίδονται στον ίδιο χώρο. Οπότε θα προστεθούν στο επιθυμητό σήμα και άλλα ανεπιθύμητα, τα οποία θεωρούνται ως θόρυβος. Οι τρόποι αντιμετώπισης του φαινομένου αυτού που αφορούν την αφαίρεση του θορύβου από ένα ηχογραφημένο σήμα

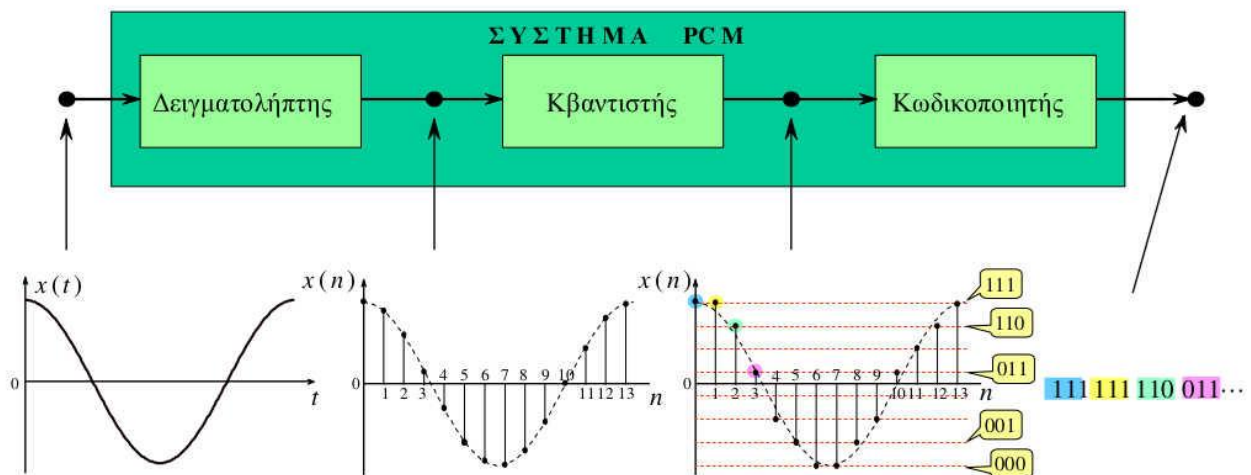
---

<sup>3</sup>Στου δίσκου βινυλίου, όπως αναφέρθηκε παραπάνω, ο ήχος δεν αποθηκεύεται ως ηλεκτρικό σήμα αλλά “κωδικοποιείται” στις αυλακώσεις του δίσκου και κατά την αναπαραγωγή του “παράγονται” τα ηλεκτρικά σήματα.

είναι το θέμα της παρούσας εργασίας και στους τρόπους αυτούς θα γίνει αναφορά στα επόμενα κεφάλαιο.

### 1.3- Ψηφιακός Ήχος

Όπως αναφέρθηκε παραπάνω, ο ήχος όταν παράγεται είναι αναλογικός. Όταν θα μετατραπεί σε ηλεκτρικό σήμα από το μικρόφωνο, αν δεν υποστεί κάποια περαιτέρω επεξεργασία, το σήμα που θα αποθηκευθεί θα είναι αναλογικό. Όμως, τα σύγχρονα συστήματα καταγραφής, αποθήκευσης και αναπαραγωγής του ήχου είναι ψηφιακά. Σχεδόν όλα τα συστήματα που χρησιμοποιούν σήματα ομιλίας και ήχου είναι ψηφιακά, με χαρακτηριστικότερα τα τηλέφωνα, το ραδιόφωνο και την τηλεόραση. Οπότε θα πρέπει να μετατραπεί ο αναλογικό ήχος σε ψηφιακό σήμα, το οποίο ονομάζεται “Ψηφιακός Ήχος”. Η πιο ευρέως χρησιμοποιούμενη μέθοδος παραγωγής ψηφιακού ήχου είναι η μέθοδος της παλμοκωδικής διαμόρφωσης (Pulse Code Modulation-PCM), της οποίας τα βασικότερα



στάδια φαίνονται στην παρακάτω εικόνα (Εικόνα 1.2).

Εικόνα 1.2: Σύστημα Παλμοκωδικής Διαμόρφωσης (PCM)

Στο σύστημα αυτό, στην είσοδο του εφαρμόζεται ο ηχογραφημένος αναλογικός ήχος και στην έξοδο παράγεται μία ακολουθία δυαδικών ψηφίων που αναπαριστά τον αρχικό αναλογικό ήχο σε ψηφιακή μορφή, δηλαδή ο ψηφιακός ήχος. Για να γίνει όμως η μετατροπή αυτή θα πρέπει το σήμα του ήχου από αναλογικό που ήταν, δηλαδή συνεχούς χρόνου και συνεχών τιμών, να γίνει διακριτού χρόνου και διακριτών τιμών και στην συνέχεια οι τιμές αυτές να κωδικοποιηθούν με κατάλληλο ψηφιακό κώδικα.

Η μετατροπή του σήματος συνεχούς χρόνου σε διακριτού γίνεται με την “δειγματοληψία” (Sampling). Με την διαδικασία αυτή, μέσω μίας κατάλληλης διάταξης που ονομάζεται “δειγματολήπτης”, λαμβάνονται τιμές του σήματος, δείγματα, ανά τακτά χρονικά διαστήματα, και από τα δείγματα αυτά μπορεί να γίνει ανακατασκευή του αρχικού σήματος και αναπαραγωγή του ήχου. Για να είναι δυνατόν αυτό, θα πρέπει η δειγματοληψία να γίνει σύμφωνα με το “θεώρημα δειγματοληψίας Nyquist-Shannon”. Σύμφωνα με αυτό, για να μπορεί να γίνει ανακατασκευή ενός δειγματοληπτημένου σήματος, χωρίς να έχουμε απώλεια πληροφορίας, πρέπει τα χρονικά διαστήματα κατά τα οποία λαμβάνονται τα δείγματα, που ονομάζεται “χρόνος δειγματοληψίας” να είναι αρκετά μικρά ώστε να μπορούν να λαμβάνονται τουλάχιστον δύο δείγματα σε κάθε περίοδο του σήματος. Για να είναι εφικτό αυτό, ο χρόνος (ή περίοδος) δειγματοληψίας πρέπει να έχει τιμή το πολύ την μισή της περιόδου του σήματος. Αντίστοιχα, η συχνότητα δειγματοληψίας πρέπει να είναι τουλάχιστον ίση με την διπλάσια συχνότητα του σήματος, και αν το σήμα περιέχει περισσότερες από μία συχνότητες, να είναι τουλάχιστον ίση με το διπλάσιο της μέγιστης συχνότητας του σήματος. Η μαθηματική έκφραση του παραπάνω είναι η παρακάτω σχέση (Σχέση 1.4).

$$f_s \geq 2 \cdot f_{max} \text{ (Σχέση 1.4)}$$

όπου  $f_s$  είναι η συχνότητα δειγματοληψίας (sampling frequency) ενώ η  $f_{max}$  είναι η μέγιστη συχνότητα που περιέχει το σήμα.

Η παραπάνω έκφραση ονομάζεται “Κριτήριο Nyquist” και η ελάχιστη τιμή συχνότητας δειγματοληψίας που υπολογίζεται από αυτήν ονομάζεται “Συχνότητα Nyquist”. Συνήθως επιλέγεται συχνότητα λίγο μεγαλύτερη από την συχνότητα Nyquist για την δειγματοληψία σημάτων. Όσον αφορά τον ήχο, επειδή το ανθρώπινο αυτί μπορεί να αντιλαμβάνεται ήχους που έχουν συχνότητα μέχρι περίπου τα 20kHz, με μέγιστη θεωρητική τιμή τα 22.5kHz, οπότε τα ηχητικά σήματα υψηλότερων συχνοτήτων δεν χρειάζεται να μπορούν να αναπαραχθούν σωστά, η συχνότητα Nyquist θα είναι ίση με 44.1kHz, συχνότητα που χρησιμοποιείται για την δειγματοληψία του ήχου που ψηφιοποιείται για να καταγραφεί στους μουσικούς οπτικούς δίσκους (CD).

Έχοντας δειγματοληψήσει το σήμα, το σήμα συνεχούς χρόνου έχει μετατραπεί σε σήμα διακριτού χρόνου. Στην συνέχεια, το σήμα αυτό μετατρέπεται από σήμα συνεχών τιμών σε σήμα διακριτών τιμών. Η διαδικασία με την οποία γίνεται αυτό ονομάζεται “κβάντωση”. Είναι η διαδικασία μετατροπής των δειγμάτων (τιμών) της εξόδου του δειγματολήπτη που αναφέρθηκε παραπάνω σε ακολουθία διακριτών τιμών, οι οποίες ανήκουν σε ένα πεπερασμένο σύνολο επιπέδων πλάτους του σήματος και ονομάζονται “στάθμες ή επίπεδα κβαντισμού”. Αν τα επίπεδα αυτά απέχουν την ίδια απόσταση τιμών μεταξύ τους, τότε η κβάντωση χαρακτηρίζεται ως “ομοιόμορφη” (uniform), ενώ στην αντίθετη περίπτωση ως “μη ομοιόμορφη” (non-uniform).

Κατά την διαδικασία την κβάντωσης, αρχικά εξετάζεται το σήμα ήχου και καθορίζεται το εύρος τιμών που καταλαμβάνει. Στην συνέχεια, οι τιμές αυτές χωρίζονται σε κατάλληλο πλήθος σταθμών και για κάθε δείγμα επιλέγεται η στάθμη στην οποία είναι πιο κοντά η τιμή του. Η επιλογή του πλήθους των σταθμών αυτών θα πρέπει να είναι κατάλληλη ώστε να μην έχουμε απώλεια πληροφορίας αλλά και να μην έχουμε υπερβολικά πολλές

στάθμες γιατί θα απαιτείται μεγαλύτερος αποθηκευτικός χώρος, οπότε γίνεται ένας συμβιβασμός μεταξύ των δύο ώστε να μπορεί ο ψηφιακός ήχος να είναι υψηλής ποιότητας αλλά και όσο δυνατόν μικρότερου μεγέθους.

Τέλος, το σήμα διακριτού χρόνου και διακριτών τιμών θα κωδικοποιηθεί για να παραχθεί το τελικό ψηφιακό σήμα. Η κωδικοποίηση αυτή θα γίνει με κατάλληλο κώδικα, ανάλογα με την εφαρμογή για την οποία προορίζεται ο παραγόμενος ψηφιακός ήχος. Έτσι, κάθε δείγμα θα αντιστοιχηθεί με μία ακολουθία δυαδικών ψηφίων και ο ψηφιακός ήχος θα είναι η αλληλουχία των ακολουθιών κάθε δείγματος. Το πόσα ψηφία θα χρειαστούν για την κωδικοποίηση κάθε δείγματος θα εξαρτηθεί, αφενός από την εφαρμογή για την οποία προορίζεται και αφετέρου από το πλήθος των σταθμών που ορίστηκαν. Αυτός είναι και ο λόγος για τον οποίο, όσο περισσότερες στάθμες ορίζονται κατά την κβάντωση ενός σήματος, τόσο περισσότερος αποθηκευτικός χώρος θα απαιτείται.

## **ΚΕΦΑΛΑΙΟ 2**

### ***ΘΟΡΥΒΟΣ και ΑΝΤΙΜΕΤΩΠΙΣΗ ΤΟΥ***

#### **2.1- Θόρυβος**

Στο προηγούμενο κεφάλαιο αναφέρθηκε η έννοια του θορύβου. Τι είναι όμως ο “θόρυβος”; Αναφορικά με την έννοια αυτή, είναι πιο εύκολη η περιγραφή των φυσικών ιδιοτήτων του θορύβου, των πηγών από τις οποίες παράγεται και τις επιδράσεις του στον άνθρωπο παρά να δοθεί ένας ακριβής ορισμός. Κάποιοι από τους ορισμούς που έχουν δοθεί και χρησιμοποιούνται για τον “θόρυβο” στα ακουστικά σήματα τον ορίζουν σύμφωνα με τις ιδιότητες που έχει, και ένας τέτοιος ορισμός καθορίζει τον θόρυβο ως το ηχητικό ερέθισμα που είναι ανεπιθύμητο στον ακροατή είτε διότι δεν είναι ευχάριστο είτε είναι ενοχλητικό και επιπλέκεται στην αντίληψη των επιθυμητών ήχων είτε, τέλος, είναι βλαβερό για τον οργανισμό.

Στον σύγχρονο κόσμο, οι πηγές τέτοιων ηχητικών ερεθισμάτων που να χαρακτηρίζονται ως “θόρυβος” είναι πολλές και μάλιστα έχει αναγνωριστεί ο θόρυβος ως μία μορφή μόλυνσης του περιβάλλοντος, όχι λιγότερο επικίνδυνη για τον άνθρωπο από τις άλλες πηγές περιβαλλοντολογικής μόλυνσης. Η μορφή αυτή μόλυνσης ονομάζεται “Ηχορύπανση” και, ανάλογα με την περίπτωση εφαρμογής, ορίζονται κάποια επιτρεπτά όρια στο επίπεδο ήχου που μπορεί να παράγει οποιοσδήποτε, από ένα μηχάνημα μέχρι και ένα εργοστάσιο έτσι ώστε να μην καθίσταται επικίνδυνη για τους κατοίκους μίας περιοχής το επίπεδο ηχορύπανσης της.

Ένας άλλος ορισμός του θορύβου θεωρεί ως “θόρυβο” κάθε ήχο που παράγεται από κάποια “τυχαία” πηγή, χωρίς να περιέχει κάποια πληροφορία και το φάσμα του δεν περιέχει καθαρούς τόνους. Ο ορισμός αυτός αναφέρεται στις ιδιότητες που έχει ο ήχος που θα χαρακτηριστεί ως θόρυβος. Μία “τυχαία” πηγή θα είναι μία πηγή ήχου η οποία μπορεί να

είναι κάποια φυσική διεργασία που απλά παράγει ηχητικά κύματα χωρίς συγκεκριμένα χαρακτηριστικά και τα οποία, φυσικά, δεν περιέχουν πληροφορία.

Ανάλογα με τον τρόπο με τον οποίο παράγεται ο θόρυβος, θα έχει κάποια ιδιαίτερα χαρακτηριστικά τα οποία θα αφορούν, κυρίως, το συχνοτικό του φάσμα, σύμφωνα με το οποίο διακρίνονται κάποια είδη ή κατηγορίες θορύβου. Μία πρώτη κατηγοριοποίηση αφορά το αν το σήμα του θορύβου είναι σταθερό στη διάρκεια του χρόνου, οπότε τότε ο θόρυβος καλείται “στάσιμος” (stationary) ενώ αν μεταβάλλεται στο χρόνο καλείται “μη στάσιμος” (non-stationary).

Μία άλλη κατηγοριοποίηση γίνεται σύμφωνα με τα χαρακτηριστικά του φάσματος του θορύβου, και διακρίνονται, κυρίως, τα παρακάτω είδη θορύβου.

- **Θόρυβος Περιβάλλοντος:** Ως θόρυβος περιβάλλοντος θεωρείται ο θόρυβος από πολλές πηγές κοντά ή σε απόσταση από τον δέκτη του ακουστικού σήματος και μπορεί να χαρακτηριστεί ως “θόρυβος χωρίς ταυτότητα”. Στην κατηγορία αυτή θα ανήκουν τα ηχητικά κύματα των υποήχων και των υπερήχων που αναφέρθηκαν στο προηγούμενο κεφάλαιο.
- **Θόρυβος φόντου ή βάθους:** Στο είδος αυτό θορύβου ανήκουν τα ηχητικά σήματα που παράγονται από πολλές πηγές στον ίδιο χώρο, διαφορετικές από την πηγή ενδιαφέροντος του ακροατή, όπως είναι παραδείγματος χάριν ο θόρυβος του εξαερισμού σε έναν συναυλιακό χώρο.
- **Τυχαίος θόρυβος:** Στο είδος αυτό θορύβου ανήκουν τα ηχητικά σήματα που παράγονται από τυχαίες πηγές και το είδος αυτό του θορύβου δεν μπορεί να προβλεφθεί χρονικά.
- **Λευκός Θόρυβος:** Ως “Λευκός Θόρυβος” ορίζεται ο θόρυβος που έχει συνεχές ακουστικό φάσμα συχνοτήτων, με ισοκατανεμημένη ενέργεια ήχου σε κάθε πλάτος ζώνης συχνοτήτων στο φάσμα που μας ενδιαφέρει. Είναι δηλαδή ένα σύνθετο



ηχητικό κύμα που περιέχει όλες τις συχνότητες του φάσματος ενδιαφέροντος όπου κάθε κύμα κάθε συχνότητας έχει την ίδια ένταση με τα υπόλοιπα.

- **Ροζ Θόρυβος:** Ως “Ροζ Θόρυβος” ορίζεται ο θόρυβος που τα βασικά του χαρακτηριστικά είναι ότι έχει ίση ποσότητα ενέργειας ανά οκτάβα από τα 20Hz μέχρι τα 20.000Hz, δηλαδή σε όλο το ακουστικό συχνοτικό φάσμα του ήχου.

Η παραπάνω κατηγοριοποίηση του θορύβου έγινε βάσει των χαρακτηριστικών που έχει ο ήχος που χαρακτηρίζεται ως “θόρυβος”. Στις κατηγορίες αυτές θα ανήκει οποιοδήποτε ηχητικό κύμα θορύβου που παράγεται από μία πραγματική πηγή, δηλαδή σε μία πηγή που παράγει ένα ηχητικό κύμα. Όμως, εκτός των πραγματικών πηγών θορύβου υπάρχουν και πηγές θορύβου που οφείλονται στα εργαλεία με τα οποία επεξεργαζόμαστε τον ήχο. Ο θόρυβος που παράγεται με τον τρόπο αυτό προστίθεται άμεσα στο ηχητικό κύμα και είναι αρκετά δύσκολη η απομάκρυνση του. Μία τέτοια πηγή θορύβου είναι τα διάφορα ηλεκτρικά και ηλεκτρονικά στοιχεία που υπάρχουν στις διατάξεις επεξεργασίας του ήχου, όπως είναι οι καταγραφείς ήχου, οι ενισχυτές και τα ηχεία, οι οποίες λόγω της κίνησης των ηλεκτρονίων μέσα στο κύκλωμα παράγουν κάποιες παραμορφώσεις στα ηλεκτρικά σήματα που αναπαριστούν κάποιο ηχητικό σήμα. Το είδος αυτό του θορύβου χαρακτηρίζεται ως “θερμικός θόρυβος” και σύμφωνα με το συχνοτικό του φάσμα χαρακτηρίζεται ως “Λευκός Θόρυβος”. Δεν μπορεί να εξουδετερωθεί πλήρως και για τον λόγο αυτό θέτει ένα άνω όριο στην απόδοση των συστημάτων αλλά σχεδιάζοντας τις κυκλωματικές διατάξεις να λειτουργούν σε μικρότερη θερμοκρασία μπορεί να μειωθεί η επίδραση του.

Ένα άλλο είδος θορύβου που οφείλεται στα κυκλώματα που χρησιμοποιούνται για την επεξεργασία του ήχου είναι ο θόρυβος που εμφανίζεται στα κυκλώματα αυτά λόγω επαγωγικών φαινομένων που προκαλούνται από τις ηλεκτρομαγνητικές διαταραχές του

περιβάλλοντος και μπορεί να αντιμετωπιστεί με καλύτερη ηλεκτρομαγνητική θωράκιση<sup>4</sup> των κυκλωματικών διατάξεων που χρησιμοποιούνται. Ακόμα, θόρυβος σε ένα σήμα μπορεί να εμφανιστεί λόγω της διέλευσης του σήματος από κάποια μη γραμμική διάταξη, ο οποίος ονομάζεται “Θόρυβος Ενδοδιαμόρφωσης”. Το είδος αυτό του θορύβου μπορεί να παραχθεί και από διάφορες διεργασίες στις οποίες υπόκειται το σήμα, όπως από την παραγωγή σημάτων σε μία συχνότητα από το άθροισμα ή την διαφορά δύο αρχικών συχνοτήτων ή και τον πολλαπλασιασμό των συχνοτήτων αυτών, ή και από την μείξη δύο σημάτων διαφορετικών συχνοτήτων, διαδικασία που θα παράξει και σήματα σε συχνότητα που είναι ίση με το άθροισμα και την διαφορά των συχνοτήτων των σημάτων αυτών αλλά και πολλαπλάσια αυτών<sup>5</sup>. Τέλος, θόρυβος μπορεί να εμφανιστεί και να οφείλεται στα ποικίλα ηλεκτρικά φαινόμενα που μπορεί να προκαλέσουν αλλοίωση στο ηλεκτρικό σήμα που αναπαριστά τον ήχο, όπως τέτοιο είδος είναι ο “θόρυβος βολής” ή “θόρυβος Schottky”, ο οποίος παρατηρείται κυρίως στην ημιαγωγική επαφή των ολοκληρωμένων κυκλωμάτων και συνδέεται με την τυχαία εκπομπή ή άφιξη των ηλεκτρονίων στην άνοδο.

Εκτός όμως των παραπάνω, υπάρχει και ένα είδος θορύβου που οφείλεται στην ψηφιοποίηση του ήχου, όπως η διαδικασία που περιγράφηκε στο προηγούμενο κεφάλαιο κατά την παρουσίαση του συστήματος PCM. Όπως προαναφέρθηκε, η ψηφιοποίηση του ήχου γίνεται με την κβάντωση. Όμως, η κβάντωση είναι η αντιστοίχιση των τιμών των δειγμάτων του σήματος σε προκαθορισμένες στάθμες πλάτους του μεγέθους του σήματος. Οι τιμές δηλαδή των δειγμάτων “στρογγυλοποιούνται” προς κάποιες προκαθορισμένες τιμές, τις στάθμες κβαντισμού. Η στρογγυλοποίηση αυτή αλλοιώνει την πραγματική τιμή των δειγμάτων και εισάγει ένα “σφάλμα” στις τιμές που κωδικοποιήθηκαν συγκριτικά με τις

---

<sup>4</sup>Με τον όρο “ηλεκτρομαγνητική θωράκιση” περιγράφεται η απομόνωση μίας κυκλωματικής διάταξης ή μίας συσκευής από το περιβάλλον και συνήθως επιτυγχάνεται με την προσθήκη κατάλληλου περιβλήματος το οποίο δεν επιτρέπει την διέλευση ηλεκτρομαγνητικών κυμάτων προς την συσκευή.

<sup>5</sup>Τα σήματα που θα παραχθούν σε συχνότητες ίσες με το άθροισμα, τη διαφορά και πολλαπλάσια αυτών θα είναι σήματα τα οποία, γενικά, δεν είναι επιθυμητά και για τον λόγο αυτό χαρακτηρίζονται ως “θόρυβος”.

πραγματικές τιμές, που ονομάζεται “σφάλμα κβαντισμού”<sup>6</sup> και οι αλλοιώσεις που προκαλούνται στον ήχο είναι ο “θόρυβος κβαντισμού”.

Μια άλλη κατηγοριοποίηση του θορύβου γίνεται σύμφωνα με τον τρόπο που αυτός θα επηρεάσει το επιθυμητό σήμα. Αν ο θόρυβος προστίθεται στο σήμα, τότε θα καλείται ως “προσθετικός θόρυβος”. Τέτοιου είδους θόρυβος θα είναι ο θόρυβος περιβάλλοντος και ο θόρυβος φόντου που περιγράφηκαν παραπάνω, όπως και οποιοσδήποτε θόρυβος που καταγράφεται μαζί με τον επιθυμητό. Αν ο θόρυβος είναι θόρυβος ενδοδιαμόρφωσης, ανάλογα με το στοιχείο ή την διάταξη που θα τον προκαλεί, τότε ο θόρυβος μπορεί να πολλαπλασιάζεται με το σήμα, οπότε να καλείται “πολλαπλασιαστικός θόρυβος”. Τέλος, αν ο θόρυβος που επηρεάζει το σήμα επιδρά σε αυτό συνελκτικά, τότε καλείται “συνελκτικός θόρυβος”.

Η συνέλιξη είναι μία πράξη που εφαρμόζεται σε δύο συναρτήσεις, συνεχείς ή διακριτές, και συμβολίζεται με το “\*”. Είναι μία πολύ χρήσιμη πράξη για τον κλάδο της επεξεργασίας σημάτων, που είναι ο κλάδος που ασχολείται με την αποθορυβοποίηση των σημάτων ομιλίας, που είναι το βασικό θέμα της παρούσας εργασίας, γιατί με αυτήν μπορεί να υπολογιστεί η έξοδος ενός γνωστού συστήματος για δεδομένη είσοδο. Αν οι δύο συνεχής συναρτήσεις είναι οι  $f(t)$  και  $g(t)$ , τότε το αποτέλεσμα της συνέλιξης των συναρτήσεων αυτών θα είναι ίσο με:

$$f * g(t) = \int_{k=-\infty}^{\infty} f(k) \cdot g(t - k) dt \text{ (Σχέση 2.1)}$$

ενώ αν οριστούν οι διακριτές συναρτήσεις  $f(n)$  και  $g(n)$ , τότε το αποτέλεσμα της συνέλιξης των συναρτήσεων αυτών θα είναι ίσο με:

---

<sup>6</sup>Το σφάλμα κβαντισμού μπορεί να πάρει τιμές από μηδέν μέχρι και το μισό την απόστασης των επιπέδων κβάντωσης.

$$f * g(n) = \sum_{k=-\infty}^{\infty} f(k) \cdot g(n - k) \text{ (Σχέση 2.2)}$$

## 2.2- Ανάλυση Fourier

Ο Γάλλος μαθηματικός Jean-Baptiste Joseph Fourier (Φουριέ) απέδειξε ότι κάθε περιοδικό σήμα  $x(t)$ , ανεξαρτήτως μορφής, με περίοδο  $T_0$ , μπορεί να αναπτυχθεί στην παρακάτω τριγωνομετρική σειρά άπειρων ημιτονικών σημάτων με μη μηδενικές γωνίες φάσης:

$$x(t) = c_0 \cos \theta_0 + c_1 \cos(2\pi f_0 t + \theta_1) + c_2 \cos(2\pi 2f_0 t + \theta_2) + \dots + c_n \cos(2\pi n f_0 t + \theta_n) \text{ (Σχέση 2.3)}$$

όπου  $c_n \geq 0$ , για  $n=0, 1, 2, \dots$ ,  $-\pi < \theta_n \leq \pi$ , για  $n=1, 2, 3, \dots$ , και  $\theta_0=0$  ή  $\pi$ . Η συχνότητα  $f_0=1/T_0$  ονομάζεται θεμελιώδης συχνότητα ή, απλώς, συχνότητα του περιοδικού σήματος και η συχνότητα  $n \cdot f_0$  ονομάζεται  $n$ -στή αρμονική του σήματος. Το  $c_n$  είναι το πλάτος της  $n$ -στής αρμονικής και η  $\theta_n$  είναι η φάση της  $n$ -στής αρμονικής.

Από την παραπάνω σχέση προκύπτει το συμπέρασμα ότι οποιοδήποτε περιοδικό σήμα  $x(t)$  μπορεί να συντεθεί από το άθροισμα άπειρων ημιτονικών όρων με συχνότητες ακέραια πολλαπλάσια της συχνότητας του περιοδικού σήματος, που θα έχουν όμως κατάλληλα πλάτη και κατάλληλες γωνίες φάσης. Οι τιμές των πλατών και των γωνιών φάσης εξαρτώνται από το εκάστοτε σήμα  $x(t)$ .

Αναπτύσσοντας τα παραπάνω, ο Fourier ανέπτυξε μία μέθοδο μέσω της οποίας γίνεται εφικτή η ανάλυση ενός σύνθετου σήματος στους επιμέρους “ημιτονικούς όρους” από

τους οποίους αποτελείται. Κάθε ημιτονικός όρος θα αντιστοιχεί σε ένα σήμα συγκεκριμένης συχνότητας που ενυπάρχει στο σύνθετο σήμα και το σύνολο των όρων αυτών θα δίνει την φασματική εικόνα του σήματος, μιας και κάθε ημιτονικός όρος θα αντιστοιχεί σε μία συγκεκριμένη συχνότητα, σε μία φασματική γραμμή. Οπότε, με τον τρόπο αυτό μπορεί να εκφραστεί ένα σήμα που είναι γνωστό στο πεδίο του χρόνου στο πεδίο των συχνοτήτων. Η μέθοδος αυτή ονομάζεται “Μετασχηματισμός Fourier” και χρησιμοποιείται για τον προσδιορισμό του φάσματος κάποιου σήματος που είναι γνωστή η έκφραση του στο πεδίο του χρόνου. Αν το σήμα είναι συνεχούς χρόνου, τότε ο μετασχηματισμός ονομάζεται απλά “Μετασχηματισμός Fourier” (FT-Fourier Transform) ενώ αν είναι διακριτού χρόνου, ονομάζεται “Μετασχηματισμός Fourier Διακριτού Χρόνου” (DTFT-Discrete Time Fourier Transform).

Ο μετασχηματισμός Fourier συμβολίζεται με  $F\{x(t)\}$  και ο DTFT συμβολίζεται με  $F\{x(n)\}$ , και δίνονται από τις σχέσεις:

$$X(f) = F(x(t)) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j2\pi ft} dt (FT) \quad (\text{Σχέσεις 2.4})$$

$$X(f) = F[x(n)] = \sum_{-\infty}^{\infty} x(n) \cdot e^{-j2\pi fn} (DTFT)$$

Από τους παραπάνω τύπους (Σχέσεις 2.4) φαίνεται αμέσως ότι, ενώ το σήμα  $x(t)$  είναι πραγματική συνάρτηση του χρόνου  $t$ , ο μετασχηματισμός Fourier αυτού  $X(f)$  είναι μιγαδική συνάρτηση της συχνότητας  $f$ , η οποία υπεισέρχεται ως παράμετρος στο ολοκλήρωμα. Ο ορισμός αυτός του μετασχηματισμού Fourier ισχύει και για μιγαδικά σήματα που συχνά χρησιμοποιούνται στις Τηλεπικοινωνίες και στις εφαρμογές που αφορούν την φωνή.

Αν το πραγματικό σήμα  $x(t)$  είναι σήμα τάσης και το εφαρμόσουμε σε μια αντίσταση  $R=1\Omega$ , η στιγμιαία ισχύς που αποδίδεται στην αντίσταση είναι ίση με  $x^2(t)/R=x^2(t)/(1\Omega)$ , η οποία, αν αγνοήσουμε το ζήτημα των μονάδων, είναι ίση με  $x^2(t)$ . Όπως έχουμε πει, αυτήν θα την ονομάζουμε στιγμιαία ισχύ του σήματος  $x(t)$ . Επομένως, σε χρονικό διάστημα  $[t_1, t_2]$  η ενέργεια που το σήμα  $x(t)$  παρέχει στην αντίσταση είναι ίση με  $E=\int_{t_1}^{t_2} x^2(t) dt$ . Έτσι, η ολική ενέργεια που παρέχει το σήμα  $x(t)$  (η ολική ενέργεια του σήματος  $x(t)$ ) είναι ίση με

$$E_{ολ} = \int_{t_1}^{t_2} x^2(t) dt \text{ (Σχέση 2.5)}$$

Τα σήματα για τα οποία το παραπάνω ολοκλήρωμα από  $-\infty$  έως  $+\infty$  έχει πεπερασμένη τιμή ονομάζονται “ενεργειακά σήματα”. Υπάρχει όμως ενδεχόμενο το ολοκλήρωμα αυτό να έχει άπειρη τιμή, οπότε το αντίστοιχο σήμα ονομάζεται “μη ενεργειακό σήμα”. Παράδειγμα ενεργειακού σήματος είναι ο ορθογωνικός παλμός, δηλαδή το σήμα που για κάποιο χρονικό διάστημα  $(T_1, T_2)$  έχει τιμή  $A$  και για τις υπόλοιπες τιμές του χρόνου  $t$  έχει τιμή  $0$ . Προφανώς, για το σήμα αυτό έχουμε  $E_{ολ}=A^2(T_2-T_1)$ . Γενικά, κάθε σήμα που έχει πεπερασμένη διάρκεια και οι τιμές που παίρνει είναι πεπερασμένες (δεν απειρίζονται) είναι ενεργειακό σήμα. Υπάρχουν και ενεργειακά σήματα άπειρης διάρκειας. Παράδειγμα είναι ο εκθετικός παλμός, δηλαδή το σήμα που για  $t<0$  έχει τιμή  $0$  και για  $t\geq 0$  έχει τιμή  $e^{-at}$ , με  $a>0$ . Η ολική ενέργεια αυτού του σήματος είναι ίση με  $1/(2a)$ .

Αντίθετα, κάθε περιοδικό σήμα είναι μη ενεργειακό. Αυτό γιατί σε κάθε περίοδο του έχει πεπερασμένη ενέργεια (ίση με το ολοκλήρωμα του τετραγώνου του σε μια περίοδο). Αυτή όμως, πολλαπλασιαζόμενη επί τις άπειρες περιόδους που διαρκεί το περιοδικό σήμα, δίνει άπειρη ενέργεια.

Αναγκαία και ικανή συνθήκη για να υπάρχει το ολοκλήρωμα της παραπάνω σχέσης (Σχέση 2.4), δηλαδή για να έχει ένα σήμα  $x(t)$  μετασχηματισμό Fourier και αυτός να μην απειρίζεται, είναι το σήμα  $x(t)$  να είναι ενεργειακό. Οπότε μόνο τα ενεργειακά σήματα έχουν μετασχηματισμό Fourier υπό την αυστηρή μαθηματική έννοια. Εμείς όμως, εργαζόμενοι λιγότερο αυστηρά, θα βρούμε μετασχηματισμό Fourier και για ορισμένα μη ενεργειακά σήματα.

Αντικαθιστώντας στη σχέση (2.1) το  $e^{-j2\pi ft}$  από τον τύπο του Euler με το ίσο του

$$e^{-j2\pi ft} = \cos(2\pi ft) - j \cdot \sin(2\pi ft) \quad (\text{Σχέση 2.6})$$

για πραγματικό σήμα  $x(t)$ , έχουμε ότι

$$X(f) = \int_{-\infty}^{\infty} x(t) \cdot \cos(j2\pi ft) dt - j \int_{-\infty}^{\infty} x(t) \cdot \sin(j2\pi ft) dt \quad (\text{Σχέση 2.7})$$

Θέτοντας κάθε ένα ολοκλήρωμα ως έναν παράγοντα  $R(f)$  και  $I(f)$  σύμφωνα με τις σχέσεις:

$$R(f) = \int_{-\infty}^{\infty} x(t) \cdot \cos(j2\pi ft) dt$$

$$I(f) = - \int_{-\infty}^{\infty} x(t) \cdot \sin(j2\pi ft) dt \quad (\text{Σχέσεις 2.8})$$

παίρνουμε ότι  $X(f)=R(f)+jI(f)$ . Το  $R(f)$  είναι το πραγματικό μέρος του μετασχηματισμού Fourier  $X(f)$  του σήματος  $x(t)$  και το  $I(f)$  είναι το φανταστικό μέρος αυτού. Στηριζόμενοι στις γνωστές ταυτότητες  $\cos(-\theta)=\cos\theta$  και  $\sin(-\theta)=-\sin\theta$ , από τις παραπάνω σχέσεις (Σχέσεις 2.8) αποδεικνύουμε αμέσως ότι ισχύουν οι σχέσεις  $R(-f)=R(f)$  και  $I(-f)=-I(f)$ . Επομένως το  $R(f)$  είναι άρτια συνάρτηση της συχνότητας  $f$  και το  $I(f)$  είναι περιττή συνάρτηση της συχνότητας  $f$ .

Θυμόμαστε ότι το γινόμενο δύο άρτιων ή δύο περιττών σημάτων είναι άρτιο σήμα, ότι το γινόμενο ενός άρτιου επί ένα περιττό σήμα είναι περιττό σήμα και ότι για ένα άρτιο σήμα  $y(t)$  έχουμε

$$\int_{-a}^a y(t)dt = 2 \cdot \int_0^a y(t)dt \quad (\text{Σχέση 2.9})$$

ενώ για ένα περιττό σήμα  $y(t)$  έχουμε

$$\int_{-a}^a y(t)dt = 0 \quad (\text{Σχέση 2.10})$$

Αν το σήμα  $x(t)$  είναι άρτιο, το σήμα  $x(t)\cos 2\pi ft$  είναι άρτιο και το σήμα  $x(t)\sin 2\pi ft$  είναι περιττό. Έτσι έχουμε  $I(f)=0$ , δηλ. ότι ο μετασχηματισμός Fourier ενός άρτιου πραγματικού σήματος είναι ίσος με  $X(f)=R(f)$ , ήτοι είναι πραγματική και, φυσικά, άρτια συνάρτηση της συχνότητας  $f$ . Μάλιστα δίνεται και από τον τύπο:

$$X(f) = R(f) = 2 \int_0^{\infty} x(t) \cdot \cos(j2\pi ft) dt \quad (\text{Σχέση 2.11})$$



Επίσης, αν το σήμα  $x(t)$  είναι περιττό, το σήμα  $x(t)\cos 2\pi ft$  είναι περιττό και το σήμα  $x(t)\sin 2\pi ft$  είναι άρτιο. Έτσι έχουμε  $R(f)=0$ , δηλ. ότι ο μετασχηματισμός Fourier ενός περιττού σήματος είναι ίσος με  $X(f)=jI(f)$ , ήτοι είναι φανταστική και, φυσικά, περιττή συνάρτηση της συχνότητας  $f$ . Μάλιστα δίνεται και από τον τύπο:

$$X(f) = j \cdot I(f) = -2j \int_0^{\infty} x(t) \cdot \sin(j2\pi ft) dt \quad (\text{Σχέση 2.12})$$

Για ένα οποιοδήποτε πραγματικό σήμα  $x(t)$ , η γραφική παράσταση του  $|X(f)|$  (μέτρου του  $X(f)$ ) είναι το φάσμα πλάτους του σήματος  $x(t)$  και η γραφική παράσταση του  $\text{Arg}\{X(f)\}$  (ορίσματος του  $X(f)$ ) είναι το φάσμα φάσης του σήματος  $x(t)$ . Όπως ένας μιγαδικός αριθμός  $z$ , που έχει μέτρο  $r$  και όρισμα  $\theta$ , γράφεται και ως  $z=re^{j\theta}$ , έτσι και ο μετασχηματισμός Fourier  $X(f)$  ενός σήματος  $x(t)$  γράφεται και ως  $X(f)=|X(f)|e^{j\text{Arg}\{X(f)\}}$ . Το φάσμα πλάτους  $|X(f)|$  παίρνει πάντα θετικές τιμές.

Αποδεικνύεται ότι από το μετασχηματισμό Fourier  $X(f)$  ενός σήματος, που αποτελεί τη μαθηματική έκφραση του σήματος στο πεδίο της συχνότητας, μπορούμε να βρούμε τη μαθηματική έκφραση  $x(t)$  του σήματος στο πεδίο του χρόνου, μέσω του αντίστροφου μετασχηματισμού Fourier  $F^{-1}\{X(f)\}$ , ο οποίος δίνεται από τη σχέση:

$$x(t) = F^{-1}[X(f)] = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} df \quad (\text{Σχέση 2.13})$$

μέσω της οποίας μπορεί να βρεθεί η έκφραση του σήματος συνεχούς χρόνου στο πεδίο του χρόνου, αν είναι γνωστή η έκφραση του στο πεδίο των συχνοτήτων, ή αλλιώς, αν είναι γνωστός ο μετασχηματισμός Fourier του σήματος συνεχούς χρόνου. Αν είναι γνωστός ο μετασχηματισμός Fourier ενός σήματος διακριτού χρόνου, τότε η σχέση που εκφράζει τον αντίστροφο DTFT είναι η παρακάτω:

$$x(n) = F^{-1}[X(\omega)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{j2\pi f t} d\omega \quad (\text{Σχέση 2.14})$$

Λόγω όμως της μεγάλης πληθώρας της ποικιλίας των σημάτων τα οποία είναι επιθυμητό να επεξεργαστούν και να αναλυθούν, σε πολλές περιπτώσεις και ο FT και ο DTFT κρίνονται ανεπαρκείς για την ανάλυση ενός συγκεκριμένου σήματος. Για τον λόγο αυτό έχουν αναπτυχθεί και άλλες μέθοδοι, όπως είναι ο μετασχηματισμός Fourier βραχέος χρόνου (STFT-Short Time Fourier Transform).

Ο STFT μπορεί να εφαρμοστεί σε πολλές περιπτώσεις για τις οποίες ο FT είναι ανεπαρκής. Χαρακτηριστική τέτοια περίπτωση είναι το σήμα να μην έχει σταθερή συχνότητα αλλά αυτή να μεταβάλλεται με τον χρόνο. Στην περίπτωση αυτή, ο μετασχηματισμός Fourier θα δώσει τις πληροφορίες για την φασματική ανάλυση του συγκεκριμένου σήματος, χωρίς όμως να δίνει πληροφορίες σχετικά με το ποιες φασματικές συνιστώσες υπάρχουν σε κάθε χρονική στιγμή, οπότε η χρήση του FT κρίνεται ανεπαρκής. Για τον λόγο αυτό, το συγκεκριμένο σήμα θα χωριστεί σε επιμέρους τμήματα, αρκετά μικρής χρονικής διάρκειας έτσι ώστε να μπορεί να θεωρηθεί, σε κάθε τέτοιο τμήμα, ότι η συχνότητα διατηρείται σταθερή. Στα τμήματα αυτά μπορεί να εφαρμοστεί ο FT και να εξαχθεί το φάσμα του σήματος. Η σχέση που δίνει τον STFT είναι η εξής:

$$STFT[x(t)] = F_x(t, \omega) = \int_{-\infty}^{\infty} x(u)w'(u-t)e^{-j\omega u} du \text{ (Σχέση 2.15)}$$

όπου το  $w(t)$  είναι μία συνάρτηση που καθορίζει τα τμήματα στα οποία θα χωριστεί το αρχικό σήμα κατά τον STFT που ονομάζεται “παραθυρική συνάρτηση”.

Όμως, λόγω του ότι οι παραπάνω μετασχηματισμοί παράγουν άπειρους όρους για ένα σήμα και τα σήματα που παράγονται από τους αντίστροφους μετασχηματισμούς είναι συνεχής συναρτήσεως, δεν μπορούν να χρησιμοποιηθούν για την επεξεργασία σημάτων από τους ηλεκτρονικούς υπολογιστές γιατί δεν μπορούν να διαχειριστούν συνεχή μεγέθη και άπειρο πλήθος όρων. Έτσι, υπάρχει μία άλλη μέθοδος Fourier με την οποία μπορεί να μεταβεί ένα σήμα από το πεδίο του χρόνου στο πεδίο των συχνοτήτων που αναπτύχθηκε για να μπορεί να χρησιμοποιείται από τα ηλεκτρονικά υπολογιστικά συστήματα, η οποία ονομάζεται “Διακριτός Μετασχηματισμός Fourier” (DFT-Discrete Fourier Transform). Αντίστοιχα, έχει αναπτυχθεί και μέθοδος του STFT που να μπορεί να χρησιμοποιηθεί από τα ηλεκτρονικά υπολογιστικά συστήματα, η οποία είναι η STDFT και μπορεί να χρησιμοποιηθεί στις αντίστοιχες φορές που δεν θα μπορεί να χρησιμοποιηθεί ο DFT.

Για τον DFT ισχύουν όσα αναφέρθηκαν παραπάνω για τον FT και αντιστοιχεί σε δείγματα του FT τα οποία ισαπέχουν κατά συχνότητα. Είναι μία πεπερασμένη ακολουθία, τα στοιχεία της οποίας δίνονται από την παρακάτω σχέση (Σχέση 2.16):

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{-k}(n) \text{ (Σχέση 2.16)}$$

όπου  $W_N^{-k}(n) = W_N^{-kn} = e^{j2\frac{\pi}{N}kn}$  και η ανεξάρτητη συνιστώσα  $k$  είναι ο δείκτης που αναφέρεται στη συχνοτική συνιστώσα με συχνότητα  $\omega_k = 2\frac{\pi k}{N}$ , με  $0 \leq k \leq N-1$ . Αν είναι γνωστός ο DFT ενός σήματος, τότε το σήμα αυτό μπορεί να βρεθεί από την σχέση του αντίστροφου DFT:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)W_N^k(n) \text{ (Σχέση 2.17)}$$

Λόγω της ευρείας χρήσης που βρήκε ο DFT στα ψηφιακά συστήματα και της ανάγκης πραγματοποίησης στο μικρότερο δυνατό χρονικό διάστημα, ερευνήθηκαν τρόποι με τους οποίους θα μπορούσε να μειωθεί ο χρόνος για τους απαιτούμενους υπολογισμούς που απαιτεί ο DFT. Για την εκπλήρωση του στόχου αυτού έχουν αναπτυχθεί αλγόριθμοι οι οποίοι στηρίζονται στον υπολογισμό πολλών μικρότερων DFT του αρχικού σήματος μικρότερου μήκους από  $N$  σημεία. Συνδυάζοντας κατάλληλα όλους αυτούς τους μικρότερους DFT παράγεται ο DFT  $N$  σημείων του σήματος. Ο τρόπος με τον οποίο θα οριστούν οι μικρότεροι DFT καθώς και ο τρόπος με τον οποίο θα συνδυαστούν μεταξύ τους για την παραγωγή του τελικού DFT διαφέρει από αλγόριθμο σε αλγόριθμο και έχουν αναπτυχθεί πολλοί τέτοιοι αλγόριθμοι. Όλοι βασίζονται στην παραπάνω ιδέα, στην εύρεση μικρότερων DFT και στον κατάλληλο συνδυασμό αυτών για την παραγωγή του τελικού DFT, και λόγω του ότι αυτό που επιτυγχάνουν είναι η ταχύτερη εφαρμογή του μετασχηματισμού Fourier χωρίς το αποτέλεσμα να αποτελεί προσέγγιση, καλούνται “Ταχύς Μετασχηματισμός Fourier” (FFT-Fast Fourier Transform).

Με την ανάλυση Fourier, η οποία χρησιμοποιείται για την φασματική ανάλυση των σημάτων συνδέεται και η Cepstrum<sup>7</sup> ανάλυση. Η πρώτη εργασία που αναφερόταν στην ανάλυση αυτή δημοσιεύθηκε το 1963, δύο χρόνια πριν την εργασία του Fourier περί της φασματικής ανάλυσης του, το 1965. Στην εργασία αυτή καθορίζεται ως “το φάσμα ισχύος του λογαρίθμου του φάσματος ισχύος” και αρχικά χρησιμοποιήθηκε από την σεισμολογία για την ανίχνευση πιθανής ηχού (αντίλαλου) στα σεισμικά κύματα. Μέχρι την δημοσίευση της εργασίας που παρουσίαζε τον FT, δεν είχε λάβει πολλή προσοχή η ανάλυση Cepstrum. Μετά από την δημοσίευση, επαναδιατυπώθηκε η θεωρία στην οποία στηρίζεται έτσι ώστε να χρησιμοποιείται η ανάλυση Fourier. Έτσι, το Cepstrum ισχύος ενός σήματος μπορεί να υπολογιστεί από την παρακάτω σχέση:

$$PowerCepstrum = |F^{-1}[\log|F(f(t))|^2]|^2 \text{ (Σχέση 2.18)}$$

όπου  $f(t)$  είναι το σήμα,  $F(f(t))$  ο μετασχηματισμός Fourier του σήματος,  $\log()$  ο δεκαδικός λογάριθμος του μετασχηματισμού Fourier και  $F^{-1}$  ο αντίστροφος μετασχηματισμός Fourier.

Εκτός των παραπάνω, υπάρχουν και άλλοι μετασχηματισμοί που είναι παρόμοιοι με τους παραπάνω μετασχηματισμούς και χρησιμοποιούνται κατά περίπτωση. Ο μετασχηματισμός, εκτός των παραπάνω, που χρησιμοποιείται συχνότερα είναι ο Διακριτός Μετασχηματισμός Συνημιτόνου που είναι παρόμοιος με τον DFT, με την ειδοποιό διαφορά ότι χρησιμοποιεί μόνο τις πραγματικές τιμές, δηλαδή την αντίστοιχη συνάρτηση της συνάρτησης  $R(f)$  που παρουσιάστηκε σε παραπάνω σχέση (Σχέση 2.8).

---

<sup>7</sup>Το Cepstrum είναι αναγραμματισμός του Spectrum, όπου τα τέσσερα πρώτα γράμματα της λέξης έχουν αντιστραφεί.

### 2.3- Αντιμετώπιση του προβλήματος του Θορύβου

Η προσθήκη θορύβου στα ηχητικά σήματα, και ειδικά στα σήματα ομιλίας, που αποτελούν το βασικό θέμα της παρούσας εργασίας, προκαλούν την αλλοίωση του αρχικού σήματος και της πληροφορίας που αυτό περιέχει και την δυσκολότερη ακρόαση του ηχητικού κύματος από τον ακροατή. Και στην περίπτωση που το ηχητικό σήμα είναι σήμα ομιλίας, η δυσκολότερη κατανόηση του ανάγεται σε σοβαρό πρόβλημα, μιας και αν δεν είναι καθαρά αντιληπτό το ηχητικό σήμα ομιλίας χάνεται η πληροφορία που αυτό περιέχει. Για τον λόγο αυτό, η μείωση του θορύβου στα σήματα ομιλίας είναι μείζονος σημασίας από τις αρχές των συστημάτων καταγραφής, μετάδοσης και αναπαραγωγής ηχητικών σημάτων.

Μία πρώτη προσέγγιση για την αντιμετώπιση του προβλήματος του θορύβου είναι η εκμηδένιση του, είτε εντοπίζοντας και “αποσιωπώντας” τις πηγές του είτε εκμηδενίζοντας τις επιπτώσεις του, ηχομονώνοντας τον χώρο στον οποίο γίνεται η καταγραφή του ήχου. Με την προσέγγιση αυτή αντιμετωπίζεται το πρόβλημα του θορύβου κατά την γέννηση του, όπου με καλή σχεδίαση του χώρου καταγραφής μπορεί να γίνει καταγραφή καθαρού ήχου. Όμως, αφενός μεν γιατί δεν είναι εύκολη η παραπάνω διαδικασία και σίγουρα δεν είναι πάντα εφικτή, και αφετέρου δε, γιατί όσο προσεκτική και αν γίνει η σχεδίαση του χώρου ηχογραφήσεων και όσο καλά και αν έχει μελετηθεί το περιβάλλον του έτσι ώστε να εξαλειφθούν όλες οι πηγές θορύβου, στο σήμα θα ενυπάρχει και θόρυβος που θα οφείλεται σε κάποια από τα αίτια που περιγράφηκαν παραπάνω και δεν οφείλονται στον περιβάλλοντα χώρο, η συγκεκριμένη προσέγγιση του προβλήματος δεν δίνει ικανοποιητική λύση. Σε πολλές περιπτώσεις εφαρμόζεται η προσέγγιση αυτή και χρησιμοποιείται σε συνδυασμό με άλλες μεθόδους αντιμετώπισης του θορύβου, αντιμετωπίζοντας έτσι αποδοτικότερα το πρόβλημα του θορύβου, όπως παραδείγματος χάριν με την ηχομόνωση συγκεκριμένων χώρων, όπως οι χώροι ηχογράφησης. Δεν θα γίνει περαιτέρω αναφορά στην προσέγγιση αυτή

στην παρούσα εργασία γιατί το συγκεκριμένο θέμα ξεφεύγει από τα πλαίσια της και το κύριο θέμα της που είναι η αποθορυβοποίηση σημάτων ομιλίας.

Η πρώτη προσέγγιση, λοιπόν, αντιμετωπίζει το πρόβλημα των ενθόρυβων σημάτων προσπαθώντας να εξαλείψει την πηγή του έτσι ώστε να μην προστεθεί στο επιθυμητό σήμα. Η δεύτερη προσέγγιση για την αντιμετώπιση του προβλήματος του θορύβου λαμβάνει υπόψιν ότι ο θόρυβος υπάρχει και προστίθεται στο σήμα και μελετά την αφαίρεση αυτού από το ενθόρυβο σήμα και τον καθαρισμό του με διάφορους τρόπους και μεθόδους. Στην προσέγγιση αυτή ανήκουν οι αλγόριθμοι αποθορυβοποίησης σημάτων ομιλίας. Οι τρόποι και οι μέθοδοι που έχουν αναπτυχθεί έχουν μεγάλη ποικιλία και έχει υλοποιηθεί ένα πλήθος αλγορίθμων οι οποίοι ακολουθούν την προσέγγιση αυτή. Λόγω των ποικίλων τρόπων και μεθόδων που έχουν αναπτυχθεί και συνεχίζουν και αναπτύσσονται αλλά και την υλοποίηση αλγορίθμων οι οποίοι είναι παραλλαγές κάποιων ήδη υπάρχοντων μεθόδων το πλήθος των αλγορίθμων που αντιμετωπίζουν το πρόβλημα του θορύβου με την δεύτερη προσέγγιση. Η προσέγγιση αυτή έχει χωριστεί σε κάποιες επιμέρους κατηγορίες όπου σε κάθε μία θα ανήκουν και οι τρόποι που έχουν τα χαρακτηριστικά που περιγράφονται παρακάτω.

Την πρώτη κατηγορία αποτελούν οι μέθοδοι που στηρίζονται στην αναζήτηση χαρακτηριστικών του σήματος ομιλίας τα οποία παρουσιάζουν υψηλότερο βαθμό αντοχής στο θόρυβο (noiseresistance) και ενίσχυση αυτών. Οι μέθοδοι που ανήκουν στην κατηγορία αυτή μελετούν τη φύση του ήχου και ανακαλύπτει κάποια χαρακτηριστικά ή παραμέτρους οι οποίες παρουσιάζουν αυξημένη ευαισθησία στον θόρυβο και η μελέτη του τρόπου επίδρασης του θορύβου σε αυτές. Μετά από την παραπάνω ανάλυση, τα χαρακτηριστικά παραμέτρων που ανακαλύφθηκε ότι είναι πιο αναισθητες στο θόρυβο χρησιμοποιούνται έτσι ώστε το σήμα συνολικά να γίνει πιο αναισθητο στο θόρυβο. Από μαθηματικής απόψεως, η όλη διαδικασία στηρίζεται στην βελτιστοποίηση των συντελεστών Cepstrum που είναι στην ουσία ο μετασχηματισμός Fourier του Spectrum. Την διαδικασία αυτή υλοποιούν οι

αλγόριθμοι Φασματικής Επέκτασης, οι οποίοι ενισχύουν κυρίως τις φασματικές κορυφές του σήματος και υποβαθμίζουν τις περιοχές χαμηλής συχνότητας.

Κατά την ανάπτυξη και την μελέτη των μεθόδων αύξησης της αντοχής που έχουν τα σήματα ομιλίας στον θόρυβο οδήγησε στην εμφάνιση αρκετών καινοτόμων ιδεών που χρησιμοποιούνται από τον συγκεκριμένο κλάδο. Μία από αυτές είναι η τεμαχιοποίηση ομιλίας. Και μία άλλη είναι η δημιουργία υπολογιστικών μοντέλων ομιλίας για ηχητικά συστήματα, τα οποία παρουσιάζουν υψηλό βαθμό υπολογιστικής ακρίβειας, συμπιέζουν την μη γραμμικότητα που εμφανίζουν τα ενθόρυβα σήματα ομιλίας προβαίνοντας σε μετασχηματισμό της περιβάλλουσας και σταδιακή προσαρμογή της συχνότητας σύμφωνα με μία ανασταλτική-αναδρομική διαδικασία.

Την δεύτερη κατηγορία αποτελούν οι μέθοδοι που αποσκοπούν στην μείωση του ποσοστού της απόκλισης του καθαρού ηχητικού σήματος από το ενθόρυβο, κάτι που μπορεί να επιτευχθεί με πολλούς τρόπους. Και η απομάκρυνση του θορύβου από το σήμα με διάφορες μεθόδους, που είναι και οι μέθοδοι αποθορυβοποίησης που αποτελούν το κύριο αντικείμενο της παρούσας εργασίας. Οι αλγόριθμοι που υλοποιούν τέτοιες μεθόδους είναι πολλοί, και στη συνέχεια της παρούσας εργασίας θα γίνει αναφορά σε αρκετούς από αυτούς. Και ως επί το πλείστον, οι αλγόριθμοι αυτοί χρησιμοποιούν κάποια από τις μεθόδους Fourier για την φασματική ανάλυση του σήματος και την επεξεργασία του στο πεδίο των συχνοτήτων για την αποδοτικότερη αφαίρεση του θορύβου.



## **ΚΕΦΑΛΑΙΟ 3**

### ***ΑΠΟΘΟΡΥΒΟΠΟΙΗΣΗ***

Όπως αναφέρθηκε παραπάνω, οι μέθοδοι αποθορυβοποίησης είναι πολλές και ποικίλες, όπου κάθε μία μία θα έχει τα πλεονεκτήματα και τα μειονεκτήματα της που θα την καθιστούν κατάλληλη ή όχι για μία εφαρμογή. Οι εφαρμογές στις οποίες χρησιμοποιείται η αποθορυβοποίηση είναι πολλές και κάθε μία θα έχει και κάποιες ιδιαίτερες απαιτήσεις, γεγονός στο οποίο, εν μέρη, οφείλεται το μεγάλο πλήθος και η μεγάλη ποικιλία αλγορίθμων αποθορυβοποίησης που αναφέρθηκε παραπάνω. Ως επί το πλείστον, οι αλγόριθμοι αυτοί χρησιμοποιούν την φασματική ανάλυση που αναφέρθηκε στο προηγούμενο κεφάλαιο για την αποθορυβοποίηση του σήματος ομιλίας και πολλοί από αυτούς στηρίζονται στις ίδιες αρχές. Οπότε θα υπάρχουν κάποιες γενικές κατηγορίες αλγορίθμων αποθορυβοποίησης οι οποίες θα περιέχουν διάφορους αλγόριθμους οι οποίοι θα στηρίζονται στις ίδιες αρχές και θα διαφέρουν σε κάποιες λεπτομέρειες. Για τον λόγο αυτό, στο παρόν κεφάλαιο θα γίνει η περιγραφή των αλγορίθμων αποθορυβοποίησης με την περιγραφή των κύριων αυτών κατηγοριών και των αρχών στις οποίες στηρίζεται η λειτουργία τους και την αναφορά των σημαντικότερων εκπροσώπων κάθε κατηγορίας, με τα ιδιαίτερα χαρακτηριστικά του

καθενός. Τέλος, στο τέλος του κεφαλαίου αναλύεται ένας από τους αλγόριθμους αποθορυβοποίησης για καλύτερη εποπτεία όσων αναφέρθηκαν στην παρούσα εργασία.

### **3.1- Γραμμικό Φιλτράρισμα Ψηφιακών Σημάτων**

Όπως αναφέρθηκε παραπάνω, η καταγραφή και επεξεργασία του ήχου πραγματοποιείται πλέον με χρήση ψηφιακών υπολογιστικών συστημάτων. Για τον λόγο αυτό ο ήχος ψηφιοποιείται με κάποια κατάλληλη διαδικασία, όπως το σύστημα PCM που περιγράφηκε στο πρώτο κεφάλαιο της παρούσας εργασίας. Ο θόρυβος κβαντισμού που εμφανίζεται, χρησιμοποιώντας συχνότητα δειγματοληψίας ίση με 8kHz και 16 μπίτ ανάλυση δείγματος, είναι αμελητέος συγκριτικά με τον θόρυβο που παράγεται από τις άλλες πηγές, πόσο μάλλον αν ψηφιοποιηθεί με διαφορετικά χαρακτηριστικά, όπως παραδείγματος χάριν με 32 μπίτ ανάλυση δείγματος, που στην περίπτωση αυτή θα είναι ακόμα μικρότερος ο θόρυβος κβάντισης.

Η ψηφιοποίηση του ήχου είναι η πρώτη διαδικασία την οποία θα υποστεί ένα ηχητικό σήμα. Στην συνέχεια μπορούν να εφαρμοστούν διάφορες τεχνικές απομάκρυνσης του θορύβου από το σήμα αυτό. Η πρώτη και βασικότερη είναι η διέλευση του ψηφιακού σήματος από ένα γραμμικό φίλτρο και η όσο το δυνατόν μέγιστη αφαίρεση των ανεπιθύμητων συνιστωσών του θορύβου από το σήμα ομιλίας. Ένα γραμμικό φίλτρο θα είναι ένα κύκλωμα ή ένα σύστημα μέσω του οποίου διέρχεται ένα σήμα και του οποίου η απόκριση, δηλαδή το πως θα επηρεαστεί το σήμα που διέρχεται από μέσα του, εξαρτάται από την συχνότητα του σήματος. Έτσι, μπορούν να κατασκευαστούν κατάλληλα φίλτρα που να εξασθενούν πολύ τα σήματα συγκεκριμένων συχνοτήτων που θεωρούνται ως θόρυβος και να

διέρχονται μέσω αυτού ανεπεηρέαστα τα επιθυμητά σήματα ομιλίας. Με τον τρόπο αυτό γίνεται μία βασική αποθορυβοποίηση των σημάτων.

Τα γραμμικά φίλτρα χωρίζονται σε κάποιες κατηγορίες, ανάλογα με τα χαρακτηριστικά του καθενός. Το φίλτρο που θα χρησιμοποιηθεί σε κάθε περίπτωση θα εξαρτάται από τα χαρακτηριστικά του ήχου που είναι επιθυμητό να αποθορυβοποιηθεί. Οι κύριες κατηγορίες είναι τα φίλτρα πεπερασμένης κρουστικής απόκρισης FIR (Finite Impulse Response) και τα φίλτρα άπειρης κρουστικής απόκρισης IIR (Infinite Impulse Response). Και αν ένα σήμα διέλθει μέσω ενός FIR φίλτρου με  $h(t)$  να είναι η γνωστή κρουστική του απόκριση, τότε η έξοδος του φίλτρου θα δίνεται από την συνέλιξη του σήματος εισόδου του με την κρουστική απόκριση αυτού, όπως φαίνεται και από την παρακάτω σχέση:

$$y[t] = x[t] * h[t] \text{ (Σχέση 3.1)}$$

Η παραπάνω σχέση εκφράζει την έξοδο του φίλτρου συναρτήσει της εισόδου του και της κρουστικής του απόκρισης, στο πεδίο του χρόνου. Το φιλτράρισμα όμως μπορεί να εκφραστεί και στο πεδίο των συχνοτήτων, παίρνοντας τον FT της παραπάνω σχέσης, με τον FT του  $y[t]$  να είναι το  $Y[f]$ , τον FT του  $x[t]$  να είναι το  $X[f]$ , τον FT της κρουστικής απόκρισης  $h[t]$  να είναι το  $H[f]$ , που ονομάζεται “Απόκριση Συχνότητας” και τέλος, τον FT την πράξη της συνέλιξης να είναι η πράξη του πολλαπλασιασμού. Έτσι, η σχέση που εκφράζει την έξοδο του φίλτρου στο πεδίο των συχνοτήτων είναι η παρακάτω (Σχέση 3.2).

$$Y[f] = X[f] \cdot H[f] \text{ (Σχέση 3.2)}$$

Τα φίλτρα έχουν κάποια συγκεκριμένα χαρακτηριστικά, όπως προαναφέρθηκε, με το σημαντικότερο αυτών να είναι η μορφή της απόκρισης συχνότητας, δηλαδή της συμπεριφοράς του φίλτρου συναρτήσει της συχνότητας του σήματος. Αν έχει σχεδιαστεί έτσι ώστε να “κόβει”, δηλαδή να εξασθενίζει σε μεγάλο βαθμό, τα σήματα που έχουν συχνότητα μεγαλύτερη από μία συγκεκριμένη τιμή, τότε το φίλτρο χαρακτηρίζεται ως “Χαμηλοπερατό” (LowPass). Αν κόβει τα σήματα που έχουν συχνότητα μέχρι ένα ανώτερο όριο, τότε χαρακτηρίζεται ως “Υψηλοπερατό” (High Pass). Αν κόβει τα σήματα των οποίων η συχνότητα είναι μέσα σε μία συγκεκριμένη ζώνη συχνοτήτων, τότε το φίλτρο χαρακτηρίζεται ως “Ζωνοφρακτικό” (BandStop) ενώ αν κόβει τα σήματα των οποίων η συχνότητα είναι εκτός μίας συγκεκριμένης ζώνης συχνοτήτων, τότε το φίλτρο χαρακτηρίζεται ως “Ζωνοπερατό” (BandPass).

Εφαρμόζοντας ένα από τα παραπάνω είδη φίλτρων σε ένα σήμα ομιλίας, ένα μέρος του θορύβου απομακρύνεται και το σήμα αποθορυβοποιείται σε έναν βαθμό. Η χρήση ενός απλού φίλτρου επιτρέπει τον καθαρισμό ενός ηχητικού σήματος το οποίο έχει στάσιμο θόρυβο (stationary) με στενό συχνοτικό εύρος. Όμως, στις περισσότερες εφαρμογές ο θόρυβος σε ένα σήμα ομιλίας δεν είναι στάσιμο και δεν έχει στενό συχνοτικό εύρος. Για τον λόγο αυτό, η χρήση μόνο ενός απλού φίλτρου δεν δίνει, στις περισσότερες των περιπτώσεων, πολύ ικανοποιητικά αποτελέσματα και έχουν αναπτυχθεί πολλές άλλες τεχνικές αποθορυβοποίησης για τις διάφορες εφαρμογές.

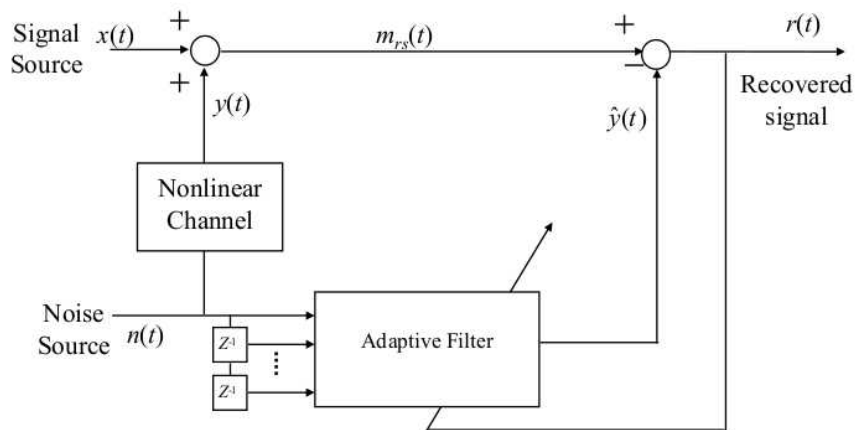
Όπως αναφέρθηκε και παραπάνω, έχουν υλοποιηθεί διάφορα είδη φίλτρων έτσι ώστε να μπορούν να χρησιμοποιηθούν φίλτρα σε πολλές και ποικίλες εφαρμογές. Ένα από τα είδη φίλτρων που έχουν υλοποιηθεί και χρησιμοποιούνται στις εφαρμογές αποθορυβοποίησης είναι τα φίλτρα Wiener. Τα φίλτρα αυτά είναι ψηφιακά φίλτρα που άρχισαν να χρησιμοποιούνται το 1942 και θεωρούνται βέλτιστα για την συγκεκριμένη εφαρμογή και στηρίζονται στην ελαχιστοποίηση του τετραγωνικού σφάλματος του σήματος από ένα σήμα

αναφοράς. Μία παραλλαγή αυτού εμφανίζεται στον αλγόριθμο που παρουσίασαν ο Satorius, E.H. etal<sup>8</sup> στην εργασία τους “Noise cancellation via linear prediction filtering” (Ακύρωση θορύβου μέσω φιλτραρίσματος γραμμικής πρόβλεψης) στην οποία δείχνουν ότι δεν είναι απαραίτητη η ύπαρξη του σήματος αναφοράς, δηλαδή δεν είναι απαραίτητο να είναι γνωστό το φασματικό περιεχόμενο του θορύβου για να επιτευχθεί η αποθορυβοποίηση αλλά μπορεί αυτό να προβλεφθεί, χρησιμοποιώντας ένα Wiener φίλτρο γραμμικής πρόβλεψης LPF (Linear Prediction Filter), αν το συχνοτικό εύρος του σήματος είναι σημαντικά μικρότερο από του θορύβου.

Στην εργασία “Adaptive Noise Cancellation with Computational-Intelligence-based Approach” (Προσαρμοστική ακύρωση θορύβου με προσέγγιση βασισμένη στην υπολογιστική ευφυΐα) από Chunshien Li et al παρουσιάζεται μία μέθοδος προσαρμοστικής ακύρωσης θορύβου. Η προσαρμοστική ακύρωση θορύβου στηρίζεται σε φίλτρα τα οποία μπορούν να μεταβάλλουν τα χαρακτηριστικά τους ανάλογα με τις ανάγκες της εφαρμογής και η απομάκρυνση του θορύβου στηρίζεται στην εκτίμηση του σε πραγματικό χρόνο έτσι ώστε να καθοριστούν τα χαρακτηριστικά του φίλτρου και να απομακρυνθεί ο θόρυβος. Τα φίλτρα αυτά χαρακτηρίζονται ως “Προσαρμοστικά Φίλτρα” (Adaptive Filters) και η λογική σύμφωνα με την οποία λειτουργούν παρατίθεται σε μπλοκ διάγραμμα στην παρακάτω εικόνα (Εικόνα 3.1).

---

<sup>8</sup>Etal: “και συνεργάτες”

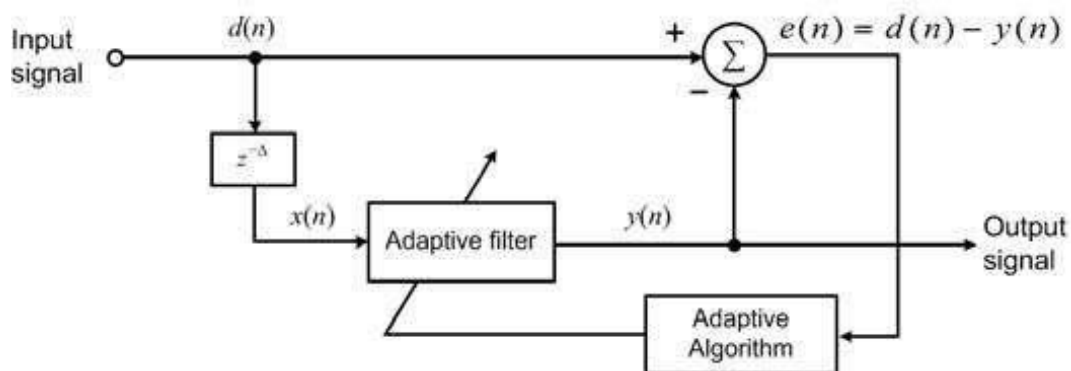


Εικόνα 3.1: Μπλοκ διάγραμμα αρχής λειτουργίας προσαρμοστικού φίλτρου [10]

Το σήμα  $x(t)$  είναι το επιθυμητό σήμα ομιλίας που παράγεται από την πηγή σήματος (Signal Source), στο οποίο προστίθεται ένα σήμα θορύβου  $y(t)$  από το μη γραμμικό κανάλι (Non-linear Channel) μέσω του οποίου διέρχεται το  $x(t)$  ενώ το σήμα  $m_{rs}(t)$  είναι το ενθόρυβο σήμα που είναι επιθυμητό να αποθορυβοποιηθεί. Το προσαρμοστικό φίλτρο (Adaptive filter) θα κάνει μία εκτίμηση του θορύβου και θα αφαιρέσει την εκτίμηση αυτή από το σήμα  $m_{rs}(t)$ , οπότε το σήμα  $r(t)$  θα είναι το αποθορυβοποιημένο σήμα που ανακτήθηκε (Recovered Signal). Το πρώτο φίλτρο του είδους αυτού παρουσιάστηκε από τους Widrow et al. Το 1975. Τα συστήματα που χρησιμοποιούν το είδος αυτό των φίλτρων, επειδή απαιτείται επεξεργασία δεδομένων πραγματικού χρόνου, έχουν αυξημένες υπολογιστικές ανάγκες με αποτέλεσμα η απόκριση των συστημάτων αυτών να είναι αργή. Για την αντιμετώπιση του προβλήματος αυτού αναπτύχθηκαν πολλές τεχνικές και μέθοδοι που στηρίζονται στην αρχή των προσαρμοστικών φίλτρων, όπως είναι τα φίλτρα Wiener, ο αλγόριθμος αναδρομικών ελαχίστων τετραγώνων (Recursive Least Squares-RLS), ο αλγόριθμος ελαχίστων μέσων τετραγώνων (Least Mean Squares-LMS) και τα φίλτρα Kalman.

Μία εναλλακτική μέθοδος εφαρμογής του προσαρμοστικού φιλτραρίσματος είναι η μέθοδος προσαρμοστικής βελτίωσης γραμμής (Adaptive Line Enhancer-ALE) που

στηρίζεται στον αλγόριθμο LMS. Ένα σύστημα ALE θα είναι ένα σύστημα το οποίο θα έχει την ικανότητα να αυτο-παραμετροποιείται κατάλληλα, ανάλογα με την κατάσταση του συστήματος και του περιβάλλοντος και τα χαρακτηριστικά του θορύβου που παράγεται σε πραγματικό χρόνο και με κατάλληλες διεργασίες να βελτιώσει το σήμα εξόδου του συστήματος. Μίας τέτοιας διαδικασίας το μπλοκ διάγραμμα παρατίθεται στην παρακάτω εικόνα (Εικόνα 3.2). Το σήμα  $d(n)$  είναι το ενθόρυβο σήμα που εισέρχεται στο ALE σύστημα (Input Signal). Από το σήμα αυτό, μέσω ενός προσαρμοστικού φίλτρου και ενός συστήματος που υλοποιεί έναν προσαρμοστικό αλγόριθμο επιτυγχάνεται μία εκτίμηση για το σήμα θορύβου το οποίο στη συνέχεια αφαιρείται από το αρχικό σήμα  $d(n)$  παράγοντας το καθαρισμένο σήμα εξόδου  $y(n)$ .



Εικόνα 3.2: Μπλοκ διάγραμμα αρχής λειτουργίας ALE συστήματος [9]

Στον αλγόριθμο των Chunshien Lietal τον ρόλο του προσαρμοστικού φίλτρου τον αναλαμβάνει ένα σύστημα τεχνητής ευφυίας “neuro-fuzzysystem (NFS)” το οποίο κάνει την εκτίμηση του χρησιμοποιώντας την τεχνική της τυχαίας βελτιστοποίησης (Random Optimization-RO) και της εκτίμησης ελαχίστων τετραγώνων (Least Square Estimation-LSE) δίνοντας πολύ καλά αποτελέσματα στην αποθορυβοποίηση σημάτων ομιλίας.

### 3.2- Αποθορυβοποίηση στο πεδίο του Χρόνου

Όταν ο θόρυβος έχει τις ίδιες φασματικές συνιστώσες με το επιθυμητό σήμα, τότε η αφαίρεση του θα πρέπει να γίνει στο πεδίο του χρόνου, με μεθόδους που στηρίζονται στις στατιστικές ιδιότητες του. Ο πιο απλός τρόπος είναι με χρήση ενός FIR φίλτρου στο πεδίο του χρόνου. Η μαθηματική αναπαράσταση του επιθυμητού σήματος θα δίνεται από την παρακάτω σχέση (Σχέση 3.3):

$$x(n) = [x(n), x(n - 1), x(n - 2), x(n - 3), \dots, x(n - N + 1)] \text{ (Σχέση 3.3)}$$

όπου είναι ο χρόνος και  $N$  το πλήθος των δειγμάτων του  $x(n)$ . Το σήμα θορύβου  $d(m)$  θα προστεθεί στο σήμα αυτό και θα παραχθεί το ενθόρυβο σήμα  $y(m)$ :

$$y(m) = x(m) + d(m) \text{ (Σχέση 3.4)}$$

όπου  $m$  είναι οι δείκτες των δειγμάτων κάθε χρονικής στιγμής. Η παραπάνω σχέση (Σχέση 3.4) θα ισχύει για τις περιπτώσεις στις οποίες ο θόρυβος είναι σταθερός, οπότε το σήμα του θορύβου και το σήμα ομιλίας θα είναι ανεξάρτητα μεταξύ τους και ο πίνακας αυτοσυσχέτισης του ενθόρυβου σήματος θα δίνεται από την παρακάτω σχέση (Σχέση 3.5):

$$R_{yy} = \langle y(m)y_{xd}(m) \rangle = R_x(m) + R_d(m) \text{ (Σχέση 3.5)}$$



όπου  $R_x(m)$  είναι ο πίνακας αυτοσυσχέτισης του σήματος ομιλίας και  $R_d(m)$  ο πίνακας αυτοσυσχέτισης του θορύβου. Όταν είναι γνωστός ο πίνακας  $R_d(m)$ , τότε μπορεί να βρεθεί το “καθαρό” σήμα από την σχέση:

$$R_x(m) = R_y(m) - R_d(m) \text{ (Σχέση 3.6)}$$

Αν ο θόρυβος είναι στάσιμος, τότε ο πίνακας αυτοσυσχέτισης του θα είναι σταθερός και θα μπορεί να χρησιμοποιηθεί η παραπάνω σχέση (Σχέση 3.6) για την εύρεση του καθαρού σήματος. Αυτό επιτυγχάνεται, στην πραγματικότητα, με την υλοποίηση ενός ψηφιακού φίλτρου σύμφωνα με τον γνωστό πίνακα αυτοσυσχέτισης. Τα παραπάνω χρίσκουν εφαρμογή στην εργασία “Noisecancellation in time and frequency domain using neural networks” (Ακύρωση θορύβου στο πεδίο του χρόνου και της συχνότητας χρησιμοποιώντας νευρωνικά δίκτυα) των Rahman, M.T. et al στο οποίο ταυτοποιείται ο θόρυβος στο πεδίο του χρόνου μέσω ενός τεχνητού νευρωνικού δικτύου το οποίο, στη συνέχεια, προσαρμόζει ή “εκπαιδεύει” το σύστημα στην βέλτιστη αφαίρεση του θορύβου.

Μία άλλη μέθοδος αποθορυβοποίησης σημάτων ομιλίας στο πεδίο του χρόνου παρουσιάστηκε από τον Lowerre, G.M. στην εργασία “Timedomain use of the EM algorithm in noise cancellation” (Χρήση του EM αλγορίθμου στο πεδίο του χρόνου για την ακύρωση θορύβου). Η μέθοδος που παρουσιάζει αναπτύχθηκε για την ακύρωση του θορύβου σε εγγραφή που γίνεται με χρήση πολλαπλών μικροφώνων και χρησιμοποιεί τον αλγόριθμο μέγιστης εκτίμησης (Estimate Maximize-EM) στο πεδίο του χρόνου. Η χρήση του αλγορίθμου αυτού γίνεται συνδυαστικά με ένα στατιστικό μοντέλο που αναπτύχθηκε

### 3.3- Αποθορυβοποίηση στο πεδίο των Συχνοτήτων

Ο βασικότερος τρόπος με τον οποίο γίνεται η αφαίρεση του θορύβου από ένα ηχητικό σήμα είναι τη διέλευση του μέσω ενός κατάλληλου φίλτρου και η διαδικασία να γίνει στο πεδίο των συχνοτήτων. Για να επιτευχθεί αυτό, αρχικά θα πρέπει το σήμα να πλαισιοποιηθεί και με χρήση ενός από τους μετασχηματισμούς να μετασχηματιστεί από το πεδίο του χρόνου στο πεδίο των συχνοτήτων. Το φίλτρο θα μειώσει την ισχύ των φασματικών συχνοτήτων που έχουν αναγνωριστεί ως “θόρυβος”, αφαιρώντας τον από το σήμα. Στη συνέχεια, με χρήση του αντίστροφου μετασχηματισμού που χρησιμοποιήθηκε, μετασχηματίζεται το αποθορυβοποιημένο σήμα από το πεδίο των συχνοτήτων στο πεδίο του χρόνου και γίνεται η ανακατασκευή του ηχητικού καθαρισμένου σήματος.

Η διαδικασία που περιγράφηκε παραπάνω μπορεί να εκφραστεί με την παρακάτω σχέση (Σχέση 3.7):

$$\hat{x}(m) = U^*(m) \cdot F(m) \cdot U(m) \cdot y(m) \text{ (Σχέση 3.7)}$$

όπου το  $\hat{x}(m)$  είναι το αποθορυβοποιημένο σήμα στο πεδίο του χρόνου,  $U(m)$  είναι ο πίνακας του μετασχηματισμού που χρησιμοποιήθηκε,  $F(m)$  ο πίνακας του κέρδους του φίλτρου,  $U^*(m)$  ο πίνακας του αντίστροφου μετασχηματισμού και  $y(m)$  είναι το ενθόρυβο σήμα. Για τους παραπάνω πίνακες, θα ισχύει ότι  $U \cdot U^* = I$ , (όπου  $I$  είναι ο μοναδιαίος πίνακας) μιας και η διαδικασία που εφαρμόζει ο κάθε ένας θα είναι η αντίστροφη του άλλου οπότε και οι δύο πίνακες μεταξύ τους θα είναι αντίστροφοι.

Η παραπάνω σχέση (Σχέση 3.7) περιγράφει συνολικά την διαδικασία της φασματικής αφαίρεσης του θορύβου. Αρχικά, από το σήμα εισόδου  $y(m)$ , με εφαρμογή ενός μετασχηματισμού παράγεται το φάσμα του σήματος εισόδου σύμφωνα με την σχέση:

$$Y(m) = U(m) \cdot y(m) \text{ (Σχέση 3.8)}$$

και το φάσμα εξόδου του συστήματος θα είναι το φάσμα εισόδου του αφού διέλθει μέσω ενός κατάλληλου φίλτρου, οπότε μαθηματικοποιημένα η διαδικασία αυτή θα δίνεται από την παρακάτω σχέση:

$$\hat{X}(m) = F(m)Y(m) \text{ (Σχέση 3.9)}$$

και το σήμα εξόδου στο πεδίο του χρόνου θα μπορεί να βρεθεί με εφαρμογή του αντίστροφου μετασχηματισμού:

$$\hat{x}(m) = U * (m)\hat{X}(m) \text{ (Σχέση 3.10)}$$

όπου σε όλες τις παραπάνω σχέσεις, τα σήματα που θεωρούνται ότι περιγράφονται στο πεδίο του χρόνου γράφονται με πεζό γράμμα ενώ τα σήματα που θεωρούνται ότι περιγράφονται στο πεδίο των συχνοτήτων γράφονται με κεφαλαίο γράμμα.

Σύμφωνα με τα παραπάνω λοιπόν, χρησιμοποιώντας την διαδικασία αυτή μπορεί να επιτευχθεί αποθορυβοποίηση στο πεδίο των συχνοτήτων ενός σήματος. Το βασικότερο αυτής

της διαδικασίας είναι ο μετασχηματισμός που θα χρησιμοποιηθεί. Συνήθως, για την αποθορυβοποίηση των σημάτων ομιλίας που τα επεξεργαζόμαστε, κατά βάση, στην ψηφιακή τους μορφή, χρησιμοποιείται ο DFT και ο DCT.

### **3.4-Καταστολή θορύβου σε σήματα ομιλίας με χρήση φίλτρου εξομάλυνσης, παραμετροποιημένο με αντιληπτικά κριτήρια (Favrot-Faller)**

Μεταξύ όλων των τεχνικών καταστολής σταθερού θορύβου για την ενίσχυση της ομιλίας από ένα μικρόφωνο, η τεχνική μετασχηματισμού φάσματος ή φασματικής αφαίρεσης παραμένει η πιο συνήθης γιατί είναι απλή και εύκολα υλοποιήσιμη. Η τεχνική αυτή βασίζεται στη μέτρηση ενός μικροφώνου του σήματος θορύβου  $y[n]$ , το οποίο θα είναι ίσο με

$$y[n] = x[n] + v[n] \text{ (Σχέση 3.11)}$$

όπου το  $x[n]$  είναι το επιθυμητό σήμα ομιλίας και το  $v[n]$  αντιπροσωπεύει το σήμα του θορύβου, με το  $v[n]$  να θεωρείται σταθερή διαδικασία, οπότε ο θόρυβος είναι στάσιμος (stationary).

Ο μετασχηματισμός φάσματος εφαρμόζεται στο πεδίο των συχνοτήτων εφαρμόζοντας έναν διακριτό μετασχηματισμό Fourier μικρής διάρκειας (discrete short-time Fourier transform (STFT)). Τα σήματα ομιλίας είναι μη σταθερά αλλά για πλαίσια μικρού χρόνου (μεταξύ 10 και 40 ms) μπορούν να θεωρούνται σταθερά. Εφαρμόζοντας STFT και στα δύο μέρη της σχέσης 3.11 έχουμε:

$$Y[k, m] = X[k, m] + V[k, m] \text{ (Σχέση 3.12)}$$

όπου το  $k$  είναι ο αριθμός του χρονικού πλαισίου και  $m$  ο δείκτης του συχνοτικού. Μιας και η ανθρώπινη διακριτική ικανότητα είναι σχετικά αναισθητή στην αλλοίωση της φάσης, η εκτιμώμενη φάση επιλέγεται να είναι ίση με την φάση του θορύβου και η εκτίμηση του  $X[k, m]$  είναι επαρκής για να ανακτηθεί το επιθυμητό σήμα ομιλίας.

Εντός ενός πλαισίου, και η ομιλία  $x$  και ο θόρυβος  $v$  μοντελοποιούνται ως ασυσχέτιστες σταθερές διαδικασίες. Έτσι, το φάσμα ισχύος του  $y$  μπορεί να προσεγγιστεί ως εξής:

$$|Y[k, m]|^2 \approx |X[k, m]|^2 + |V[k, m]|^2 \text{ (Σχέση 3.13)}$$

Τότε, το φάσμα ισχύος  $|X[k, m]|^2$  του επιθυμητού σήματος ομιλίας μπορεί να ανακτηθεί αφαιρώντας μία εκτίμηση του φάσματος ισχύος του θορύβου  $|V[k, m]|^2$  από το  $|Y[k, m]|^2$  και το φάσμα ισχύος να υπολογιστεί ως εξής:

$$|(\tilde{X})[k, m]|^2 = |Y[k, m]|^2 - |(\tilde{V})[k, m]|^2 \text{ (Σχέση 3.14)}$$

Αντίστοιχα με το φιλτράρισμα, το φασματικό μέγεθος του σήματος ομιλίας ανακτάται από τη σχέση:

$$|\tilde{X}[k, m]| = G[k, m]|Y[k, m]| \text{ (Σχέση 3.15)}$$

όπου το  $G[k, m]$  συμβολίζεται ως φίλτρο κέρδους:

$$G[k, m] = \left( \frac{|Y[k, m]|^2 - |\tilde{V}[k, m]|^2}{|Y[k, m]|^2} \right)^{\frac{1}{2}} (\Sigma \chi \acute{\epsilon} \sigma \eta 3.16)$$

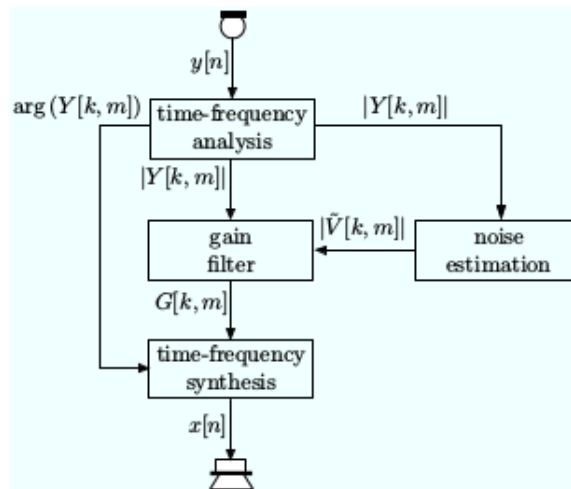
Μία πιο γενική μορφή, που ονομάζεται παραμετρικό φίλτρο κέρδους, μπορεί να προέλθει εισάγοντας τις παραμέτρους  $\alpha$  και  $\beta$ :

$$G_P[k, m] = \left( \frac{|Y[k, m]|^\alpha - \beta |\tilde{V}[k, m]|^\alpha}{|Y[k, m]|^\alpha} \right)^{\frac{1}{\alpha}} (\Sigma \chi \acute{\epsilon} \sigma \eta 3.17)$$

όπου το  $\alpha$  αναπαριστά έναν εκθέτη και το  $\beta$  μία παράμετρο ελέγχου της ποσότητας του θορύβου που πρέπει να κατασταλεί στην περίπτωση χαμηλής (ή υψηλής) εκτίμησης, που μπορεί να επιλεγεί σύμφωνα με το λόγο σήματος προς θόρυβο (SNR) και του δείκτη συχνότητας  $m$ . Μία από τις πρώτες υλοποιήσεις του μετασχηματισμού φάσματος προτάθηκε από τον Boll στην εργασία του “Suppression of acoustic noise in speech using spectral subtraction” (Καταστολή του ακουστικού θορύβου στην ομιλία με φασματική αφαίρεση), υλοποιεί μία απλή αφαίρεση μεγέθους ( $\alpha=1$ ,  $\beta=1$ ). Μία καλύτερη παραλλαγή είναι η εκθετική αφαίρεση ( $\alpha=2$ ), όπου το κέρδος φίλτρου που ορίζεται από την σχέση 3.17 έχει περιορισμό ώστε να μην ποτέ μικρότερο του  $-L$  dB:

$$G_{PL}[k, m] = \max(10^{\frac{-L}{20}}, G_P) (\Sigma \chi \acute{\epsilon} \sigma \eta 3.18)$$

Μετά το φιλτράρισμα, το εκτιμώμενο φάσμα ομιλίας μετατρέπεται πίσω στο πεδίο του χρόνου με την μέθοδο “πρόσθεση επικάλυψης σήματος” (overlap add signal reconstruction method). Το σχήμα της παρακάτω εικόνας (Εικόνα 3.3) δείχνει το διάγραμμα βαθμίδων που υλοποιούν τον μετασχηματισμό φάσματος.



Εικόνα 3.3: Διάγραμμα βαθμίδων που υλοποιούν τον μετασχηματισμό φάσματος [6]

Ένα κρίσιμο κομμάτι του μετασχηματισμού φάσματος είναι η σταθερή εκτίμηση θορύβου. Εφαρμόζεται με το να ενημερώνεται το εκτιμώμενο μικρού χρόνου φάσμα θορύβου  $|V \sim [k, m]|$  με τον χρόνο. Τις περισσότερες φορές, η εκτίμηση εκτελείται σε δύο διαδοχικά στάδια:

1. Θεωρούμε ότι η ανάλυση των πρώτων πλαισίων περιέχουν μόνο θόρυβο. Έτσι υπολογίζεται στα τυφλά μία εκτίμηση θορύβου εντός αυτών των πλαισίων.
2. Σε κάθε πλαίσιο, διακρίνουμε αν αυτό περιέχει μόνο θόρυβο (χρησιμοποιώντας έναν ανιχνευτή ομιλίας, συμβολίζεται με VAD (Voice Activity Detection) και μόνο σε αυτήν την περίπτωση ενημερώνουμε την εκτίμηση του θορύβου.

Το πρώτο στάδιο είναι απλό στην υλοποίηση και οδηγεί σε μία αρχική εκτίμηση θορύβου. Το δεύτερο στάδιο συνεχώς προσαρμόζεται στο σήμα των πλαισίων και είναι πιο περίπλοκο στην υλοποίηση.

Ο μετασχηματισμός φάσματος είναι αποδοτικός αναφορικά με την αφαίρεση πρόσθετου σταθερού θορύβου. Η τεχνική προσφέρει ένα καλό δίλημμα μεταξύ απλότητας υλοποίησης και αποδοτικότητας. Παρόλα αυτά το μειονέκτημα της μεθόδου είναι το γνωστό τεχνούργημα που ονομάζεται “μουσικός θόρυβος” (musical noise). Οι μεγάλες αποκλίσεις

μεταξύ του φάσματος του θορύβου και των σφαλμάτων εκτίμησης έχουν σαν αποτέλεσμα διαφορές μεταξύ του πραγματικού θορύβου και της εκτίμησης αυτού. Αυτό έχει σαν αποτέλεσμα κάποιος θόρυβος να παραμένει στο τελικό φάσμα σε περίπτωση που η εκτίμηση ήταν υπό (υποτιμήθηκε). Από την άλλη, όταν η εκτίμηση του θορύβου είναι υπέρ (υπερεκτίμηση), αρνητικές τιμές του φάσματος ισχύος μπορούν να προκύψουν οι οποίες κόβονται. Ο υπολειπόμενος θόρυβος μεταξύ των κομμένων ποσοτήτων του φάσματος δημιουργεί τον μουσικό θόρυβο. Οι ποικίλες υλοποιήσεις της μεθόδου επικεντρώνονται στην μείωση του τεχνουργήματος αυτού.

Όπως διατυπώθηκε παραπάνω στην σχέση 3.16, η αφαίρεση του φάσματος ισχύος αντιστοιχεί στην εκτίμηση ελαχίστων τετραγώνων του επιθυμητού σήματος ομιλίας. Παρόλα αυτά, καμία υπόθεση δεν έγινε για την κατανομή των τιμών του φάσματος ισχύος της ομιλίας και του θορύβου. Περαιτέρω υλοποιήσεις του μετασχηματισμού φάσματος εισάγουν μία μοντελοποίηση των στατιστικών κατανομών των φασματικών τιμών της ομιλίας και του θορύβου. Αυτές ορίζουν κυρίως δύο στατιστικές οντότητες:

- Το SNR ορίζεται ως μία τοπική εκτίμηση του SNR στο παρόν πλαίσιο, εκ των υστέρων.
- Το SNR ορίζεται ως μία τοπική εκτίμηση του SNR στο παρόν πλαίσιο, εκ των προτέρων, λαμβάνοντας υπόψιν αρκετά προηγούμενα πλαίσια.

Ένα καινούριο κέρδος φίλτρου υπολογίζεται ως συνάρτηση των “εκ των προτέρων” και “εκ των υστέρων” SNR. Παρόλο που η υπολογιστική πολυπλοκότητα μεγαλώνει, τέτοιες μέθοδοι οδηγούν σε καλά αποτελέσματα όσον αφορά την εξάλειψη του μουσικού θορύβου. Έχει δειχθεί ότι η μέθοδος καταστολής θορύβου των Ephraim και Malah που παρουσιάζεται στην εργασία τους “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator” (Ενίσχυση ομιλίας χρησιμοποιώντας έναν ελαχίστου μέσου τετραγωνικού σφάλματος εκτιμητή του βραχούς φασματικού εύρους) εξαλείφει καλά τον



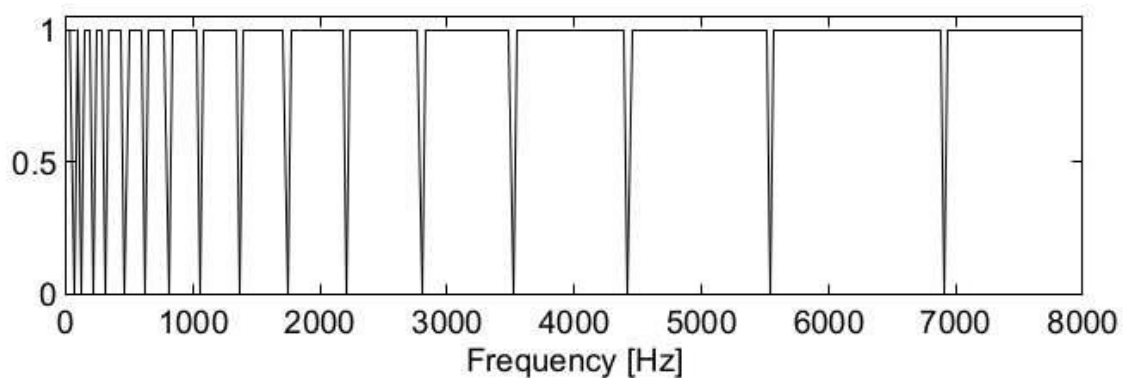
μουσικό θόρυβο με την μη γραμμική χρονική εξομάλυνση που εισάγει. Οι ίδιοι, στη μετεγενέστερη εργασία τους “Speech enhancement using a minimum mean square error log-Spectral amplitude estimator” (Ενίσχυση ομιλίας χρησιμοποιώντας έναν ελαχίστου μέσου τετραγωνικού σφάλματος εκτιμητή του λογαριθμικού φασματικού εύρους) βελτίωσαν κι άλλο την μέθοδο τους ελαχιστοποιώντας το σφάλμα εκτίμησης στο λογαριθμικό φασματικό πεδίο, καθοδηγούμενοι από την ανθρώπινη αντίληψη. Όπως αναφέρθηκε παραπάνω, άλλες τεχνικές εφαρμόζουν αποτελέσματα από αντιληπτική κωδικοποίηση ήχου. Αυτοί υπολογίζουν ένα κατώφλι απόκρυψης (masked threshold) εκτίμησης και ελέγχουν την καταστολή θορύβου ως μία συνάρτηση εκτίμησης (function thereof). Μόνο ο θόρυβος πάνω από το κατώφλι απόκρυψης αφαιρείται, ενώ ο θόρυβος κάτω από το κατώφλι απόκρυψης, θεωρώντας ότι είναι μη ακουστός, διατηρείται. Το επιχείρημα για βελτίωση είναι ότι συνολικά λιγότερος θόρυβος αφαιρείται και οπότε αναμένουμε να εισαχθούν λιγότερα τεχνουργήματα.

Η καταστολή σταθερού θορύβου με μετασχηματισμό φάσματος εφαρμόζει ένα κέρδος φίλτρου στο σήμα του θορύβου ώστε να ανακτήσει το επιθυμητό σήμα φωνής. Μην κάνοντας καμία υπόθεση σχετικά με την κατανομή των τιμών των φασμάτων ισχύος φωνής και θορύβου και εφαρμόζοντας ένα απλό κέρδος φίλτρου, όπως στην παραπάνω σχέση (Σχέση 3.18), η προτεινόμενη μέθοδος προσεγγίζει το πρόβλημα του μουσικού θορύβου θεωρώντας ότι η εξομάλυνση καθορίζεται από ιδιότητες της ανθρώπινης αντίληψης. Ο μουσικός θόρυβος είναι κυρίως μία συνέπεια των μεταβολών του χρόνου και της συχνότητας στην εκτίμηση του θορύβου που έχει άμεση επίπτωση στο κέρδος φίλτρου. Επομένως για να μειώσουμε τις μεταβολές αυτές και τα τεχνουργήματα μουσικού θορύβου που έχουν σαν αποτέλεσμα, εφαρμόζεται εξομάλυνση συχνότητας και χρόνου στο κέρδος φίλτρου.

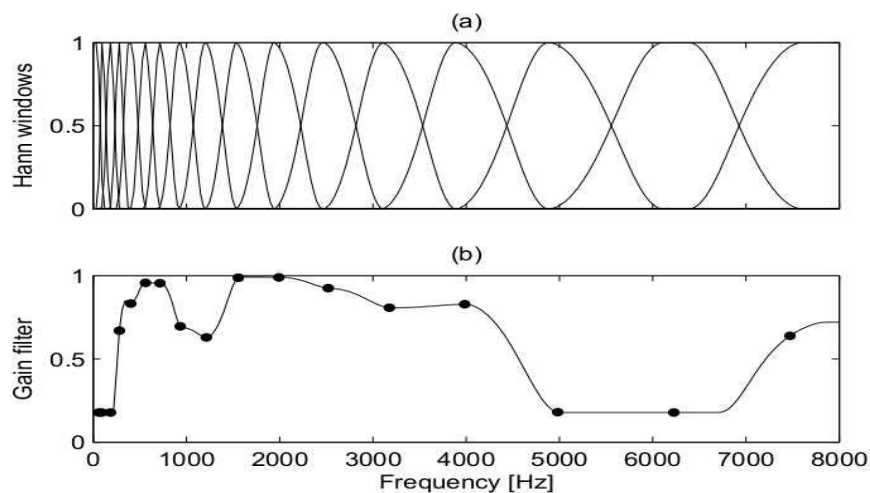
Η ομοιόμορφη ανάλυση φάσματος του STFT δεν είναι καλά προσαρμοσμένη στην ανθρώπινη αντίληψη. Οι ομοιόμορφα διανεμημένοι φασματικοί συντελεστές (uniformly

spaced spectral coefficients)  $Y[k,m]$  ομαδοποιούνται σε ένα πλήθος μη επικαλυπτόμενων διαιρέσεων (partitions) με εύρη ζώνης που μιμούνται την ανάλυση συχνότητας του ανθρώπινου ακουστικού συστήματος.

Για συχνότητα δειγματοληψίας 16kHz, χρησιμοποιούμε μήκος DFT 512 και χρησιμοποιούνται 15 διαιρέσεις, που η κάθε μία έχει εύρος ζώνης περίπου το διπλάσιο του ισοδύναμου ορθογώνιου εύρους ζώνης (ERB). Οι ζώνες που αντιστοιχούν στις διαιρέσεις φαίνονται στην παρακάτω εικόνα (Εικόνα 3.4).



Εικόνα 3.4: Ζώνες διαιρέσεων του φάσματος [6]



Εικόνα 3.5: a) φίλτρα παρεμβολής Hann b) τιμές του κέρδους φίλτρου πριν (έντονες κουκκίδες) και μετά (συνεχής γραμμή) την παρεμβολή. [6]

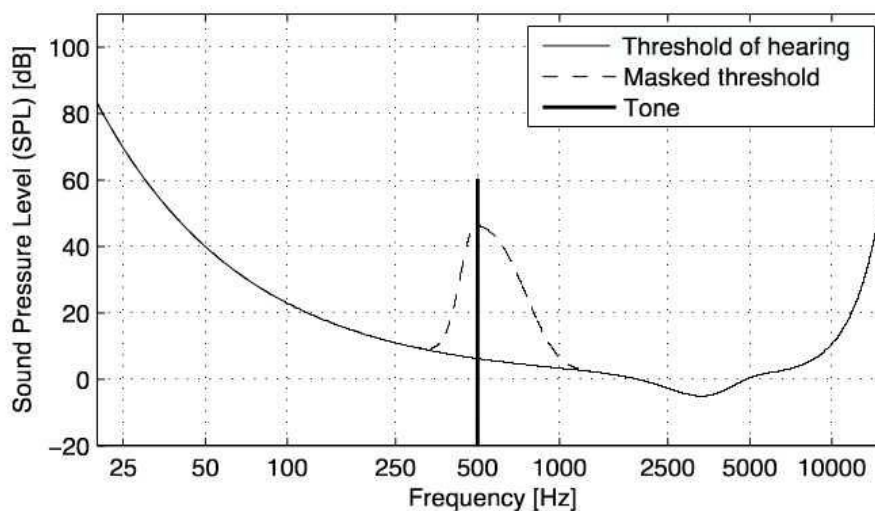
Το κέρδος φίλτρου (Σχέση 3.18) υπολογίζεται μόνο για την κεντρική συχνότητα της κάθε διαίρεσης, με αποτέλεσμα την χαμηλή πολυπλοκότητα υπολογισμών. Πριν εφαρμοστεί αυτό το κέρδος φίλτρου διαίρεσης στο ομοιόμορφο STFT φάσμα του σήματος, παρεμβάλλεται χρησιμοποιώντας φίλτρα παρεμβολής Hann. Το μέρος (a) του σχήματος 3 απεικονίζει τα φίλτρα Hann και το μέρος (b) δείχνει ένα παράδειγμα των τιμών του κέρδους φίλτρου πριν (έντονες κουκκίδες) και μετά (συνεχής γραμμή) την παρεμβολή. Η εξομάλυνση συχνότητας του κέρδους φίλτρου οδηγεί σε ομαλότερες μεταβολές του φάσματος που παράγεται ως συνάρτηση της συχνότητας και, έτσι, μειώνει τον μουσικό θόρυβο και άλλα τεχνουργήματα.

Όπως αναφέρθηκε παραπάνω, η εξομάλυνση χρόνου είναι ελεγχόμενα συνδεδεμένη με το κατώφλι απόκρυψης. Όσο χαμηλότερο είναι το κατώφλι απόκρυψης, δηλαδή όσο περισσότερες παραμορφώσεις είναι ακουστές, τόσο μεγαλύτερη είναι η χρονική σταθερά εξομάλυνσης που επιλέγεται για να μειώσει τις παραμορφώσεις. Αυτό λειτουργεί καλά γιατί συνήθως το κατώφλι απόκρυψης είναι χαμηλά για σταθερό και τονικό θόρυβο, δηλαδή μία μεγάλη χρονική σταθερά εξομάλυνσης είναι επιτρεπτή. Αντιστρόφως, όσο μεγαλύτερο είναι το κατώφλι απόκρυψης, τόσο μικρότερη είναι η χρονική σταθερά που επιλέγεται, και έτσι το κέρδος φίλτρου προσαρμόζεται αρκετά γρήγορα ώστε να ακολουθεί το σήμα.

**Masking:** Πριν συζητήσουμε την εξομάλυνση χρόνου, επανεξετάζουμε και εξηγούμε το masking (συγκάλυψη). Ταυτόχρονη συγκάλυψη συμβαίνει όταν η αντίληψη ενός ήχου επηρεάζεται από την παρουσία ενός άλλου ήχου. Θεωρείστε το απόλυτο κατώφλι της ακοής που απεικονίζεται στην παρακάτω εικόνα (Εικόνα 3.6). Όταν εκπέμπεται ένας ήχος αυτό το απόλυτο κατώφλι μπορεί να παραμορφωθεί τοπικά σε ένα υψηλότερο κατώφλι απόκρυψης το οποίο πλέον αντιστοιχεί σε ένα πιο ήσυχο επίπεδο για έναν δεύτερο ήχο για να γίνει αντιληπτός σε αυτό το όχι ήσυχο περιβάλλον. Αν ο θόρυβος εμφανίζεται ταυτόχρονα με την

επιθυμητή ομιλία, ένα ακροατής δεν θα ακούσει τον θόρυβο αυτό, όσο αυτός παραμένει χαμηλότερα από το κατώφλι απόκρυψης που καθορίζεται από το σήμα ομιλίας.

Σε αντίθεση με την ταυτόχρονη συγκάλυψη, προσωρινή συγκάλυψη εμφανίζεται όταν ένας ήχος κάνει άλλους ήχους μη ακουστους (αντιληπτούς) οι οποίοι εμφανίζονται αμέσως πριν ή αμέσως μετά από τον πρώτο ήχο. Η προ-συγκάλυψη είναι σχετικά σύντομη και μπορεί να διαρκέσει μέχρι και 20ms. Μετά-συγκάλυψη μπορεί να έχει επίδραση μέχρι και 200ms.



Εικόνα 3.6: Το κατώφλι ακοής (Συνεχής γραμμή), το αντίστοιχο κατώφλι μάσκας (διακεκομένη) για έναν τόνο με  $f=500\text{Hz}$  και ισχύ  $60\text{dB}$  [6]

**Εξομάλυνση χρόνου:** Αρκετές τεχνικές χρησιμοποιούν το κατώφλι απόκρυψης για να προσαρμόσουν την δύναμη της καταστολής θορύβου έτσι ώστε αν το κατώφλι απόκρυψης είναι υψηλό, ο υπολειπόμενος θόρυβος μπορεί να συγκαλυφθεί φυσικά και να μην είναι ακουστός. Έτσι, δεν χρειάζεται να τον μειώσουμε ώστε να κρατήσουμε την παραμόρφωση όσον το δυνατόν χαμηλότερη. Παρόλα αυτά, αν το κατώφλι απόκρυψης είναι χαμηλά, ο υπολειπόμενος θόρυβος θα είναι ενοχλητικός για τον ανθρώπινο ακροατή και θα χρειάζεται να τον μειώσουμε.

Όπως αναφέρθηκε παραπάνω, αυτό έχει σαν αποτέλεσμα να εισάγεται περισσότερη παραμόρφωση όταν το κατώφλι είναι χαμηλό και λιγότερη όταν είναι υψηλό. Ελέγχοντας την εξομάλυνση χρόνου σε αντίθεση με το ποσό του θορύβου που αφαιρείται ως συνάρτηση του κατωφλίου απόκρυψης, επιτυγχάνεται το επιθυμητό αποτέλεσμα της εισαγωγής μίας ποσότητας παραμόρφωσης που ακολουθεί το κατώφλι απόκρυψης.

Το κέρδος φίλτρου που περιγράφεται στην παραπάνω σχέση (Σχέση 3.17) εξομαλύνεται στον χρόνο χρησιμοποιώντας μονοπολική εξαγωγή μέσου όρου:

$$\widetilde{G}_P[k, b] = \mu G_P[k, b] + (1 - \mu) \widetilde{G}_P[k - 1, b] \text{ (Σχέση 3.19)}$$

όπου ο θετικός αριθμός  $\mu$  καθορίζει το βαθμό της εξομάλυνσης ως προς το χρόνο. Η χρονική σταθερά της εκθετικής μείωσης σε δευτερόλεπτα είναι περίπου:

$$T_\mu = \frac{1}{\mu f_s} \text{ (Σχέση 3.20)}$$

όπου το  $f_s$  συμβολίζει τον ρυθμό φασματικής δειγματοληψίας. Η χρονική σταθερά, και ισοδύναμα το  $\mu$ , για κάθε διαίρεση ελέγχονται ανάλογα με το κατώφλι απόκρυψης.

Όσο περισσότερο είναι τονικό και σταθερό το σήμα, τόσο λιγότερη συγκάλυψη εμφανίζεται και όσο περισσότερο ασταθής ή παροδικός είναι, τόσο περισσότερη συγκάλυψη εμφανίζεται. Έτσι, για να εκτιμήσουμε στο περίπου το σχετικό ποσό συγκάλυψης, ένα μέτρο διακύμανσης υπολογίζεται για κάθε διαίρεση:

$$\Phi[k, b] = \frac{\sigma_y^2[k, b] - \sigma_y^2[k - 1, b]}{\max(\sigma_y^2[k, b], \sigma_y^2[k - 1, b])} \text{ (Σχέση 3.21)}$$

όπου

$$\sigma_y^2[k, b] = E(Y[k, m]Y^* [k, m]) \text{ (Σχέση 3.22)}$$

και το  $E\{\dots\}$  αντιπροσωπεύει την μέση τιμή μικρής διάρκειας. Το μέτρο διακύμανσης  $\Phi[k, b]$  παίρνει τιμές στο διάστημα  $[-1, 1]$ . Για διαφορετικούς τύπους συνιστωσών σήματος, διαφορετικές χρονικές σταθερές επιλέγονται, ανάλογα με την συγκάλυψη και την ανάγκη της προσαρμογής του κέρδους φίλτρου αρκετά γρήγορα ώστε να ακολουθεί το σήμα:

**Τόνος:** Για συνιστώσες τονικών σημάτων, η χρονική σταθερά επιλέγεται μεγαλύτερη, μιας και αυτά είναι οι πιο σταθερές συνιστώσες σήματος και εμφανίζεται μικρότερη συγκάλυψη.

**Θόρυβος:** Για συνιστώσες σημάτων που μοιάζουν με θόρυβο, η χρονική σταθερά μπορεί να επιλεγεί λιγότερο μεγάλη από αυτή των τονικών συνιστωσών, μιας και εμφανίζεται περισσότερη συγκάλυψη.

**TransientOnset:** Για παροδική έναρξη είναι σημαντικό το κέρδος φίλτρου να ακολουθεί τον χρονικό φάκελο του σήματος, που οδηγεί σε επιλογή μικρότερης χρονικής σταθεράς από ότι στις συνιστώσες σήματος που μοιάζουν με θόρυβο.

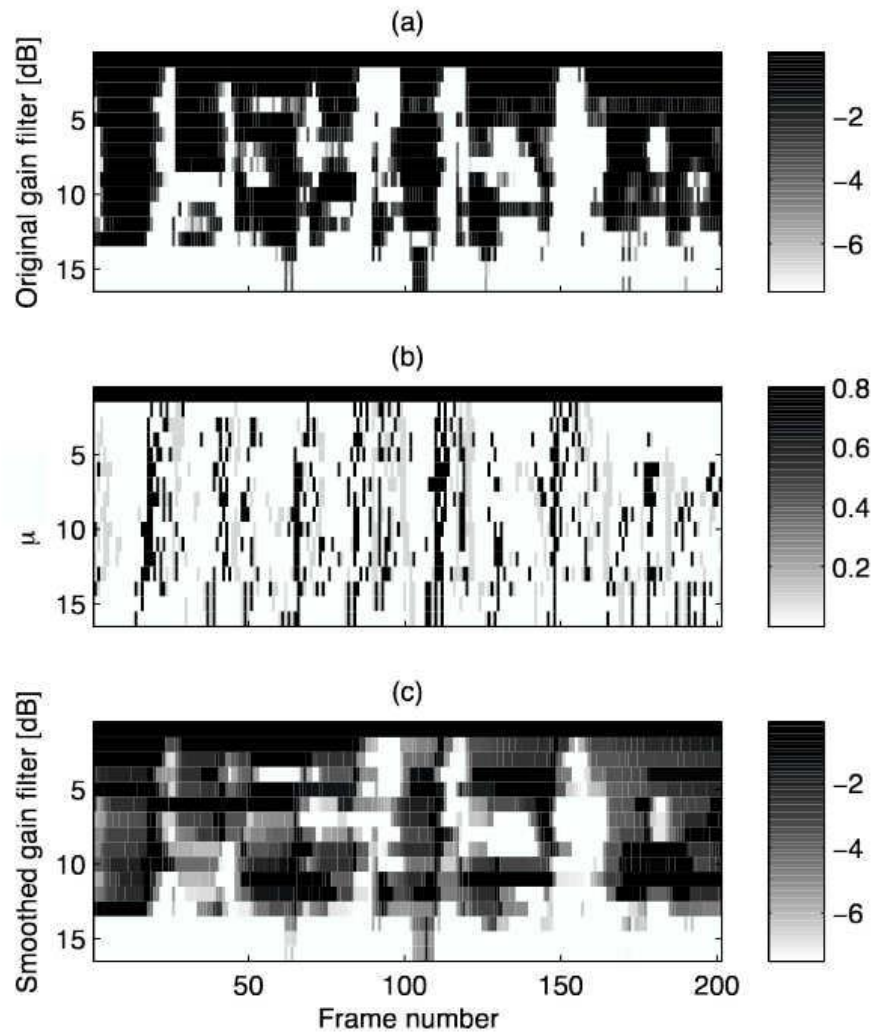
**TransientOffset:** Η μικρότερη χρονική σταθερά εξομάλυνσης επιλέγεται για την παροδική λήξη ώστε να αποφευχθεί το να υπάρξει μετά-ηχώ μετά την παροδική λήξη.

Στον παρακάτω πίνακα (Πίνακας 3.1) προσδιορίζονται οι τέσσερις χρονικές σταθερές που χρησιμοποιούνται για τους τέσσερις διαφορετικούς τύπους σήματος.

Πίνακας 1: Χρονική σταθερά  $T_\mu$  για κάθε είδος σήματος που ορίζεται στο αντίστοιχο διάστημα του μέτρου διακύμανσης  $\Phi[k, b]$ .

<b>Signalproperty</b>	<b><math>T_\mu</math> (sec.)</b>	<b><math>\Phi[k, b]</math></b>
Tone	0.0125	$ \Phi[k, b]  < 0.1$
Noise	0.1	$0.1 \leq  \Phi[k, b]  \leq 0.5$
Transientonset	1	$\Phi[k, b] > 0.5$
Transientoffset	10	$\Phi[k, b] < -0.5$

Στην παρακάτω εικόνα (Εικόνα 3.6) το διάγραμμα (a) δείχνει το αρχικό, μη εξομαλυμένο κέρδος φίλτρου στο διαιρεμένο πεδίο συχνοτήτων για ένα παράδειγμα σήματος θορυβώδους φωνής. Το διάγραμμα (b) δείχνει την σταθερά εξομάλυνσης  $\mu$  η οποία σχετίζεται με την χρονική σταθερά που υπολογίζεται από τον παραπάνω τύπο (Σχέση 3.20). Το εξομαλυμένο κέρδος φίλτρου φαίνεται στο διάγραμμα (c) της παρακάτω εικόνας. Παροδικότητα οδηγεί σε μικρή εξομάλυνση, ενώ κατά τη διάρκεια των θορυβωδών τμημάτων το κέρδος φίλτρου είναι πιο εξομαλυμένο. Οπότε, ο αλγόριθμος αυτός των AlexisFavrot και ChristoffFaller που παρουσίασαν το 2007 έχει αρκετά καλά αποτελέσματα για την αποθορυβοποίηση των σημάτων ομιλίας.



Εικόνα 3.7: a) Το αρχικό, μη εξομαλυσμένο κέρδος φίλτρου στο διαιρεμένο πεδίο συχνοτήτων, b) Σταθερά εξομάλυνσης  $\mu$ , c) Το εξομαλυσμένο κέρδος φίλτρου



## **ΚΕΦΑΛΑΙΟ 4**

### ***ΑΛΓΟΡΙΘΜΟΙ ΑΠΟΘΟΥΒΟΠΟΙΗΣΗΣ***

Στα πλαίσια της παρούσας εργασίας, υλοποιήθηκαν δύο εφαρμογές στο προγραμματιστικό περιβάλλον MatLab για καλύτερη εποπτεία των εννοιών και των μεθόδων που αναφέρθηκαν παραπάνω. Στο παρόν κεφάλαιο θα γίνει η παρουσίαση των εφαρμογών αυτών. Η πρώτη εφαρμογή χρησιμοποιεί ένα χαμηλοπερατό φίλτρο για την αφαίρεση του θορύβου και η δεύτερη είναι η υλοποίηση της μεθόδου φασματικής αφαίρεσης θορύβου που παρουσιάστηκε μέσω του αλγορίθμου των AlexisFavrot και ChristoffFaller στο προηγούμενο κεφάλαιο (παράγραφο 3.4). Για τις εφαρμογές αυτές, αρχικά γίνεται μία σύντομη σύνοψη της θεωρίας που αναφέρθηκε στα προηγούμενα κεφάλαια και στην οποία στηρίζεται ο αλγόριθμος του προγράμματος που χρησιμοποιεί το χαμηλοπερατό φίλτρο και ο αλγόριθμος του προγράμματος που χρησιμοποιεί την φασματική αφαίρεση, και στην συνέχεια αναλύεται ο τρόπος με τον οποίο επιτυγχάνεται η θορυβοποίηση και η αποθορυβοποίηση ενός σήματος ομιλίας με τις δύο μεθόδους που υλοποιήθηκαν.

#### **4.1-Πρόσθεση θορύβου**

Με τον όρο “σήμα” ορίζεται ένα φυσικό μέγεθος το οποίο μεταβάλλεται σε σχέση με κάποια ανεξάρτητη μεταβλητή (ή μεταβλητές), όπως παραδείγματος χάριν ο χρόνος (χρονική μεταβλητή) ή ο χώρος (χωρικές μεταβλητές). Τα σήματα χρησιμοποιούνται για την μετάδοση και την αποθήκευση πληροφοριών και ένα σήμα θα μεταφέρει ενέργεια ή ισχύ και κάποιο μήνυμα ή πληροφορία.

Ανάλογα με την πληροφορία που περιέχουν ή το φυσικό μέγεθος το οποίο θα μεταβάλλεται, χωρίζονται σε διάφορες κατηγορίες. Κάποια παραδείγματα σημάτων είναι τα

σήματα ομιλίας και ήχου (με τα οποία ασχολούμαστε στην παρούσα εργασία), τα σήματα εικόνας, τα σεισμικά σήματα και τα ιατρικά σήματα.

Ανάλογα με τις τιμές που μπορεί να πάρει ο χρόνος και το φυσικό μέγεθος που μεταβάλλεται, τα σήματα χωρίζονται σε σήματα συνεχούς και διακριτού χρόνου και σε σήματα συνεχών και διακριτών τιμών (κβαντισμένα). Τα σήματα που είναι διακριτού χρόνου και διακριτών τιμών, των οποίων οι τιμές έχουν κωδικοποιηθεί, ονομάζονται ψηφιακά.

Στα σήματα ομιλίας, το μέγεθος που μεταβάλλεται είναι τα ηχητικά κύματα, ή αλλιώς ο ήχος, που είναι διαμήκη μηχανικά κύματα στα οποία το μέγεθος που μεταβάλλεται είναι η πίεση του αέρα. Λόγω της προσθετικής ιδιότητας που έχει η πίεση, σε ένα σήμα ομιλίας το οποίο θα ηχογραφηθεί (θα μετατραπεί σε ηλεκτρικό σήμα) θα προστεθούν και τα υπόλοιπα σήματα ομιλίας που διαδίδονται στον ίδιο χώρο. Έτσι, το σήμα που θα ηχογραφηθεί θα είναι το άθροισμα όλων αυτών των σημάτων, δηλαδή του επιθυμητού σήματος και των υπόλοιπων σημάτων που διαδίδονται στον ίδιο χώρο. Οπότε θα προστεθούν στο επιθυμητό σήμα και άλλα ανεπιθύμητα, τα οποία θεωρούνται ως θόρυβος.

Και το φαινόμενο αυτό προσομοιώνεται στο πρόγραμμα (script) που υλοποιήθηκε στο MatLab(έκδοση R2016a), του οποίου ο κώδικας παρατίθεται στο παράρτημα Α.

#### **4.1.1-Ανάλυση του αλγορίθμου του προγράμματος προσθήκης θορύβου**

Το πρόγραμμα αυτό κάνει μίξη καθαρής ηχογράφησης με θόρυβο σε επιθυμητή στάθμη SNR δηλαδή κάνει τα εξής:

Διαβάζει ένα αρχείο ήχου (καθαρή ηχογράφηση) και ένα δεύτερο αρχείο ήχου (τον θόρυβο). Το αποτέλεσμα θα είναι δύο διανύσματα π.χ. Speech και Noise στα οποία φορτώνονται τα δεδομένα των αρχείων του σήματος ομιλίας και ενός σήματος λευκού θορύβου.

Επειδή τα σήματα που φορτώθηκαν θα πρέπει να προστεθούν, εφαρμόζεται ένας έλεγχος για την ισότητα του πλήθους των δειγμάτων του κάθε διανύσματος και, σε περίπτωση που κάποιο έχει περισσότερα δείγματα, “κόβονται” τα πλεονάζοντα δείγματα και πλέον τα σήματα μπορούν να προστεθούν.

Στην συνέχεια, υπολογίζεται η ενέργεια του Speech ( $E_{\text{speech}}$ ) και του Noise ( $E_{\text{noise}}$ ). (Η ενέργεια είναι το άθροισμα των δειγμάτων του ήχου υψωμένα στο τετράγωνο. Δηλαδή αν ο ήχος είναι η ακολουθία 2,3,5 η ενέργεια είναι  $4+9+25=38$ )

Μετά τον υπολογισμό των ενεργειών των σημάτων, υπολογίζεται ένας συντελεστής με τον οποίο να πολλαπλασιάζεται ο θόρυβος και προκύπτει σήμα με επιθυμητή ενέργεια σε συγκεκριμένο SNR. Ο SNR είναι ο λόγος της ενέργειας της καθαρής ηχογράφησης προς την ενέργεια του θορύβου σε dB και δίνεται από την παρακάτω σχέση:

$$SNR = 10 * \log(E_{\text{speech}}/E_{\text{noise}}) \quad (\text{Σχέση 4.1})$$

Έστω  $E_{\text{noise}}$  η ενέργεια του θορύβου σε επιθυμητή στάθμη (-3dB).

Η ενέργεια αυτή θα υπολογίζεται από τον τύπο:

$$\frac{E_{\text{speech}}}{E_{\text{noise}}} = 10^{(SNR/10)} \rightarrow E_{\text{noise}} = \frac{E_{\text{speech}}}{10^{(SNR/10)}} \quad (\text{Σχέση 4.2})$$

Υπολογίζει τον συντελεστή με τον οποίο θα πολλαπλασιαστεί το διάνυσμα του θορύβου:

$$E_{\text{noise}} = a^2 * E_{\text{noise}} \rightarrow a = \left(\frac{E_{\text{noise}}}{E_{\text{noise}}}\right)^{1/2} = \sqrt{\frac{E_{\text{noise}}}{E_{\text{noise}}}} \quad (\text{Σχέση 4.3})$$

Τέλος, κατασκευάζεται το ενθόρυβο σήμα με χρήση του τύπου:

$$NoisySound = Speech + Noise * a \text{ (Σχέση 4.4)}$$

και το ενθόρυβο σήμα αποθηκεύεται σε ένα wav αρχείο ήχου.

## 4.2-Αφαίρεση θορύβου με χρήση χαμηλοπερατού φίλτρου

Κάθε σήμα ήχου, ως κύμα, θα έχει κάποιες φυσικές ιδιότητες. Οι σημαντικότερες είναι η συχνότητα και το πλάτος του (ή αλλιώς η ένταση), και ως κύμα μπορεί να παρασταθεί με την παρακάτω εξίσωση:

$$y = A_0 \sin(\omega t - kx) \text{ (Σχέση 4.5)}$$

η οποία εκφράζει την απομάκρυνση από την θέση ισορροπίας,  $y$ , ενός μορίου της ύλης εντός της οποίας διαδίδεται το ηχητικό κύμα, συναρτήσει του πλάτους του κύματος,  $A_0$ , της κυκλικής του συχνότητας,  $\omega$ , και του όρου  $kx$  που αφορά τη θέση του σημείου του οποίου η θέση περιγράφεται από την παραπάνω σχέση (Σχέση 4.5).

Η σχέση 4.5 αφορά την περιγραφή του ηχητικού κύματος συναρτήσει του χρόνου, περιγράφοντας έτσι την χρονική μεταβολή της θέσης του κάθε μορίου της ύλης εντός της οποίας μεταδίδεται το συγκεκριμένο ηχητικό κύμα. Για τον λόγο αυτό αναφέρεται ως “περιγραφή του κύματος στο πεδίο του χρόνου”.

Ο θόρυβος, λοιπόν, είναι τα σήματα που προστίθενται στο επιθυμητό σήμα ήχου και το αλλοιώνουν, με αποτέλεσμα να είναι δυσκολότερη η ανάκτηση της πληροφορίας (του ήχου) που περιέχει το σήμα αυτό. Για τον λόγο αυτό είναι επιθυμητή η αφαίρεση του θορύβου από τα ενθόρυβα σήματα. Όμως αυτό είναι αρκετά δύσκολο, λόγω της προσθετικής ιδιότητας που έχουν τα σήματα ήχου. Έτσι, σύμφωνα με την παραπάνω σχέση (Σχέση 4.5), το μέγεθος  $y$  που θα έχει το ηχητικό, ενθόρυβο, σήμα, θα οφείλεται και στον επιθυμητό ήχο αλλά και στον θόρυβο και για τον λόγο αυτό δεν μπορεί να γίνει διάκριση αυτών.

Η διάκριση είναι λοιπόν αδύνατη όταν η περιγραφή του σήματος γίνεται στο πεδίο του χρόνου, με την παραπάνω σχέση (Σχέση 4.5) να είναι μία τέτοια απλή περιγραφή. Όμως, όπως αναφέρθηκε και παραπάνω, μία σημαντική ιδιότητα των ηχητικών κυμάτων είναι η συχνότητα τους. Κάθε κύμα της απλής μορφής που περιγράφεται από την σχέση 4.5 θα έχει μία συγκεκριμένη συχνότητα (απλός τόνος). Τα ηχητικά κύματα όμως, συνήθως, είναι πιο πολύπλοκα και αποτελούνται από πολλά άλλα κύματα διαφορετικών συχνοτήτων. Έτσι, μπορεί να γίνει η ανάλυση ενός τέτοιου κύματος στα κύματα διαφορετικών συχνοτήτων από τα οποία δημιουργείται. Η ανάλυση αυτή ονομάζεται “συχνοτική ή φασματική ανάλυση” και δίνει την περιγραφή του συγκεκριμένου σήματος στο πεδίο των συχνοτήτων. Για την “μετάβαση” ενός σήματος από το πεδίο του χρόνου στο πεδίο των συχνοτήτων υπάρχουν αρκετές μέθοδοι, όπου μία από αυτές είναι η “ανάλυση Fourier”.

Η ανάλυση Fourier είναι στην ουσία ο μαθηματικός αλγόριθμος που μπορεί να χρησιμοποιηθεί για την μετάβαση του σήματος από το πεδίο του χρόνου στο πεδίο των συχνοτήτων και περιέχει διάφορες επιμέρους μεθόδους, όπως είναι ο μετασχηματισμός Fourier (Fourier Transform-FT) που χρησιμοποιείται για την φασματική ανάλυση αναλογικών σημάτων, ο διακριτός μετασχηματισμός Fourier (Discrete Fourier Transform-DFT) που χρησιμοποιείται για την φασματική ανάλυση διακριτών σημάτων, ο μετασχηματισμός Fourier βραχέος χρόνου (STFT-Short Time Fourier Transform) καθώς και

ο ταχύς μετασχηματισμός Fourier (Fast Fourier Transform-FFT) που είναι ένας πιο αποδοτικός αλγόριθμος για εφαρμογή σε διακριτά σήματα. Οπότε, χρησιμοποιώντας κάποιο από τις διαθέσιμες μεθόδους φασματικής (ή συχνοτικής) ανάλυσης ενός σήματος μπορεί να περιγραφεί ένα σήμα του οποίου είναι γνωστή η μορφή στο πεδίο του χρόνου, στο πεδίο των συχνοτήτων.

Η συχνοτική ανάλυση ενός σήματος χρησιμοποιείται κατά κόρον για την αφαίρεση το θορύβου από ένα σήμα. Λόγω του ότι ο θόρυβος και το επιθυμητό σήμα δεν θα έχουν, συνήθως, την ίδια συχνότητα, κάνοντας την συχνοτική (ή φασματική) ανάλυση του σήματος είναι εφικτή η αφαίρεση σημάτων συγκεκριμένων συχνοτήτων, “καθαρίζοντας” έτσι το σήμα από τον θόρυβο. Ακόμα και αν δεν είναι εφικτή η πλήρης αφαίρεση του θορύβου, μπορεί να αφαιρεθεί ένα σημαντικό κομμάτι του, μειώνοντας έτσι την αλλοίωση του αρχικού σήματος. Η αφαίρεση αυτή γίνεται με χρήση κατάλληλων διατάξεων, που ονομάζονται “φίλτρα”, τα οποία είναι στην ουσία κάποια κυκλώματα των οποίων η σχέση εξόδου/εισόδου εξαρτάται από την συχνότητα της εισόδου.

Υπάρχουν διάφορα είδη φίλτρων, όπου τα βασικότερα είναι τρία. Τα χαμηλοπερατά φίλτρα είναι κυκλώματα τα οποία, όταν το σήμα εισόδου τους έχει μικρή συχνότητα, το σήμα της εξόδου τους θα είναι το σχεδόν ίδιο με το σήμα της εισόδου τους, δηλαδή το σήμα της εισόδου “περνάει” ανεπηρέαστο από το φίλτρο, ενώ όταν το σήμα εισόδου έχει μεγάλη συχνότητα, μεγαλύτερη από μία συγκεκριμένη συχνότητα, χαρακτηριστική του κάθε φίλτρου, που ονομάζεται συχνότητα αποκοπής, το σήμα αυτό θα υπόκειται σε απόσβεση και η έξοδος του φίλτρου θα είναι η είσοδος του μειωμένη κατά ένα μεγάλο ποσοστό. Έτσι, αν ένα σήμα που αποτελείται από διάφορα επιμέρους σήματα διαφορετικών συχνοτήτων περάσει μέσα από ένα φίλτρο, στην έξοδο του φίλτρου αυτού θα πάρουμε το σήμα εισόδου, στο οποίο όμως η ένταση των σημάτων με συχνότητες μεγαλύτερες της συχνότητας αποκοπής θα έχουν πολύ μικρότερη ένταση και θα επηρεάζουν το ηχητικό σήμα πολύ

λιγότερο, μειώνοντας έτσι τις αλλοιώσεις του επιθυμητού σήματος. Έτσι, με την χρήση κατάλληλου φίλτρου μπορεί να αφαιρεθούν σε μεγάλο βαθμό τα σήματα του θορύβου με συχνότητες μεγαλύτερες της συχνότητας αποκοπής.

Τα άλλα βασικά είδη φίλτρων είναι τα υψηλοπερατά, τα οποία θα αφήνουν να περάσουν ανεπηρέαστα τα σήματα με συχνότητες μεγαλύτερες της συχνότητας αποκοπής και θα “κόβουν” τα σήματα με συχνότητες μικρότερες της συχνότητας αποκοπής, και τα ζωνοπερατά, τα οποία θα αφήνουν να περάσουν ανεπηρέαστα τα σήματα συχνοτήτων εντός μίας ζώνης συχνοτήτων και θα κόβουν τα σήματα των υπόλοιπων συχνοτήτων.

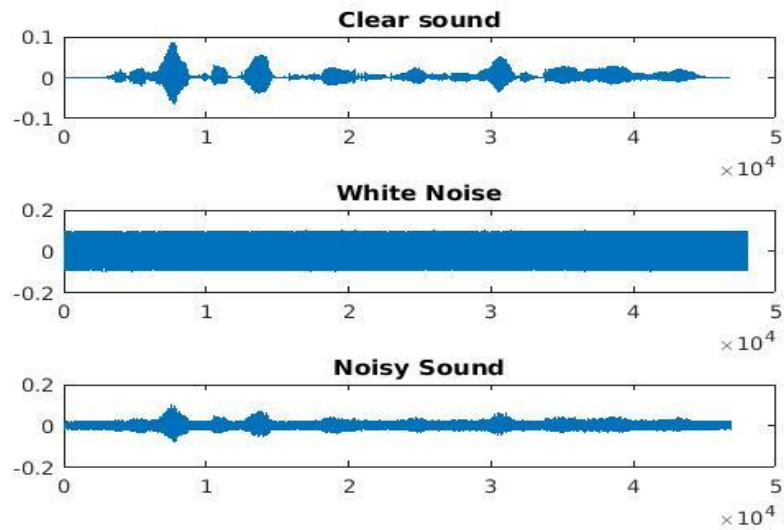
Μία απλή υλοποίηση της αφαίρεσης θορύβου από το ενθόρυβο σήμα που παράχθηκε από το script που περιγράφηκε παραπάνω γίνεται με το παρακάτω script που υλοποιήθηκε στο MatLab (έκδοση R2016a), του οποίου ο κώδικας παρατίθεται στο παράρτημα Β.

#### **4.2.1-Εφαρμογή του αλγορίθμου του προγράμματος αφαίρεσης θορύβου με χρήση χαμηλοπερατού φίλτρου**

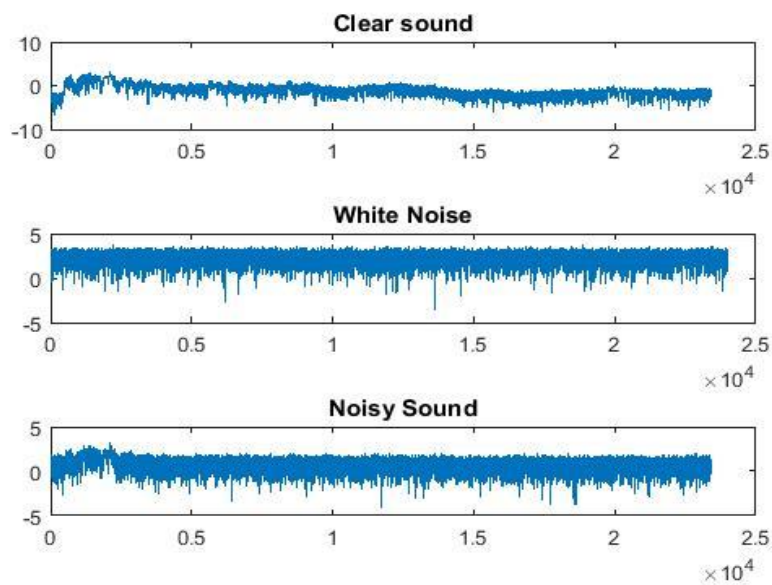
Αρχικά φορτώνονται τα δεδομένα των αρχείων ήχου, όπου το ένα αρχείο περιέχει τον “καθαρό” ήχο, το δεύτερο τον λευκό θόρυβο και το τρίτο τον ενθόρυβο ήχο που δημιουργήθηκε με το προηγούμενο script. Στην συνέχεια δημιουργείται η γραφική παράσταση και των τριών σημάτων στο πεδίο του χρόνου, η οποία παρατίθεται στην παρακάτω εικόνα.

Στις παρακάτω γραφικές παραστάσεις φαίνονται τα δεδομένα των τριών αρχείων ήχου που παράχθηκαν με χρήση του προηγούμενου προγράμματος. Σε αυτό φαίνεται η επίδραση που έχει ο θόρυβος στον καθαρό ήχο. Όμως, για να επιτευχθεί η αφαίρεση του ήχου με χρήση ενός φίλτρου, θα πρέπει να γίνει η φασματική ανάλυση των παραπάνω δεδομένων, έτσι ώστε να είναι γνωστό το φασματικό περιεχόμενο κάθε σήματος και να

σχεδιαστεί κατάλληλα το φίλτρο. Για τον λόγο αυτό, σχεδιάζονται και οι γραφικές παραστάσεις των παραπάνω σημάτων στο πεδίο των συχνοτήτων, χρησιμοποιώντας την μέθοδο FFT, οι οποίες παρατίθενται στην παρακάτω εικόνα (Εικόνα 4.2).



Εικόνα 4.1: Κυματομορφές στο πεδίο του χρόνου των δεδομένων του καθαρού ήχου, του θορύβου και του ενθόρυβου σήματος

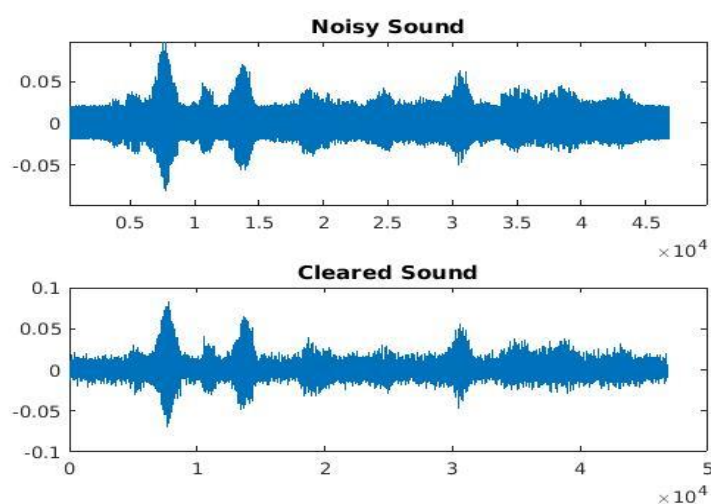


Εικόνα 4.2: Φάσμα του καθαρού ήχου, του θορύβου και του ενθόρυβου σήματος



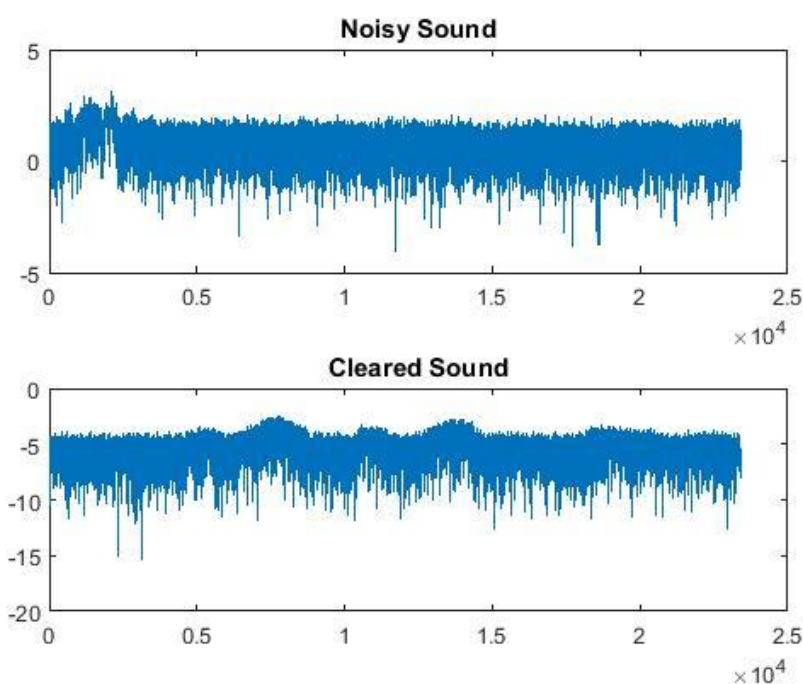
Από την παραπάνω εικόνα (Εικόνα 4.2), στην οποία περιέχονται οι γραφικές παραστάσεις των σημάτων στο πεδίο των συχνοτήτων, έχοντας στον οριζόντιο άξονα τις συχνότητες και στον κάθετο την ένταση του σήματος, γίνεται φανερό ότι ο λευκός θόρυβος περιέχει όλες τις συχνότητες, ενώ το αρχικό σήμα περιέχει συχνότητες μέχρι περίπου τα 4kHz. Το ενθόρυβο σήμα θα περιέχει όλες τις συχνότητες των σημάτων από τα οποία προήλθε, δηλαδή θα έχει τιμή σε όλο το συχνοτικό φάσμα λόγω του λευκού θορύβου, όπου όμως στις συχνότητες του καθαρού ήχου θα έχει μεγαλύτερη ένταση. Έτσι, χρησιμοποιώντας ένα χαμηλοπερατό φίλτρο με κατάλληλη συχνότητα αποκοπή μπορεί να αφαιρεθεί ένα μεγάλο ποσοστό του θορύβου από το σήμα. Έτσι, σχεδιάζεται ένα φίλτρο Butterworth, έκτης τάξης με συχνότητα αποκοπής 4800 Hz.

Το φίλτρο που σχεδιάστηκε, στην συνέχεια εφαρμόζεται στα δεδομένα του ενθόρυβου αρχείου, φιλτράροντας τα κατάλληλα. Στο σημείο αυτό έχει “καθαριστεί” ο ήχος από ένα μεγάλο ποσοστό του θορύβου που αυτός περιείχε. Για να γίνει φανερό αυτό, σχεδιάζονται τα δεδομένα του ενθόρυβου ήχου στο πεδίο του χρόνου, οι γραφικές παραστάσεις των οποίων παρατίθενται στο παρακάτω σχήμα.



Εικόνα 4.3: Κυματομορφές στο πεδίο του χρόνου του ενθόρυβου και του καθαρισμένου σήματος

Από τις γραφικές παραστάσεις της παραπάνω εικόνας (Εικόνα 4.3) φαίνεται ότι έχει μειωθεί η ένταση του ήχου, κάτι που οφείλεται στην αφαίρεση του θορύβου από αυτόν. Για καλύτερη εποπτεία των παραπάνω, σχεδιάζονται τα δεδομένα του ενθόρυβου και του καθαρισμένου ήχου στο πεδίο των συχνοτήτων, οι γραφικές παραστάσεις των οποίων παρατίθενται στην παρακάτω εικόνα (Εικόνα 4.4).



Εικόνα 4.4: Φάσμα του ενθόρυβου και του καθαρισμένου σήματος

Από την παραπάνω εικόνα (Εικόνα 4.4) γίνεται φανερό ότι ο ενθόρυβος ήχος καθαρίστηκε από τον θόρυβο με συχνότητες μεγαλύτερες της συχνότητας αποκοπής που καθορίστηκε, η οποία ήταν 4800Hz. Έτσι, τα σήματα με συχνότητες μικρότερες της συχνότητας αποκοπής δεν επηρεάστηκαν από το φίλτρο ενώ τα σήματα με συχνότητες μεγαλύτερες της συχνότητας αποκοπής έχουν υποστεί μεγάλη απόσβεση, η οποία μεγαλώνει όσο πιο μεγάλη γίνεται η συχνότητα του σήματος, και όπως διακρίνεται και στο παραπάνω διάγραμμα, για συχνότητες μεγαλύτερες των 10kHz έχει αφαιρεθεί πλήρως ο θόρυβος.

### 4.3-Αφαίρεση θορύβου με χρήση φασματικής αφαίρεσης

Το πρόγραμμα που υλοποιήθηκε βασίστηκε στην εργασία των Alexis Favrot και Christof Faller όπως περιγράφεται στο paper “Perceptually Motivated Gain Filter Smoothing for Noise Suppression” το οποίο παρουσιάστηκε στο 123<sup>ο</sup> Συνέδριο (5–8 Οκτώβρη 2007, Νέα Υόρκη) της “Audio Engineering Society”.

Η τεχνική που χρησιμοποιείται, σύμφωνα με την παραπάνω εργασία, βασίζεται στη θεώρηση ότι ένα θορυβώδες σήμα ομιλίας μπορεί να αναπαρασταθεί από τον παρακάτω τύπο

$$y[n] = x[n] + v[n] \text{ ( Σχέση 4.6 )}$$

όπου το  $x$  είναι το επιθυμητό σήμα ομιλίας και το  $v$  αντιπροσωπεύει το σήμα του θορύβου, με το  $v$  να θεωρείται σταθερή διαδικασία. Οπότε, αρχικά φορτώνονται τα δεδομένα του ήχου από το πρόγραμμα ώστε να γίνει η επεξεργασία τους και ο καθαρισμός του ήχου από τον θόρυβο. Στα δεδομένα τα οποία θα επεξεργαστεί το πρόγραμμα, τα οποία είναι στο πεδίο των συχνοτήτων, εφαρμόζεται ένας μετασχηματισμός φάσματος, εφαρμόζοντας έναν διακριτό μετασχηματισμό Fourier μικρής διάρκειας (discreteshort-timeFouriertransform (DSTFT)). Τα σήματα ομιλίας είναι μη σταθερά αλλά αν χωριστούν σε πλαίσια μικρού χρόνου μπορούν να θεωρηθούν σταθερά. Εφαρμόζοντας DSTFT και στα δύο μέρη της παραπάνω σχέσης (Σχέση 4.6) έχουμε:

$$Y [k, m] = X[k, m] + V [k, m] , \text{ (Σχέση 4.7)}$$

όπου το  $k$  είναι ο αριθμός του χρονικού πλαισίου και  $m$  ο δείκτης του συχνοτικού. Μιας και η ανθρώπινη διακριτική ικανότητα είναι σχετικά αναισθητή στην αλλοίωση της φάσης, η

εκτιμώμενη φάση επιλέγεται να είναι ίση με την φάση του θορύβου και η εκτίμηση του  $X[k,m]$  είναι επαρκής για να ανακτηθεί το επιθυμητό σήμα ομιλίας.

Εντός ενός πλαισίου, και η ομιλία  $x$  και ο θόρυβος  $v$  μοντελοποιούνται ως ασυσχέτιστες σταθερές διαδικασίες. Έτσι, το φάσμα ισχύος του  $y$  μπορεί να προσεγγιστεί ως εξής:

$$|Y[k, m]|^2 \approx |X[k, m]|^2 + |V[k, m]|^2 \quad (\text{Σχέση 4.8})$$

Τότε, το φάσμα ισχύος  $|X[k, m]|^2$  του επιθυμητού σήματος ομιλίας μπορεί να ανακτηθεί αφαιρώντας μία εκτίμηση του φάσματος ισχύος του θορύβου  $|V[k, m]|^2$  από το  $|Y[k, m]|^2$

$$|\tilde{X}[k, m]|^2 = |Y[k, m]|^2 - |\tilde{V}[k, m]|^2 \quad (\text{Σχέση 4.9})$$

Αντίστοιχα με το φιλτράρισμα, το φασματικό μέγεθος του σήματος ομιλίας ανακτάται από τη σχέση:

$$|\tilde{X}[k, m]| = G[k, m]|Y[k, m]| \quad (\text{Σχέση 4.10})$$

όπου το  $G[k, m]$  συμβολίζεται ως “κέρδος φίλτρου” (η “Συνάρτηση Μεταφοράς” του συστήματος):

$$G[k, m] = \left( \frac{|Y[k, m]|^2 - |\tilde{V}[k, m]|^2}{|Y[k, m]|^2} \right)^{\frac{1}{2}} \quad (\text{Σχέση 4.11})$$

Μία πιο γενική μορφή της παραπάνω ποσότητας, που ονομάζεται παραμετρικό κέρδος φίλτρου, μπορεί να προέλθει εισάγοντας τις παραμέτρους  $\alpha$  και  $\beta$ :

$$G_p[k, m] = \left( \frac{|Y[k, m]|^\alpha - \beta |\tilde{V}[k, m]|^\alpha}{|Y[k, m]|^\alpha} \right)^{\frac{1}{\alpha}} \text{(Σχέση 4.12)}$$

όπου το  $\alpha$  αναπαριστά έναν εκθέτη και το  $\beta$  μία παράμετρο ελέγχου της ποσότητας του θορύβου που πρέπει να κατασταλεί στην περίπτωση χαμηλής (ή υψηλής) εκτίμησης, που μπορεί να επιλεγθεί σύμφωνα με το λόγο σήματος προς θόρυβο (SNR) και του δείκτη συχνότητας  $m$ .

Μία καλή υλοποίηση των παραπάνω παρουσιάζεται στην εργασία των M. Berouti, R. Schwartz, και J. Makhoul, “Enhancement of speech corrupted by musical noise,” στην οποία, εκτός του ότι χρησιμοποιείται η τιμή 2 για το  $\alpha$ , στο κέρδος του φίλτρου καθορίζεται ένα ελάχιστο όριο, το όριο των  $-L$  dB, και αν η τιμή του  $G_p$  πάρει τιμή μικρότερη του ορίου αυτού, αντικαθίσταται από το όριο, σύμφωνα με τον παρακάτω τύπο :

$$G_p[k, m] = \max(10^{\frac{-L}{20}}, G_p) \quad \text{(Σχέση 4.13)}$$

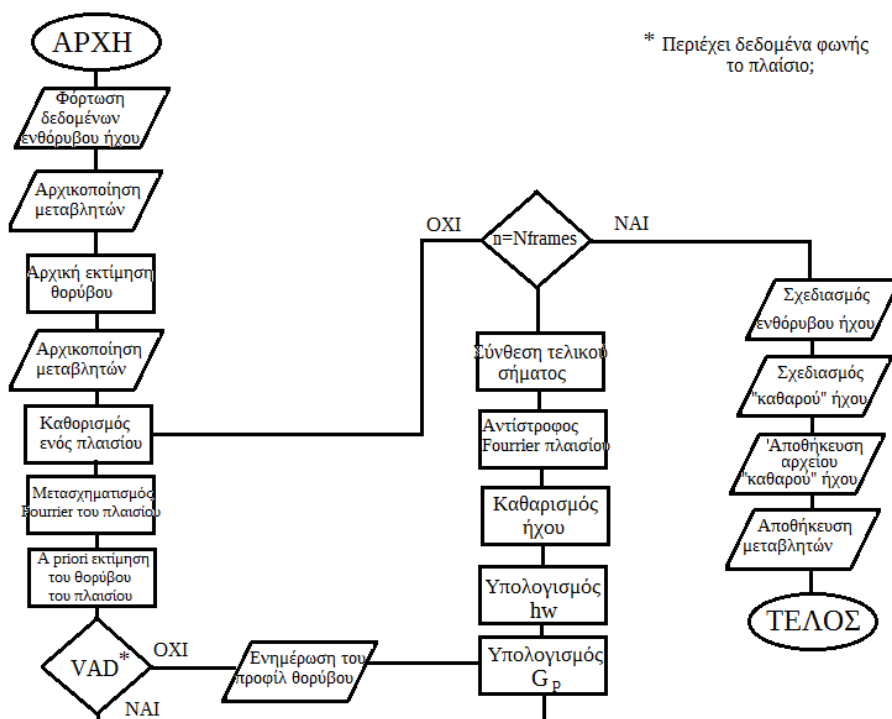
Οπότε, ο “καθαρισμένος” ήχος υπολογίζεται βάσει της σχέσης 4.10 αφού έχει υπολογιστεί το κέρδος φίλτρου από την σχέση 4.12 όπου η τιμή του  $\alpha$  ορίζεται ίση με 2 και του  $\beta$  ίση με 1. Για τον υπολογισμό όμως αυτό απαιτείται μία εκτίμηση του θορύβου  $|\tilde{V}[k, m]|$ . Η εκτίμηση αυτή εκτελείται σε δύο διαδοχικά στάδια:

1. Θεωρούμε ότι η ανάλυση των πρώτων πλαισίων περιέχουν μόνο θόρυβο. Έτσι υπολογίζεται στα τυφλά μία εκτίμηση θορύβου εντός αυτών των πλαισίων το πλήθος των οποίων καθορίζεται από τον χρήστη.
2. Σε κάθε πλαίσιο διακρίνουμε αν αυτό περιέχει μόνο θόρυβο (χρησιμοποιώντας έναν ανιχνευτή ομιλίας, συμβολίζεται με VAD) και στην περίπτωση που εκτιμηθεί ότι είναι “μόνο θόρυβος” ενημερώνουμε την εκτίμηση του θορύβου.

Αφού λοιπόν έχει εκτιμηθεί ο θόρυβος (με τον πρώτο τρόπο αρχικά και κατά την εκτέλεση του προγράμματος με τον δεύτερο) εφαρμόζεται η σχέση 4.10 σε συνδυασμό με την σχέση 4.12 (όπου έχει γίνει η απλοποίηση του  $|Y [k, m]|$  από τον συνδυασμό των δύο τύπων), διαδοχικά, σε κάθε πλαίσιο του σήματος του ήχου και το τελικό σήμα (το σύνολο όλων των πλαισίων μετά την ολοκλήρωση της επεξεργασίας τους) επαναφέρεται στο πεδίο των συχνοτήτων με αντίστροφο μετασχηματισμό Fourier και γίνεται η επανασύσταση του από τα πλαίσια και αποτελεί το αποτέλεσμα (την έξοδο) του προγράμματος.

Η έξοδος του προγράμματος αυτού σχεδιάζεται σε γραφική παράσταση ώστε να συγκριθεί με το αρχικό σήμα που έχει σχεδιαστεί και εξάγεται σε ένα .wav αρχείο, το όνομα του οποίου έχει καθορίσει ο χρήστης κατά την κλήση του.

Η παραπάνω διαδικασία εμφανίζεται στην παρακάτω εικόνα (Εικόνα 4.5) όπου

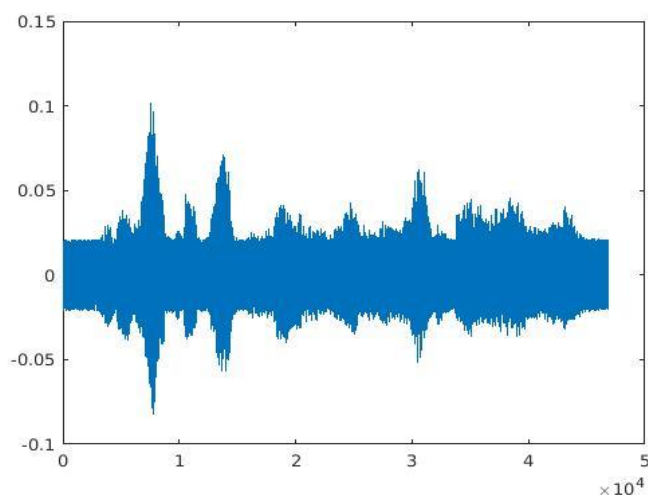


φαίνεται το αναλυτικό διάγραμμα ροής του αλγορίθμου της.

Εικόνα 4.5: Διάγραμμα ροής του αλγορίθμου

### 4.3.1-Εφαρμογή του αλγορίθμου του προγράμματος αφαίρεσης θορύβου με χρήση φασματικής αφαίρεσης

Στην εφαρμογή του παραπάνω κώδικα, χρησιμοποιήθηκε το “θορυβώδες” αρχείο ήχου “noisy\_sound.wav” που παράχθηκε από την εφαρμογή που περιγράφηκε στην παράγραφο 4.1, του οποίου η διάρκεια είναι  $\approx 2$ sec. Το διάγραμμα των δεδομένων αυτών φαίνεται στην παρακάτω εικόνα (Εικόνα 4.6):

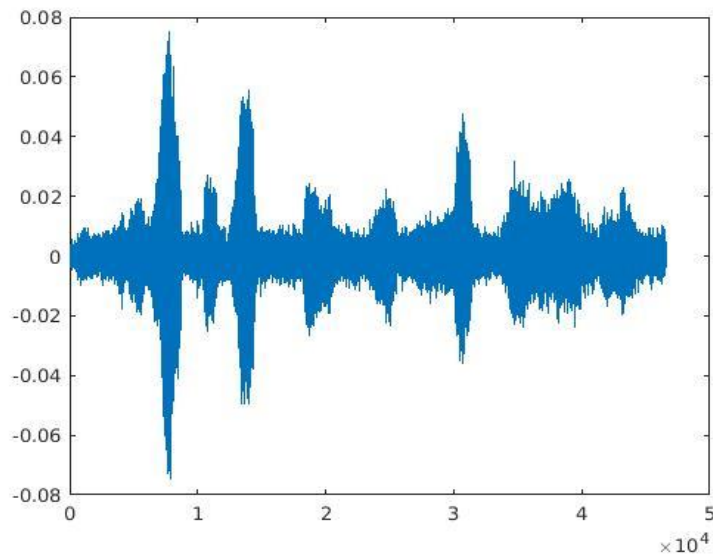


Εικόνα 4.6: Αρχικός θορυβώδης ήχος

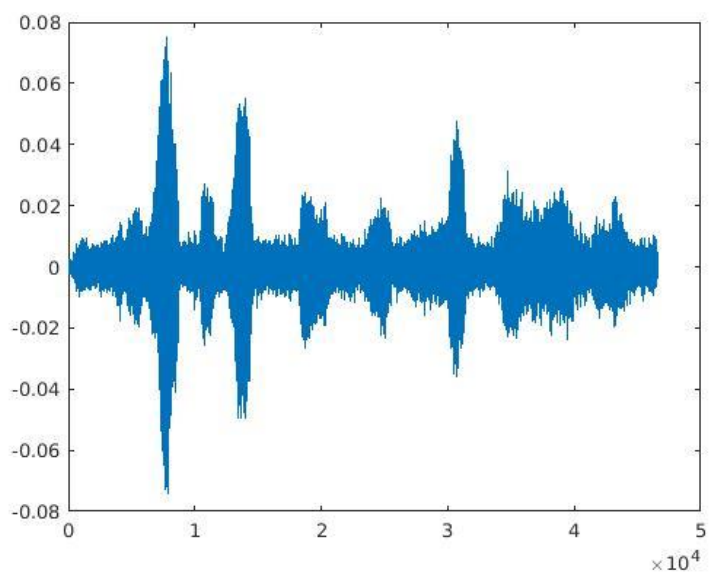
Οι τιμές που εισήχθησαν για την εκτέλεση του προγράμματος είναι:

- 10 πλαίσια θεωρούνται ως “καθαρός θόρυβος”
- $\alpha=2$
- $\beta=1$

Το αποτέλεσμα του προγράμματος, με χρήση της σχέσης 4.12 για τον υπολογισμό του  $G_p$  και ορισμός του  $L=8$  (Σχέση 4.13) είναι ο καθαρισμένος ήχος που φαίνεται στο διάγραμμα της παρακάτω εικόνας (Εικόνα 4.8), όπου παρατηρούμε ότι έχει περιοριστεί η ένταση του ήχου, οπότε έχει αφαιρεθεί μεγάλο μέρος του θορύβου. Με αναπαραγωγή του διαπιστώνουμε την μείωση του θορύβου συγκριτικά με την ομιλία, δηλαδή τον καθαρισμό του από τον θόρυβο.



Εικόνα 4.7: “Καθαρός” ήχος (με χρήση των εξισώσεων των σχέσεων 4.12 και 4.13)



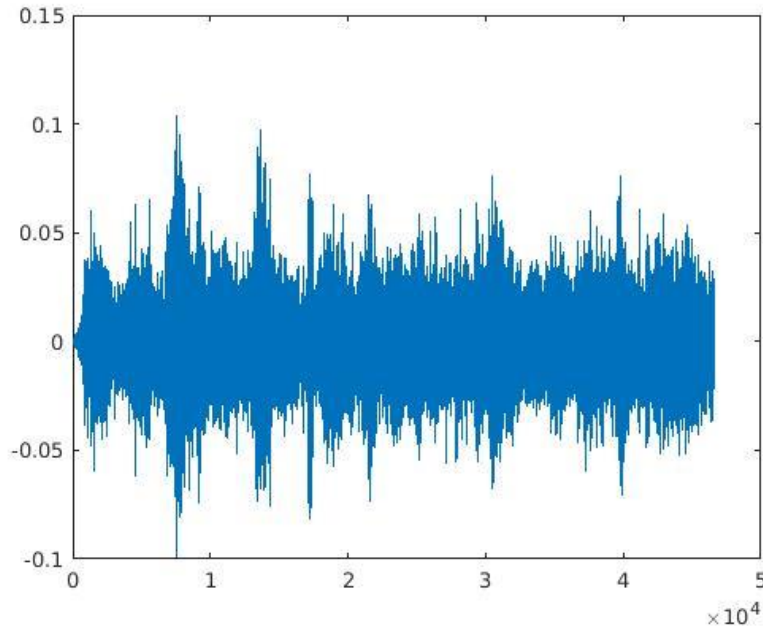
Εικόνα 4.8: “Καθαρός” ήχος (με χρήση της εξίσωσης της σχέσης 4.12 αλλά όχι της 4.13)



Στην συνέχεια, “απενεργοποιήθηκε” η εξίσωση της σχέσης 4.13, οπότε το κέρδος  $G_p$  δεν περιορίζεται από το  $L$ , και εκτελέστηκε ξανά το πρόγραμμα με το ίδιο αρχικό αρχείο ήχου.

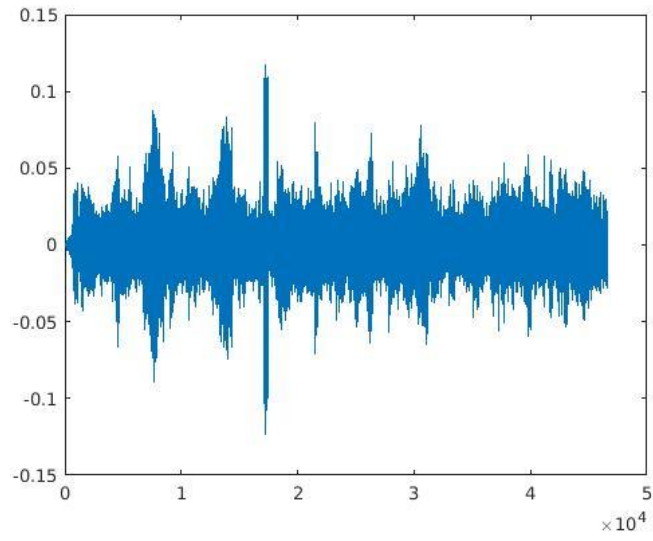
Το διάγραμμα του “καθαρού” ήχου που παράχθηκε από την εκτέλεση αυτή παρουσιάζεται στην παραπάνω εικόνα (Εικόνα 4.8). Όπως γίνεται φανερό από το διάγραμμα αυτό, δεν υπάρχει κάποια αισθητή διαφορά από την προηγούμενη περίπτωση. Αυτό οφείλεται στο ότι το αρχείο δεν ήταν πολύ “θορυβώδες” και ο θόρυβος που περιείχε ήταν λευκός θόρυβος.

Στην συνέχεια, εκτελέστηκε το πρόγραμμα και με χρήση της εξίσωσης της σχέσης 4.11 για τον υπολογισμό του  $G_p$ . Το αποτέλεσμα της εκτέλεσης του αλγορίθμου, αντικαθιστώντας την σχέση 4.11 με την 4.10 και χρησιμοποιώντας την τιμή 10 για την εξομάλυνση (χρησιμοποιώντας έναν μέσο όρο 10 σημείων) παρουσιάζεται στην παρακάτω εικόνα (Εικόνα 4.9), όπου από το διάγραμμα του ήχου φαίνεται ότι έχει αφαιρεθεί μικρότερο ποσοστό του θορύβου, όπως εύκολα γίνεται αντιληπτό και από την αναπαραγωγή του αρχείου του “καθαρού” ήχου.

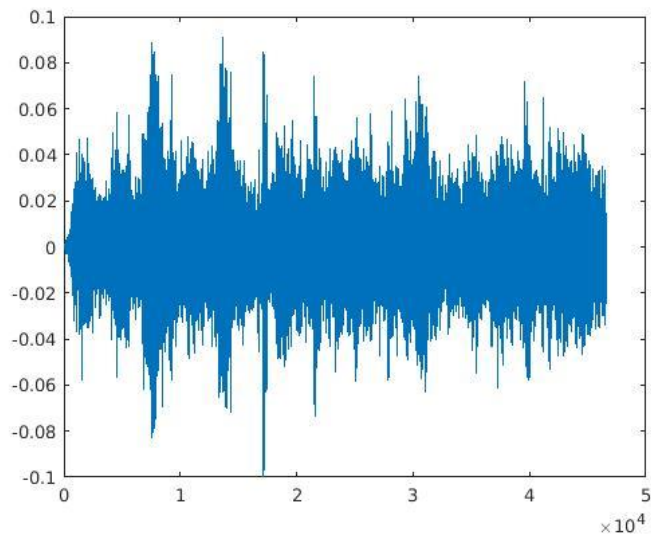


Εικόνα 4.9: “Καθαρός” ήχος (με χρήση της σχέσης 4.11, με εξομάλυνση 10 σημείων)

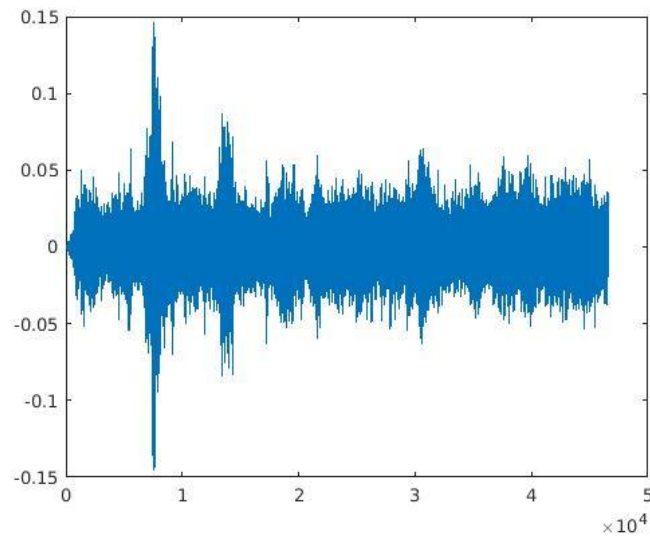
Τέλος, έγινε η εκτέλεση του προγράμματος για την περίπτωση που χρησιμοποιείται απευθείας η σχέση 4.11 χωρίς εξομάλυνση, καθώς και για τις περιπτώσεις όπου χρησιμοποιείται εξομάλυνση 5, 20 και 40 σημείων. Τα διαγράμματα των παραγόμενων αρχείων ήχου παρουσιάζονται στις παρακάτω εικόνες (Εικόνες 4.10, 4.11, 4.12 και 4.13 αντίστοιχα).



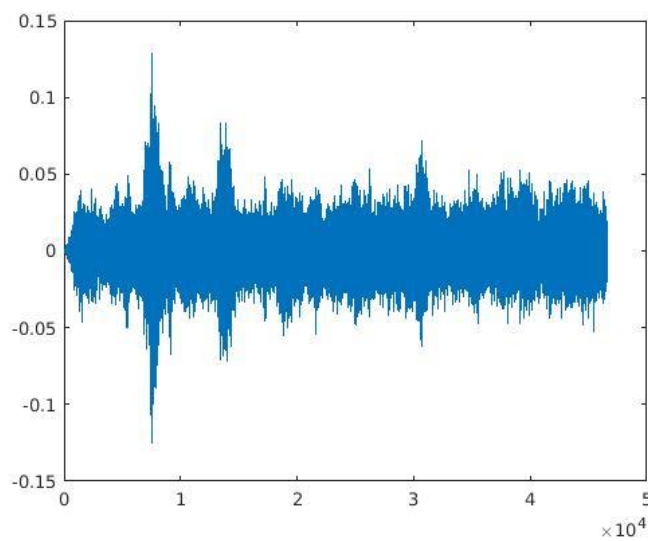
Εικόνα 4.10: “Καθαρός” ήχος (με χρήση της σχέσης 4.11 χωρίς εξομάλυνση)



Εικόνα 4.11: “Καθαρός” ήχος (με χρήση της σχέσης 4.11 με εξομάλυνση 5 σημείων)



Εικόνα 4.12: “Καθαρός” ήχος (με χρήση της σχέσης 4.11 με εξομάλυνση 20 σημείων)



Εικόνα 4.13: “Καθαρός” ήχος (με χρήση της σχέσης 4.11 με εξομάλυνση 40 σημείων)

Από τα παραπάνω διαγράμματα, φαίνεται ότι η εξομάλυνση του κέρδους του φίλτρου βελτιώνει τον καθαρισμό του ήχου, μειώνοντας περισσότερο τον θόρυβο στον τελικό ήχο, συγκριτικά με την περίπτωση που δεν χρησιμοποιείται η smooth. Αυτό γίνεται αντιληπτό και με αναπαραγωγή του τελικού αρχείου, όπου στην περίπτωση όπου δεν χρησιμοποιείται η

smooth, ο ήχος της φωνής είναι πιο θορυβώδης από τις περιπτώσεις όπου χρησιμοποιείται. Τέλος, το καλύτερο αποτέλεσμα, αναφορικά με τον αναπαραγόμενο ήχο, εμφανίστηκε στις περιπτώσεις όπου έγινε εξομάλυνση 10 και 20 σημείων (Εικόνα 4.9 και 4.12 αντίστοιχα).

## **ΚΕΦΑΛΑΙΟ 5**

### **5.1-Συμπεράσματα**

Στα προηγούμενα κεφάλαια της παρούσας εργασίας έγινε η περιγραφή των μεθόδων στις οποίες στηρίζεται η αποθορυβοποίηση των σημάτων ομιλίας. Λόγω όμως της μεγάλης ποικιλίας των εφαρμογών στις οποίες καλούνται να εφαρμοστούν οι αλγόριθμοι αποθορυβοποίησης, και το πλήθος των αλγορίθμων αυτών είναι μεγάλο, με πολλούς τέτοιους αλγόριθμους να αποτελούν απλά παραλλαγή κάποιας συγκεκριμένης μεθόδου.

Ως επί το πλείστον, είναι απαραίτητη μία μέθοδος μετασχηματισμού του σήματος ομιλίας από το πεδίο του χρόνου στο πεδίο των συχνοτήτων έτσι ώστε να ανακαλυφθεί το φασματικό του περιεχόμενο και να υποστεί την κατάλληλη επεξεργασία. Για την καλύτερη εποπτεία αυτού, αναλύθηκε ένας συγκεκριμένος αλγόριθμος, ο αλγόριθμος των AlexisFavrot και ChristoffFaller, οι οποίοι χρησιμοποιώντας έναν συνδυασμό των τεχνικών που αναφέρθηκαν, επιτυγχάνουν την κατασκευή ενός κατάλληλου φίλτρου για την αποθορυβοποίηση σημάτων ομιλίας.

Από τον αλγόριθμο που παρουσιάστηκε καθώς και την απλή εφαρμογή που υλοποιήθηκε γίνεται φανερό ότι η αφαίρεση του θορύβου όταν αυτός είναι γνωστός είναι μία αρκετά απλή διαδικασία. Όταν όμως η εφαρμογή έχει μεγαλύτερες απαιτήσεις, η λύση που θα δοθεί θα εξαρτάται από τα ιδιαίτερα χαρακτηριστικά της εφαρμογής και θα χρησιμοποιεί περισσότερες από μία από τις μεθόδους που αναφέρθηκαν στην παρούσα εργασία.

## ΠΑΡΑΡΤΗΜΑ Α΄

### Κώδικας Προγράμματος Προσθήκης Θορύβου

```
clear all

close all

%Load data
[S,Sfs]=audioread('LDC93S1.wav');
[WN, WNfs]=audioread('white_noise.wav');

if length(S)<length(WN)
    WN=WN(1:length(S),:);
elseif length(S)>length(WN)
    S=S(1:length(WN),:);
end

%calculate energy
Espeech=sum(S.^2);
Enoise=sum(WN.^2);

%calculate SNR
SNR=10*log2(Espeech/Enoise');

%calculate a
SNR_des=-3;
Enoise_des=Espeech/(10^(SNR_des/10));
a=sqrt(Enoise_des/Enoise);

%construct final noisy sound
sig_n=S+WN*a;

%save noisy sound
wavwrite(sig_n,Sfs,'noisy_sound.wav');
```

## ΠΑΡΑΡΤΗΜΑ Β΄

### Κώδικας Προγράμματος Αφαίρεσης Θορύβου

```
close all

%load sound data
[s,fs]=audioread('LDC93S1.wav');
[n,fn]=audioread('white_noise.wav');
[s_n,fs_n]=audioread('noisy_sound.wav');

%plot data (time domain)
figure(1)
subplot(3,1,1)
plot(s)
title('Clear sound')
subplot(3,1,2)
plot(n)
title('White Noise')
subplot(3,1,3)
plot(s_n)
title('Noisy Sound')

%plot data (frequency domain)
figure(2)
subplot(3,1,1)
plot(real(fft(s)))
title('Clear sound')
subplot(3,1,2)
plot(real(fft(n)))
title('White Noise')
subplot(3,1,3)
plot(real(fft(s_n)))
```



```

title('Noisy Sound')

%design filter
[b,a]=butter(6,0.3,'low');

%filter noisy sound
c_f=filter(b,a,s_n);

%plot noisy and cleared data

figure(3)
subplot(2,1,1)
plot(s_n)
title('Noisy Sound')
subplot(2,1,2)
plot(c_f)
title('Cleared Sound')

%plot noisy and cleared data (frequency domain)

figure(4)
subplot(2,1,1)
plot(real(fft(s_n)))
title('Noisy Sound')
subplot(2,1,2)
plot(real(fft(c_f)))
title('Cleared Sound')

%create audio file
audiowrite('cleared_sound.wav',c_f,fs);

```

## ΠΑΡΑΡΤΗΜΑ Γ΄

### Κώδικας Προγράμματος Αφαίρεσης Θορύβου με χρήση φασματικής αφαίρεσης

```
function [x]=remove_noise(filename,outfile)

%Step 1
if nargin<2
fprintf('Usage: remove_noise(noisyfile.wav,outFile.wav) \n\n');
return;
end

close all

[x,Srate]=audioread(filename);

%Step 2
len = floor(20*Srate/1000); % Frame size in samples
if rem(len,2)==1, len=len+1; end;
PERC = 50; % window overlap in percent of frame size
len1 = floor(len*PERC/100);
len2 = len-len1;

alpha = 2.0; % power exponent
beta = 1;
L = 8;
noise_frames = 3;

win = hamming(len); % define window
```

```

%Step 3

nFFT=2*2^nextpow2(len);

noise_mean = zeros(nFFT,1);

j = 1;

for j=1:noise_frames

noise_mean = noise_mean+abs(fft(win.*x(j:j+len-1),nFFT));

    j = j+len;

end

noise_mu = noise_mean/noise_frames;

noise_mu2 = noise_mu.^2;

```

```

%Step 4

k = 1;

img = sqrt(-1);

aa = 0.98;

mu = 0.98;

eta = 0.15;

ksi_min = 10^(-25/10);

x_old = zeros(len1,1);

Nframes = floor(length(x)/len2)-1;

xfinal = zeros(Nframes*len2,1);

for n=1:Nframes

```

```

%Step 5

insign = win.*x(k:k+len-1);      %Windowing

spec = fft(insign,nFFT);        %compute fourier transform of a frame

sig = abs(spec); % compute the magnitude

sig2 = sig.^2;

```

```

theta = angle(spec);

%Step 6
gammak = min(sig2./noise_mu2,40); % limit post SNR to avoid overflows
if n==1
ksi = aa+(1-aa)*max(gammak-1,0);
else
ksi = aa*Xk_prev./noise_mu2 + (1-aa)*max(gammak-1,0); % a priori SNR
ksi = max(ksi_min,ksi); % limit ksi to -25 dB
end

%Step 7
log_sigma_k = gammak.*ksi./ (1+ ksi)- log(1+ ksi);
vad_decision = sum(log_sigma_k)/ len;
if (vad_decision< eta)
% noise only frame found
noise_mu2 = mu* noise_mu2+ (1- mu)* sig2;
noise_mu = sqrt(noise_mu2);
end

%Step 8
A = ksi./(1+ksi); % Log Minimum mean-square estimator (Log-MMSE) estimator
vk = A.*gammak;
ei_vk = 0.5*expint(vk);
hw = A.*exp(ei_vk);

%Step 9
Gp(n,:) = smooth(((abs(sig.^2 -
noise_mu.^2))./(sig.^2)).^(1/2),10);%equation 6 with smooth

```

```
%Gp(n,:) = ((abs(sig.^2 - noise_mu.^2))./(sig.^2)).^(1/2);%equation 6 with-  
out smooth
```

```
%Gp(n,:) = ((sig.^alpha - beta*noise_mu.^alpha)./(sig.^alpha)).^(1/alpha);  
%equation 7
```

```
%Gp(n,:)=max(10^(-L/20),Gp(n,:));%equation 8
```

```
%Step 10
```

```
sig = sig.*hw.*Gp(n,:);
```

```
Xk_prev = sig.^2;
```

```
%Step 11
```

```
xi_w = ifft( sig .* exp(img*theta),nFFT);
```

```
xi_w = real( xi_w);
```

```
%Step 12
```

```
xfinal(k:k+len2-1) = x_old+xi_w(1:len1);
```

```
x_old = xi_w(1+len1:len);
```

```
%Step 13
```

```
k = k+len2;
```

```
end
```

```
%Step 14
```

```
figure(1)
```

```
plot(x)
```

```
%Step 15
```

```
figure(2)
plot(xfinal)
```

```
%Step 16
```

```
audiowrite(outfile,xfinal,Srate);
```

```
savevalues.mat
```

## Πηγές

1. Randall, R. B, “A History of Cepstrum Analysis and its Application to Mechanical Problems”, School of Mechanical and Manufacturing Engineering University of New South Wales,
2. B.P. Bogert, M.J.R.Healy, and J.W. Tukey, The Quefrency Alanysis of Time series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and SapheCracking,in Proc. of theSymp. On Time Series Analysis , by M. Rosenblatt (Ed.),. Wiley, NY, (1963), pp. 209-243.
3. S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE trans. Acoust. SpeechSig. Processing, vol. 27, no. 2, pp. 113–120, Nov. 1979.
4. M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by musical noise,” in Proc. ICASSP, 1979, pp. 208–211.
5. C. Favrot, Alexis, Faller, "Perceptually motivated gain filter smoothing for noise suppression", Proc. AudioEng. Soc. Conv. 123, 2007-Oct.
6. Vaseghi, S. V. , [2008] “Advanced Digital Signal Processing and Noise Reduction”, John Wiley & Sons
7. Satorius, E.H., “Noise cancellation via linear prediction filtering, Acoustics”, Speech, and Signal Processing, IEEE International Conference on ICASSP '79.
8. Lakshmikanth.S ,Natraj.K.R , Rekha.K.R , “Noise Cancellation in Speech Signal Processing-A Review” International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 1, January 2014
9. Chunshien Li, “Soft computing approach to adaptive noise filtering”, Cybernetics and Intelligent Systems, 2004 IEEE Conference
10. Rahman, M.T., “Noise cancellation in time and frequency domain using neural networks”, System Theory, 1994, Proceedings of the 26th Southeastern Symposium.

11. Lowerre, J.M., “Time domain use of the EM algorithm in noise cancellation”, *Signal Processing, IEEE Transactions on* (Volume: 39, Issue: 4), 1991.
12. Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” *IEEE Trans. Acoust., Speech, SignalProcessing*, vol. ASSP-32, pp. 1109–1121, Dec. 1984.
13. Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean square error log-spectral amplitude estimator,” *IEEE Trans. Acoust., Speech, SignalProcessing*, vol. ASSP-33, pp. 443–445, Apr. 1985.
14. Loizou, P. (2007). *Speech Enhancement: Theory and Practice*, CRC Press, Boca Raton – Florida.
15. Saeed V. Vaseghi. (2007). *Multimedia signal processing: theory and applications in speech, music and communications*, Wiley, England.