

**Τμήμα  
Μηχανικών  
Πληροφορικής τ.ε.**

Τεχνολογικό Εκπαιδευτικό Ίδρυμα  
Δυτικής Ελλάδας

**Πτυχιακή Εργασία**

Στο πλαίσιο του Προγράμματος Μεταπτυχιακών Σπουδών

«Τεχνολογίες και Συστήματα Ευρυζωνικών Εφαρμογών και Υπηρεσιών»

**«Πρόγραμμα καταστολής θορύβου σε σήματα ομιλίας με χρήση φίλτρου εξομάλυνσης,  
υλοποιημένο στο προγραμματιστικό περιβάλλον MatLab»**

**Θεοδώρα -Διαμάντω Τσιαντή**

**A.M.:M7**

**Επιβλέποντες:**

**Επίκ. Καθ. Μιχάλης Παρασκευάς**

**Δρ. Θωμάς Ζαρούχας**

**Αντίρριο, 2017**

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή  
Αντίρριο, Ημερομηνία

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Ονοματεπώνυμο, Υπογραφή
2. Ονοματεπώνυμο, Υπογραφή
3. Ονοματεπώνυμο, Υπογραφή

## ΕΥΧΑΡΙΣΤΙΕΣ

Η διπλωματική εργασία πραγματοποιήθηκε υπό την επίβλεψη των καθηγητών κ. Μιχάλη Παρασκευά και τον Δρ. Θωμά Ζαρούχα τους οποίους και ευχαριστώ θερμά για την καθοδήγησή τους, τις υποδείξεις και τον χρόνο που προσέφεραν για την προσπάθεια αυτή που έπαιξαν καθοριστικό ρόλο στην πορεία αυτής της εργασίας.

**Αντίρριο, Ιανουάριος 2017**  
**Θεοδώρα-Διαμάντω Τσιαντή**

## Πίνακας Περιεχομένων

<b>Κεφάλαιο 1-Παρουσίαση αλγορίθμων ενίσχυσης ομιλίας</b> .....	1
1.1 Εφαρμογές αλγορίθμων ενίσχυσης ομιλίας.....	14
1.2 Θόρυβος .....	14
1.2.1 Πηγή Θορύβου.....	14
1.2.2 Επίπεδα Θορύβου.....	15
1.3 Κατηγορίες Τεχνικών αποθορυβοποίησης – Αλγόριθμοι ενίσχυσης ομιλίας.....	15
<b>Κεφάλαιο 2 -Βασικές έννοιες Ακουστικής &amp; το ανθρώπινο ακουστικό σύστημα</b> .....	18
2.1 Ηχος.....	18
2.1.1 Ηχηρότητα (Loudness).....	23
2.1.2 Μουσικό ύψος του ήχου (Pitch).....	25
2.1.3 Χροιά ή Ηχόχρωμα (Timbre or Tone colour).....	26
2.2 Το ανθρώπινο ακουστικό σύστημα.....	27
2.2.1 Ευαισθησία του αυτιού.....	28
2.2.2 Η ανατομία του αυτιού.....	28
2.2.2.1 Το περύγιο: Κατευθυντικός κωδικοποιητής ήχου.....	29
2.2.2.2 Το κανάλι του αυτιού.....	30
2.2.2.3 Το μέσο αυτί.....	31
2.2.2.4 Το έσω αυτί.....	32
<b>Κεφάλαιο 3 Αλγόριθμοι Φασματικής Αφαίρεσης (Spectral Subtractive Algorithms)</b> .....	34
3.1 Βασικές αρχές της φασματικής αφαίρεσης.....	35
<b>Κεφάλαιο 4 -ΨΥΧΟΑΚΟΥΣΤΙΚΗ (PSYCOACOUSTICS)</b> .....	38
4.1 Βασικά στοιχεία.....	38
4.2 Ορια αντίληψης.....	39
4.3 MASKING EFFECT.....	39
<b>Κεφάλαιο 5-</b>	
5.1 Υλοποίηση Αλγορίθμων Αποθορυβοποίησης .....	40
5.2 Εκτίμηση θορύβου και ανίχνευση ομιλίας / ησυχίας.....	41
5.2.1 Ανιχνευτής δραστηριότητας ομιλίας φασματικής απόστασης (Spectral Distance Voice Activity Detector).....	41
Ο αλγόριθμος του Boll.....	42
Ο αλγόριθμος του Berouti.....	43
<b>Κεφάλαιο 6-ΥΛΟΠΟΙΗΣΗ ΣΕ MATLAB</b> .....	46

BIBΛΙΟΓΡΑΦΙΑ.....	71
<b>Κατάλογος Εικόνων</b>	
Σχήμα 2.1 Αραιώσεις και πυκνώσεις κατά τη διάδοση ενός κύματος.....	18
Σχήμα 2.2 Ο ήχος ως κύμα πίεσης.....	19
Σχήμα 2.3 Συχνότητες νοτών στο πιάνο.....	19
Σχήμα 2.4 Κυματομορφή στο πεδίο του χρόνου και στο πεδίο της συχνότητας .....	20
Σχήμα 2.5 Σύγκριση αρμονικών και οκτάβων. Οι αρμονικές είναι γραμμικά συσχετισμένες ενώ οι οκτάβες λογαριθμικά.....	22
Σχήμα 2.6 Μήκος κύματος και πλάτος κυματομορφής. ....	23
Σχήμα 2.7 Καμπύλες ίσης ηχηρότητας. ....	24
Σχήμα 2.8 Μηχανισμός αντιστάθμισης του αυτιού.....	25
Σχήμα 2.9 Δεσμοί και Αντί-δεσμοί.....	26
Σχήμα 2.10 Η ανατομία του ανθρώπινου αυτιού.....	29
Σχήμα 2.11 Το ακουστικό κανάλι, κλειστό στο ένα άκρο από το τύμπανο, λειτουργεί σαν “σωλήνας μουσικού οργάνου” ενός τετάρτου μήκους κύματος. Ο συντονισμός δίνει ακουστική ενίσχυση για τις σημαντικές συχνότητες φωνής.....	30
Σχήμα 2.12 Η συνάρτηση μεταφοράς (απόκριση συχνότητας) του καναλιού του αυτιού. Πρόκειται για μια σταθερή συνιστώσα που προστίθεται με κάθε κατευθυντικά κωδικοποιημένο ήχο που φτάνει στο τύμπανο.....	31
Σχήμα 2.13 Αριστερά: Τα μικρά οστά (σφύρα, άκμονας, αναβολέας) του μέσου αυτιού, που μεταδίδουν μηχανικές ταλαντώσεις του τυμπάνου στο ελλειψοειδές παράθυρο του κοχλίου. Δεξιά: Μηχανικό ανάλογο της λειτουργίας προσαρμογής αντιστάσεων στο μέσο αυτί. Η διαφορά επιφάνειας μεταξύ τυμπάνου και ελλειψοειδούς παραθύρου, μαζί με την μηχανική σύνδεση υποβιβασμού, προσαρμόζουν την κίνηση του τυμπάνου που ενεργοποιείται από τον αέρα στο ελλειψοειδές παράθυρο με υγρό .....	31
Σχήμα 2.14 Εξδανικευμένο σχέδιο του ανθρώπινου αυτιού όπου φαίνεται ξετυλιγμένος ο γεμάτος υγρό κοχλίας. Ο ήχος που μπαίνει στο κανάλι του αυτιού προκαλεί την ταλάντωση του τυμπάνου. Αυτή η ταλάντωση μεταδίδεται στον κοχλίο μέσω της μηχανικής σύνδεσης του μέσου αυτιού. Ο ήχος αναχίμα Εξιδιαιρείται μέσω στάσιμων κυμάτων που δημιουργούνται στην βασιλική (βασική) μεμβράνη.....	32
Σχήμα 3.1 Ενίσχυση ήχων με υπολογισμό του μέτρου του φάσματος .....	34
Σχήμα 3.2 Γενική μορφή αλγορίθμου φασματικής αφαίρεσης .....	37
Σχήμα 5.1 Γενικό Λογικό διάγραμμα αλγορίθμων αποθορυβοποίησης .....	40
Σχήμα 5.2 Αλγόριθμος Boll.....	43
Σχήμα 5.3 Μέθοδος Berouti (Φασματική Αφαίρεση με χρήση Φασματικού Δαπέδου) .....	44
Σχήμα 1: Σχηματικό διάγραμμα του αλγορίθμου.....	48
Σχήμα 2: Διάγραμμα ροής του αλγορίθμου του προγράμματος που υλοποιήθηκε .....	53
Σχήμα 3: Αρχικός ήχος.....	54

Σχήμα 4: “Καθαρός” ήχος (με χρήση των εξισώσεων (7) και (8) ).....	55
Σχήμα 5: “Καθαρός” ήχος (με χρήση της εξίσωσης (7) αλλά όχι της (8) ).....	56
Σχήμα 6: “Καθαρός” ήχος (με χρήση της εξίσωσης (6) με smooth 10 σημείων).....	57
Σχήμα 7: “Καθαρός” ήχος (με χρήση της εξίσωσης (6) χωρίς εξομάλυνση).....	57
Σχήμα 8: “Καθαρός” ήχος (με χρήση της εξίσωσης (6) με smooth 5 σημείων).....	58
Σχήμα 9: “Καθαρός” ήχος (με χρήση της εξίσωσης (6) με smooth 20 σημείων).....	58
Σχήμα 10: “Καθαρός” ήχος (με χρήση της εξίσωσης (6) με smooth 40 σημείων).....	59



## Κατάλογος Πινάκων

Πίνακας 2.1 Αρμονικές, δεσμοί και αντί-δεσμοί.....	27
--	----

## **Συντμήσεις Όρων**

**SNR - Signal to Noise Ratio**

**NN -Neural Networks**

**HMM-Hidden Markof Models**

**MMSE algorithms -αλγόριθμοι ελάχιστου μέσου τετραγωνικού σφάλματος**

**MLE -μέγιστης πιθανοφάνειας**

**Singular Value Decomposition- διάσπαση ιδιόμορφων τιμών**

**Hz-συχνότητα**

**Noise suppression- Απομάκρυνση του θορύβου από το σήμα**

**MCN - model compensation for noise**

**CF- Comb Filters**

**SValD-Singular Value Decomposition)**

**SpecSA-Spectral Subtractive Algorithms)**

**VAD (Voice Activity Detector)**

**Noise Counter- Μετρητής Θορύβου**

**Hangover period- Περίοδο Απόλυσης**





## ΚΕΦΑΛΑΙΟ 1

### ΠΑΡΟΥΣΙΑΣΗ ΑΛΓΟΡΙΘΜΩΝ ΕΝΙΣΧΥΣΗΣ ΟΜΙΛΙΑΣ

---

Ο τομέας της Ενίσχυσης Ομιλίας αρχικά είχε επικεντρώσει το ερευνητικό του ενδιαφέρον στην προσπάθεια αύξησης του ποσοστού απόδοσης των συστημάτων επικοινωνίας και ομιλίας, στα οποία ο θόρυβος εμφανιζόταν είτε ως σήμα εισόδου είτε ως σήμα εξόδου. Με την πάροδο του χρόνου όμως και την ραγδαία εξέλιξη της ψηφιακής τεχνολογίας εμφανίστηκε μία πληθώρα εφαρμογών οι οποίες υλοποιούσαν

τις διάφορες τεχνικές αποθορυβοποίησης, κάνοντας επιτακτική την ανάγκη για περαιτέρω ανάπτυξη της έρευνας στον συγκεκριμένο τομέα.

Η ανάγκη αυτή ικανοποιήθηκε με αποτέλεσμα τις τελευταίες τρεις δεκαετίες να έχουν παρουσιαστεί διάφοροι αποδοτικοί αλγόριθμοι επεξεργασίας ομιλίας για την βελτίωση της απόδοσης των σύγχρονων τηλεπικοινωνιακών συσκευών σε περιβάλλοντα θορύβου.

Στις περισσότερες των περιπτώσεων ο στόχος είναι να μειωθεί ή να αποκοπεί μέχρι σε κάποιο βαθμό ο προσθετικός περιβαλλοντικός θόρυβος βελτιώνοντας έτσι την ποιότητα του υποβαθμισμένου σήματος και καθιστώντας ευκολότερη την επικοινωνία. Όπως θα φανεί και παρακάτω ο στόχος της επεξεργασίας και ενίσχυσης ομιλίας ποικίλει εξαρτώμενος από την εφαρμογή στην οποία χρησιμοποιείται. Ιδανικά, θα θέλαμε οι αλγόριθμοι ενίσχυσης ομιλίας να βελτιώνουν τόσο την ποιότητα όσο και την καταληπτότητα του σήματος, γεγονός που δεν μπορεί να επιτευχθεί στην πράξη καθώς μειώνοντας τον περιβαλλοντικό θόρυβο εισάγεται παραμόρφωση της ομιλίας καθιστώντας την λιγότερο κατανοητή. Αυτή είναι η κύρια πρόκληση στην σχεδίαση αποτελεσματικών αλγορίθμων ενίσχυσης ομιλίας, πως δηλαδή θα αποκόψεις τον θόρυβο χωρίς ταυτόχρονα να εισάγεις επιπλέον παραμόρφωση στο χρήσιμο σήμα. Η λύση στο γενικό αυτό πρόβλημα της ενίσχυσης ομιλίας εξαρτάται σημαντικά από τις ιδιαιτερότητες της κάθε εφαρμογής, τα χαρακτηριστικά της πηγής θορύβου, την σχέση (αν υπάρχει) μεταξύ του θορύβου και του καθαρού σήματος και τον αριθμό των μικροφώνων και των αισθητήρων που είναι διαθέσιμοι. Συμπερασματικά μπορούμε να θεωρήσουμε ότι οι βασικοί στόχοι της αποθορυβοποίησης είναι:

- Η βελτίωση τόσο των υποκειμενικών όσο και των αντικειμενικών μεγεθών.
- Η αύξηση της σταθερότητας των συστημάτων συμπίεσης ομιλίας.
- Η αύξηση του ποσοστού απόδοσης των συστημάτων αναγνώρισης ομιλίας.

Οι δύο πρώτες κατηγορίες έχουν βασικό σκοπό την βελτίωση των αντιληπτών μεγεθών που καθορίζουν ένα σήμα ομιλίας, δηλαδή την ποιότητα και την καταληπτότητα αυτού. Τα δύο αυτά μεγέθη στην ουσία αποτελούν τα κριτήρια αξιολόγησης ενός συστήματος αποθορυβοποίησης. Η ποιότητα μετράει την καθαρότητα ενός επεξεργασμένου σήματος καθώς και τα επίπεδα του εναπομείναντος θορύβου μέσα σε αυτό. Αξίζει να αναφέρουμε ότι η ποιότητα αποτελεί μία υποκειμενική μέτρηση η οποία φανερώνει το σημείο το οποίο αποθορυβοποιημένο σήμα είναι υποφερτό στον εκάστοτε ακροατή. Η καταληπτότητα αναφέρεται στο ποσοστό των μονάδων ομιλίας που μπορεί να αναγνωρίσει ένας ακροατής. Φυσικά αυτή η μέτρηση χαρακτηρίζεται από αντικειμενικότητα

## 1.1 Εφαρμογές αλγορίθμων ενίσχυσης ομιλίας

Η ανάγκη για ενίσχυση σημάτων ομιλίας προκύπτει σε πολλές περιπτώσεις κατά τις οποίες το σήμα ομιλίας προέρχεται από μια θορυβώδη περιοχή ή είναι επηρεασμένο από θόρυβο μέσα από ένα επικοινωνιακό κανάλι. Υπάρχει μια ποικιλία από περιπτώσεις στις οποίες είναι επιθυμητό να εφαρμόσουμε κάποια μέθοδο αποθορυβοποίησης:

- Στην επικοινωνία μέσω συστημάτων κινητής τηλεφωνίας, για παράδειγμα, συνήθως υποφέρει από περιβαλλοντικό θόρυβο όπως αυτός παρουσιάζεται μέσα σε ένα αυτοκίνητο ή ένα εστιατόριο κτλ., στο άκρο εκπομπής. Οι αλγόριθμοι ενίσχυσης ομιλίας σε αυτήν την περίπτωση μπορούν να χρησιμοποιηθούν για την βελτίωση της ποιότητας ομιλίας στο άκρο λήψης.
- Στα συστήματα στενής ζώνης επικοινωνιών τα οποία λειτουργούν σε συνθήκες όπου ο θόρυβος είναι ιδιαίτερος έντονος.

- Στην επικοινωνία γη-αέρα οι τεχνικές ενίσχυσης ομιλίας είναι απαραίτητες για να ενισχύσουν την ποιότητα και κυρίως την καταληπτότητα της ομιλίας του πιλότου η οποία υπόκειται στα πολύ υψηλά επίπεδα θορύβου που επικρατούν στο πιλοτήριο. Σε αυτήν και άλλες παρόμοιες τηλεπικοινωνιακές εφαρμογές στρατιωτικής φύσεως δίνεται προτεραιότητα στην ενίσχυση της καταληπτότητας της ομιλίας.

- Στην ασύρματη και ενσύρματη επικοινωνία και ιδιαίτερος στις περιπτώσεις εκείνες που ο δέκτης βρίσκεται σε θορυβώδες περιβάλλον ή όταν το μέσο μετάδοσης περιέχει υψηλά ποσοστά θορύβου οποιασδήποτε μορφής.

- Σε ένα σύστημα τηλεδιασκέψεων θα διαπιστώσουμε μεγάλη βελτίωση στην ποιότητα του συστήματος εάν ενισχύσουμε το σήμα πριν την μετάδοση του.

- Οι άνθρωποι που χρησιμοποιούν βοηθήματα ακοής αντιμετωπίζουν σημαντικά προβλήματα σε θορυβώδη περιβάλλοντα, σε αυτές τις περιπτώσεις μπορούν να χρησιμοποιηθούν οι αλγόριθμοι ενίσχυσης ομιλίας και να προεπεξεργαστούν ή να “καθαρίσουν” το ενθόρυβο σήμα πριν αυτό ενισχυθεί.

- Σε συστήματα αναγνώρισης ομιλίας σε θορυβώδες περιβάλλον

## 1.2 Θόρυβος

Για να μπορέσουμε να σχεδιάσουμε αλγόριθμους ικανούς να μειώσουν τον θόρυβο είναι ουσιώδες να κατανοήσουμε την συμπεριφορά διάφορων τύπων θορύβου, τις πηγές θορύβου όπως αυτές περιγράφονται στα χρονικά και φασματικά χαρακτηριστικά τους, καθώς και το εύρος των επιπέδων θορύβου που μπορεί να συναντήσουμε στην πραγματική ζωή.

### 1.2.1 Πηγή Θορύβου

Ο θόρυβος μπορεί να είναι περιβαλλοντικός, προερχόμενος από παρεμβαλλόμενες ομιλίες ή από ανακλάσεις δωματίων, μπορεί επίσης να εισάγεται από το σύστημα κβαντισμού, ο θερμικός ή ακόμα και ο ηλεκτρονικός θόρυβος του καναλιού, ενώ μπορεί να είναι σχετιζόμενος ή ασυσχέτιστος με το σήμα ομιλίας. Τέλος ο θόρυβος εμφανίζεται ως προσθετικός, πολλαπλασιαστικός ή συνελκτικός.

Σε κάθε καθημερινή μας δραστηριότητα ο θόρυβος είναι παρών έχοντας διάφορες μορφές και σχήματα γεγονός που καταδεικνύεται και στα παρακάτω παραδείγματα. Στον δρόμο (π.χ. κυκλοφοριακή συμφόρηση, εργασίες κατασκευής), στο αυτοκίνητο (π.χ. θόρυβος μηχανής, άνεμος), στο γραφείο (π.χ. ανεμιστήρας υπολογιστή, αεραγωγός), στα εστιατόρια (π.χ. ομιλία από το διπλανό τραπέζι), στα καταστήματα (π.χ. κουδούνισμα τηλεφώνου, ομιλία πωλητή) και σε πολλές ακόμα περιπτώσεις ο θόρυβος συνοδεύει τις δραστηριότητές μας.

Ο θόρυβος μπορεί να είναι στάσιμο σήμα που σημαίνει ότι παραμένει αμετάβλητος κατά την πάροδο του χρόνου όπως για παράδειγμα ο θόρυβος που προέρχεται από τον ανεμιστήρα του υπολογιστή. Επίσης ο θόρυβος μπορεί να εμφανιστεί και ως μη στάσιμο σήμα, όπως ο θόρυβος του εστιατορίου, στον οποίο ομιλίες από το παρασκήνιο μπλέκονται με τον θόρυβο της κουζίνας. Τα φασματικά χαρακτηριστικά του θορύβου του εστιατορίου συνεχώς αλλάζουν καθώς οι άνθρωποι συνεχίζουν να συζητούν στα γειτονικά τραπέζια και οι σερβιτόροι αλληλεπιδρούν με τους πελάτες. Είναι προφανές ότι η διαδικασία της μείωσης του θορύβου όταν αυτός συνεχώς μεταβάλλεται (μη στάσιμος) είναι πολύ πιο δύσκολη από την διαδικασία μείωσης στάσιμου θορύβου.

Άλλο ένα χαρακτηριστικό που ξεχωρίζει τους διάφορους τύπους θορύβου είναι η μορφή του φάσματος και πιο συγκεκριμένα η κατανομή της ενέργειας του θορύβου στο πεδίο της συχνότητας.

## 1.2.2 Επίπεδα Θορύβου

Σημαντική για την σχεδίαση αλγορίθμων ενίσχυσης ομιλίας είναι η γνώση του εύρους των επιπέδων έντασης της ομιλίας και του θορύβου σε πραγματικές συνθήκες. Με αυτόν τον τρόπο μπορούμε να εκτιμήσουμε το εύρος των επιπέδων του λόγου σήματος προς θόρυβο (Signal to Noise Ratio, SNR) που συναντάμε σε πραγματικά περιβάλλοντα. Αυτό είναι σημαντικό καθώς οι αλγόριθμοι ενίσχυσης ομιλίας πρέπει να είναι αποτελεσματικοί στην μείωση του θορύβου και την βελτίωση της ποιότητας της ομιλίας μέσα στο εύρος των επιπέδων του SNR όπως αυτό έχει καθοριστεί.

## 1.3 Κατηγορίες Τεχνικών αποθορυβοποίησης – Αλγόριθμοι ενίσχυσης ομιλίας

Η εισαγωγή θορύβου στην ομιλία έχει ως αποτέλεσμα την εμφάνιση δύο βασικών φαινομένων: (α) την αλλοίωση του καθαρού σήματος η οποία συμβαίνει κυρίως στις περιπτώσεις προσθετικού ή συνελκτικού θορύβου και (β) το φαινόμενο Lombart, δηλαδή την αρνητική επίδραση του θορύβου στο κομμάτι της καταληπτότητας. Το ζήτημα αυτό προσεγγίστηκε από τις ακόλουθες οπτικές προκειμένου να επιλυθεί

- Αναζήτηση χαρακτηριστικών του σήματος ομιλίας τα οποία παρουσιάζουν υψηλότερο βαθμό αντοχής στο θόρυβο (noise resistance). Χρήση σθεναρού ελέγχου.
- Μείωση του ποσοστού απόκλισης του καθαρού σήματος ομιλίας από το ενθόρυβο, αυτή η εκδοχή οδήγησε σε δύο κατευθύνσεις
- Απομάκρυνση του θορύβου από το σήμα (Noise suppression)
- Δημιουργία μοντέλων ομιλίας τα οποία παρουσιάζουν προσαρμοστικότητα στον θόρυβο (model compensation for noise). Χρήση προσαρμοστικού ελέγχου.

Η λογική που διακατέχει το μοντέλο των σημάτων ομιλίας που παρουσιάζουν ανθεκτικότητα στον θόρυβο έχει ως εξής: αρχικά η θέσπιση παραμέτρων που παρουσιάζουν αυξημένη ευαισθησία στον θόρυβο. Στην συνέχεια η μελέτη του τρόπου με τον οποίο επιδρά ο θόρυβος σε αυτές. Τέλος χρήση χαρακτηριστικών παραμέτρων ή παραπλήσιων μετρήσεων οι οποίες είναι όσο το δυνατόν αναισθητες στον θόρυβο. Από μαθηματικής απόψεως η όλη διαδικασία στηρίζεται στη βελτιστοποίηση των συντελεστών Cepstrum που είναι στην ουσία ο αντίστροφος μετασχηματισμός Fourier του λογαριθμικού φάσματος του σήματος

Υλοποίηση αυτής της διαδικασίας αποτελούν οι αλγόριθμοι Φασματικής Επέκτασης οι οποίοι ενισχύουν κυρίως τις φασματικές κορυφές του σήματος και υποβαθμίζουν τις περιοχές χαμηλής συχνότητας. Η μελέτη του τρόπου ενίσχυσης της αντοχής των σημάτων ομιλίας έναντι στον θόρυβο αποτέλεσε το έναυσμα για μία σειρά από καινοτόμες ιδέες, όπως η τεμαχιοποίηση ομιλίας, η δημιουργία υπολογιστικών μοντέλων ομιλίας για ηχητικά συστήματα τα οποία παρουσιάζουν υψηλό βαθμό υπολογιστικής ακρίβειας, συμπερίζουν 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.

Η δεύτερη κατηγορία η οποία αποτελεί και το αντικείμενο της παρούσας εργασίας, είναι η Αναβάθμιση Ενθόρυβης Ομιλίας (Ενίσχυση Ομιλίας) και βασίζεται στην βελτίωση της ποιότητας της ομιλίας. Βρίσκει κατά κύριο λόγο εφαρμογή σε συστήματα αναγνώρισης ομιλίας αποτελώντας στάδιο προεπεξεργασίας. Βέβαια δεν είναι απόλυτο ότι ο ενισχυμένος λόγος σήματος προς θόρυβο (SNR) αυξάνει τον βαθμό απόδοσης των συστημάτων αναγνώρισης ανθρώπινης ομιλίας. Σε αυτόν τον τομέα επικεντρώθηκαν οι προσπάθειες της έρευνας η οποία οδήγησε σε εξελίξεις που άλλαξαν ριζικά τον τρόπο αντιμετώπισης της

αποθορυβοποίησης. Οι πιο σημαντικές από αυτές είναι τα παραμετρικά μοντέλα τα οποία βασίζονται σε νευρωνικά δίκτυα (Neural Networks -NN), η φασματική αφαίρεση όπου η απομάκρυνση του θορύβου γίνεται από το φάσμα ισχύος, τα Φίλτρα- Χτένα (Comb Filters) και πολλά άλλα.

Η τελευταία κατηγορία είναι αυτή των μοντέλων που προσαρμόζονται στον θόρυβο, τα γνωστά και ως HMM Hidden Markof Models), καθώς λαμβάνουν υπ' όψιν την παρουσία αυτού. Είναι στην ουσία παραμετρικά μοντέλα τα οποία βελτιστοποιούνται κατά την διάρκεια της εκάστοτε λειτουργίας που εκτελούν. Στην κατηγορία αυτή ανήκουν τα εξαρτώμενα-οδηγούμενα από τις καταστάσεις Wiener φίλτρα τα οποία χρησιμοποιούν τον λόγο φασματικής ισχύος του καθαρού σήματος ομιλίας προς το ενθόρυβο.

Παρακάτω Έχουμε ταξινομήσει τους αλγορίθμους σε τρεις κατηγορίες που επικεντρώνονται στην ενίσχυση ομιλίας έχοντας ως στόχο την ενίσχυση του σήματος :

1. **Spectral Subtractive Algorithms:** Ιστορικά είναι από τους πρώτους αλγορίθμους που έχουν χρησιμοποιηθεί για την μείωση του θορύβου, ενώ αποτελούν με διαφορά τους πιο απλούς αλγόριθμους που μπορούμε να εκτελέσουμε. Βασίζονται στο γεγονός ότι ο θόρυβος είναι προσθετικός και κατά συνέπεια μπορεί να υπολογιστεί το φάσμα του όταν δεν εμφανίζεται το σήμα ομιλίας και στην συνέχεια να το αφαιρέσουμε από το ενθόρυβο σήμα.

2. **Statistical-model-based Algorithms:** Το πρόβλημα της ενίσχυσης ομιλίας τοποθετείται στο πλαίσιο της στατιστικής εκτίμησης. Για παράδειγμα έστω ότι έχουμε ένα σετ μετρήσεων οι οποίες αντιστοιχούν στις παραμέτρους του μετασχηματισμού Fourier του σήματος θορύβου, Αυτό που επιθυμούμε να βρούμε είναι ένας γραμμικός (ή μη γραμμικός) εκτιμητής των παραμέτρων που μας ενδιαφέρουν, δηλαδή, τις παραμέτρους του μετασχηματισμού για το καθαρό σήμα. Οι αλγόριθμοι ελάχιστου μέσου τετραγωνικού σφάλματος (MMSE algorithms) και μέγιστης πιθανοφάνειας (MLE) υπάγονται σε αυτήν την κατηγορία, όπως επίσης και Wiener φίλτρα

3. **Subspace Algorithms:** Σε αντίθεση με τους αλγόριθμους των δυο προηγούμενων κατηγοριών, οι subspace αλγόριθμοι βασίζονται κυρίως στη θεωρία της γραμμικής άλγεβρας. Πιο συγκεκριμένα, αυτοί οι αλγόριθμοι βασίζονται στην αρχή ότι το καθαρό σήμα μπορεί να περιορίζεται σε έναν υποχώρο του ευκλείδειου χώρου όπως αυτός ορίζεται από το ενθόρυβο σήμα. Κατά συνέπεια αν μπορούσαμε να αναπτύξουμε μια μέθοδο αποσύνθεσης του διανύσματος του χώρου του ενθόρυβου σήματος σε έναν υποχώρο ο οποίος καταλαμβάνεται κυρίως από το καθαρό σήμα και αντίστοιχα έναν υποχώρο που καταλαμβάνεται από το σήμα του θορύβου, θα μπορούσαμε να εκτιμήσουμε το καθαρό σήμα απλά ακυρώνοντας την συνιστώσα του ενθόρυβου διανύσματος η οποία αντιστοιχεί στον “υποχώρο του θορύβου”. Η αποσύνθεση του διανύσματος του χώρου σε υποχώρους “σήματος” και “θορύβου” μπορεί να πραγματοποιηθεί χρησιμοποιώντας ευρέως γνωστές τεχνικές παραγοντοποίησης μητρών από την γραμμική άλγεβρα, και πιο συγκεκριμένα την διάσπαση ιδιόμορφων τιμών (Singular Value Decomposition) ή την παραγοντοποίηση ιδιοδιανύσματος-ιδιοτιμής

## Κεφάλαιο 2

### Βασικές έννοιες Ακουστικής & το ανθρώπινο ακουστικό σύστημα

Στο σημείο αυτό πριν ασχοληθούμε με το κύριο αντικείμενο της παρούσας εργασίας παραθέτουμε μερικές βασικές έννοιες μέσω των οποίων θα οδηγηθούμε σε καλύτερη κατανόηση στην συνέχεια της εργασίας.

#### 2.1 Ηχος

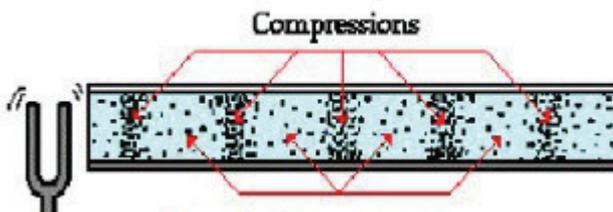
Ως ήχος μπορεί να οριστεί μια διάδοση κυμάτων στον αέρα ή σε άλλο ελαστικό μέσο ή ως εκείνη η

διέγερση του μηχανισμού ακρόασης που οδηγεί στην αντίληψη ενός ήχου (αίσθηση). Το ποιος ορισμός ισχύει εξαρτάται από το εάν η προσέγγιση είναι φυσική ή ψυχοακουστική. Ο τύπος του προβλήματος υπαγορεύει και την προσέγγιση στον ήχο. Εάν το ενδιαφέρον είναι στη διαταραχή του αέρα που δημιουργείται από ένα μεγάφωνο, τότε η μελέτη του ήχου ανάγεται σε ένα φυσικό πρόβλημα. Εάν το ενδιαφέρον είναι πώς ηχεί σε ένα πρόσωπο πλησίον στο μεγάφωνο, τότε θα πρέπει να χρησιμοποιηθούν ψυχοακουστικές μέθοδοι.

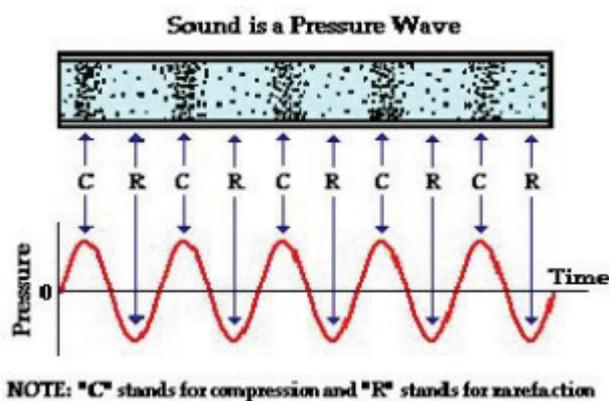
Οποιοσδήποτε ήχος, οποιουδήποτε είδους, προκαλείται από μια δόνηση. Χωρίς δόνηση δεν μπορεί να υπάρξει κανένας ήχος. Το δονούμενο σώμα αναγκάζει τα μόρια αέρα γύρω του να δονηθούν και εκείνα με τη σειρά τους μεταδίδουν την δόνηση στα διπλανά τους μόρια. Κατά αυτόν τον τρόπο μια διαταραχή του αέρα κινείται από την πηγή του ήχου και μπορεί τελικά να φθάσει στα αυτιά ενός ακροατή. Όταν ακούμε έναν ήχο, ο αέρας που φτάνει στα τύμπανά μας δονείται αναγκάζοντάς και αυτά να δονηθούν. Αυτές οι δονήσεις ανιχνεύονται και αναλύονται από τον εγκέφαλό μας. Αν και είναι συνήθως ο αέρας που ενεργεί ως διαβιβάζον μέσο, ο ήχος μπορεί να διαβιβαστεί και από άλλα μέσα, π.χ. στερεά, υγρά.

Ο ήχος δεν αποτελείται από τον αέρα που κινείται προς εμάς, αλλά ταξιδεύει μέσω του αέρα ως ένα ηχητικό κύμα. Ένα ηχητικό κύμα αποτελείται από μια διαταραχή που κινείται από μια πηγή προς τις περιβάλλουσες θέσεις με αποτέλεσμα η ενέργεια να μεταφέρεται από μια θέση σε άλλη.

Καθώς το κύμα μεταδίδεται, η διαταραχή των μορίων είναι στην κατεύθυνση διάδοσης των κυμάτων. Η μετατόπιση των μορίων του μέσου οδηγεί σε περιοχές με εναλλασσόμενη υψηλή πυκνότητα μορίων και χαμηλή πυκνότητα μορίων. Οι περιοχές υψηλής πυκνότητας μορίων καλούνται πυκνώσεις (compressions). Οι περιοχές χαμηλής πυκνότητας μορίων καλούνται αραιώσεις (rarefactions).



Σχήμα 2.1 Αραιώσεις και πυκνώσεις κατά τη διάδοση ενός κύματος.

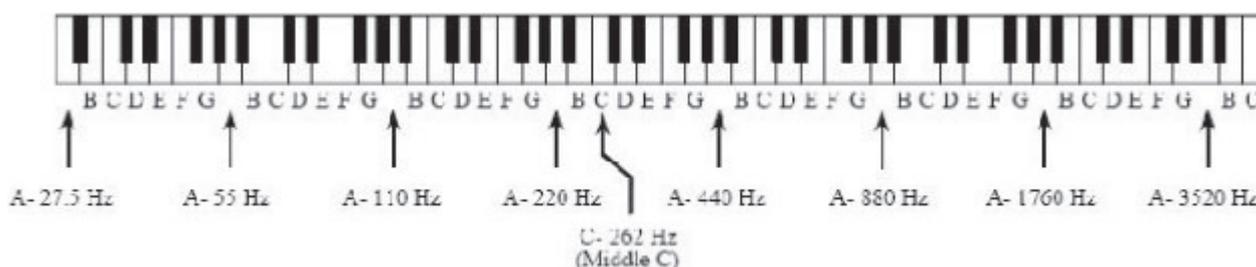


Σχήμα 2.2 Περιοδική μεταβολή ακουστικής πίεσης

Οι αραιώσεις και οι πυκνώσεις κινούνται στην κατεύθυνση διάδοσης των κυμάτων. Τα μόρια του μέσου

δεν κινούνται ομαδικά στην κατεύθυνση της μετακίνησης κυμάτων, δονούνται γύρω από τις κανονικές θέσεις τους. Κάθε πλήρης δόνηση ενός μορίου καλείται κύκλος (δηλ. από την αρχική θέση του, σε μια μέγιστη απόσταση σε μια κατεύθυνση, πίσω στην αρχική θέση, έπειτα σε μια μέγιστη μετατόπιση στην αντίθετη κατεύθυνση και πίσω στην αρχική θέση).

Ο αριθμός κύκλων που ολοκληρώνονται σε ένα δευτερόλεπτο καλείται συχνότητα (Hz) της δόνησης. Το φάσμα της ανθρώπινης ακοής εκτείνεται προσεγγιστικά από τα 16Hz έως τα 16kHz. Αυτό το φάσμα συχνοτήτων έχει εύρος ίσο με περίπου 10 οκτάβες από τον χαμηλότερο αντιληπτό τόνο ως τον υψηλότερο. Οι συχνότητες των νοτών που μπορούν να παιχτούν σε ένα πιάνο κυμαίνονται από 27.5 Hz και φτάνουν πάνω από τα 4 kHz



Σχήμα 2.3 Συχνότητες νοτών στο πιάνο.

Οποιαδήποτε νότα παίζεται σε ένα πιάνο θα ηχήσει διαφορετική σε μια νότα του ίδιου τόνου που παράγεται από έναν άλλο τύπο οργάνου.

Η μουσική νότα που παράγεται με έναν συντονισμό καλείται καθαρός τόνος (pure tone) επειδή αποτελείται από έναν τόνο μιας συχνότητας. Μια νότα που παίζεται σε ένα πιάνο, ή στα περισσότερα άλλα όργανα, αποτελείται από διάφορους τέτοιους τόνους που ηχούν μαζί σε διαφορετικές συχνότητες. Αυτές οι συχνότητες συσχετίζονται με τη συχνότητα που δίνει στη νότα το χαρακτηριστικό τόνο της, η οποία είναι συνήθως η χαμηλότερη.

Οι κυματομορφές μουσικής ή ομιλίας απέχουν πολύ από τη μορφή του απλού ημιτόνου. Είναι όμως πολύ ενδιαφέρον το γεγονός ότι όσο πολύπλοκο κι αν είναι ένα κύμα, αν είναι περιοδικό, μπορεί να περιγραφεί ως άθροισμα ημιτονικών όρων διαφορετικών συχνοτήτων, φάσεων και πλάτους.

Ο ημιτονικός όρος με τη χαμηλότερη συχνότητα καλείται θεμελιώδης (fundamental). Οι άλλοι όροι καλούνται partials (μερικές). Εάν οι αρμονικοί ήχοι έχουν συχνότητες που είναι ακέραια πολλαπλάσια ( $\times 2$ ,  $\times 3$ ... μέχρι  $\times 14$ ) της θεμελιώδους συχνότητας καλούνται αρμονικές (harmonics). Οι μουσικοί συχνά χρησιμοποιούν τον όρο «μερική» αντί του όρου «αρμονική», αλλά θα πρέπει σ' αυτό το σημείο να γίνει μία διάκριση μεταξύ των όρων γιατί οι μερικές πολλών οργάνων δεν είναι αρμονικά συσχετιζόμενες με την θεμελιώδη συχνότητα. Δηλαδή μπορεί να μην είναι ακέραια πολλαπλάσια της θεμελιώδους συχνότητας, όπως για παράδειγμα οι μερικές του πιάνου.

Επομένως, για την περίπτωση μονοφωνικών ήχων, το σήμα εισόδου περιλαμβάνει μία μοναδική αρμονική δομή, κάποια μη αρμονικά συστατικά και μία επιπρόσθετη ποσότητα ανεπιθύμητου θορύβου. Άρα, το σήμα μπορεί προσεγγιστικά να περιγραφεί ως:

$$x(t) = \sum_k k A_k(t) \sin(k\omega t + \varphi_k) + e(t)$$

όπου:  $x(t)$  είναι το σήμα εισόδου στο πεδίο του χρόνου

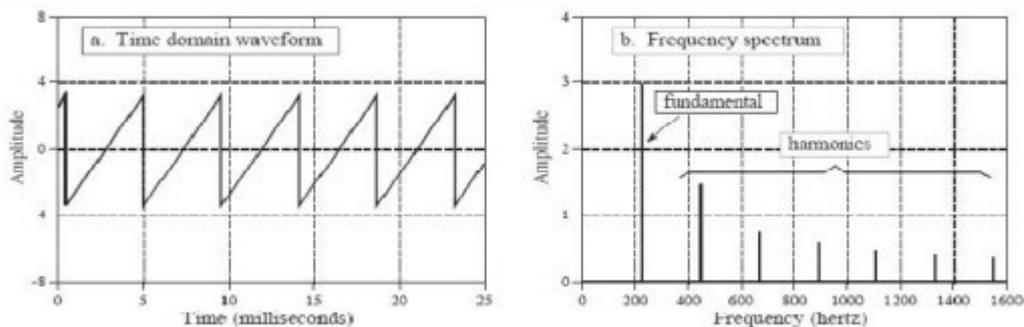
$\omega$  είναι η θεμελιώδης συχνότητα της τρέχουσας νότας

$A_k(t)$  είναι το πλάτος της  $k$ -αρμονικής τη χρονική στιγμή  $t$

$\varphi_k$  είναι η φάση της  $k$ -αρμονικής

$e(t)$  είναι οι λοιποί όροι (μη αρμονικά στοιχεία και θόρυβος)

Είναι η διαφορά στο αρμονικό περιεχόμενο των νοτών που δίνει σε κάθε μουσικό όργανο το χαρακτηριστικό ήχο του ή χροιά (timbre). Επομένως αν και η υψηλότερη νότα ενός πιάνου έχει μια θεμελιώδη συχνότητα μεγαλύτερη από 4kHz, ο εξοπλισμός που χρησιμοποιείται για να καταγράψει τη μουσική πρέπει να είναι σε θέση να χειριστεί πολύ υψηλότερες συχνότητες για να διατηρήσει τις αρμονικές που συνδέονται με κάθε νότα.



Σχήμα 2.4 Κυματομορφή στο πεδίο του χρόνου και στο πεδίο της συχνότητας

Οι ήχοι που παράγονται από κρουστικά αποτελέσματα είναι ιδιαίτερα πλούσιοι σε υψηλές αρμονικές. Αυτοί εμφανίζονται κυρίως στην έναρξη ενός ήχου, π.χ. όταν ακουμπάμε μια χορδή ή όταν χτυπάμε ένα κύμβαλο. Ο εξοπλισμός ηχογράφησης πρέπει να είναι σε θέση να αντιμετωπίσει αυτές τις υψηλές συχνότητες ειδάλλως η τονική ποιότητα των ήχων θα αλλάξει. Τα κύμβαλα, παραδείγματος χάριν, μπορούν να παραγάγουν τις συχνότητες μεταξύ 20kHz και 25kHz.

Υπάρχουν δύο τύποι ήχων: ήχοι καθορισμένου τόνου, την οποία οι μουσικοί καλούν νότα και ήχος χωρίς καθορισμένο τόνο, ο θόρυβος. “Η μουσική” περιλαμβάνει όχι μόνο τις νότες (καθορισμένου τόνου) αλλά επίσης και πολύ θόρυβο, π.χ. κρουστά. Ο θόρυβος είναι ένα αναπόσπαστο τμήμα της μουσικής. Η διαφορά μεταξύ μιας νότας και ενός θορύβου είναι προφανής στο αυτί αλλά τι προκαλεί τη διαφορά; Μια νότα αποτελείται από τις περιοδικές δονήσεις ενώ ένας θόρυβος αποτελείται από μη περιοδικές δονήσεις.

**Περιοδική κίνηση** είναι η μετακίνηση που είναι επαναλαμβανόμενη. Μια μουσική νότα παράγεται από

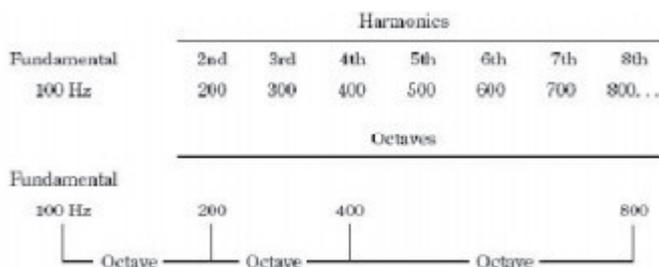
την περιοδική κίνηση σε μια ευδιάκριτη συχνότητα. Η μη περιοδική κίνηση θεωρείται ως θόρυβος.

Ως θόρυβος μπορεί να οριστεί ένας ανεπιθύμητος ήχος. Ένας μηχανικός ήχος ενδιαφέρεται για δύο είδη του θορύβου:

1. συστημικός θόρυβος, εισαγόμενος από τα μικρόφωνα, τους ενισχυτές, τα όργανα καταγραφής ταινιών, κ.λπ.
2. περιβαλλοντικός θόρυβος σε μια αίθουσα, ένα στούντιο, ένα δωμάτιο, κ.λπ.

Ένας μηχανικός ήχος ενδιαφέρεται επίσης για την αντήχηση. Όλα τα αντικείμενα που μπορούν να δονηθούν έχουν μια ορισμένη συχνότητα στην οποία θα δονηθούν εντονότερα (δηλ. με το μέγιστο εύρος). Εάν ένα σώμα διεγείρεται σε ένα ολόκληρο φάσμα συχνοτήτων θα δονηθεί περίπου εξίσου σε όλες τις συχνότητες εκτός από εκείνες τις συχνότητες που είναι πλησιέστερα στη φυσική συχνότητά του. Σε μια συχνότητα θα δονηθεί εντονότερα. Αυτή η συχνότητα καλείται συχνότητα συντονισμού και ο όρος καλείται αντήχηση. Μια χορδή ενός μουσικού οργάνου θα δονηθεί σε μια συχνότητα, την συχνότητα συντονισμού της. Τα μόρια αέρα έχουν μάζα έτσι οποιοδήποτε εσωκλειόμενο σώμα του αέρα (π.χ. ο σωλήνας ή ο «σωλήνας» ενός κλαρινέτου) έχει μια συχνότητα συντονισμού. Ο αέρας που περιλαμβάνεται σε μια αίθουσα ή ένα στούντιο μπορεί επίσης να έχει την αντήχηση, με συνέπεια ορισμένοι ήχοι να ενισχύονται αδικαιολόγητα.

Στη μουσική, η σειρά των ήχων διαιρείται σε τμήματα γνωστά ως οκτάβες. Η οκτάβα είναι μία λογαριθμική σύλληψη που είναι στενά συνδεδεμένη με τις μουσικές κλίμακες και ορολογίες, εξαιτίας της σχέσης της με τα χαρακτηριστικά της ανθρώπινης ακοής.



**Σχήμα 2.5** Σύγκριση αρμονικών και οκτάβων. Οι αρμονικές είναι γραμμικά συσχετισμένες ενώ οι οκτάβες λογαριθμικά

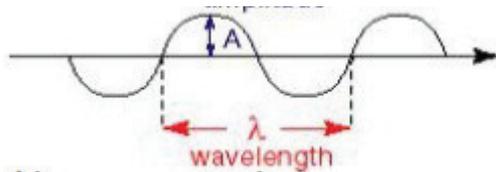
Στο σημείο αυτό, στο πλαίσιο των ημιτονικών σημάτων, των αρμονικών κ.τ.λ. πρέπει να ορίσουμε την έννοια του φάσματος (spectrum). Αντίστοιχα με το ορατό φάσμα του φωτός, υπάρχει και το φάσμα του ήχου, το σύνολο των συχνοτήτων που γίνονται αντιληπτά από το ανθρώπινο αυτί. Δεν μπορούμε να αντιληφθούμε την υπεριώδη ακτινοβολία, γιατί η συχνότητα της ηλεκτρομαγνητικής του ενέργειας είναι πέρα από τα όρια που μπορεί να αντιληφθεί το ανθρώπινο μάτι. Επίσης, δεν μπορούμε να δούμε την υπέρυθη γιατί η συχνότητά του είναι πολύ χαμηλή. Αντίστοιχα, υπάρχουν ήχοι πολύ χαμηλής και πολύ υψηλής συχνότητας που το ανθρώπινο αυτί δεν μπορεί να συλλάβει. Το φάσμα μας δείχνει πώς η ενέργεια του σήματος διαμοιράζεται στις διάφορες συχνότητες.

Παρατηρώντας τα φάσματα ημιτονικών σημάτων, τετραγωνικών και τριγωνικών παλμών, παρατηρούμε συγκέντρωση της ενέργειας στις αρμονικές συχνότητες και τίποτα στο ενδιάμεσο αυτών των συχνοτήτων. Τα παραπάνω σήματα είναι χαρακτηριστικά περιοδικών σημάτων. Στην περίπτωση του τυχαίου θορύβου η ενέργεια είναι κατανομημένη σε όλες τις συχνότητες. Ακόμα όμως και ο θόρυβος μπορεί να θεωρηθεί ως

άθροισμα ημιτονικών σημάτων με συνεχώς μεταβαλλόμενη συχνότητα, πλάτος και φάση.

Τα μόρια που δονούνται λόγω της μετάβασης ενός κύματος θεωρούνται σε φάση εάν κινούνται προς την ίδια κατεύθυνση και έχουν την ίδια μετατόπιση (δηλ. είναι οι ίδιες αποστάσεις από τις αρχικές θέσεις τους).

Το μέτωπο του κύματος είναι μια επιφάνεια στην οποία όλα τα μόρια είναι στην ίδια φάση της δόνησης. Η απόσταση μεταξύ των διαδοχικών κυματομορφών καλείται μήκος κύματος του ήχου. Το μήκος κύματος αναπαριστάται με το ελληνικό σύμβολο λάμδα. Σε επίπεδο πυκνώσεων και αραιώσεων το μήκος κύματος είναι η απόσταση μεταξύ των παρακείμενων κέντρων της πύκνωσης ή των παρακείμενων κέντρων αραιώσης



Σχήμα 2.6 Μήκος κύματος και πλάτος κυματομορφής.

Ο ήχος διαδίδεται με σταθερή ταχύτητα (ταχύτητα του ήχου) εάν η θερμοκρασία και η πίεση του αέρα είναι σταθερές. Αυτή είναι περίπου 330 μέτρα ανά δευτερόλεπτο. Το μήκος κύματος και η συχνότητα συσχετίζονται με την ταχύτητα του ήχου με την παρακάτω σχέση:

$$v=f*\lambda$$

Υπάρχουν τρία κύρια χαρακτηριστικά του ήχου:

1. Ηχηρότητα που συσχετίζεται με το πλάτος
2. Τόνος που συσχετίζεται με τη συχνότητα
3. Χροιά που συσχετίζεται με το φάσμα.

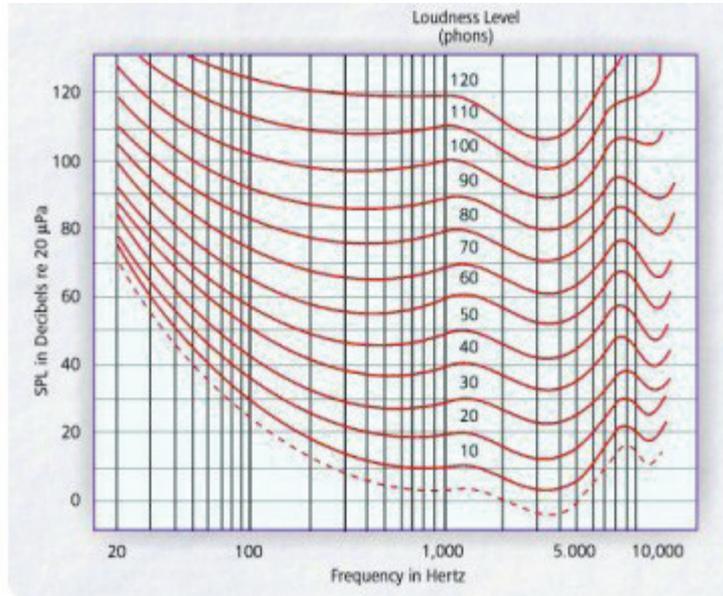
### 2.1.1 Ηχηρότητα (Loudness)

Ένας δυνατός ήχος (νότα ή θόρυβος) παράγεται από δονήσεις ισχυρότερες από εκείνες που παράγουν έναν λιγότερο δυνατό ήχο. Εύρος ενός κύματος είναι η μέγιστη μετατόπιση των δονούμενων μορίων από τις θέσεις ηρεμίας τους.

Για να παραχθεί δόνηση, πρέπει να καταναλωθεί ενέργεια. Επομένως, για να παράγουμε ένα πιο δυνατό ήχο περισσότερη ενέργεια πρέπει να δαπανηθεί. Το ποσό της ενέργειας για να παραχθεί ένας ήχος ονομάζεται ισχύς του ήχου και μετριέται σε Watt. Ενώ ο ήχος εκπέμπεται από μια πηγή, η συγκέντρωση ισχύος γίνεται όλο και μικρότερη όσο η απόσταση από την πηγή μεγαλώνει αφού το ίδιο ποσό ισχύος απλώνεται σε μεγαλύτερη επιφάνεια. Το ποσό ισχύος ανά τετραγωνικό μέτρο ονομάζεται ένταση του ήχου (intensity).

Οι άνθρωποι δεν αντιλαμβάνονται την ένταση του ήχου γραμμικά. Για έναν ήχο που γίνεται αντιληπτός ως δύο φορές πιο δυνατός, η ένταση πρέπει να είναι δέκα φορές μεγαλύτερη. Εξαιτίας αυτού, το αντιληπτό επίπεδο έντασης ενός ήχου μετριέται σε λογαριθμική κλίμακα χρησιμοποιώντας τη μονάδα του decibel (ή dB). Το κατώτατο όριο της ακρόασης ορίζεται ως 0 dB. Ένας ήχος που είναι τόσο δυνατός που προκαλεί

πραγματικά τον πόνο - το “κατώτατο όριο του πόνου” - είναι περίπου 120 dB. Με τη λογαριθμική κλίμακα στο μυαλό, το κατώτατο όριο του πόνου (120 dB) αντιπροσωπεύει μια ένταση 1.000.000.000.000 (1 τρισεκατομμύριο) φορές μεγαλύτερες από το κατώτατο όριο της ακρόασης (0 dB)!



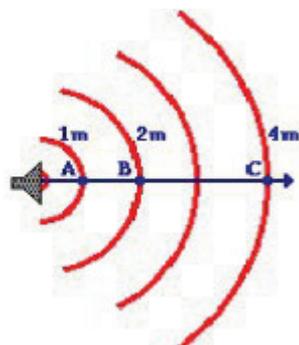
Σχήμα 2.7 Καμπύλες ίσης ηχηρότητας.

Οι ακόλουθες “εμπειροτεχνικές μέθοδοι” για την ηχηρότητα αποδεικνύονται χρήσιμες:

- Για να διπλασιαστεί η αντιληπτή ηχηρότητα ενός ήχου, η ισχύς (σε Watt) πρέπει να αυξηθεί δέκα φορές και το επίπεδο πίεσης του ήχου θα αυξηθεί κατά 10 dB. Παράδειγμα: Ένα επίπεδο πίεσης (SPL) 90 dB που παράγονται με ισχύ 10 Watt θα διπλασιαστεί σε αντιληπτή ηχηρότητα όταν αυξάνεται η δύναμη σε 100 Watt και η SPL αυξηθεί σε 100 dB.
- Ο διπλασιασμός της ισχύος ενός ήχου πιάνει ένα κέρδος 3 dB στην αντιληπτή ηχηρότητα. Παράδειγμα: Ένα επίπεδο πίεσης (SPL) 90 dB παράγεται με ισχύ 10 Watt. Εάν η ισχύς διπλασιαστεί στα 20 Watt, το SPL θα αυξηθεί σε 93 dB.
- Κινούμενοι από μία θέση ως προς μια δεδομένη ηχητική πηγή προς μια άλλη θέση που βρίσκεται σε διπλάσια απόσταση από την πηγή, η ηχηρότητα θα μειωθεί στο  $\frac{1}{4}$  (δηλ., μια απώλεια 6 dB). Παράδειγμα: Σε μια απόσταση 20 ποδιών από μια πηγή, ένα επίπεδο πίεσης (SPL) 90 dB παράγεται. Ο διπλασιασμός της απόστασης από την πηγή σε 40 πόδια θα οδηγήσει σε ένα SPL 84 dB. Όσο μεγαλύτερο το εύρος ενός κύματος, τόσο μεγαλύτερη θα είναι η ενέργεια των δονούμενων μορίων και ο ήχος θα είναι περισσότεροι έντονος (intensity). Καθώς ένα ηχητικό κύμα ταξιδεύει έξω από την πηγή, η ενέργεια μεταφέρεται από ένα δονούμενο μόριο στο επόμενο. Η ενέργεια απελευθερώνεται βαθμιαία υπό μορφή θερμότητας έτσι η ένταση του ήχου μειώνεται όσο η απόσταση από την πηγή αυξάνεται.

Η ένταση του ήχου λέγεται ότι είναι αντιστρόφως ανάλογη προς την απόσταση από την πηγή. Η μείωση υπακούει έναν αντίστροφο νόμο (ο αντίστροφος τετραγωνικός νόμος). Όταν ένας ακροατής απομακρύνεται από μια ηχητική πηγή το ηχητικό επίπεδο δεν εμφανίζεται να μειώνεται υπό τέτοιες αναλογίες. Αυτό συμβαίνει επειδή το αντί διαθέτει ένα μηχανισμό αντιστάθμισης. Το τύμπανο του αυτιού συνδέεται με το εσωτερικό αυτί με ένα σύστημα τριών ακουστικών οσταρίων που παίζουν τον ρόλο μοχλών. Οι θέσεις των αξόνων αυτών των μοχλών μπορούν να αλλάξουν για να παρέχουν μεγαλύτερη ή μικρότερη δύναμη. Η

ρύθμιση αυτή επιτρέπει στους αδύνατους ήχους να φθάσουν στο εσωτερικό αυτί με τη μέγιστη δύναμη ενώ οι δυνατοί ήχοι μειώνονται για να αποτρέψουν πιθανή ζημία στο εσωτερικό αυτί. Ένας ακροατής που κινείται δύο φορές μακρύτερα από ένα μεγάφωνο δεν θα αντιλαμβανόταν μια μείωση στην ηχηρότητα στο ένα τέταρτο. Εντούτοις η (μετρήσιμη) ένταση του ήχου θα είχε μειωθεί σε εκείνη την έκταση (Σχήμα 2.8).



Σχήμα 2.8 Μηχανισμός αντιστάθμισης του αυτιού.

**Ηχηρότητα (loudness)** κατόπιν, είναι το μέγεθος της αίσθησης από κάποιον που ακούει έναν ήχο. Η ένταση είναι μια μετρήσιμη, φυσική ποσότητα. Η ηχηρότητα εξαρτάται όχι μόνο από την ένταση ενός ήχου αλλά και από την ευαισθησία των αυτιών του ακροατή.

Η συσσώρευση των μορίων προκαλεί την πίεση του αέρα να είναι μεγαλύτερη από ότι κανονικά. Η πίεση είναι επομένως μέγιστη στις περιοχές της πύκνωσης και ελάχιστη στις περιοχές αραιώσης.

Εάν δύο ίδια ηχητικά κύματα φθάσουν στην ίδια θέση συγχρόνως οι περιοχές υψηλής πίεσης θα συμπέσουν και το ίδιο και οι περιοχές χαμηλής πίεσης. Αυτό θα οδηγήσει σε ένα ενιαίο κύμα μεγαλύτερης έντασης. Τα δύο κύματα θα ενισχύσουν το ένα το άλλο και το φαινόμενο (αισθητό γεγονός) καλείται **ενίσχυση**.

Εάν ένα κύμα φθάσει μισό μήκος κύματος αργότερα, η περιοχή υψηλής πίεσης του ενός κύματος θα ακυρωθεί από την περιοχή χαμηλής πίεσης του άλλου. Σαν αποτέλεσμα δεν ακούγεται κανένας ήχος. Αυτό καλείται ακύρωση. Όταν εμφανίζεται ενίσχυση, τα κύματα θεωρούνται σε φάση το ένα με το άλλο. Σε οποιαδήποτε άλλη περίπτωση τα κύματα είναι εκτός φάσης και εμφανίζεται μερική ή πλήρης ακύρωση.

### 2.1.2 Μουσικό ύψος του ήχου (Pitch)

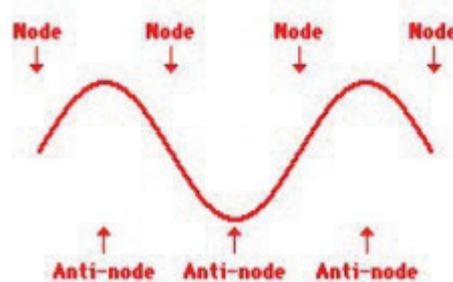
Η συχνότητα ενός ήχου γίνεται αντιληπτή ως ύψος (pitch). Δεν μπορούμε να ισχυριστούμε ότι συχνότητα και το ύψος είναι το ίδιο. Συσχετίζονται, αλλά όχι με γραμμικό τρόπο. Το ύψος είναι ένας υποκειμενικός όρος, ο οποίος εξαρτάται εκτός από την συχνότητα και από το επίπεδο της πίεσης του ήχου. Η μονάδα μέτρησης του είναι το mel, ένα ύψος αναφοράς 1000 mels ορίζεται ως το ύψος ενός ήχου στα 1000 Hz με πίεση 60 dB. Οι ήχοι με υψηλή συχνότητα γίνονται αντιληπτοί ως ήχοι με μεγάλο ύψος ενώ τους ήχους με λίγους κύκλους ανά δευτερόλεπτο τους αντιλαμβανόμαστε ως ήχους με χαμηλό ύψος. Όταν η συχνότητα ενός ήχου διπλασιάζεται την αντιλαμβανόμαστε ως διπλασιασμό του ύψους.

Η συχνότητα μετριέται στα Χέρτς. Ένα **Χέρτζ (Hz)** αντιστοιχεί σε ένα κύκλο ανά δευτερόλεπτο. Μια δόνηση υψηλής συχνότητας παράγει μια νότα με υψηλό τόνο, ενώ μια χαμηλής συχνότητας δόνηση δίνει μια νότα με χαμηλό τόνο.

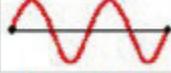
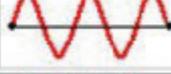
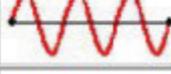
### 2.1.3 Χροιά ή Ηχώχρωμα (Timbre or Tone colour)

Η ιδιότητα του ήχου που επιτρέπει σε μας να καθορίσουμε τη διαφορά μεταξύ ενός σαξόφωνου και ενός φλάουτου καλείται χροιά ή ηχώχρωμα. Οι διαφορετικές χροιές εμφανίζονται επειδή το αντιληπτό ύψος των περισσότερων ήχων περιέχει στην πραγματικότητα πολλές συχνότητες. Κατά συνέπεια, όταν ακούει κάποιος τη νότα “μέσο C” για το πιάνο, πολλές άλλες συχνότητες είναι επίσης παρούσες. Το κυρίαρχο ύψος (σε αυτήν την περίπτωση το μέσο C), καλείται θεμελιώδης συχνότητα του ήχου. Οι άλλες παρούσες συχνότητες εμφανίζονται σε μια μαθηματική σειρά αποκαλούμενη αρμονική σειρά (harmonic series) ή σειρά αρμονικού ήχου (overtone series). Η συχνότητα κάθε αρμονικής είναι ακέραιο πολλαπλάσιο της θεμελιώδους συχνότητας (δηλ.,  $2X$ ,  $3X$ ,  $4X$ , κ.λπ.).

Η χροιά είναι άλλος ένας υποκειμενικός όρος. Ο ανάλογος φυσικός όρος είναι το φάσμα. Ακούγοντας μία ορχήστρα σε μία αίθουσα συναυλιών η χροιά που ακούμε είναι διαφορετική σε διαφορετικές θέσεις μέσα στο χώρο. Η μουσική αποτελείται από ένα ευρύ σύνολο συχνοτήτων, το πλάτος και η φάση των οποίων επηρεάζονται από ανακλάσεις στις διάφορες επιφάνειες του δωματίου. Ένας τρόπος ανάλυσης θα ήταν να λάβουμε το φάσμα του ήχου από διαφορετικές θέσεις. Παρολαυτά, αυτές θα ήταν φυσικές μετρήσεις και η υποκειμενική χροιά ακόμα θα μας διέφευγε. Το σημαντικό είναι να συνειδητοποιήσουμε τη διαφορά μεταξύ χροιάς και φάσματος. Η χροιά καθορίζεται από το πλήθος και τη σχετική δύναμη των μερικών του οργάνου και είναι συνώνυμη με την ποιότητα του ήχου. Ενώ οι περισσότεροι μουσικοί ήχοι έχουν τους αρμονικούς ήχους που είναι ακέραια πολλαπλάσια της θεμελιώδους συχνότητας, με μερικούς μουσικούς ήχους δεν συμβαίνει το ίδιο.



Σχήμα 2.9 Δεσμοί και Αντί-δεσμοί.

Harmonic	# of Nodes	# of Antinodes	Pattern
1st	2	1	
2nd	3	2	
3rd	4	3	
4th	5	4	
5th	6	5	
6th	7	6	
nth	n + 1	n	--

**Πίνακας 2.1** Αρμονικές, δεσμοί και αντί-δεσμοί.

## 2.2 Το ανθρώπινο ακουστικό σύστημα

Σε αυτήν ενότητα γίνεται μια περιγραφή του ανθρώπινου ακουστικού συστήματος, και παρατίθενται έρευνες της ψυχοακουστικής σχετικά με την ανθρώπινη αντίληψη του ήχου.

Η μελέτη της κατασκευής του αυτιού είναι μια μελέτη φυσιολογίας. Η μελέτη της ανθρώπινης αντίληψης του ήχου αποτελεί όμως κεφάλαιο της ψυχολογίας. Η ψυχοακουστική είναι ένας συνολικός όρος που περιλαμβάνει την φυσική κατασκευή του αυτιού, τις διαδρομές κίνησης του ήχου, την αντίληψη του ήχου, και τις σχέσεις μεταξύ τους. Το ηχητικό κύμα-ερέθισμα που φτάνει στο αυτί προκαλεί μηχανικές κινήσεις που έχουν σαν αποτέλεσμα λειτουργίες των νεύρων που καταλήγουν στον εγκέφαλο και δημιουργούν μια αίσθηση.

### 2.2.1 Ευαισθησία του αυτιού

Η λεπτότητα και η ευαισθησία της ακοής μας μπορεί να φαίνεται έντονα με ένα μικρό πείραμα: “Ανοίγεται αργά η ογκώδης πόρτα ενός ανηχοϊκού θαλάμου, και φαίνονται οι εξαιρετικά παχείς τοίχοι, και οι σφήνες από υαλόνημα μήκους ενός μέτρου, με τις κορυφές τους προς τα μέσα, που καλύπτουν όλους τους τοίχους, την οροφή, και αυτό που θα μπορούσε να ονομαστεί πάτωμα, εκτός από το ότι περπατούμε πάνω σε ένα ανοικτό ατσάλινο πλέγμα. Φέρνουμε μια καρέκλα και καθόμαστε. Το πείραμα χρειάζεται χρόνο, και επειδή έχουμε ενημερωθεί προηγούμενα, περιμένουμε, μετρώντας τις σφήνες από υαλόνημα για να περάσει η ώρα. Ο χώρος είναι αλλόκοτος. Η θάλασσα του ήχου και των θορύβων της ζωής και των δραστηριοτήτων όπου συνήθως βρισκόμαστε, και την οποία κανονικά σπάνια συνειδητοποιούμε, είναι τώρα εμφανής με την απουσία της. Η σιωπή μας πιέζει (σχεδόν απόλυτα), για 10 λεπτά, και περνά μισή ώρα. Ανακαλύπτονται νέοι ήχοι, ήχοι που προέρχονται μέσα από το ίδιο μας το σώμα. Στην αρχή, το δυνατό χτύπημα της καρδιάς μας, που μόλις συνέρχεται από την νέα κατάσταση. Περνά μια ώρα. Ακούγεται το αίμα μέσα στις φλέβες μας. Τέλος, αν τα αυτιά μας είναι τεντωμένα, η υπομονή μας ανταμείβεται από έναν παράξενο σφυριχτό ήχο μεταξύ των χτυπημάτων της καρδιάς και την ροή του

αίματος. Τι είναι αυτό; Είναι ο ήχος των σωματιδίων του αέρα που χτυπούν στα τύμπανα μας. Η κίνηση των τύμπανων που είναι αποτέλεσμα αυτού του σφυριχτού ήχου είναι απίστευτα μικρή, μόνο το 1/100 του εκατομμυριοστού του εκατοστού του μέτρου, ή το 1/10 της διαμέτρου του μορίου του υδρογόνου!”

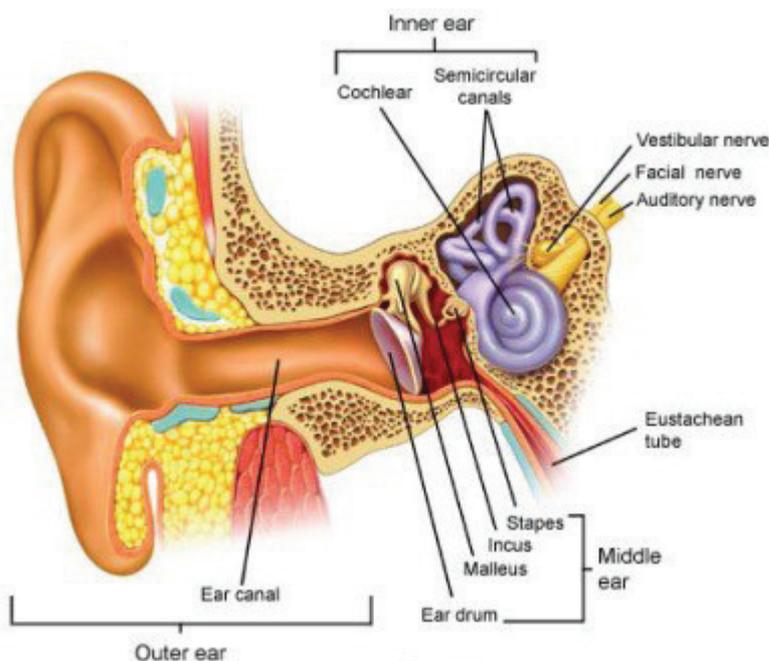
Το ανθρώπινο αυτί δεν μπορεί να ανιχνεύσει ήχους απαλότερους από την ροή των μορίων του αέρα στο τύμπανο. Αυτό είναι και το κατώφλι ακουστότητας. Δεν υπάρχει λόγος να έχουμε περισσότερο ευαίσθητα αυτιά, επειδή οποιοσδήποτε ήχος μικρότερης στάθμης θα πνιγόταν από τον θόρυβο των σωματιδίων του αέρα. Αυτό σημαίνει ότι η μεγαλύτερη ευαισθησία της ακοής μας μόλις και ταιριάζει με τους απαλότερους ήχους που είναι δυνατοί σε ένα αέριο μέσο. Είναι τυχαίο; Είναι θέμα προσαρμογής; Είναι θέμα σχεδίασης;

Στο άλλο άκρο, τα αυτιά μας μπορούν να αντιδράσουν στον βρυχηθμό του κανονιού, στον θόρυβο της εκτόξευσης πυραύλου, ή σε ένα αεριωθούμενο που απογειώνεται. Ειδικά προστατευτικά χαρακτηριστικά του αυτιού προστατεύουν τον ευαίσθητο μηχανισμό από ζημιά από όλους τους θορύβους εκτός από τους εξαιρετικά έντονους.

### 2.2.2 Η ανατομία του αυτιού

Τα τρία κύρια τμήματα του ανθρώπινου ακουστικού συστήματος, που φαίνεται στο Σχήμα 2.10, είναι το έξω αυτί (outer ear), το μέσο αυτί (middle ear), και το έσω αυτί (inner ear). Το έξω αυτί

αποτελείται από το περύγιο (pinna) και από το ακουστικό κανάλι (auditory meatus). Το ακουστικό κανάλι τελειώνει στην ακουστική μεμβράνη ή τύμπανο (tympanic membrane). Το μέσο αυτί είναι ένας χώρος γεμάτος αέρα που διασχίζεται από τα τρία μικρά οστά (ossicles) που ονομάζονται σφύρα (malleus), άκμονας (incus), και αναβολέας (stapes). Η σφύρα είναι κολλημένη στο τύμπανο και ο αναβολέας κολλημένος στο ελλειψοειδές παράθυρο (oval window) του έσω αυτιού. Όλα μαζί αυτά τα τρία οστά σχηματίζουν μια μηχανική σύνδεση μοχλού μεταξύ του τύμπανου το οποίο ενεργοποιείται από τον αέρα και τον κοχλία του έσω αυτιού ο οποίος είναι γεμάτος υγρό. Το έσω αυτί τελειώνει στο ακουστικό νεύρο (auditory nerve), το οποίο στέλνει ερεθίσματα στον εγκέφαλο.



Σχήμα 2.10 Η ανατομία του ανθρώπινου αυτιού.

### 2.2.2.1 Το περύγιο: Κατευθυντικός κωδικοποιητής ήχου

Στην αρχαιότητα, το περύγιο θεωρούνταν είτε σαν υποτυπώδες όργανο είτε σαν απλή συσκευή συλλογής ήχων. Και πράγματι, είναι συσκευή συλλογής ήχων. Το περύγιο προσφέρει διαφοροποίηση, των ήχων που προέρχονται από εμπρός σε σχέση με τους ήχους που προέρχονται από πίσω. Αν βάλουμε την παλάμη μας πίσω από το αυτί αυξάνεται το ενεργό μέγεθος του περυγίου και με τον τρόπο αυτό αυξάνεται και η φαινομενική ηχηρότητα κατά ποσό που εξαρτάται από τη συχνότητα. Για τις σημαντικές συχνότητες ομιλίας (2.000 μέχρι 3.000 Hz), η πίεση ήχου στο τύμπανο αυξάνεται κατά περίπου 5 dB. Αυτή η διαφοροποίηση εμπρός/πίσω είναι η ηπιότερη συνεισφορά του περυγίου.

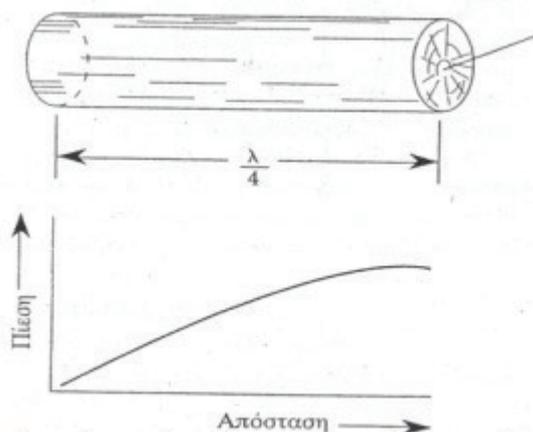
Οι πρόσφατες έρευνες έδειξαν ότι το περύγιο εκτελεί μια πολύ αποφασιστική λειτουργία χαράσσοντας κατευθυντικές πληροφορίες σε όλους τους ήχους που συλλαμβάνονται από το αυτί. Αυτό σημαίνει ότι πληροφορίες σχετικές με την πηγή του ήχου προστίθενται στο περιεχόμενο του ήχου έτσι ώστε η τελική ακουστική πίεση στο τύμπανο να δίνει την δυνατότητα στον εγκέφαλο να ερμηνεύει και το περιεχόμενο του ήχου και την κατεύθυνση από την οποία αυτός προέρχεται.

Ένα απλό ψυχοακουστικό πείραμα μπορεί να δείξει τον τρόπο με τον οποίο από τις απλές μεταβολές του ήχου που πέφτει στο αυτί έχουμε σαν αποτέλεσμα υποκειμενικές εντυπώσεις κατεύθυνσης. Με ένα ακουστικό στο ένα αυτί, ακούμε με ένα φίλτρο ρυθμιζόμενου ανοίγματος έναν τυχαίο θόρυβο με πλάτος ζώνης μιας οκτάβας και με κεντρική συχνότητα τα 8 kHz. Αν ρυθμίσουμε το φίλτρο στα 7.2 kHz ο ήχος θα φαίνεται ότι προέρχεται από πηγή στο ύψος του

παρατηρητή. Όταν το άνοιγμα ρυθμιστεί στα 8 kHz ο ήχος φαίνεται ότι έρχεται από επάνω. Όταν το άνοιγμα είναι στα 6.3 kHz ο ήχος φαίνεται να έρχεται από κάτω. Το πείραμα αυτό δείχνει ότι το ανθρώπινο ακουστικό σύστημα εξαγεί πληροφορίες κατεύθυνσης από τα φασματικά χαρακτηριστικά του ήχου στο τύμπανο.

### 2.2.2.2 Το κανάλι του αυτιού

Το κανάλι του αυτιού αυξάνει και αυτό την ηχηρότητα των ήχων που το διαπερνούν. Στο σχήμα 2.11, το κανάλι του αυτιού, που έχει μέση διάμετρο περίπου 0.7 cm και μήκος περίπου 3 cm, φαίνεται σε σχηματική μορφή σαν ευθύ και με ομογενή διάμετρο σε όλο το μήκος του. Από ακουστικής πλευράς, η προσέγγιση είναι λογική. Έχουμε ένα αγωγό σε σχήμα σωλήνα που στο μέσα άκρο του κλείνει με το τύμπανο.

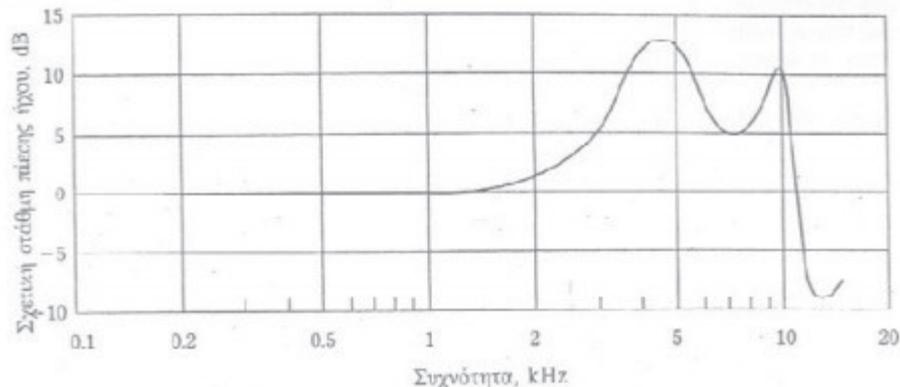


Σχήμα 2.11 Το ακουστικό κανάλι, κλειστό στο ένα άκρο από το τύμπανο, λειτουργεί σαν “σωλήνας μουσικού οργάνου” ενός τετάρτου μήκους κύματος. Ο συντονισμός δίνει ακουστική ενίσχυση για τις σημαντικές

### συχνότητες φωνής.

Όταν η επιστήμη της ακουστικής βρισκόταν στα πρώτα της βήματα, οι πρώτοι ερευνητές μελέτησαν έντονα τους σωλήνες των μουσικών οργάνων. Η ακουστική ομοιότητα του καναλιού του αυτιού με σωλήνα μουσικού οργάνου δεν διέφυγε από τους πρώτους ερευνητές στον τομέα αυτό. Το φαινόμενο του συντονισμού στο κανάλι του αυτιού αυξάνει την πίεση ήχου στο τύμπανο σε ορισμένες συχνότητες. Το μέγιστο βρίσκεται κοντά στην συχνότητα στην οποία ο σωλήνας των 3 cm έχει ένα τέταρτο μήκους κύματος, περίπου 3.000 Hz.

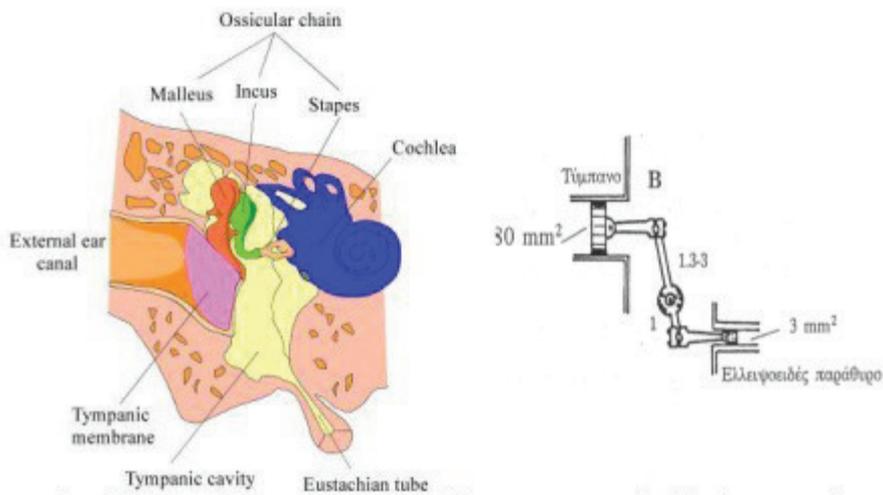
Το ακουστικό κανάλι, κλειστό στο ένα άκρο από το τύμπανο (σχήμα 2.11), λειτουργεί σαν “σωλήνας μουσικού οργάνου” ενός τετάρτου μήκους κύματος. Ο συντονισμός δίνει ακουστική ενίσχυση για τις σημαντικές συχνότητες φωνής. Στο σχήμα 2.12 φαίνεται η αύξηση της πίεσης ήχου στο τύμπανο σε σχέση με την πίεση στο άνοιγμα του καναλιού του αυτιού. Παρατηρούμε μια πρωτεύουσα κορυφή περίπου στα 4kHz που προκαλείται από το φαινόμενο συντονισμού του σωλήνα στο τέταρτο του μήκους κύματος. Ο πρωτεύων συντονισμός του σωλήνα αυξάνει την πίεση ήχου στο τύμπανο περίπου κατά 12 dB στον κύριο συντονισμό περίπου στα 4kHz . Υπάρχει ένας δευτερεύων συντονισμός κοντύτερα στα 10kHz. με μικρότερη μέγιστη πίεση.



Σχήμα 2.12 Η συνάρτηση μεταφοράς (απόκριση συχνότητας) του καναλιού του αυτιού. Πρόκειται για μια σταθερή συνιστώσα που προστίθεται με κάθε κατευθυντικά κωδικοποιημένο ήχο που φτάνει στο τύμπανο.

#### 2.2.2.3 Το μέσο αυτί

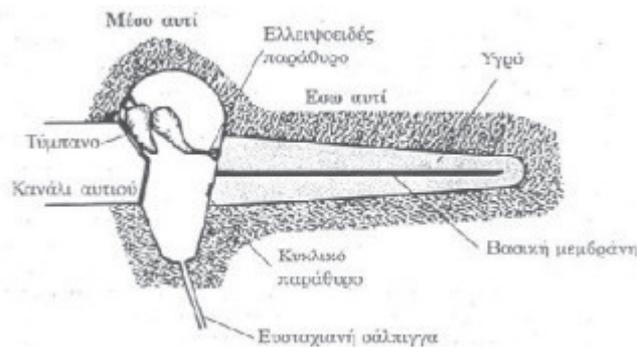
Ἡ μετάδοση της ενέργειας του ήχου από ένα αραιό μέσο όπως είναι ο αέρας σε πυκνό μέσο όπως το νερό αποτελεί σοβαρό πρόβλημα. Αν δεν υπάρχει πολύ εξειδικευμένος εξοπλισμός, ο ήχος από τον αέρα ανακλάται στο νερό όπως το φως σε καθρέπτη. Το ζήτημα καταλήγει σε προσαρμογή αντιστάσεων, και στην περίπτωση μας ο λόγος αντιστάσεων είναι περίπου 4.000 : 1. Ας σκεφτούμε πόσο ικανοποιητικό θα ήταν να οδηγούμε το πηνίο φωνής ενός μεγαφώνου 1 Ohm με έναν ενισχυτή που έχει αντίσταση εξόδου 4.000 Ohm. Είναι φανερό ότι δεν πρόκειται να μεταδοθεί αρκετή ισχύς.



**Σχήμα 2.13** Αριστερά: Τα μικρά οστά (σφύρα, άκμονας, αναβολέας) του μέσου αυτιού, που μεταδίδουν μηχανικές ταλαντώσεις του τυμπάνου στο ελλειψοειδές παράθυρο του κοχλίου. Δεξιά: Μηχανικό ανάλογο της λειτουργίας προσαρμογής αντιστάσεων στο μέσο αυτί. Η διαφορά επιφάνειας μεταξύ τύμπανου και ελλειψοειδούς παραθύρου, μαζί με την μηχανική σύνδεση υποβιβασμού, προσαρμόζουν την κίνηση του τυμπάνου που ενεργοποιείται από τον αέρα στο ελλειψοειδές παράθυρο με υγρό

Το θέμα είναι να μεταφερθεί με μέγιστη απόδοση η ασθενική ενέργεια που παριστάνεται από την κίνηση ταλάντωσης ενός μάλλον εύθραυστου διαφράγματος, στο υγρό του έσω αυτιού. Στο Σχήμα 2.13 προτείνεται η διπλή λύση. Τα τρία μικρά οστά σχηματίζουν μια μηχανική σύνδεση μεταξύ του τυμπάνου και του ελλειψοειδούς παραθύρου, το οποίο βρίσκεται σε απαλή επαφή με το υγρό του έσω αυτιού. Το πρώτο από τα τρία οστά, η σφύρα, είναι συνδεδεμένο με το τύμπανο. Το τρίτο, ο αναβολέας, στην πραγματικότητα αποτελεί τμήμα του ελλειψοειδούς παραθύρου. Στη σύνδεση αυτή υπάρχει κίνηση μοχλού με αναλογία από 1.3 : 1 μέχρι 3 : 1. Δηλαδή, η κίνηση του τυμπάνου ελαττώνεται κατά το μέγεθος αυτό στο ελλειψοειδές παράθυρο του έσω αυτιού.

Το πρόβλημα της προσαρμογής ήχου στον αέρα σε ήχο στο υγρό του έσω αυτιού λύνεται όμορφα με την μηχανική του μέσου αυτιού. Απόδειξη ότι η προσαρμογή αντιστάσεων μαζί με την ενίσχυση συντονισμού στο σχήμα 2.12 λειτουργούν αποτελεσματικά, αποτελεί το γεγονός ότι μια κίνηση διαφράγματος μικρή όσο οι μοριακές διαστάσεις δίνει την αντίληψη κατωφλίου.



Σχήμα 2.14 Εξιδανικευμένο σχέδιο του ανθρώπινου αυτιού όπου φαίνεται ξετυλιγμένος ο γεμάτος υγρό κοχλίας. Ο ήχος που μπαίνει στο κανάλι του αυτιού προκαλεί την ταλάντωση του τύμπανου. Αυτή η ταλάντωση μεταδίδεται στον κοχλία μέσω της μηχανικής σύνδεσης του μέσου αυτιού. Ο ήχος αναχίμα Εξιδαλύεται μέσω στάσιμων κυμάτων που δημιουργούνται στην (βασική) μεμβράνη.

#### 2.2.2.4 Το έσω αυτί

Μέχρι το σημείο αυτό έχουμε εξετάσει μόνο τους ακουστικούς ενισχυτές και τα χαρακτηριστικά μηχανικής προσαρμογής της αντίστασης του μέσου αυτιού. Όλα αυτά είναι σχεδόν καλά κατανοητά. Η πολύπλοκη λειτουργία του κοχλίας εξακολουθεί ακόμη να βρίσκεται μέσα σε μυστήριο, αλλά η εκτενής έρευνα μας δίνει σταθερά νέες γνώσεις.

Στο σχήμα 2.10 φαίνεται πόσο κοντά βρίσκονται τα τρία, κάθετα μεταξύ τους, ημικυκλικά κανάλια του μηχανισμού του προθαλάμου, του οργάνου ισορρόπησης, και του κοχλίας, δηλαδή του οργάνου που αναλύει τον ήχο. Και τα τρία περιέχουν το ίδιο υγρό, αλλά οι λειτουργίες τους είναι ανεξάντλητες. Ο κοχλίας, που έχει μέγεθος όσο ένα μπιζέλι, βρίσκεται μέσα σε στερεό κόκαλο. Είναι στριμμένος σαν σαλιγκάρι απ' όπου παίρνει και το όνομα του. Στο σχήμα 2.14, για περιγραφικούς λόγους, αυτό το στρίψιμο των 2 στροφών τεντώνεται στο πλήρες μήκος του, που είναι περίπου 2.5 εκατοστά. Το γεμάτο υγρό έσω αυτί χωρίζεται κατά μήκος με δύο μεμβράνες, την μεμβράνη του Reissner και την βασική (ή βασική) μεμβράνη. Άμεσα μας ενδιαφέρει η βασική μεμβράνη και η απόκριση της στις ταλαντώσεις ήχου μέσα στο υγρό.

Η ταλάντωση του τύμπανου ενεργοποιεί τα μικρά οστά. Η κίνηση του αναβολέα, που είναι κολλημένος στο ελλειψοειδές παράθυρο, προκαλεί την ταλάντωση του υγρού του έσω αυτιού. Μια κίνηση του ελλειψοειδούς παραθύρου προς τα μέσα έχει σαν αποτέλεσμα την ροή του υγρού γύρω από το πέρα άκρο της βασικής μεμβράνης, με αποτέλεσμα κίνηση προς τα έξω του κυκλικού παραθύρου. Ο ήχος που ενεργοποιεί το ελλειψοειδές παράθυρο έχει σαν αποτέλεσμα την δημιουργία στάσιμων κυμάτων στην βασική μεμβράνη. Η θέση των μέγιστων πλατών των στάσιμων κυμάτων στην βασική μεμβράνη μεταβάλλεται καθώς αλλάζει η συχνότητα του ήχου που προκαλεί την διέγερση.

Ο ήχος χαμηλής συχνότητας έχει σαν αποτέλεσμα μέγιστο πλάτος κοντά στο πέρα άκρο της βασικής μεμβράνης. Ο ήχος υψηλής συχνότητας δημιουργεί κορυφές κοντά στο ελλειψοειδές παράθυρο. Σε περίπτωση πολύπλοκου σήματος όπως είναι η μουσική ή η ομιλία, δημιουργούνται πολλές στιγμιαίες κορυφές, οι οποίες μετακινούνται διαρκώς σε πλάτος και θέση κατά μήκος της βασικής μεμβράνης. Αυτές οι κορυφές συντονισμού της βασικής μεμβράνης αρχικά θεωρούνταν ότι ήταν πολύ πλατειές ώστε να μη μπορεί να εξηγηθεί η οξύτητα του διαχωρισμού συχνοτήτων που εμφανίζεται στο ανθρώπινο αυτί.

Η πρόσφατη έρευνα έδειξε ότι σε χαμηλές εντάσεις ήχου, οι καμπύλες συντονισμού της βασικής

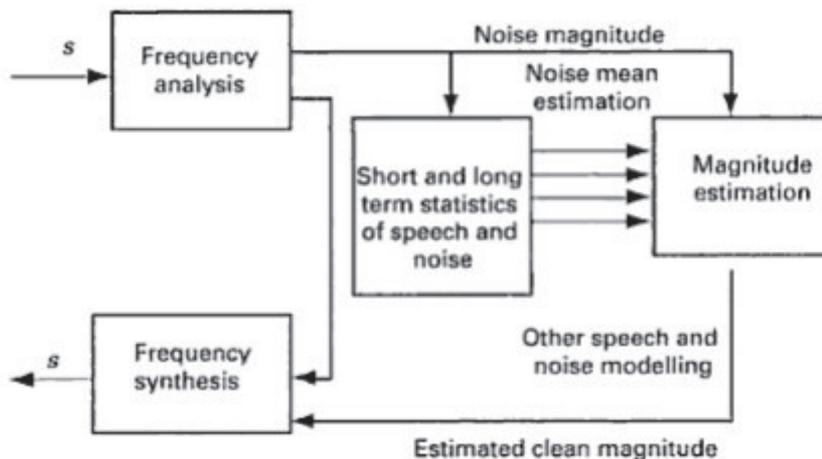
μεμβράνης είναι πολύ οξείες, και φαρδαινούν μόνο για έντονο ήχο. Σήμερα φαίνεται ότι η οξύτητα των καμπύλων μηχανικού συντονισμού της βασικής μεμβράνης μπορεί να συγκριθεί με την οξύτητα των απλών ινών του ακουστικού νεύρου οι οποίες την γεμίζουν.

### Κεφάλαιο 3

#### Αλγόριθμοι Φασματικής Αφαίρεσης (Spectral Subtractive Algorithms)

Θεωρώντας ότι έχουμε προσθετικό θόρυβο μπορούμε να εξασφαλίσουμε μια εκτίμηση του φάσματος του καθαρού σήματος αφαιρώντας μια εκτίμηση του φάσματος θορύβου από το φάσμα του ενθόρυβου σήματος. Το φάσμα του θορύβου μπορεί να υπολογιστεί και να επανεκτιμηθεί κατά την διάρκεια περιόδων όπου εμφανίζεται μόνο ο θόρυβος και όχι το χρήσιμο σήμα. Η βασική υπόθεση που κάναμε για να είναι η παραπάνω διαδικασία εφικτή είναι το σήμα του θορύβου να είναι στάσιμο ή αργά μεταβαλλόμενο, και το φάσμα του θορύβου να μην αλλάζει σημαντικά μεταξύ των περιόδων επανεκτίμησης του. Το ενισχυμένο σήμα προκύπτει μετά τον υπολογισμό του αντίστροφου μετασχηματισμού Fourier του φάσματος του σήματος που εκτιμήσαμε χρησιμοποιώντας την φάση του ενθόρυβου σήματος. Από τα παραπάνω διαπιστώνουμε την απλότητα της διαδικασίας η οποία όμως επιτυγχάνεται με ένα τίμημα. Η διαδικασία αφαίρεσης πρέπει να γίνεται με μεγάλη προσοχή για να αποφύγουμε τυχόν παραμόρφωση στο χρήσιμο σήμα. Αν αφαιρέσουμε μεγάλο κομμάτι του φάσματος ίσως να αφαιρέσουμε και κομμάτι χρήσιμου σήματος, ενώ αν αφαιρέσουμε μικρό κομμάτι ο θόρυβος που θα παραμείνει θα μειώσει την ποιότητα του εκτιμώμενου σήματος.

Το γεγονός ότι η φάση δεν συμβάλει σημαντικά στην ποιότητα του ήχου οδήγησε στην ανάπτυξη αυτής της κατηγορίας αλγορίθμων που βασίζονται στην εκτίμηση του μέτρου του φάσματος. Αποτελούν εκτιμητές του πεδίου της συχνότητας όπου κάθε εκτίμηση του μέτρου του φάσματος του καθαρού σήματος επανασυνδυάζεται με την φάση του ενθόρυβου σήματος πριν αυτό επανασυντεθεί με την διαδικασία επικάλυψης – άθροισης.



Σχήμα 3.1 Ενίσχυση ήχων με υπολογισμό του μέτρου του φάσματος

### 3.1 Βασικές αρχές της φασματικής αφαίρεσης

Αρχικά υποθέτουμε ότι το ενθόρυβο σήμα εισόδου  $y(n)$  αποτελείται από το καθαρό (χρήσιμο) σήμα  $x(n)$  και το σήμα προσθετικού θορύβου  $d(n)$ , δηλαδή έχουμε

$$y(n) = x(n) + d(n)$$

Παίρνοντας τον μετασχηματισμό Fourier και στους δύο όρους της παραπάνω εξίσωσης προκύπτει

$$Y(\omega) = X(\omega) + D(\omega)$$

Η πολική μορφή των παραπάνω μεγεθών είναι :

$$Y(\omega) = |Y(\omega)|e^{j\varphi_y(\omega)}$$

όπου  $|Y(\omega)|$  το μέτρο και  $\varphi_y(\omega)$  η φάση του φάσματος του σήματος που έχει αλλοιωθεί από την παρουσία θορύβου και

$$D(\omega) = |D(\omega)|e^{j\varphi_d(\omega)}$$

όπου  $|D(\omega)|$  το μέτρο και  $\varphi_d(\omega)$  η φάση του φάσματος. Το μέτρο του φάσματος θορύβου  $|D(\omega)|$  είναι άγνωστο, αλλά μπορούμε να το αντικαταστήσουμε με την μέση τιμή όπως αυτή υπολογίζεται κατά την διάρκεια των παύσεων του χρήσιμου σήματος (δηλαδή όταν θα υπάρχει μόνο θόρυβος). Ομοίως και η φάση του θορύβου θα  $\varphi_d(\omega)$  τασταθεί από την αντίστοιχη του ενθόρυβου  $\varphi_y(\omega)$  ως καθώς η φάση δεν επηρεάζει σημαντικά την καταληπτότητα του χρήσιμου σήματος, αλλά σε μικρό βαθμό την ποιότητα αυτού (περισσότερα για αυτό παρακάτω). Μετά την πραγματοποίηση των παραπάνω αλλαγών στην αρχική μας εξίσωση προκύπτει μια εκτίμηση για το φάσμα του καθαρού σήματος:

$$\hat{X}(\omega) = [|Y(\omega)| - |\hat{D}(\omega)|]e^{j\varphi_y(\omega)}$$

Η παραπάνω εξίσωση αποτυπώνει και την βασική αρχή που πραγματεύεται η φασματική αφαίρεση, δηλαδή: υπολογίζουμε μέσω FFT το μέτρο του φάσματος του ενθόρυβου σήματος και κρατάμε μια εκτίμηση για το φάσμα του θορύβου όταν δεν εμφανίζεται το χρήσιμο σήμα. Εν συνεχεία, αφαιρούμε από το φάσμα μέτρου του ενθόρυβου σήματος το φάσμα μέτρου του θορύβου και τέλος με αντίστροφο μετασχηματισμό Fourier του φάσματος της διαφοράς (με χρήση της φάσης του ενθόρυβου σήματος) καταλήγουμε στο ενισχυμένο χρήσιμο σήμα

Σε αυτό το σημείο πρέπει να σημειώσουμε ότι το μέτρο του φάσματος του ενισχυμένου σήματος μπορεί να προκύψει αρνητικό

$$|\hat{X}(\omega)| (= |Y(\omega)| - |\hat{D}(\omega)|),$$

εξαιτίας ανακρίβειών στην εκτίμηση του φάσματος θορύβου. Το μέτρο του φάσματος, βέβαια, δεν μπορεί να προκύπτει αρνητικό και γι' αυτό πρέπει να ήμαστε ιδιαίτερος προσεκτικοί στην αφαίρεση των δύο φασμάτων. Μια λύση για αυτό το πρόβλημα προκύπτει με την χρησιμοποίηση ανορθωτή ημίσειως κύματος όπως φαίνεται παρακάτω

$$|\hat{X}(\omega)| = \begin{cases} |Y(\omega)| - |\hat{D}(\omega)| & \text{αν } |Y(\omega)| > |\hat{D}(\omega)| \\ 0 & \text{αλλιώς} \end{cases}$$

Η παραπάνω διαδικασία της αφαίρεσης του φάσματος μπορεί να επεκταθεί και στον τομέα του φάσματος ισχύος, η χρήση του οποίου μπορεί σε ορισμένες περιπτώσεις να μας διευκολύνει. Για να φτάσουμε στο φάσμα ισχύος του ενθόρυβου σήματος πολλαπλασιάζουμε το  $Y(\omega)$  στην εξίσωση

$Y(\omega) = X(\omega) + D(\omega)$  με το συζυγές του  $Y^*(\omega)$ . Η εξίσωση που προκύπτει είναι:

$$|Y(\omega)|^2 = |X(\omega)|^2 + |D(\omega)|^2 + X(\omega)D^*(\omega) + X^*(\omega)D(\omega)$$

Στην παραπάνω εξίσωση οι όροι  $|D(\omega)|^2$ ,  $X(\omega)D^*(\omega)$  και  $X^*(\omega)D(\omega)$  δεν μπορούν να υπολογιστούν άμεσα αλλά παίρνουμε προσέγγιση αυτών. Επίσης αν θεωρήσουμε ότι το  $d(n)$  είναι μηδενικού μέσου όρου και είναι ασυσχέτιστο με το καθαρό σήμα  $x(n)$  τότε οι εκτιμήσεις των διαγώνιων όρων  $E\{X(\omega)D^*(\omega)\}$  και  $E\{X^*(\omega)D(\omega)\}$  ίνονται μηδέν. Με όλες τις παραπάνω υποθέσεις καταλήγουμε στην εξίσωση που περιγράφει τον αλγόριθμο αφαίρεσης φάσματος ισχύος:

$$|\hat{X}(\omega)|^2 = |Y(\omega)|^2 - |\hat{D}(\omega)|^2$$

Όπως συνέβη και παραπάνω με τον κλασικό αλγόριθμο φασματικής αφαίρεσης, για να εγγυηθούμε ότι το φάσμα ισχύος  $|\hat{X}(\omega)|^2$  ναί θετικό μπορούμε να χρησιμοποιήσουμε ανορθωτή ημίσεως κύματος. Η παραπάνω εξίσωση μπορεί να γραφτεί και στην παρακάτω μορφή:

$$|\hat{X}(\omega)|^2 = H^2(\omega)|Y(\omega)|^2, \text{ όπου } H(\omega) = \sqrt{1 - \frac{|\hat{D}(\omega)|^2}{|Y(\omega)|^2}}$$

Ο όρος  $H(\omega)$  γνωστός απο την θεωρία γραμμικών συστημάτων ως συνάρτηση μεταφοράς αποτελεί στο αντικείμενο που μελετάμε την συνάρτηση κέρδους ή συνάρτηση απόσβεσης. Παρατηρούμε ότι το

$$|\hat{X}(\omega)|^2$$

$H(\omega)$  εξ ορισμού είναι θετικό και παίρνει τιμές στο εύρος  $0 \leq H(\omega) \leq 1$ , αν και εξαιτίας ανακριβών μετρήσεων του φάσματος θορύβου μπορεί να καταλήξουμε σε αρνητικές τιμές. Το  $H(\omega)$  καλείται συνάρτηση απόσβεσης καθώς μας δείχνει το μέγεθος της απόσβεσης που εφαρμόζεται στο φάσμα ισχύος του ενθόρυβου  $|Y(\omega)|^2$  σε μια συγκεκριμένη συχνότητα ώστε να επιτύχουμε το ενισχυμένο φάσμα ισχύος  $|\hat{X}(\omega)|^2$ . Η μορφή της συνάρτησης απόσβεσης είναι μοναδική για κάθε αλγόριθμο ενίσχυσης ήχων και για αυτό συχνά αποτελεί μέτρο σύγκρισης μεταξύ διαφορετικών αλγορίθμων.

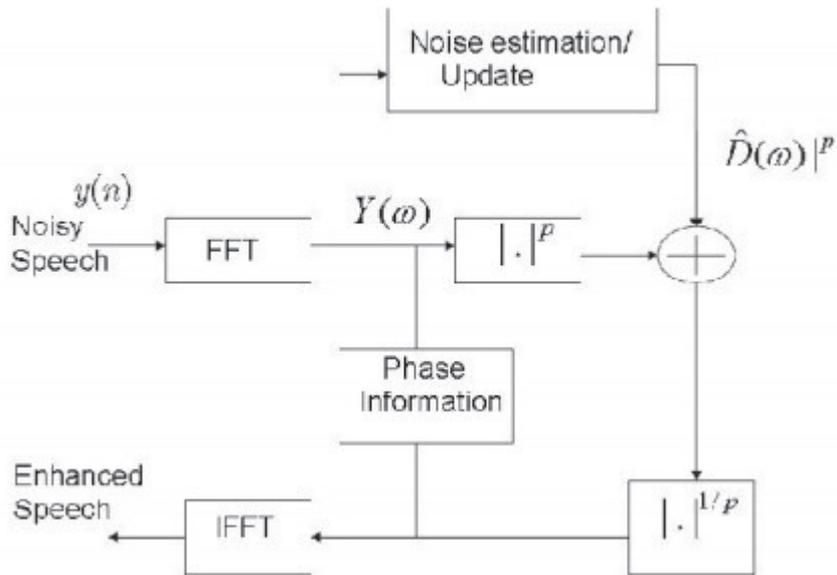
Τέλος θα δώσουμε μια πιο γενική μορφή του αλγόριθμου φασματικής αφαίρεσης όπως αυτή εκφράζεται από την παρακάτω εξίσωση:

$$|\hat{X}(\omega)|^p = |Y(\omega)|^p - |\hat{D}(\omega)|^p$$

όπου  $p$  είναι ο εκθέτης δύναμης και για  $p=1$  προκύπτει ο αρχικός αλγόριθμος κλασματικής αφαίρεσης ενώ για  $p=2$  έχουμε τον αλγόριθμο αφαίρεσης φάσματος ισχύος. Στο σχήμα 2.1 φαίνεται η γενική μορφή του αλγόριθμου φασματικής αφαίρεσης.

Είναι σημαντικό σε αυτό το σημείο να σημειώσουμε ότι η τελευταία εξίσωση όπως και η εξίσωση

αλγόριθμοι αφαίρεσης φασματικής ισχύος αποτελούν προσεγγίσεις εξαιτίας της παρουσίας των διαγώνιων όρων. Οι διαγώνιοι όροι είναι μηδενικοί μόνο υπό στατιστική έννοια, αν υποθέσουμε ότι υπάρχουν επαρκή δεδομένα και ότι τα σήματα είναι στάσιμα. Στις περισσότερες εφαρμογές οι διαγώνιοι όροι δεν είναι αμελητέοι, παρόλα αυτά όμως οι περισσότεροι αλγόριθμοι φασματικής αφαίρεσης τους θεωρούν μηδενικούς για λόγους απλότητας.



Σχήμα 3.2 Γενική μορφή αλγορίθμου φασματικής αφαίρεσης

## ΚΕΦΑΛΑΙΟ 4

### ΨΥΧΟΑΚΟΥΣΤΙΚΗ

Η ψυχοακουστική (Psychoacoustics) αποτελεί τον κλάδο της ακουστικής, ο οποίος μελετά τον υποκειμενικό τρόπο με τον οποίο το ανθρώπινο σύστημα ακοής αντιλαμβάνεται τους διάφορους ήχους, συμπεριλαμβανόμενης και της μουσικής (συνειδητοποίηση της ηχητικής πληροφορίας σε γνωστικό επίπεδο).

#### 4.1 ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ

Η ακρόαση δεν είναι ένα καθαρώς μηχανικό φαινόμενο της διάδοσης κυμάτων, αλλά είναι επίσης ένα αισθητήριο και αντιληπτικό γεγονός. Όταν ένα πρόσωπο ακούει κάτι, κάτι φθάνει στο αυτί του ως μηχανικό κύμα που ταξιδεύει μέσω του αέρα, αλλά μέσα στο αυτί μετασχηματίζεται σε νευρική δράση. Αυτοί οι σφυγμοί των νεύρων ταξιδεύουν έπειτα στον εγκέφαλο όπου γίνονται αντιληπτοί. Ως εκ τούτου, σε πολλά προβλήματα στην ακουστική είναι συμφέρον να ληφθούν υπόψη όχι μόνο οι μηχανισμοί του περιβάλλοντος, αλλά και το γεγονός ότι και το αυτί και ο εγκέφαλος περιλαμβάνονται στην εμπειρία ακούσματος ενός προσώπου.

Το αυτί παραδείγματος χάριν, παίρνει μια φασματική αποσύνθεση του ήχου ως τμήμα της διαδικασίας της μετατροπής του ήχου σε νευρικό ερέθισμα, έτσι ορισμένα αποτελέσματα χρονικών περιοχών είναι μη ακουστά. Η τεχνολογία απωλεστικής κωδικοποίησης ηχητικών δεδομένων χρησιμοποιεί αυτό το γεγονός. Επιπλέον το αυτί έχει μια λογαριθμική δυναμική απόκριση. Μια άλλη παρενέργεια της μη γραμμικής λογαριθμικής απόκρισης του αυτιού είναι ότι οι ήχοι που εμφανίζονται στο τύμπανο των αυτιών και έχουν μεγάλη φασματική συσχέτιση προκαλούν τονική παραίσθηση. Τέτοιες ψυχολογικές επιδράσεις που οφείλονται στην ανατομία του αυτιού καλούνται ψυχολογικές-ακουστικές επιδράσεις.

Υπάρχουν αληθινά ψυχοακουστικά αποτελέσματα που εισάγονται από τον εγκέφαλο. Παραδείγματος χάριν, όταν ακούει ένα πρόσωπο κάποιον παλαιό δίσκο μουσικής, σταματά σύντομα να προσέχει τον παρασιτικό θόρυβο, και απολαμβάνει τη μουσική. Ένα πρόσωπο που το κάνει αυτό εμφανίζεται συνήθως να ξεχνά για το θόρυβο συνολικά, και μπορεί να μην είναι σε θέση να προσδιορίσει μετά εάν όντως υπήρξε θόρυβος. Αυτή η επίδραση καλείται ψυχοακουστική κάλυψη (psycho-acoustical masking). Η δυνατότητα του εγκεφάλου να εκτελεί τέτοια κάλυψη είναι σημαντική για την εξέλιξη διάφορων τεχνολογιών. Σαν ένα άλλο παράδειγμα μιας ψυχοακουστικής επίδρασης, ο εγκέφαλος εμφανίζεται να χρησιμοποιεί μια συσχετική διαδικασία για την αναγνώριση προτύπων. Όταν ο άνθρωπος ακούει ένα συγκεκριμένο ηχητικό ερέθισμα μεγάλης σχετικά διάρκειας σε ένα όμως αρκετά θορυβώδες περιβάλλον κάποια μέρη του ηχητικού σήματος επικαλύπτονται από το θόρυβο. Ο εγκέφαλος έχει την τάση να συμπληρώνει αυτά τα μέρη που λείπουν, χωρίς αυτό να σημαίνει ότι τα συμπληρώνει σωστά. Αυτό είναι μια ψυχοακουστική παραίσθηση (psycho-acoustic phantom effect).

#### 4.2 ΟΡΙΑ ΑΝΤΙΛΗΨΗΣ

Το ανθρώπινο αυτί μπορεί να ακούσει ήχους από 20 Hz έως 20.000 Hz (20 kHz). Αυτό το ανώτερο όριο τείνει να μειωθεί με την ηλικία. Οι περισσότεροι ενήλικοι είναι ανίκανοι να ακούσουν επάνω από 16 kHz. Το ίδιο το αυτί δεν αποκρίνεται στις συχνότητες κάτω από 20 Hz, αλλά αυτές μπορούν να γίνουν

αντιληπτές μέσω της αίσθησης της αφής. Ο άνθρωπος αντιλαμβάνεται αλλαγές στο τονικό ύψος μεγαλύτερες από 2 Hz . Εντούτοις, ακόμα και μικρότερες διαφορές στο ύψος μπορούν να γίνουν αντιληπτές μέσω άλλων μέσων. Πχ, η παρεμβολή δύο τονικών υψών μπορεί συχνά να ακουστεί ως ένας τόνος διαφορετικής συχνότητας. Αυτή η επίδραση ονομάζεται «beating». Εντούτοις, η επίδραση της συχνότητας στο ανθρώπινο αυτί έχει μια λογαριθμική βάση. Με άλλα λόγια, το αντιλαμβανόμενο τονικό ύψος ενός ήχου συσχετίζεται με τη συχνότητα εκθετικά.

Ο αντίκτυπος αυτού είναι ότι η κλίμακα που χρησιμοποιείται στη δυτική μουσική δεν είναι μια γραμμική κλίμακα συχνότητας αλλά λογαριθμική. Το φάσμα ακουστών εντάσεων του ήχου είναι τεράστιο. Τα τύμπανα των αυτιών μας είναι ευαίσθητα στις μεταβολές της ηχητικής πίεσης.

Το χαμηλότερο όριο της ακουστικότητας καθορίζεται σε 0 dB, αλλά το ανώτερο όριο δεν είναι σαφώς καθορισμένο. Το ανώτερο όριο έχει περισσότερο να κάνει με το όριο τραυματισμού του αυτιού. Αυτό το όριο εξαρτάται επίσης από το χρόνο που εκτίθεται το αυτί μας στον ήχο.

Το αυτί μπορεί να εκτεθεί μόνο για μικρές χρονικές περιόδους σε εντάσεις άνω των 120 dB χωρίς να υποστεί μόνιμη βλάβη. Μια εξερεύνηση των χαμηλότερων ορίων της ακουστικότητας καθορίζει ότι το κατώτατο όριο στο οποίο ένας ήχος μπορεί να ακουστεί είναι εξαρτώμενο από τη συχνότητα.

Με τη μέτρηση αυτής της ελάχιστης έντασης μέσω πειραματικών δοκιμών προέκυψε το εξαρτώμενο από τη συχνότητα απόλυτο κατώτατο όριο της ακοής (Absolute Threshold of Hearing (ATH)). Το αυτί παρουσιάζει μια αιχμή της ευαισθησίας μεταξύ 1 kHz και 5 kHz, ενώ ηλικιωμένα άτομα παρουσιάζουν μειωμένη ευαισθησία σε συχνότητες άνω των 2 kHz.

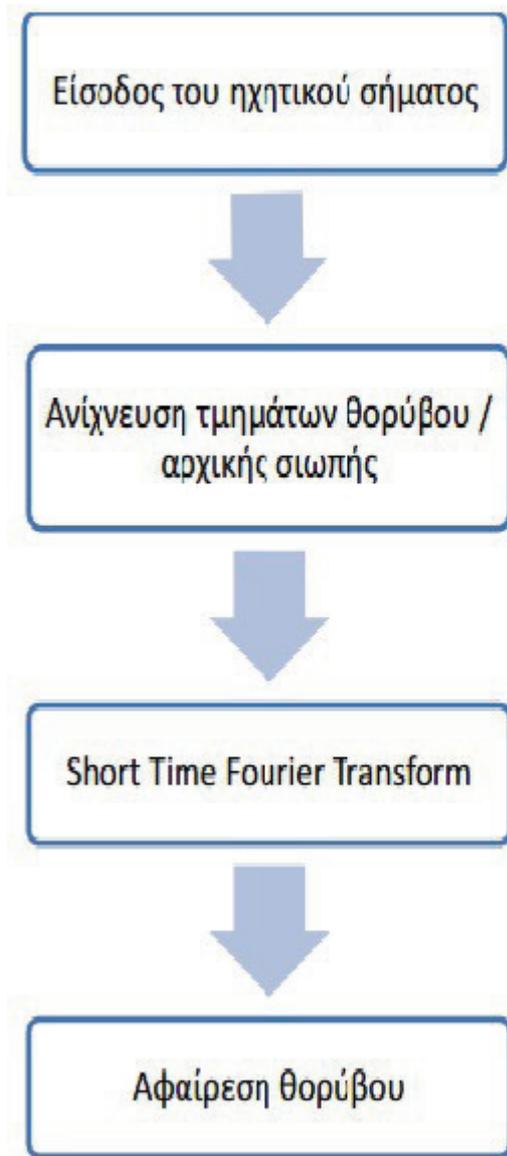
### 4.3 Το φαινόμενο της ακουστικής επικάλυψης

Σε μερικές καταστάσεις ένας ειδάλλως σαφώς ευδιάκριτος ήχος μπορεί να καλυφθεί από έναν άλλο ήχο. Παραδείγματος χάριν, η συνομιλία σε μια στάση λεωφορείου μπορεί να είναι απολύτως αδύνατη εάν ένα λεωφορείο περάσει από μπροστά. Αυτό το φαινόμενο καλείται επικάλυψη. Ένας πιο αδύνατος ήχος είναι καλυμμένος εάν γίνεται μη ακουόμενος παρουσία ενός δυνατότερου ήχου. Το φαινόμενο επικάλυψης εμφανίζεται επειδή οποιοσδήποτε δυνατός ήχος θα διαστρεβλώσει το απόλυτο κατώτατο όριο της ακρόασης, που καθιστά τους πιο ήρεμους ήχους μη ακουόμενους. Εάν δύο ήχοι εμφανίζονται ταυτόχρονα και κάποιος καλύπτεται από τον άλλον, αυτό αναφέρεται ως ταυτόχρονη επικάλυψη. Η ταυτόχρονη επικάλυψη επίσης μερικές φορές καλείται επικάλυψη συχνότητας. Η τονικότητα ενός ήχου καθορίζει μερικώς τη δυνατότητά του να καλύψει άλλους ήχους.

## ΚΕΦΑΛΑΙΟ 5

### 5.1 ΥΛΟΠΟΙΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΑΠΟΘΟΥΡΥΒΟΠΟΙΗΣΗΣ

Τα βασικά βήματα που ακολουθούνται από τους αλγορίθμους αποθορυβοποίησης φαίνονται στο παρακάτω διάγραμμα.



Σχ. 5.1 Γενικό Λογικό διάγραμμα αλγορίθμων αποθορυβοποίησης

## 5.2 Εκτίμηση θορύβου και ανίχνευση ομιλίας / ησυχίας

Μία από τις βασικότερες μεθόδους βελτίωσης ενός αλγορίθμου αποθορυβοποίησης είναι η εύρεση των περιοχών στις οποίες υπάρχει ή απουσιάζει ομιλία (και συνεπώς υπάρχει μόνο θόρυβος). Για να γίνει αυτό απαιτείται ένας αξιόπιστος ανιχνευτής ομιλίας/ησυχίας ο οποίος μπορεί να είναι καθοριστικός παράγοντας για την απόδοση του όλου συστήματος. Ο ανιχνευτής βρίσκει τα πλαίσια (frames) της ενθόρυβης ομιλίας που περιέχουν μόνο θόρυβο. Τα πλαίσια με παύσεις της ομιλίας ή σκέτου θορύβου είναι βασικά για την εκτίμηση του θορύβου. Αν δεν είναι ακριβής η ανίχνευση ομιλίας/ησυχίας τότε η ομιλία επιβαρύνεται με ηχώ και απομένων θόρυβος συνήθως παρουσιάζεται στο βελτιωμένο σήμα.

Προτείνεται μια άλλη πολύ ισχυρή και ακριβής μέθοδος που δεν απαιτεί ρητή ταξινόμηση ομιλίας/παύσης, και στην οποία η εκτίμηση του θορύβου λαμβάνεται από την ομαλοποιημένη εκτίμηση ισχύος του ενθόρυβου σήματος. Το μόνο μειονέκτημα της μεθόδου είναι η πολυπλοκότητα και η σχετικά μεγάλη απαίτηση μνήμης. Έτσι η προσέγγιση αυτή είναι ελαφρώς τροποποιημένη στην συνέχεια για να γίνει λιγότερο πολύπλοκη.

Στην αναθεωρημένη μέθοδο, για κάθε υπο-ζώνη, ένα παράθυρο χρόνου ορίζεται πάνω στο οποίο προκύπτουν τα ελάχιστα στατιστικά. Το παράθυρο είναι αρκετά μεγάλο για να γεφυρώσει τις κορυφές στο φάσμα ισχύος της ομιλίας αλλά μικρό αρκετά ώστε να ακολουθεί τον μη στατικό τύπο του θορύβου. Η υπόθεση εδώ είναι ότι οι κοιλάδες του φάσματος εξαρτώνται από το επίπεδο του θορύβου και μπορούν να χρησιμοποιηθούν για την απόκτηση της εκτίμησης του θορύβου. Αφού αυτές οι μέθοδοι δεν απαιτούν καμία ανίχνευση δραστηριότητας ομιλίας, παρουσιάζουν καλύτερη απόδοση στην μείωση του θορύβου σε ρευστό λόγο (χωρίς πολλές παύσεις).

### 5.2.1 Ανιχνευτής δραστηριότητας ομιλίας φασματικής απόστασης (Spectral Distance Voice Activity Detector)

Ο ανιχνευτής δραστηριότητας ομιλίας φασματικής απόστασης χρησιμοποιείται από τους αλγόριθμους Boll και Berouti .

Οι όροι που εμπλέκονται εδώ επεξηγούνται στην συνέχεια:

- “Σήμα (Signal)” είναι το μέτρο του φάσματος των τρεχόντων πλαισίων, που φέρει την ετικέτα του θορύβου ή ομιλίας.
- “Θόρυβος (Noise)” είναι η εκτίμηση του μέτρου φάσματος του θορύβου.
- “Μετρητής θορύβου (Noise counter)” είναι ο αριθμός των αμέσως προηγούμενων πλαισίων θορύβου.
- “Περιθώριο θορύβου (Noise margin)” (προεπιλεγμένη τιμή 3) είναι το κατώτατο όριο φασματικής απόστασης
- “Απόλυση (Hangover)” (προεπιλεγμένη τιμή 8) είναι ο αριθμός των τμημάτων θορύβου μετά από τα οποία η Σημαία Ομιλίας (Speech Flag) γίνεται μηδέν (reset).

- “Σημαία Θορύβου (Noise Flag)” τίθεται ένα αν το τμήμα έχει την ετικέτα θορύβου.
- “Απόσταση (Dist)” είναι η μέση φασματική απόσταση.

Η ανάλυση κάθε όρου δίνεται παρακάτω:

“Σήμα” είναι το φάσμα των τρεχόντων πλαισίων του σήματος εισόδου, που δίνεται σαν είσοδος στο VAD (Voice Activity Detector), και το οποίο παίρνει την ετικέτα της ομιλίας ή του θορύβου. Το περιθώριο θορύβου είναι ρυθμισμένο και σταθερό στο 3, που είναι η κατώτατη τιμή για σύγκριση με το SNR του τρέχοντος πλαισίου. Η φασματική απόσταση υπολογίζεται χρησιμοποιώντας τον τύπο του SNR

$$\text{Φασματική Απόσταση} = \log_{10}(\text{signal}) - \log_{10}(\text{noise})$$

Ο μέσος όρος της φασματικής απόστασης είναι η τιμή “Απόσταση (Dist)”. Αυτή η Dist τιμή είναι η πραγματική SNR τιμή του τρέχοντος πλαισίου. Αυτή η τιμή συγκρίνεται με το περιθώριο θορύβου και αν αυτή η τιμή είναι μικρότερη από το περιθώριο θορύβου τότε η “Σημαία Θορύβου – Noise Flag” τίθεται 1 και ο “Μετρητής Θορύβου – Noise Counter) αυξάνεται κατά 1. Αν η τιμή Dist είναι μεγαλύτερη από το περιθώριο θορύβου τότε η Σημαία Θορύβου τίθεται 0 και ο Μετρητής Θορύβου γίνεται μηδέν (reset). Αν η τιμή του Μετρητή Θορύβου είναι μεγαλύτερη από την “Περίοδο Απόλυσης - Hangover period” τότε η Σημαία ομιλίας γίνεται μηδέν και αν έχουμε το αντίστροφο τότε τίθεται ένα η Σημαία Ομιλίας.

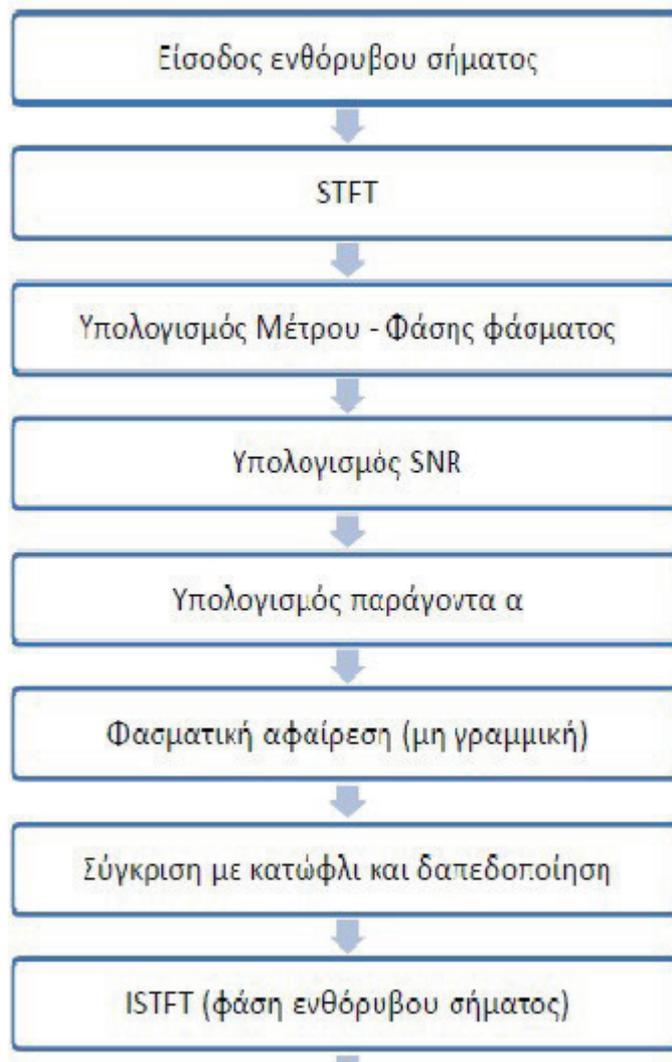
## Ο αλγόριθμος του Boll

Ο αλγόριθμος του Boll είναι ένας από τους πρώτους αλγορίθμους που χρησιμοποιήθηκαν για φασματική αφαίρεση (προτάθηκε τον Απρίλη του 1979). Μειώνει τον στατικό θόρυβο από σήματα ομιλίας αφαιρώντας τον φασματικό θόρυβο που υπολογίστηκε κατά την διάρκεια απουσίας ομιλίας. Συμπληρωματικές διεργασίες εφαρμόζονται για να μειωθεί ο απομένον θόρυβος μετά την αφαίρεση οι οποίες είναι:

- Υπολογισμός Μέσου Όρου Πλατών
- Half-wave διόρθωση
- Μείωση απομένων (μουσικού) θορύβου
- Επιπλέον μείωση απομένων θορύβου κατά τη διάρκεια περιόδων που δεν υπάρχει ομιλία

Ο αλγόριθμος του Boll δίνει καλύτερα αποτελέσματα για Φασματική Αφαίρεση Ενέργειας, αντί Ισχύος, μιας και στη δεύτερη το ποσό της παραμόρφωσης είναι πολύ μεγαλύτερο, και ο μουσικός θόρυβος ενοχλητικός.

Το λογικό διάγραμμα του αλγόριθμου του Boll που αποτελεί ουσιαστικά την υλοποίηση της γραμμικής φασματικής αφαίρεσης με κάποιες μικροβελτιώσεις παρουσιάζεται στην συνέχεια:



Σχ. 5.2 Αλγόριθμος Boll

### Ο αλγόριθμος του Berouti

Στον αλγόριθμο Berouti η αφαίρεση του φάσματος θορύβου γίνεται χρησιμοποιώντας ένα παράγοντα υπεραφαίρεσης ( $\alpha$ ), όπου  $\alpha > 1$  και ποικίλει από πλαίσιο σε πλαίσιο. Το δεύτερο στοιχείο αυτού του αλγορίθμου είναι το φασματικό δάπεδο (spectral floor) το οποίο λειτουργεί ως κατώτατο όριο για τα φασματικά στοιχεία του σήματος προς επεξεργασία. Ο αλγόριθμος Berouti, όπως και ο αλγόριθμος Boll, δίνει καλύτερα αποτελέσματα για Φασματική Αφαίρεση Ενέργειας, αντί Ισχύος, μιας και στη δεύτερη ο απομένον θόρυβος είναι περισσότερος. Το λογικό διάγραμμα της μεθόδου Berouti παρουσιάζεται παρακάτω:

Το λογικό διάγραμμα της μεθόδου Berouti παρουσιάζεται παρακάτω:



Σχ. 5.3 Μέθοδος Berouti (Φασματική Αφαίρεση με χρήση Φασματικού Δαπέδου)



## Κεφάλαιο 6

### Υλοποίηση Προγράμματος καταστολής θορύβου σε σήματα ομιλίας με χρήση φίλτρου εξομάλυνσης, στο προγραμματιστικό περιβάλλον MatLab

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

$$y[n] = x[n] + v[n] \quad (1)$$

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, εφαρμόζεται ένας διακριτός μετασχηματισμός Fourier μικρής διάρκειας (discrete short-time Fourier transform (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 (1) έχουμε:

$$Y [k, m] = X[k, m] + V [k, m] , \quad (2)$$

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:

$$|Y [k, m]|^2 \approx |X[k, m]|^2 + |V [k, m]|^2 \quad (3)$$

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  $|Y [k, m]|^2$

$$|(X\sim)[k, m]|^2 = |Y [k, m]|^2 - |(V\sim)[k, m]|^2 .$$

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

$$|X\sim[k, m]| = G[k, m]|Y [k, m]| \quad (5)$$

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

συστήματος):

$$G[k, m] = \left( \frac{|Y[k, m]|^2 - |\tilde{V}[k, m]|^2}{|Y[k, m]|^2} \right)^{\frac{1}{2}} \quad (6)$$

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

$$G_p[k, m] = \left( \frac{|Y[k, m]|^\alpha - \beta |\tilde{V}[k, m]|^\alpha}{|Y[k, m]|^\alpha} \right)^{\frac{1}{\alpha}} \quad (7)$$

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 τύπο (8) :

$$G_p[k, m] = \max\left(10^{\frac{-L}{20}}, G_p\right) \quad (8)$$

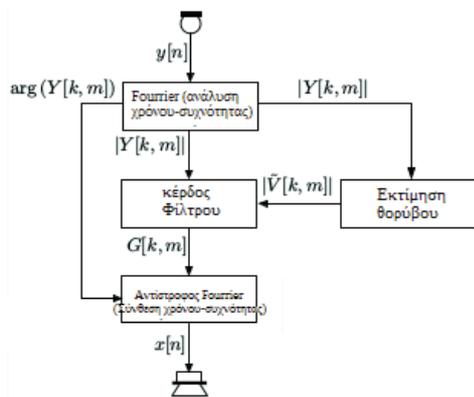
Οπότε, ο "καθαρισμένος" ήχος υπολογίζεται βάσει της σχέσης (5) αφού έχει υπολογιστεί το κέρδος φίλτρου από την (7) όπου η τιμή του  $\alpha$  ορίζεται ίση με 2 και του  $\beta$  ίση με 1. Για τον υπολογισμό όμως αυτό απαιτείται μία εκτίμηση του θορύβου  $|\tilde{V}[k, m]|$ . Η εκτίμηση αυτή εκτελείται 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 (βήμα 3) 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 (βήμα 7).

Αφού λοιπόν έχει εκτιμηθεί ο θόρυβος (με τον πρώτο τρόπο αρχικά και κατά την εκτέλεση του προγράμματος με τον δεύτερο) εφαρμόζεται η σχέση (5) (βήμα 10) σε συνδυασμό με την (7) (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 του χρόνου με αντίστροφο μετασχηματισμό Fourier (βήμα 11) και γίνεται η επανασύσταση του από τα πλαίσια (βήμα 12) και αποτελεί το αποτέλεσμα (την έξοδο) του προγράμματος.

Η έξοδος του προγράμματος αυτού σχεδιάζεται σε γραφική παράσταση (βήμα 15) ώστε να συγκριθεί με το αρχικό σήμα που έχει σχεδιαστεί (βήμα 14) και εξάγεται σε ένα .wav αρχείο (βήμα 16), το όνομα του οποίου έχει καθορίσει ο χρήστης κατά την κλήση του.

Η παραπάνω διαδικασία εμφανίζεται στο παρακάτω σχήμα (σχήμα 1) όπου φαίνεται το διάγραμμα του αλγορίθμου της.



Σχήμα 1: Σχηματικό διάγραμμα του αλγόριθμου

## Κώδικας Matlab

Χρήση αρχείου noisesub.m:

```
noisesub(filename,outfile);
```

**Βήμα 1:** Έλεγχος για την σωστή κλήση της συνάρτησης και αν καλείται λανθασμένα, εκτύπωση του μηνύματος εντός της παρένθεσης:

```
if nargin<2
    fprintf('Usage: noisesub_SNR(noisyfile.wav,outFile.wav) \n\n');
    return;
end
```

Στην συνέχεια, γίνεται η “φόρτωση” των δεδομένων του αρχείου ήχου που ορίστηκε κατά την κλήση της συνάρτησης όπου, στην μεταβλητή y αποθηκεύονται τα δεδομένα του ήχου, στην μεταβλητή Srate η συχνότητα του ήχου (ρυθμός δειγματοληψίας) και στην μεταβλητή nbits το πλήθος των ψηφίων (bits) που χρησιμοποιείται για το συγκεκριμένο αρχείο ήχου.:

```
close all
[y,Srate,nbits]=wavread(filename);
```

**Βήμα 2:** Αρχικοποίηση και ορισμός των μεταβλητών που θα χρησιμοποιηθούν. Ορίζεται η μεταβλητή `len` ως το μέγεθος των πλαισίων στα οποία θα χωριστεί ο αρχικός ήχος, η μεταβλητή `PERC` ως το ποσοστό της επικάλυψης μεταξύ των πλαισίων και οι μεταβλητές `len1` και `len2`, με την πρώτη να καθορίζει το πλήθος των δειγμάτων επικάλυψης (μετατροπή, στην ουσία του ποσοστού `PERC` σε πλήθος δειγμάτων) και την δεύτερη να καθορίζεται από το πλήθος των δειγμάτων από τα οποία αποτελείται κάθε πλαίσιο από το οποίο έχει αφαιρεθεί το πλήθος των δειγμάτων επικάλυψης:

```
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;
```

Στην συνέχεια, καθορίζονται οι μεταβλητές  $\alpha$  (α) και  $\beta$  (β) της εξίσωσης (7), η μεταβλητή `L` που καθορίζει το ελάχιστο όριο της  $G_p$  της εξίσωσης 8, και η μεταβλητή `noise_frames` η οποία καθορίζει το πλήθος των αρχικών πλαισίων που θα θεωρηθούν σαν “καθαρός θόρυβος”:

```
alpha = 2.0; % power exponent
beta = 1;
L = 8;
noise_frames = 10;
```

Τέλος, καθορίζεται ένα συμμετρικό παράθυρο `Hamming`, `len` σημείων, που θα χρησιμοποιηθεί στην επεξεργασία των δεδομένων

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

**Βήμα 3:** Έχοντας ορίσει το πλήθος των αρχικών πλαισίων τα οποία θα θεωρηθούν ως “καθαρός θόρυβος”, γίνεται ο υπολογισμός του επιπέδου του θορύβου αυτού αθροίζοντας τον θόρυβο κάθε πλαισίου με την `for`, με το `k` να παίρνει τιμές από 1 μέχρι `noise_frames` (10 στην προκειμένη περίπτωση). Η μεταβλητή `noise_mu` είναι η μέση αυτή τιμή του θορύβου και η μεταβλητή `noise_mu2` είναι το τετράγωνο της.

```
nFFT=2*2^nextpow2(len);
noise_mean = zeros(nFFT,1);
j = 1;
for k=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;
```

**Βήμα 4:** Στο βήμα αυτό καθορίζονται κάποιες άλλες μεταβλητές που θα χρησιμοποιηθούν στην επεξεργασία των δεδομένων του ήχου,  $k$ ,  $img$ ,  $aa$ ,  $\mu$ ,  $\eta$  και  $ksi\_min$  και “δεσμεύεται” η απαραίτητη μνήμη που θα χρειαστεί το πρόγραμμα στην συνέχεια, ορίζοντας τα σήματα  $x\_old$  και  $xfinal$ , οι τιμές των οποίων αρχικοποιούνται ως μηδενικά και θα υπολογιστούν στην συνέχεια, καθώς και η μεταβλητή  $Nframes$  η οποία είναι το πλήθος των πλαισίων στα οποία θα χωριστεί το σήμα.

```
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);
```

Με το βήμα 4 ολοκληρώνεται η διαδικασία αρχικοποίησης του αλγορίθμου, έχοντας ορίσει τις απαραίτητες μεταβλητές και έχοντας υπολογίσει μία πρώτη εκτίμηση του επιπέδου θορύβου του σήματος. Οπότε, το πρόγραμμα είναι έτοιμο να αρχίσει την επεξεργασία των δεδομένων του θορυβώδη ήχου.

**Βήμα 5:** Στο βήμα αυτό ξεκινάει η επεξεργασία του θορυβώδη ήχου. Η διαδικασία που ακολουθείται αφορά ένα πλαίσιο κάθε φορά και επαναλαμβάνεται, με χρήση της `for`, για  $n$  από 1 μέχρι το πλήθος των πλαισίων στα οποία θα χωριστεί το σήμα του ήχου, για όλα τα πλαίσια που έχουν καθοριστεί, το πλήθος των οποίων θα είναι  $Nframes$ , όπως προαναφέρθηκε. Αρχικά ορίζεται το πλαίσιο στο οποίο θα γίνει επεξεργασία, με χρήση του παραθύρου `Hamming` που ορίστηκε παραπάνω, και δημιουργούνται τα δεδομένα `insign`. Στα δεδομένα αυτά, εφαρμόζεται ο μετασχηματισμός `Fourier`, και αποθηκεύονται στην μεταβλητή `spec`, από την οποία στην συνέχεια καθορίζεται η μεταβλητή `sig`, η οποία περιέχει το πλάτος των `spec`, με το `sig2` να είναι το τετράγωνο του παραπάνω πλάτους, και η μεταβλητή `theta`, η οποία περιέχει πληροφορίες για την φάση του `spec`.

```
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);
```

**Βήμα 6:** Στο βήμα αυτό γίνεται μία αρχική εκτίμηση του θορύβου του πλαισίου. Αρχικά καθορίζεται ένα ελάχιστο `SNR` (`Signal to Noise ratio`) το οποίο αποθηκεύεται στην μεταβλητή `gamma_k`, και αν το `SNR` του πλαισίου είναι μικρότερο του 40, τίθεται ως 40, αλλιώς έχει την τιμή του `SNR` (`sig2./noise_mu2`). Στην συνέχεια υπολογίζεται η τιμή της μεταβλητής `ksi`, η οποία θα είναι ίση με για το πρώτο πλαίσιο και για τα επόμενα, η μεταβλητή `aa` αντικαθίσταται από το , όπου η τιμή του `noise_mu2` είναι η εκτίμηση θορύβου με βάση τα αρχικά πλαίσια και η τιμή του `Xk_prev` θα υπολογιστεί στο βήμα 10. Τέλος, τίθεται το ελάχιστο όριο του `ksi`, σύμφωνα με την τιμή του `ksi_min` που καθορίστηκε στο βήμα 4.

```

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

```

**Βήμα 7:** Στο βήμα αυτό, υλοποιείται ένας ανιχνευτής φωνής (VAD:Voice activity detector) ως εξής: Αρχικά καθορίζεται η μεταβλητή  $\log\_sigma\_k$ , σύμφωνα με την τιμή της οποίας θα καθοριστεί η μεταβλητή  $vad\_decision$ , η οποία χρησιμοποιείται από το VAD για την λήψη της απόφασης αν το πλαίσιο περιέχει μόνο θόρυβο, συγκρίνοντας την με την τιμή της  $\eta$  που καθορίστηκε στο βήμα 4. Αν η τιμή της  $\eta$  είναι μεγαλύτερη, τότε το πλαίσιο θεωρείται ότι περιέχει μόνο θόρυβο και αλλάζει την τιμή του  $noise\_mu2$  (ενημερώνει το “προφίλ” θορύβου που εκτιμήθηκε στο βήμα 3).

```

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

```

**Βήμα 8:** Στο βήμα αυτό γίνεται υπολογισμός της τιμής της συνάρτησης  $hw$ , η οποία, σε συνδυασμό με την  $G_p$  που θα υπολογισθεί στο βήμα 9, θα χρησιμοποιηθεί για την κατάλληλη μορφοποίηση του θορυβώδους σήματος ώστε αυτό να “καθαρίσει”.

```

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);

```

**Βήμα 9:** Στο βήμα αυτό γίνεται υπολογισμός της τιμής της συνάρτησης  $G_p$ . Ο υπολογισμός αυτός μπορεί να γίνει με δύο τρόπους:

- με χρήση της εξίσωσης (6)<sup>1</sup>:

```

Gp(n,:) = smooth(((abs(sig.^2 - noise_mu.^2))./(sig.^2)).^(1/2),10);

```

<sup>1</sup>

Στην περίπτωση αυτή, γίνεται χρήση της εξίσωσης (6) για τον υπολογισμό του  $G_p$ , η τιμή της οποίας εξομαλύνεται με χρήση της συνάρτησης  $smooth(\alpha,\beta)$ , η οποία εξομαλύνει τα δεδομένα του διανύσματος  $\alpha$  χρησιμοποιώντας έναν κινούμενο μέσο όρο  $\beta$  σημείων.

και με χρήση της εξίσωσης (7)

```
Gp(n,:) = ((sig.^alpha - beta*noise_mu.^alpha)/(sig.^alpha)).^(1/alpha);
Gp(n,:)=max(10^(-L/20),Gp(n,:));
```

Οι δύο αυτοί τρόποι θα συγκριθούν μεταξύ τους στην επόμενη παράγραφο.

**Βήμα 10:** Αφού έχουν υπολογισθεί οι τιμές των  $hw$  και  $Gp$ , στο βήμα αυτό γίνεται το φιλτράρισμα του ηχητικού σήματος, σύμφωνα με την εξίσωση (5) και ορίζεται και η τιμή της  $Xk\_prev$  που χρησιμοποιείται στο βήμα 6, όταν αυτό εκτελεστεί για το επόμενο πλαίσιο:

```
sig = sig.*hw.*Gp(n,:);
Xk_prev = sig.^2;
```

**Βήμα 11:** Αντίστροφος μετασχηματισμός Fourier του φιλτραρισμένου πλαισίου του σήματος:

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

**Βήμα 12:** Επανασύσταση του ηχητικού σήματος από τα επιμέρους πλαίσια, όπου σε κάθε επανάληψη προσθέτονται τα δεδομένα του φιλτραρισμένου πλαισίου στο  $xfinal$ , το οποίο θα περιέχει, μετά την ολοκλήρωση του προγράμματος, τα δεδομένα του επεξεργασμένου ηχητικού σήματος:

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

**Βήμα 13:** Στο βήμα αυτό αλλάζει η τιμή της μεταβλητής  $k$ , η οποία ορίστηκε στο βήμα 4 ίση με 1. Η νέα τιμή της θα είναι ίση με την παλιά με πρόσθεση της μεταβλητής  $len2$ , αποτελώντας στην ουσία έναν “δείκτη”, καθορίζοντας την αρχή του επόμενου πλαισίου.

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

Μετά από το βήμα αυτό, η διαδικασία και ο αλγόριθμος είτε επιστρέφει στο βήμα 5 (αν δεν έχουν επεξεργαστεί όλα τα πλαίσια) είτε εξέρχεται του βρόχου επανάληψης και συνεχίζει με το βήμα 14.

**Βήμα 14:** Σχεδιάζεται ο αρχικός “θορυβώδης” ήχος:

```
figure(1)
plot(x)
```

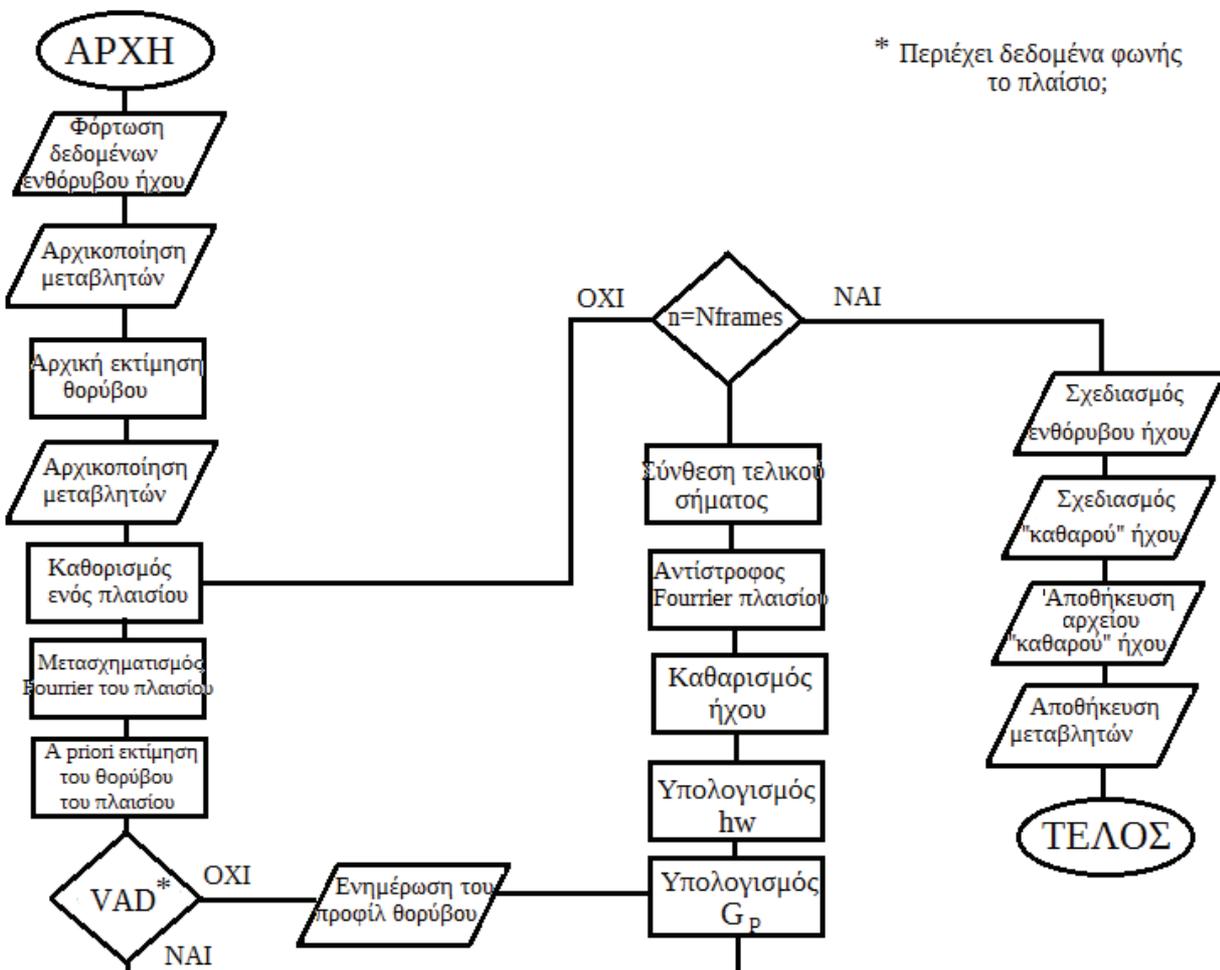
**Βήμα 15:** Σχεδιάζεται ο τελικός “καθαρός” ήχος:

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

**Βήμα 16** (Τερματισμός του προγράμματος): Αποθηκεύεται το τελικό επεξεργασμένο ηχητικό σήμα στο αρχείο που δηλώθηκε στην κλήση της συνάρτησης, δημιουργώντας στην ουσία ένα αρχείο ήχου με τα ίδια χαρακτηριστικά με το αρχικό (αναφορικά με τις τιμές των Srate και nbits) και αποθηκεύονται οι τιμές των μεταβλητών που χρησιμοποιήθηκαν στο αρχείο “values.mat”:

```
wavwrite(xfinal,Srate,nbits,outfile);
save values.mat
```

Η παραπάνω διαδικασία παρουσιάζεται σχηματικά στο διάγραμμα ροής που παρουσιάζεται στο παρακάτω σχήμα (σχήμα 2):



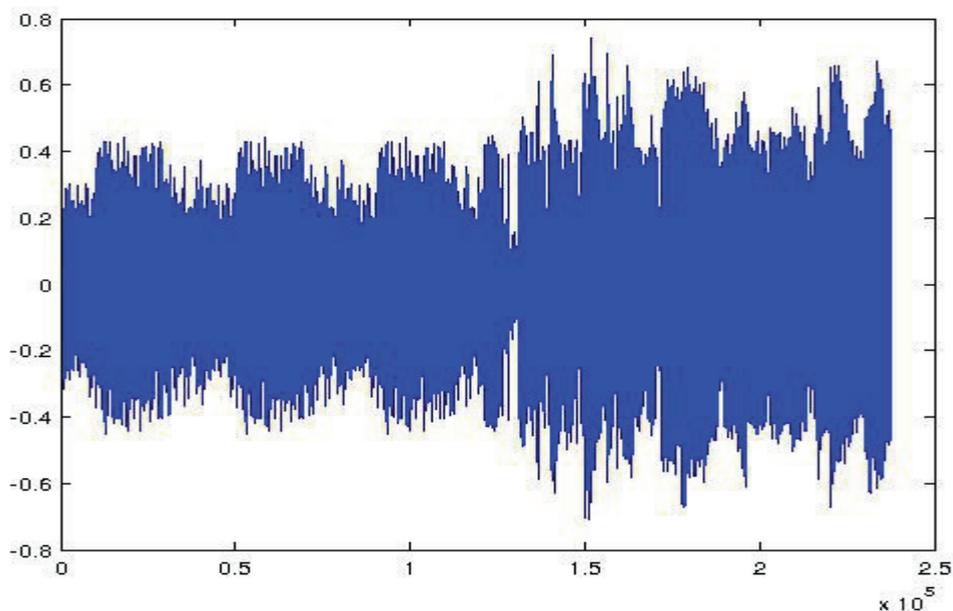
Σχήμα 2: Διάγραμμα ροής του αλγόριθμου του προγράμματος που υλοποιήθηκε.

## Εκτέλεση του προγράμματος

Στην εφαρμογή του παραπάνω κώδικα, χρησιμοποιήθηκε το “θορυβώδες” αρχείο ήχου “wavsnd.wav” του οποίου η διάρκεια είναι  $\approx 5$  sec, και τα δεδομένα του αποθηκεύονται στην μεταβλητή  $y$  (Sample rate=44100 Hz, Bits per Sample=16 bit) (βήμα 1):

Name	Size	Bytes	Class	Attributes
$y$	237469x1	1899752	double	

Το διάγραμμα των δεδομένων αυτών φαίνεται στο παρακάτω σχήμα (βήμα 13):

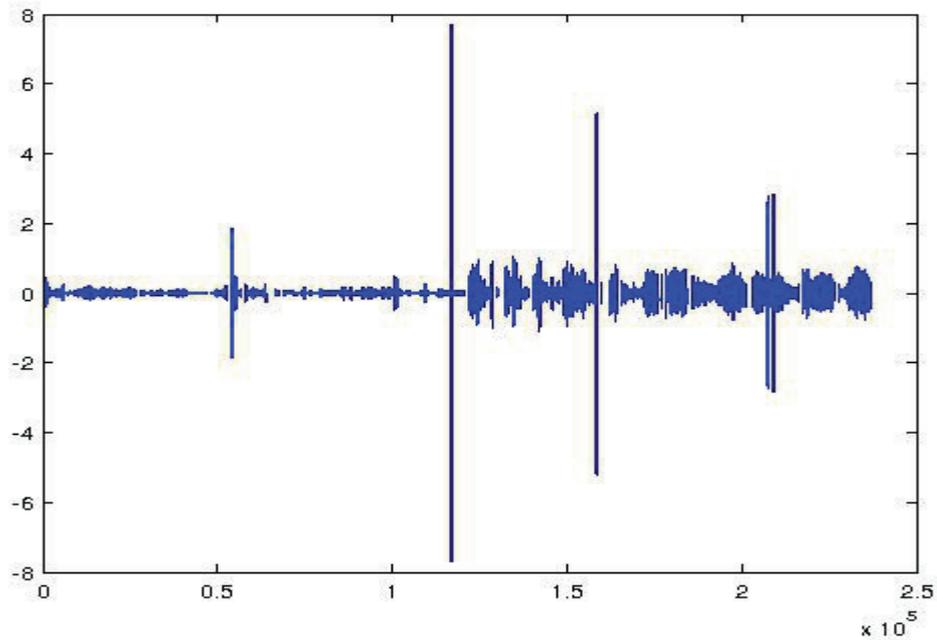


Σχήμα 3: Αρχικός ήχος

Οι τιμές που εισήχθησαν για την εκτέλεση του προγράμματος (βήμα 2) είναι:

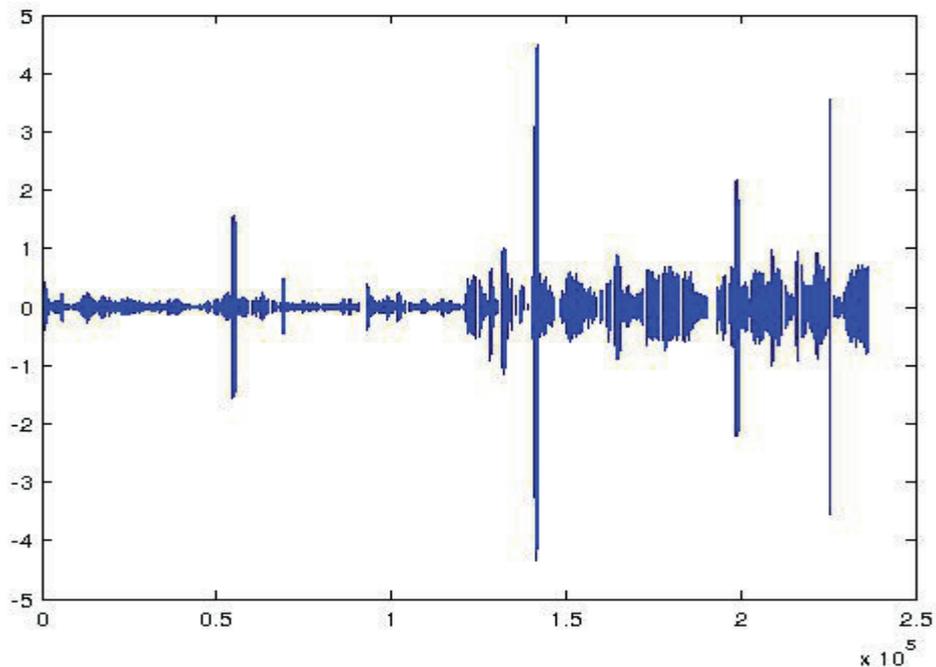
- 10 πλαίσια θεωρούνται ως “καθαρός θόρυβος”
- $\alpha=2$
- $\beta=1$

Το αποτέλεσμα του προγράμματος, με χρήση της σχέσης (7) για τον υπολογισμό του  $G_p$  (βήμα 9) και ορισμός του  $L=8$  (εξίσωση 8) είναι ο καθαρισμένος ήχος που φαίνεται στο σχήμα 4, όπου παρατηρούμε ότι έχει περιοριστεί πολύ 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.



Σχήμα 4: “Καθαρός” ήχος (με χρήση των εξισώσεων (7) και (8) )

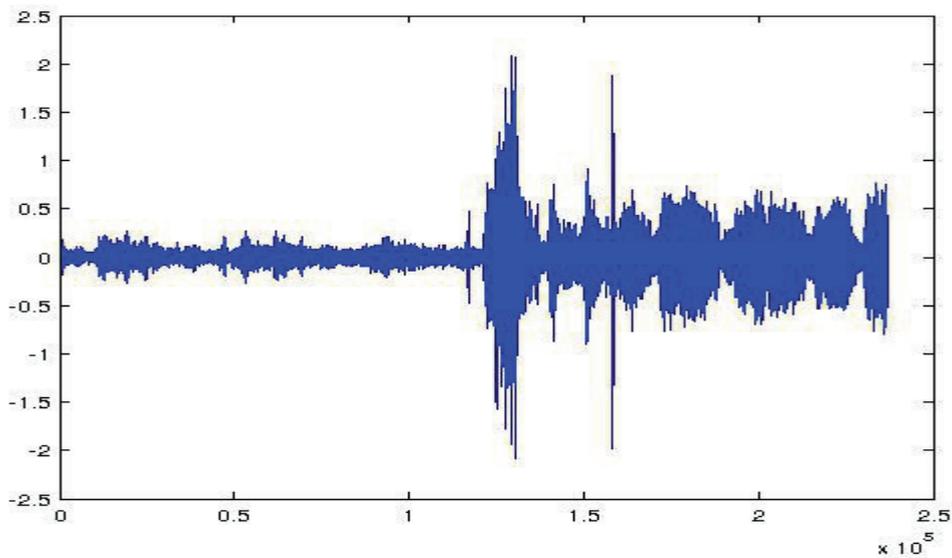
Στην συνέχεια, “απενεργοποιήθηκε” η εξίσωση (8), οπότε το κέρδος  $G_p$  δεν περιορίζεται από το  $L$ , και εκτελέστηκε ξανά το πρόγραμμα με το ίδιο αρχικό αρχείο ήχου.



Σχήμα 5: “Καθαρός” ήχος (με χρήση της εξίσωσης (7) αλλά όχι της (8) )

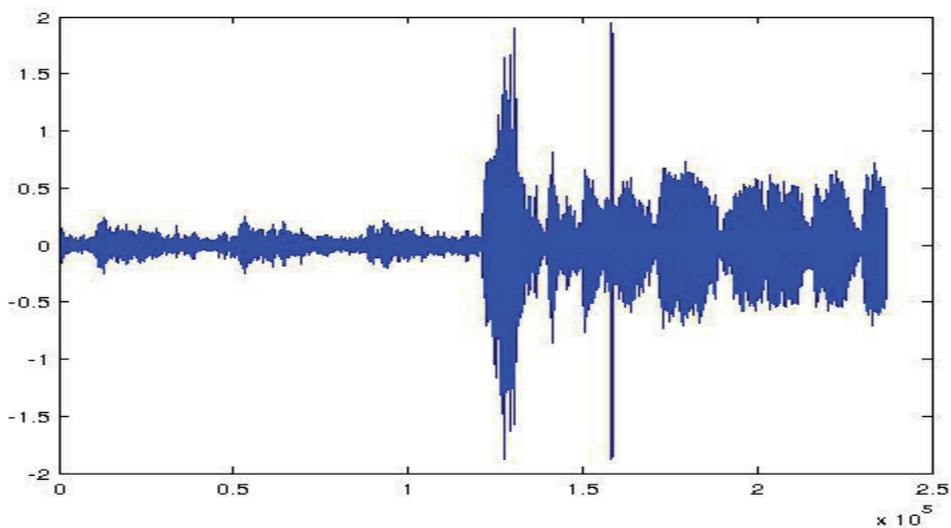
Η κυματομορφή του “καθαρού” ήχου που παράχθηκε από την εκτέλεση αυτή παρουσιάζεται στο παραπάνω σχήμα (σχήμα 5). Όπως γίνεται φανερό από το διάγραμμα αυτό, στην προκειμένη περίπτωση ο καθαρισμός ο του ήχου δεν είναι τόσο αποδοτικός όπως στην προηγούμενη περίπτωση. Αυτό γίνεται φανερό και με την αναπαραγωγή του αρχείου που δημιουργήθηκε, όπου, συγκριτικά με το αρχείο της προηγούμενης περίπτωσης, το ακουστικό αποτέλεσμα είναι χειρότερο από την προηγούμενη περίπτωση.

\_\_\_\_Στη συνέχεια, εκτελέστηκε το πρόγραμμα και με χρήση της εξίσωσης (6) για τονυπολογισμό του  $G_p$ . Το αποτέλεσμα της εκτέλεσης του αλγορίθμου, αντικαθιστώντας την εξίσωση (7) με την (6) και χρησιμοποιώντας την τιμή 10 για την εξομάλυνση (χρησιμοποιώντας έναν μέσο όρο 10 σημείων) παρουσιάζεται στο σχήμα 6, 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.

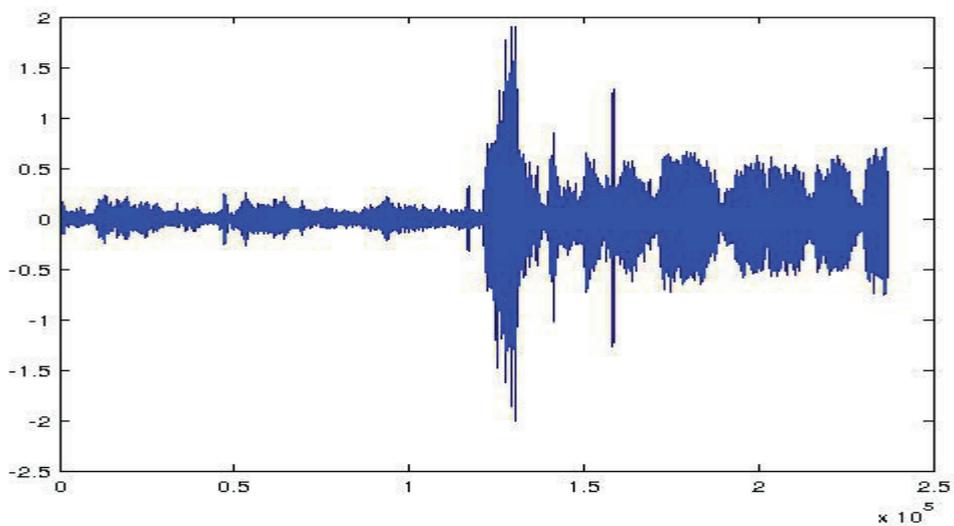


Σχήμα 6: “Καθαρός” ήχος (με χρήση της εξίσωσης (6) με smooth 10 σημείων)

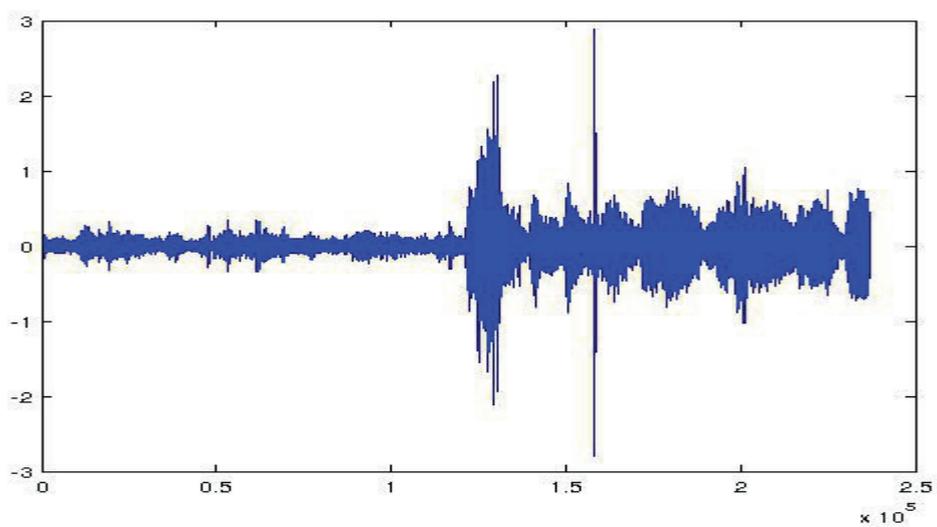
Τέλος, έγινε η εκτέλεση του προγράμματος για την περίπτωση που χρησιμοποιείται απευθείας η εξίσωση (6) χωρίς εξομάλυνση, καθώς και για τις περιπτώσεις όπου χρησιμοποιείται εξομάλυνση 5, 20 και 40 σημείων. Τα διαγράμματα των παραγόμενων αρχείων ήχου παρουσιάζονται στα παρακάτω σχήματα (σχήματα 7, 8, 9 και 10 αντίστοιχα).



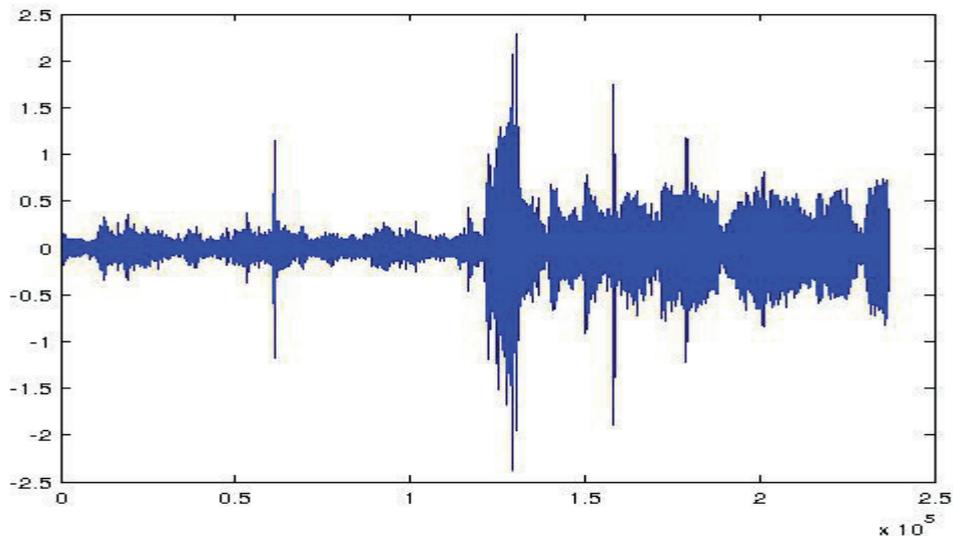
Σχήμα 7: Επεξεργασμένο ηχητικό σήμα χωρίς εξομάλυνση



Σχήμα 8: Επεξεργασμένο ηχητικό σήμα με εξομάλυνση 5 σημείων



Σχήμα 9: Επεξεργασμένο ηχητικό σήμα με εξομάλυνση 20 σημείων

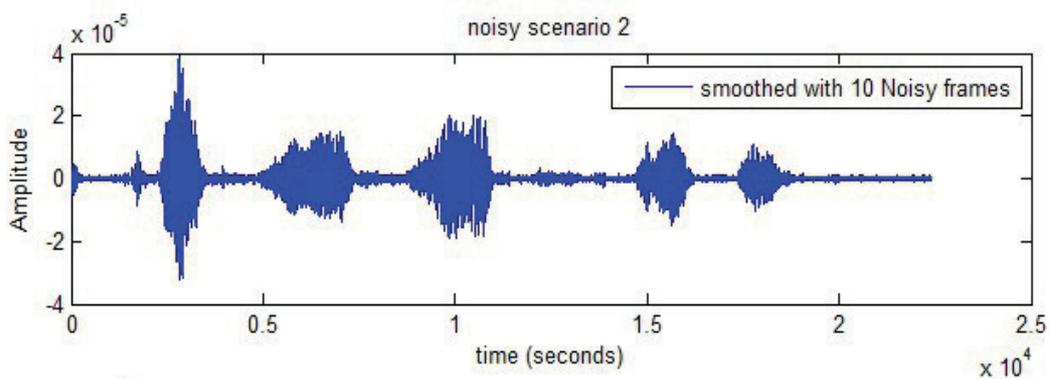
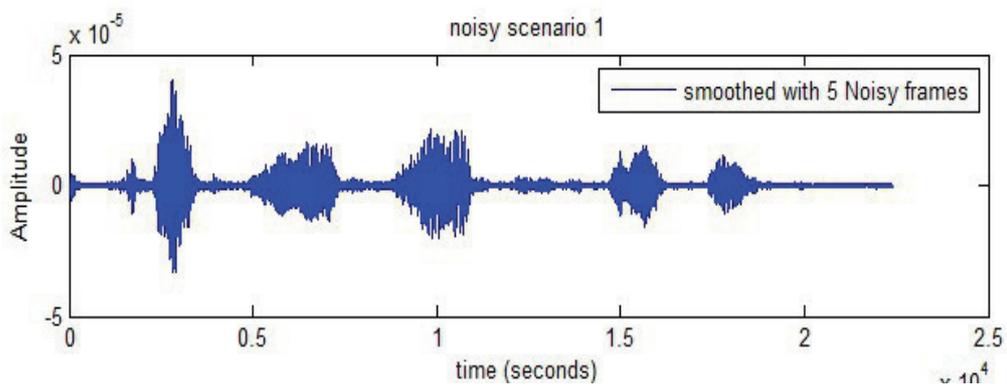
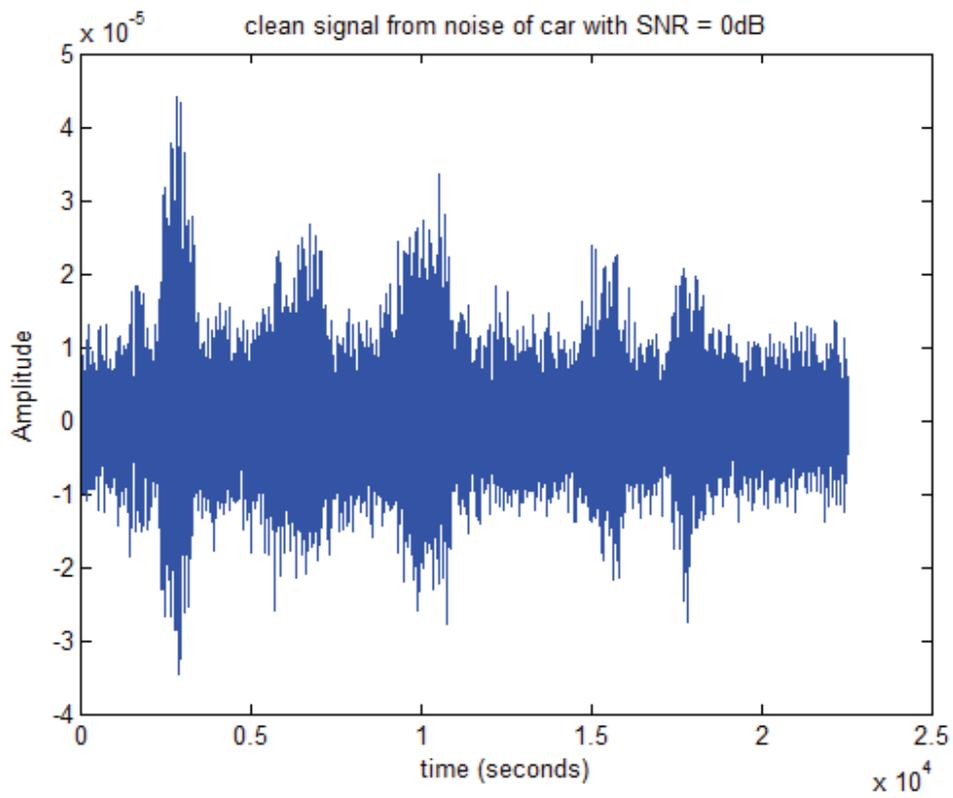


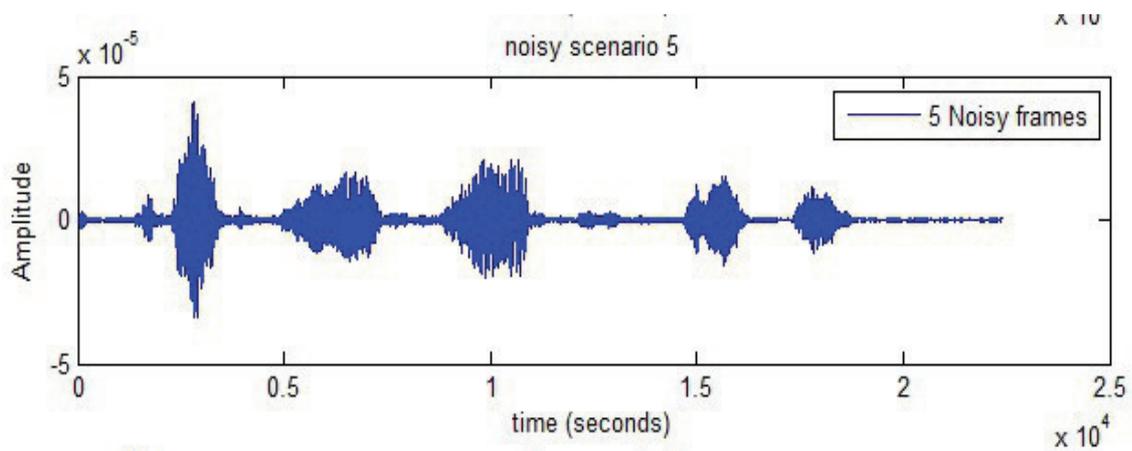
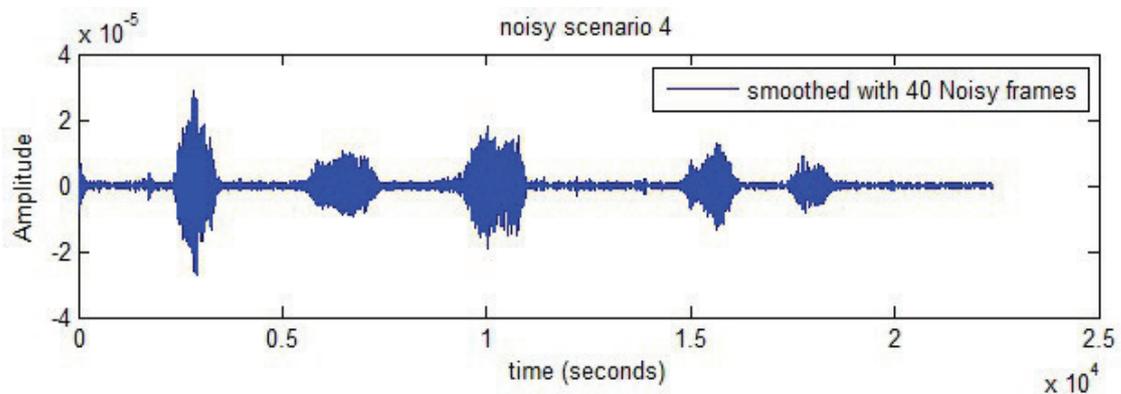
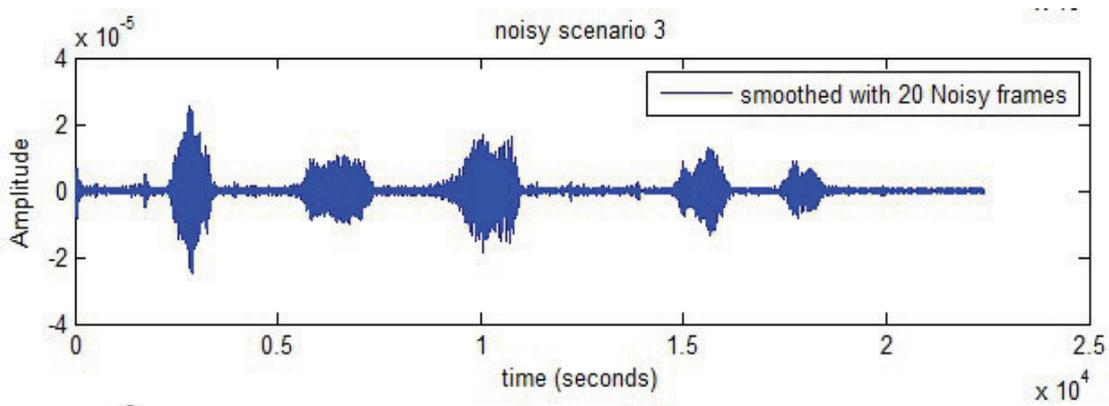
Σχήμα 10: “Επεξεργασμένο ηχητικό σήμα με εξομάλυνση 5 σημείων 40 σημείων)

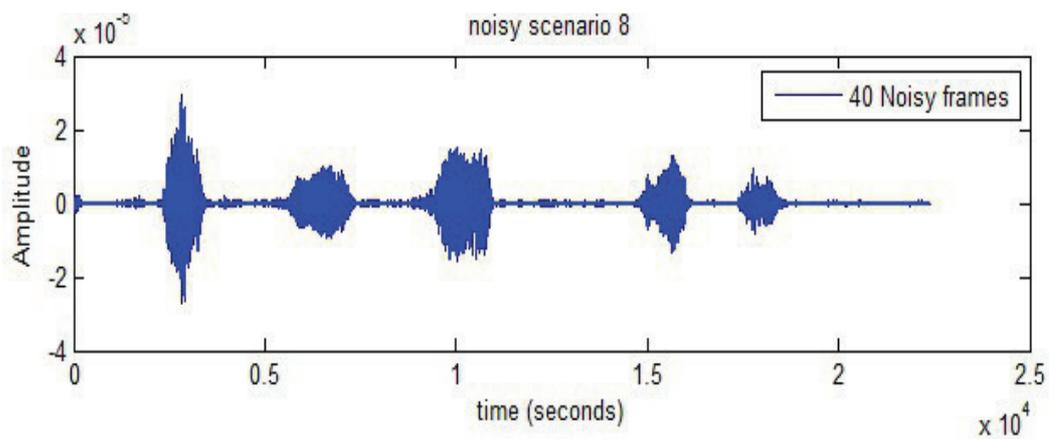
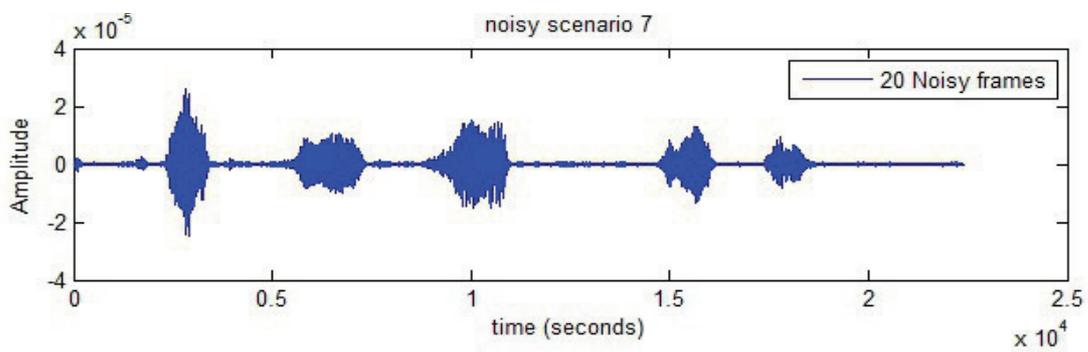
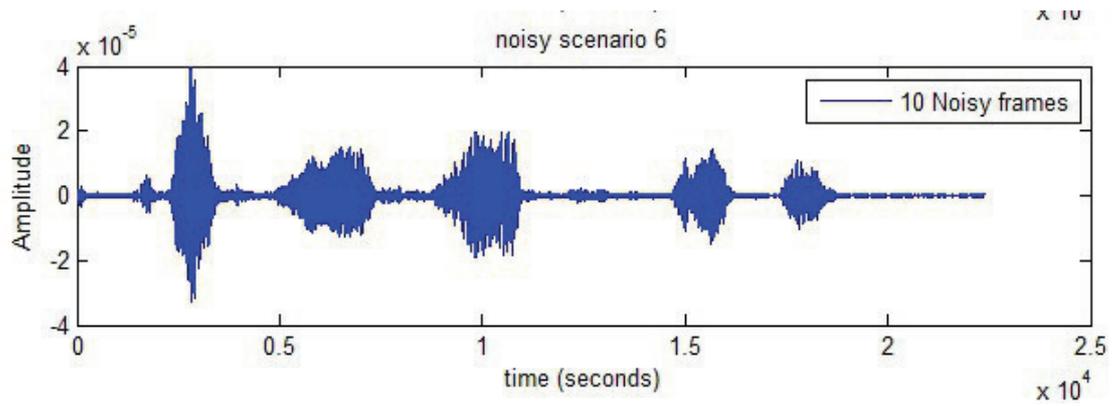
Από τα παραπάνω διαγράμματα, φαίνεται ότι η εξομάλυνση του κέρδους του φίλτρου βελτιώνει την ηχητική ποιότητα, μειώνοντας περισσότερο τον θόρυβο στον τελικό ήχο, συγκριτικά με την περίπτωση που δεν χρησιμοποιείται η συνάρτηση εξομάλυνσης. Αυτό γίνεται αντιληπτό και με αναπαραγωγή του τελικού αρχείου, όπου στην περίπτωση όπου δεν χρησιμοποιείται η συνάρτηση εξομάλυνσης, 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 (σχήμα 6 και 9 αντίστοιχα).

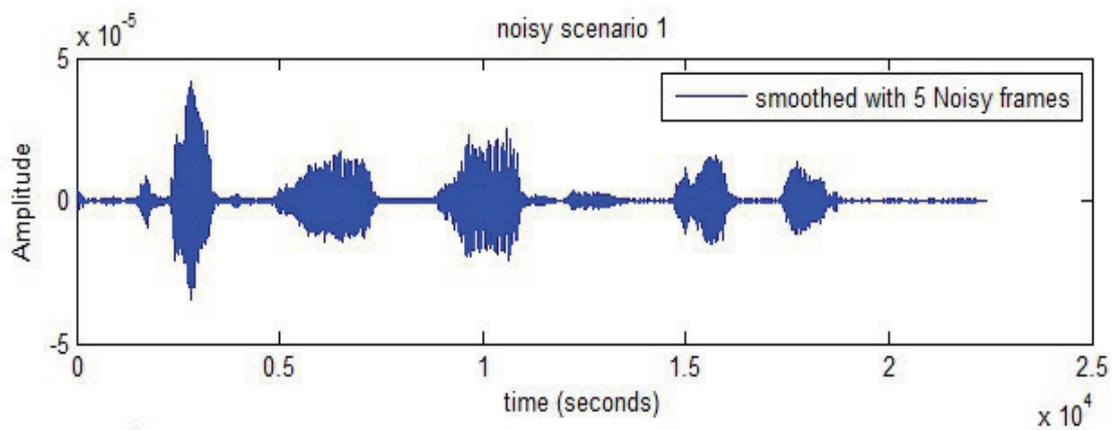
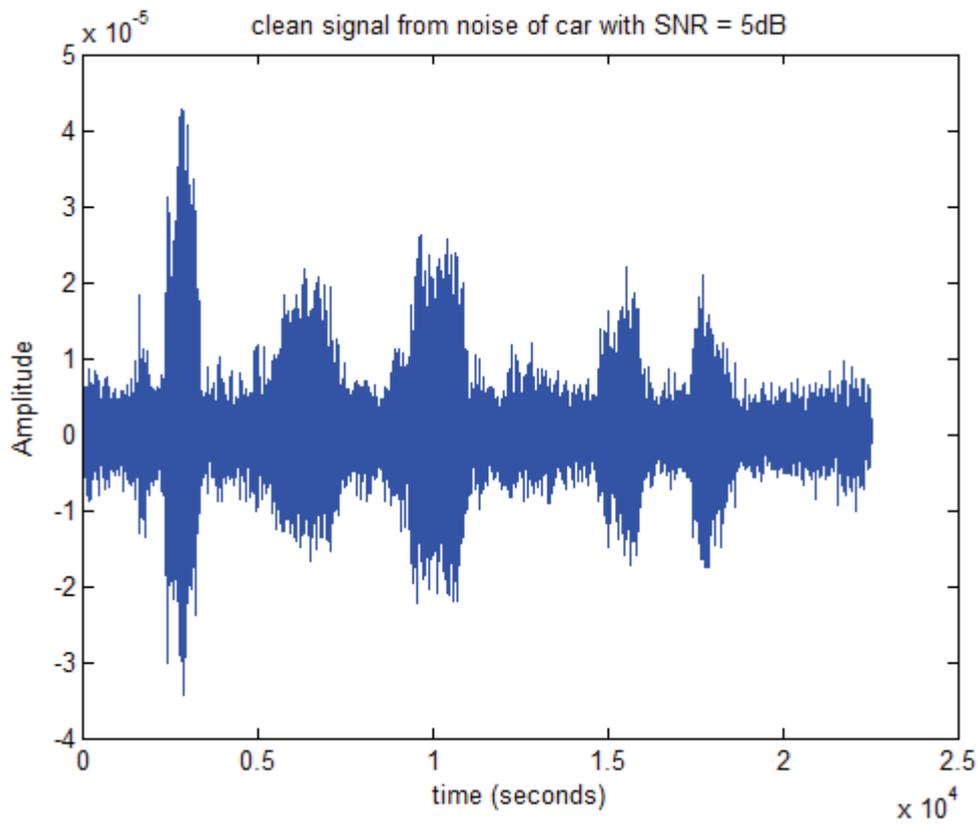
Προκειμένου να επαληθεύσουμε τα παραπάνω αποτελέσματα και σε περισσότερα σενάρια, θα χρησιμοποιηθούν δύο διαφορετικές εκδοχές ενός σήματος ομιλίας. Κατά την πρώτη, έγινε εισαγωγή θορύβου από περιβάλλον ενός αυτοκινήτου και στο άλλο από ενός αεροδρομίου. Και στις δύο περιπτώσεις δημιουργήθηκαν τρία εναλλακτικά αρχεία ήχου, με 0, 5 και 10 dB SNR . Με τη βοήθεια των subplots αναπαριστούμε τα 8 διαφορετικά σενάρια, 4 με τη χρήση εξομάλυνσης και 4 χωρίς, κάθε ένα για τις τιμές 5, 10, 20 και 40 που αντιστοιχούν στο μήκος του παραθύρου που εκτιμάει το θόρυβο, ενώ προηγουμένως προβάλλουμε σε ένα ξεχωριστό σχήμα το θορυβώδες σήμα με κατάλληλο τίτλο.

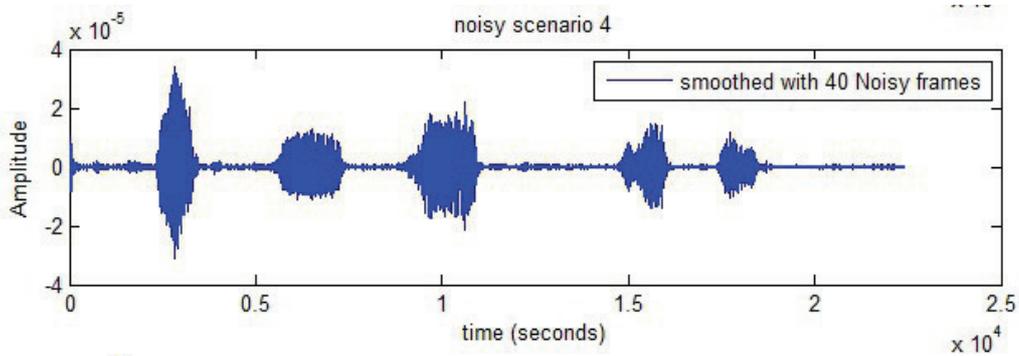
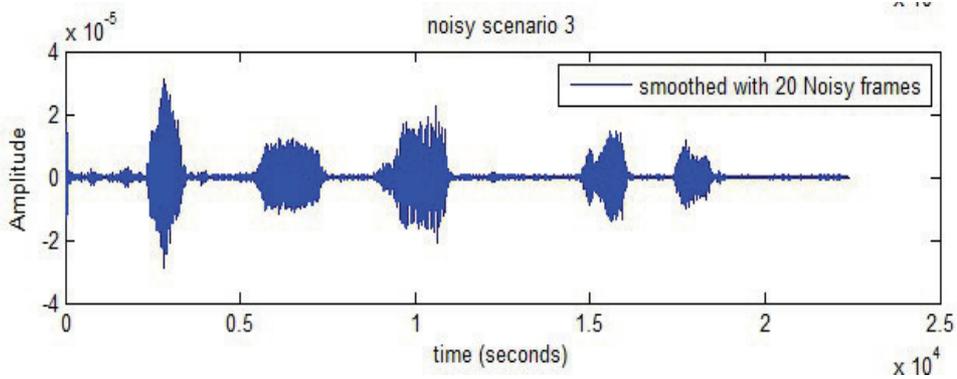
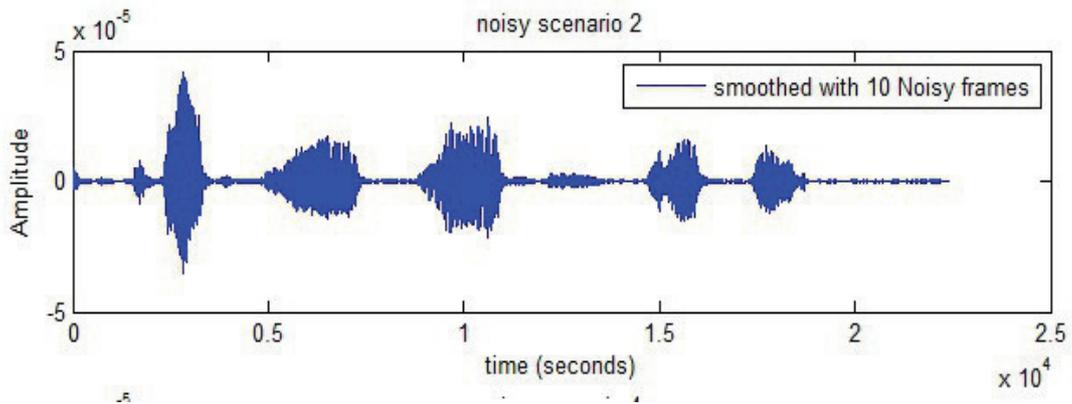
θα εμφανίσουμε για κάθε μία εκ των 6 περιπτώσεων τις 6 προηγούμενες περιπτώσεις αποθορυβοποίησης και το καθαρό σήμα στην αρχή:

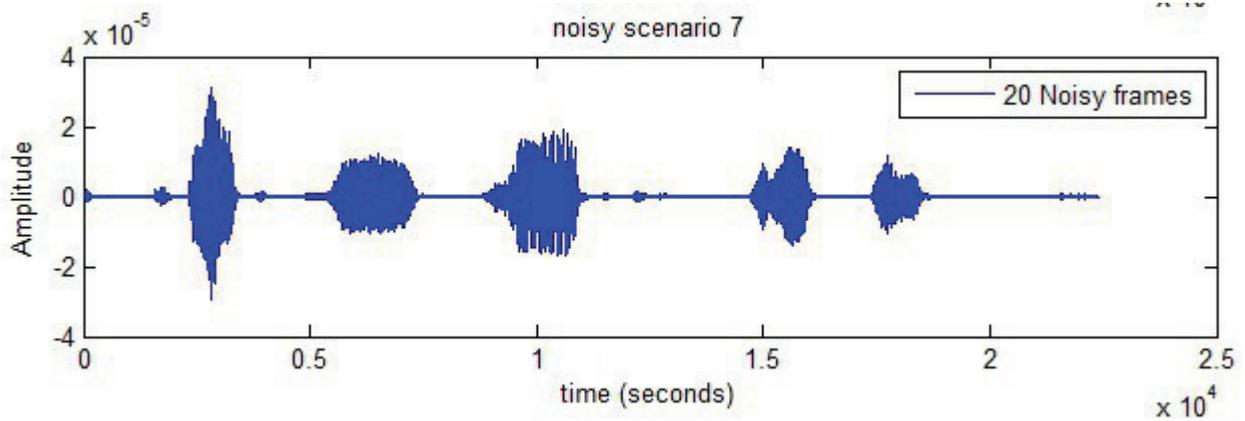
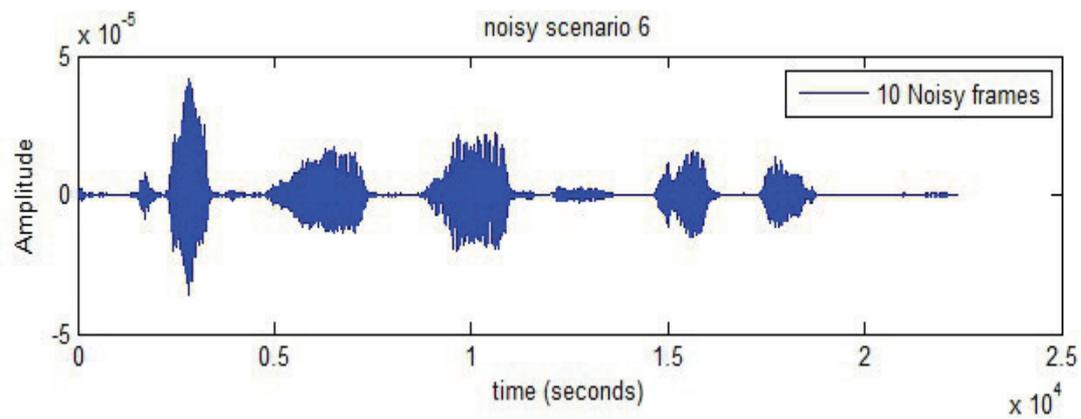
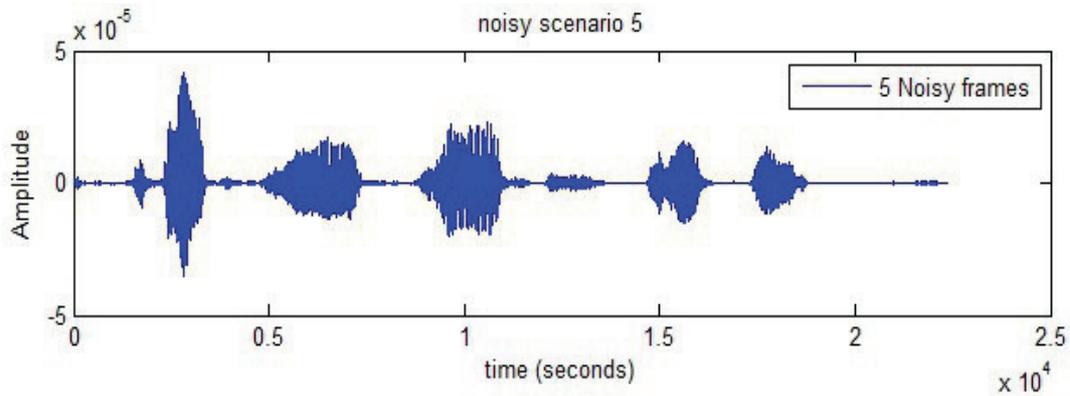


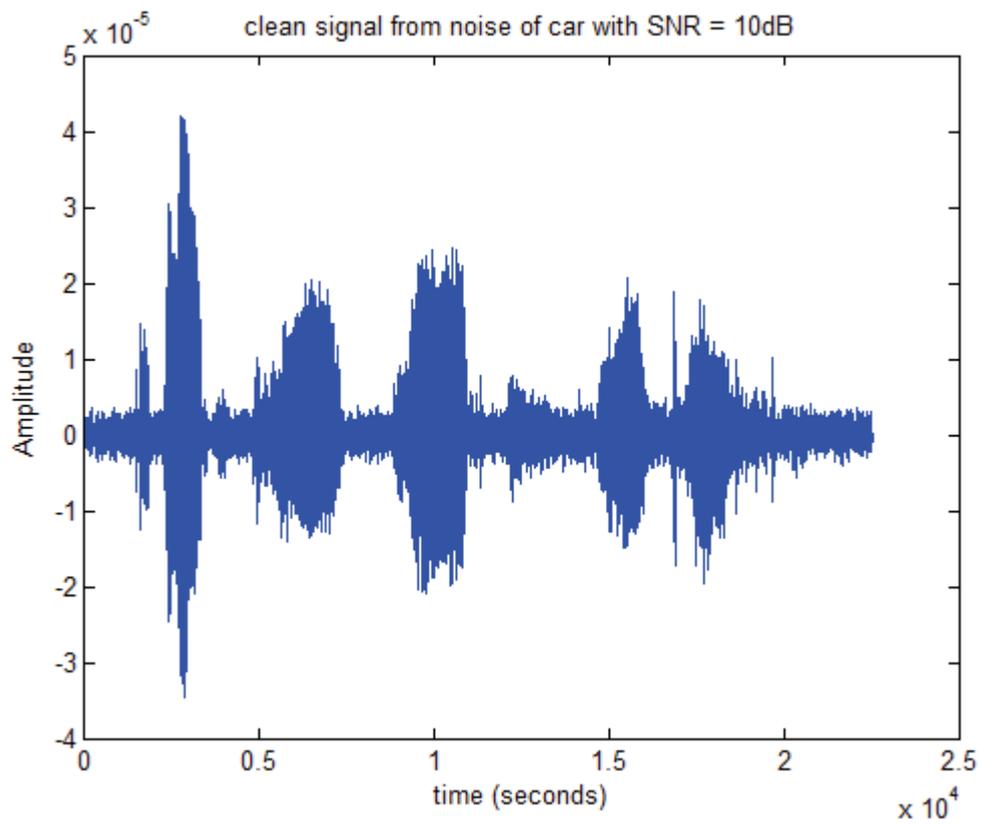
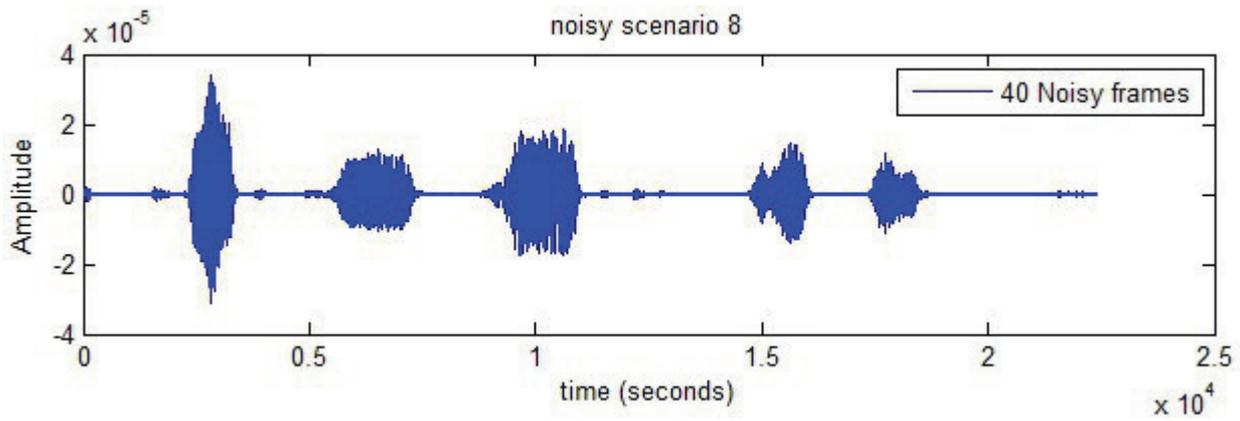


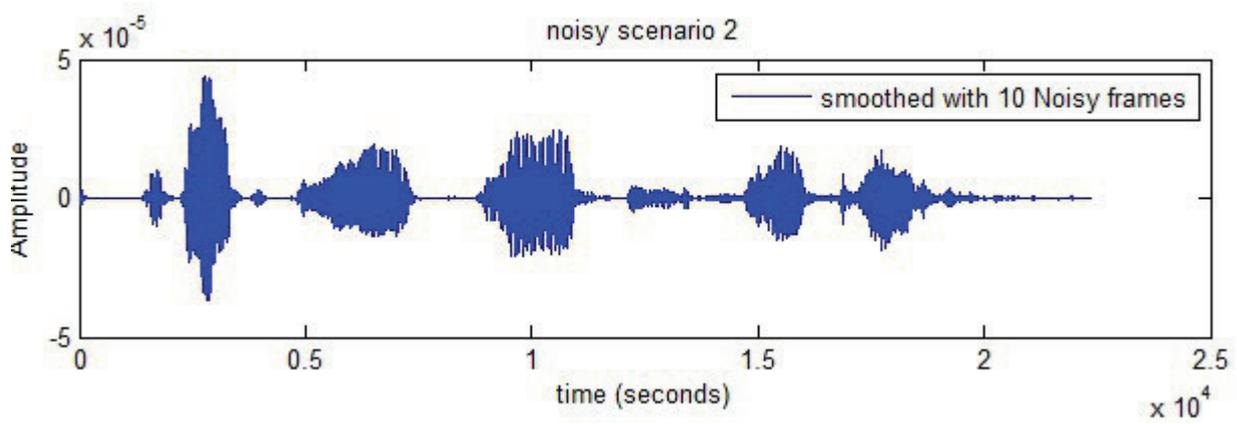
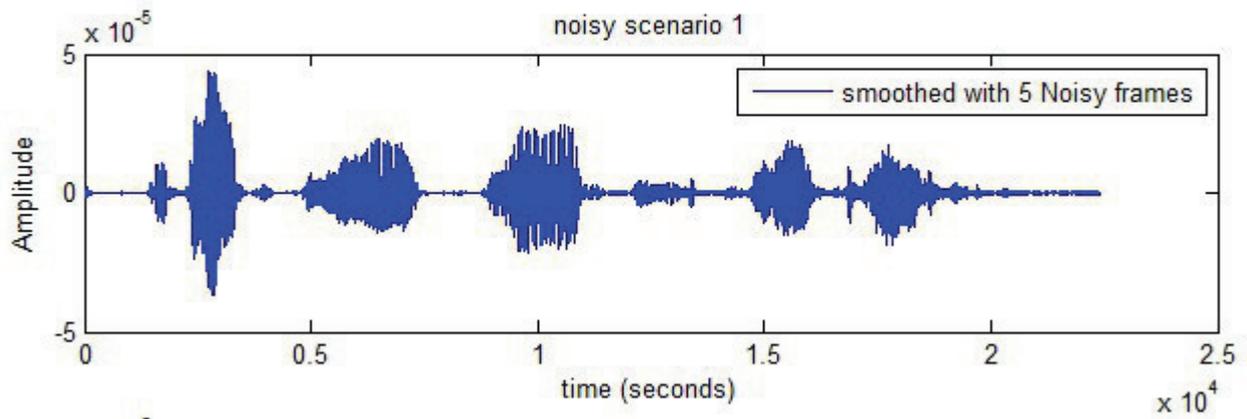


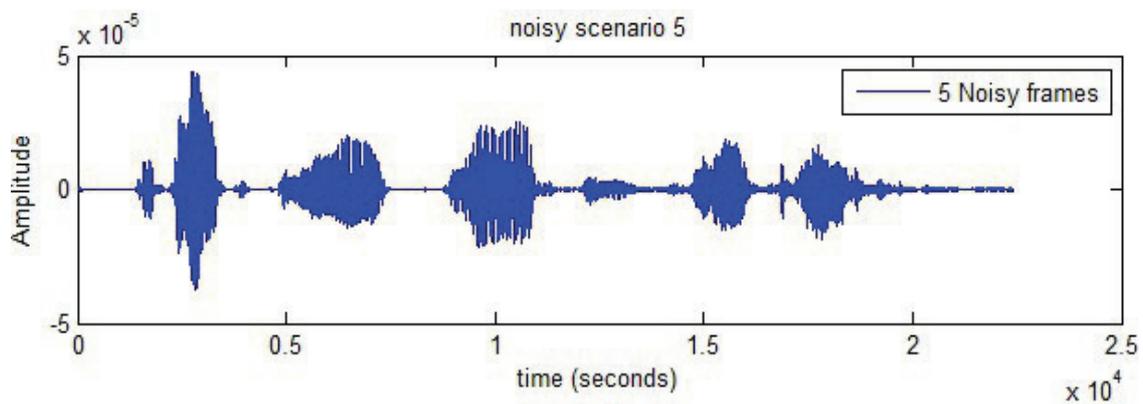
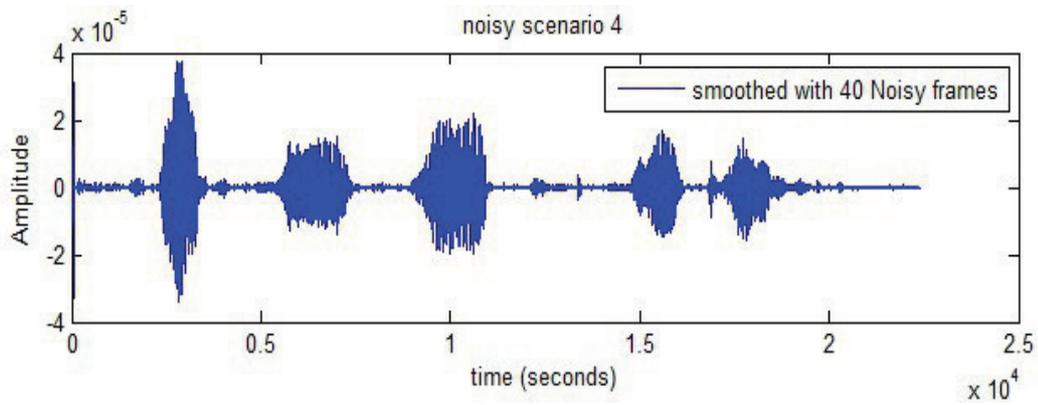
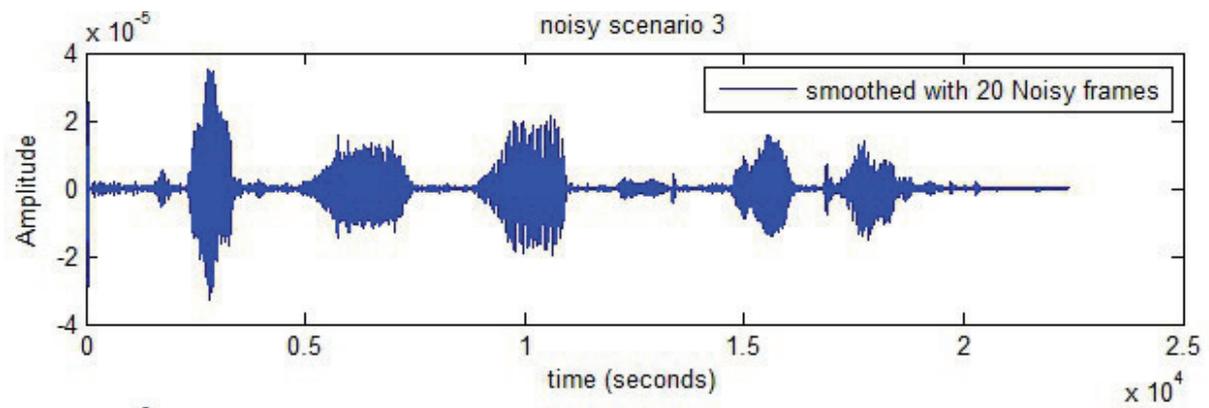


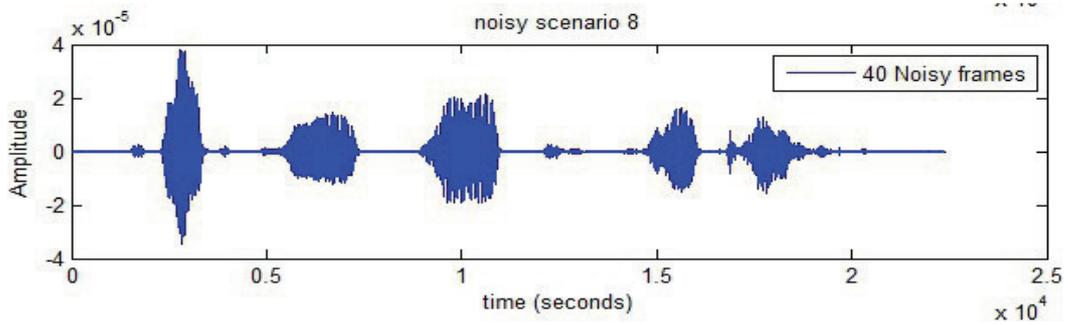
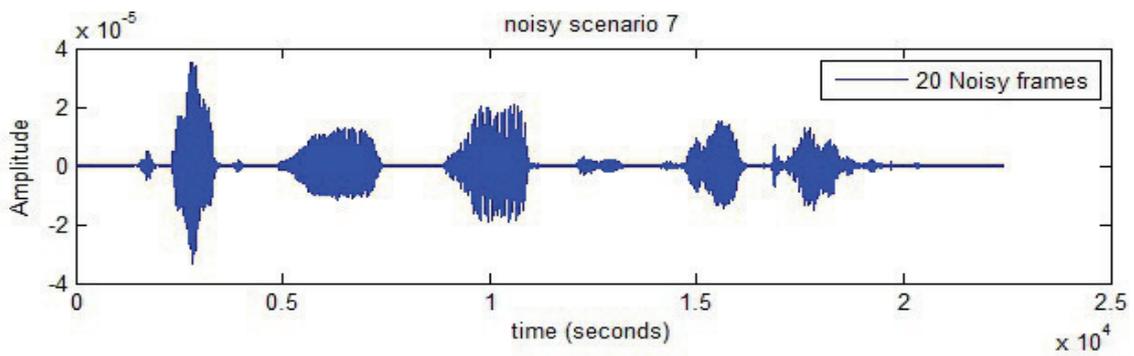
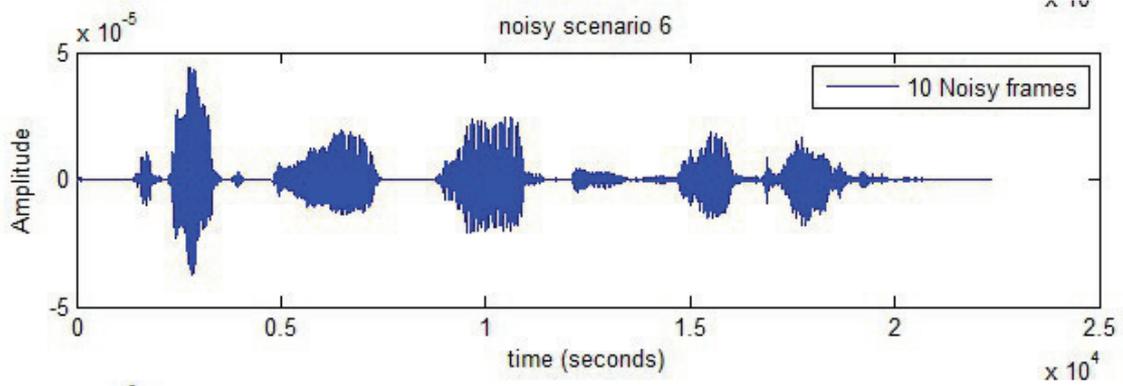


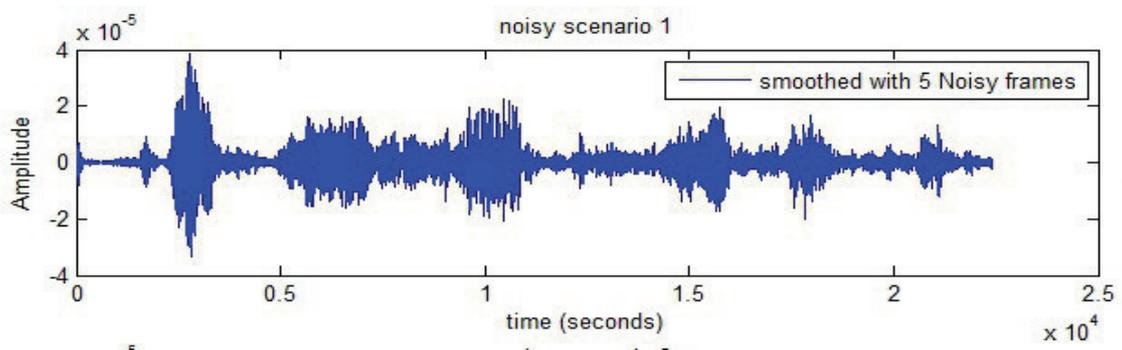
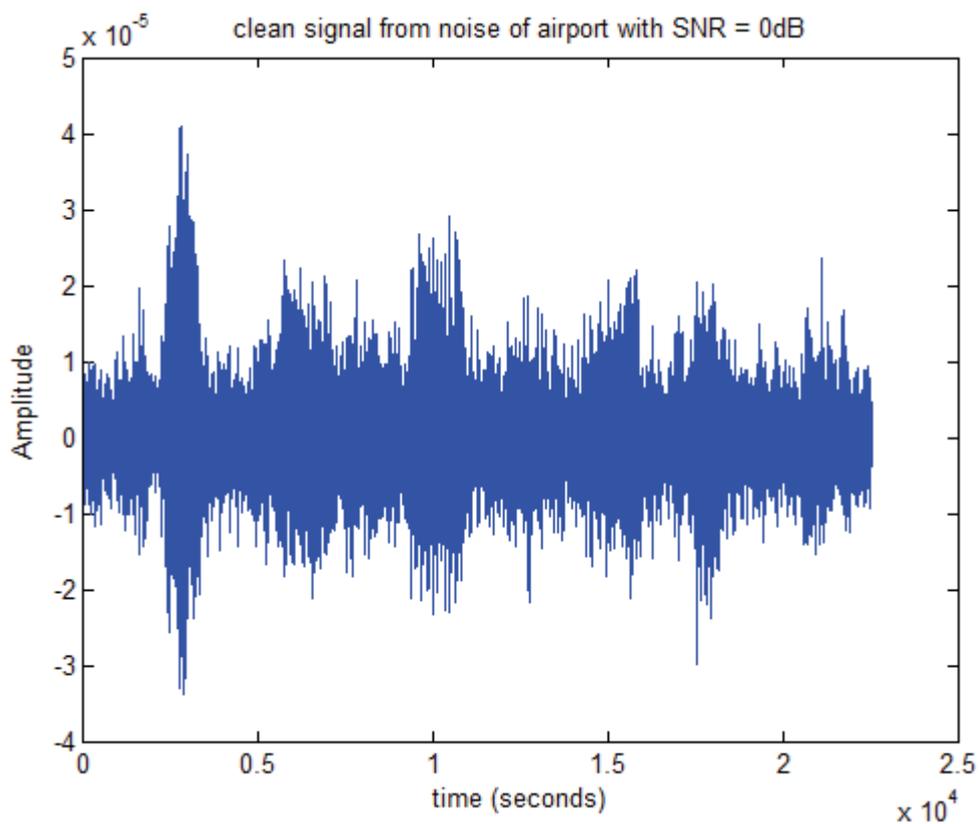


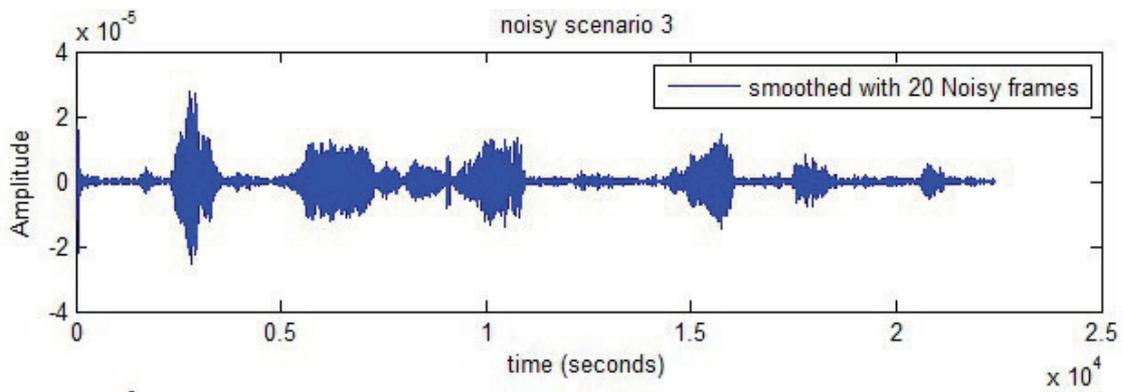
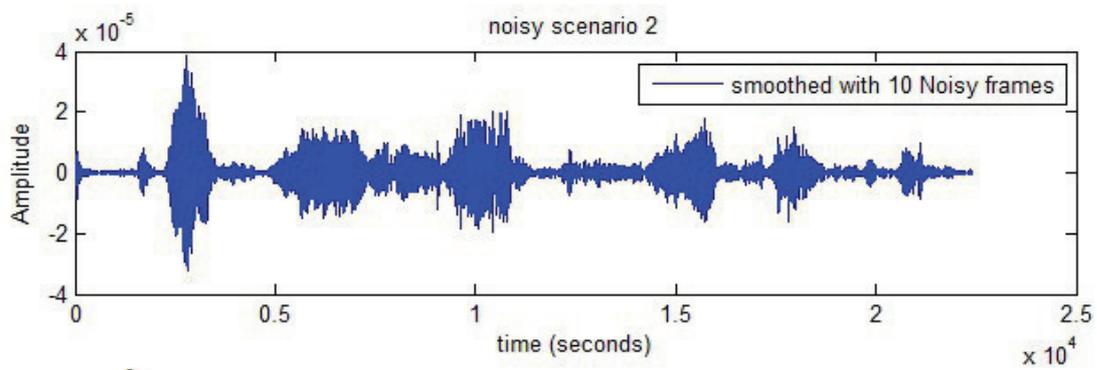


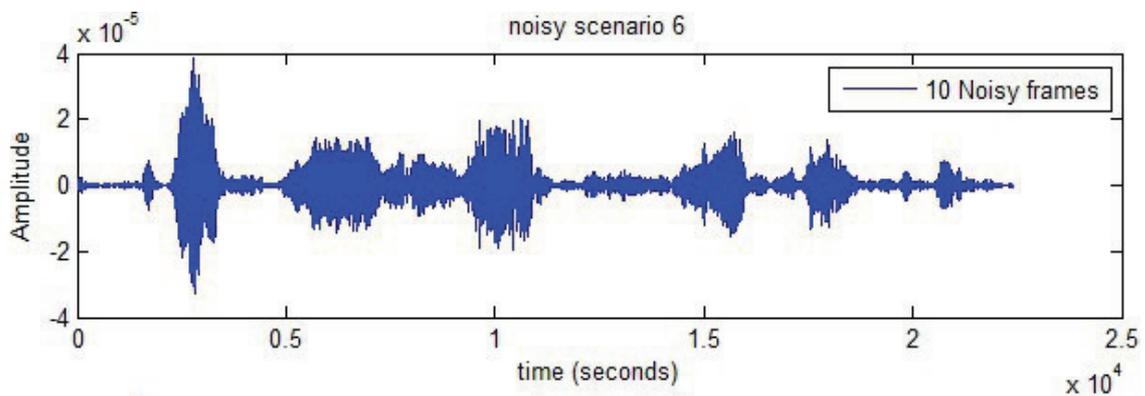
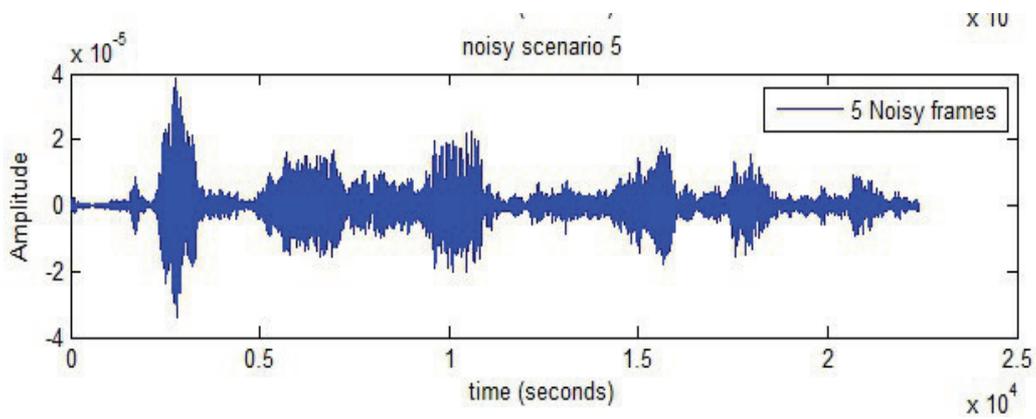
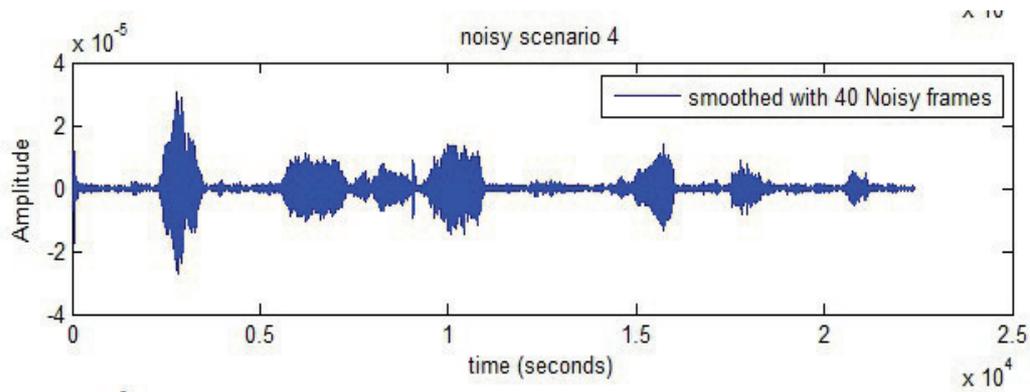


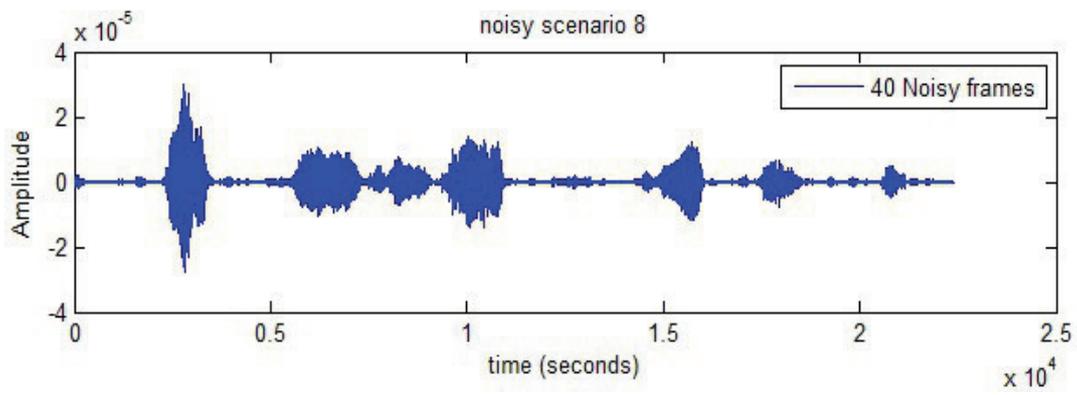
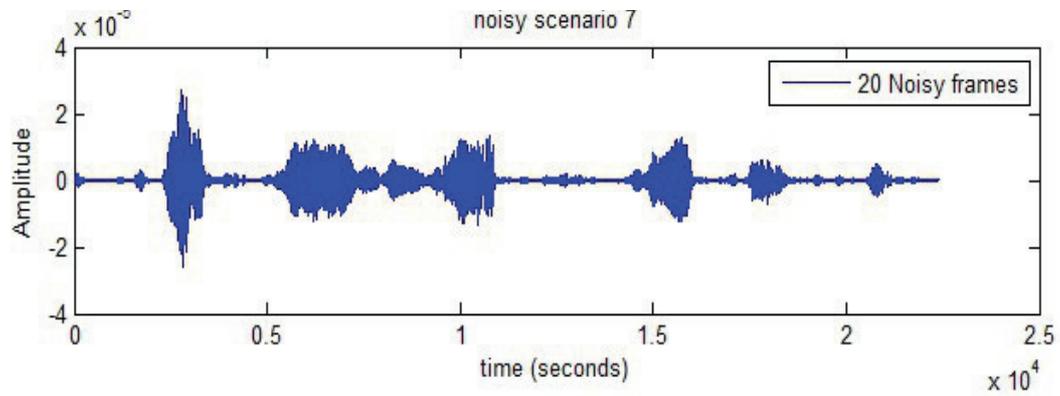


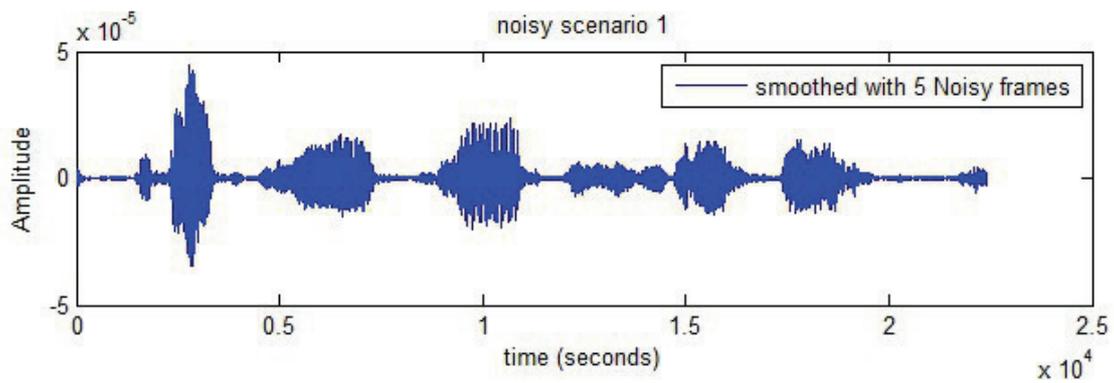
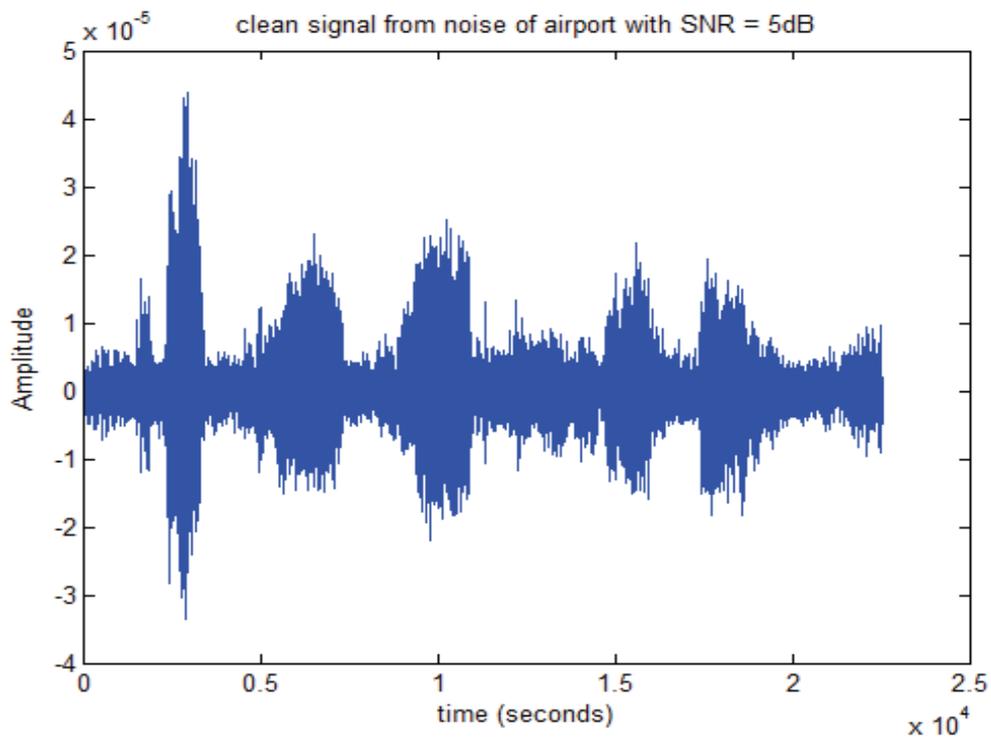


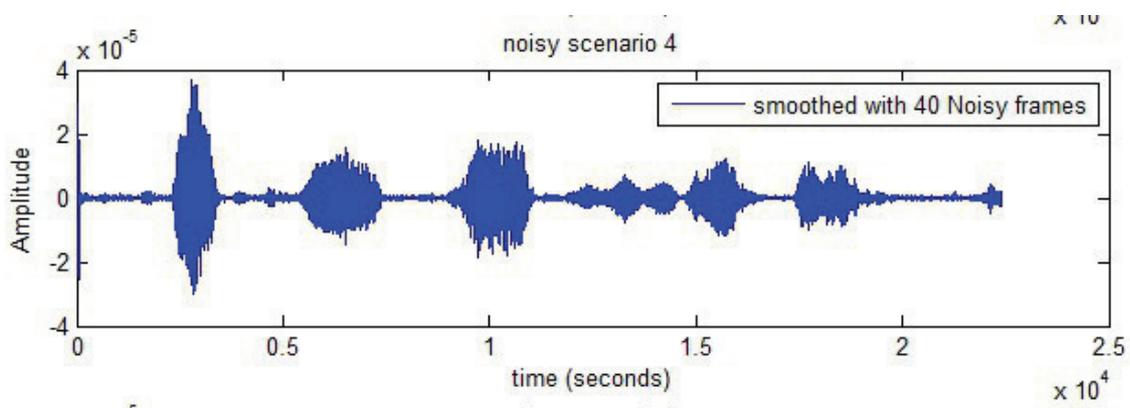
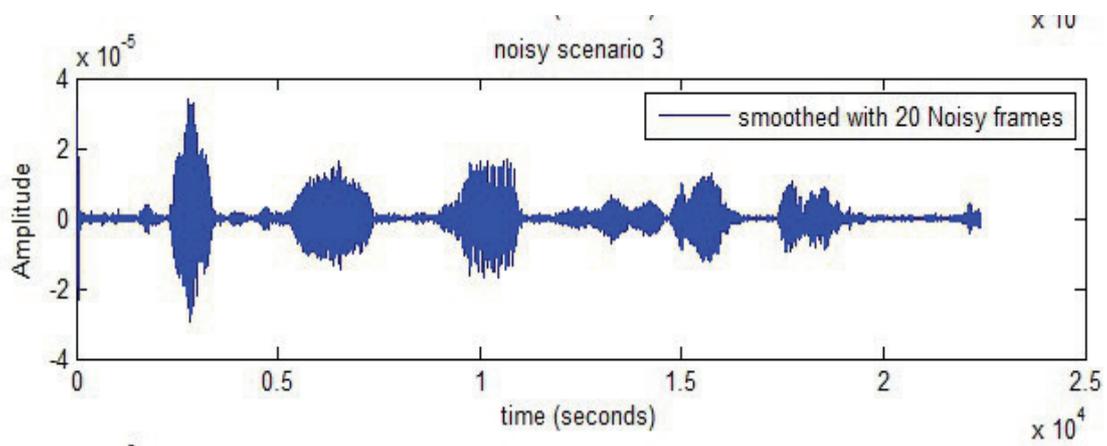
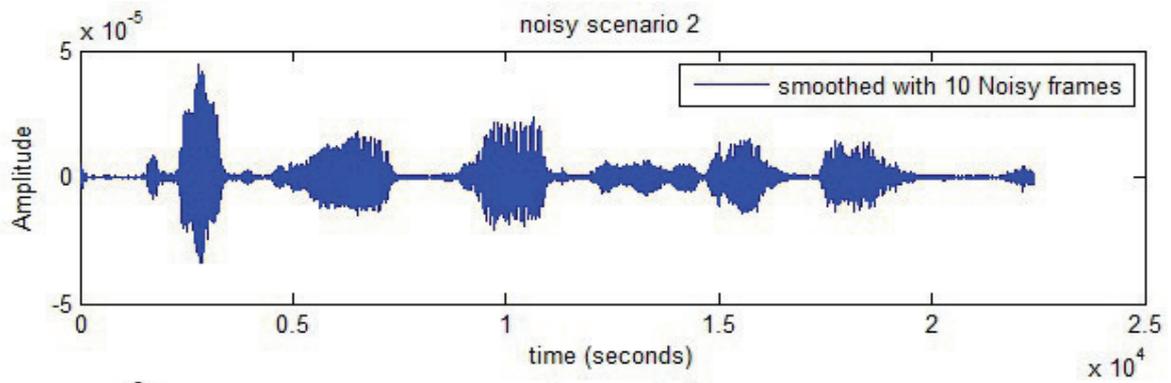


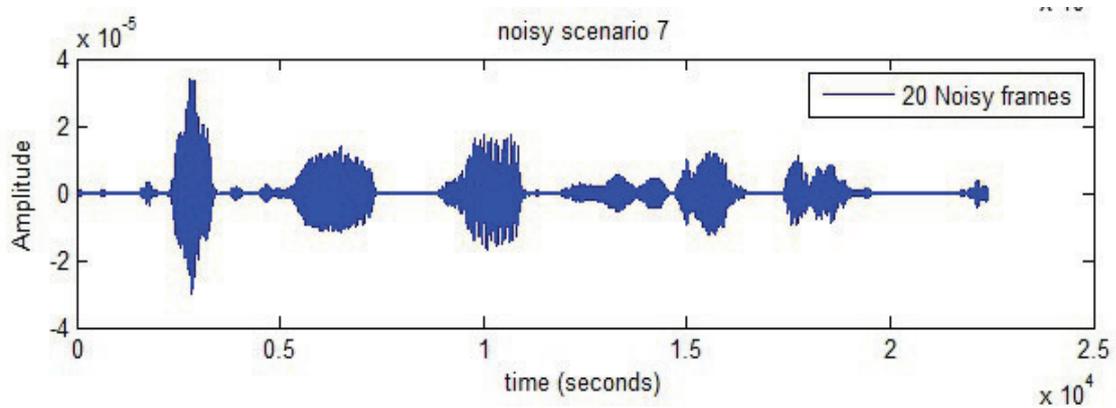
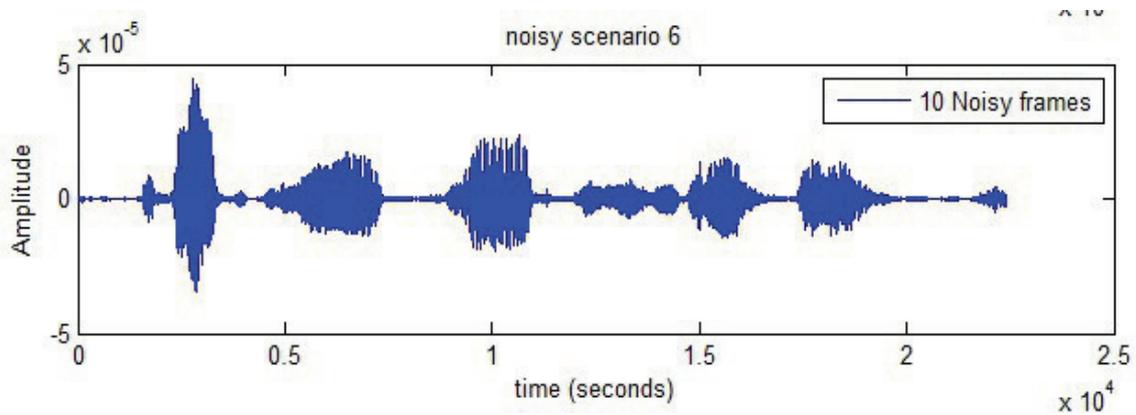
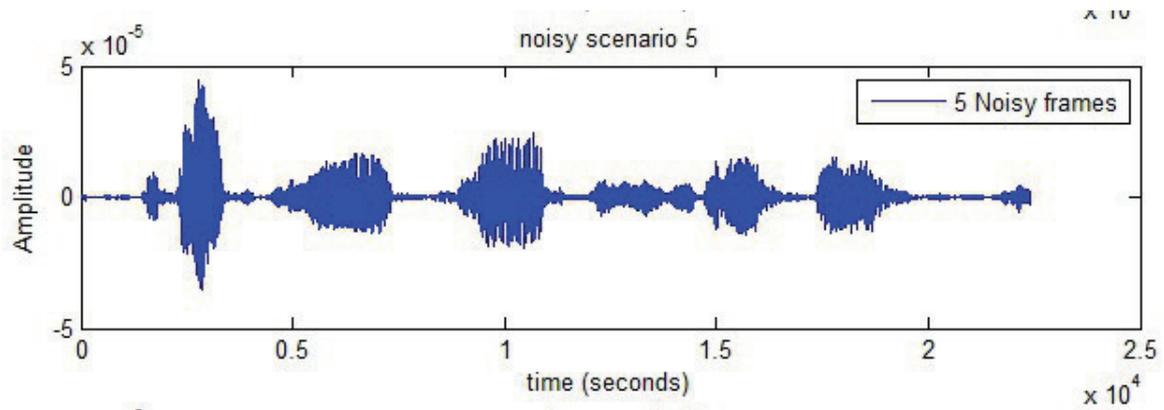


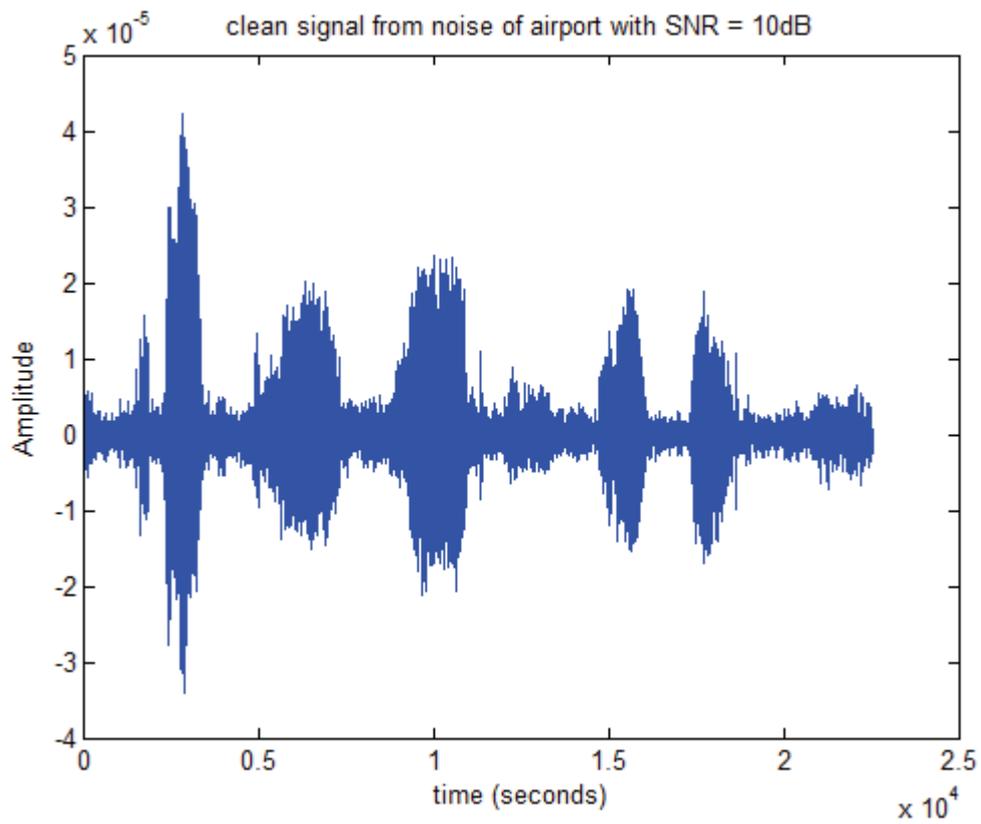
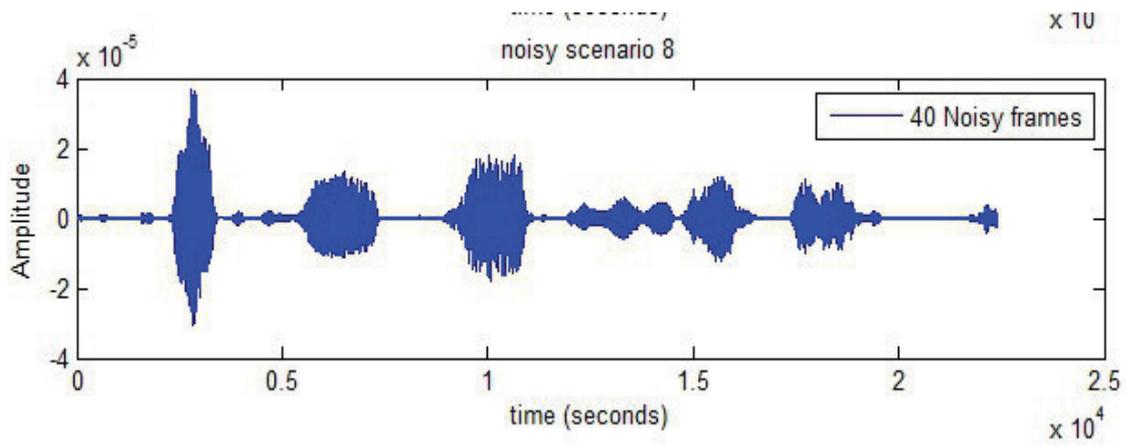


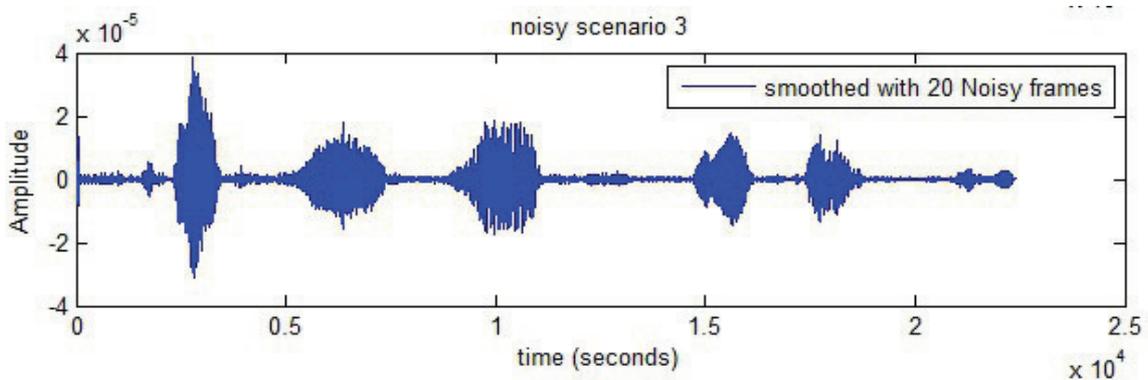
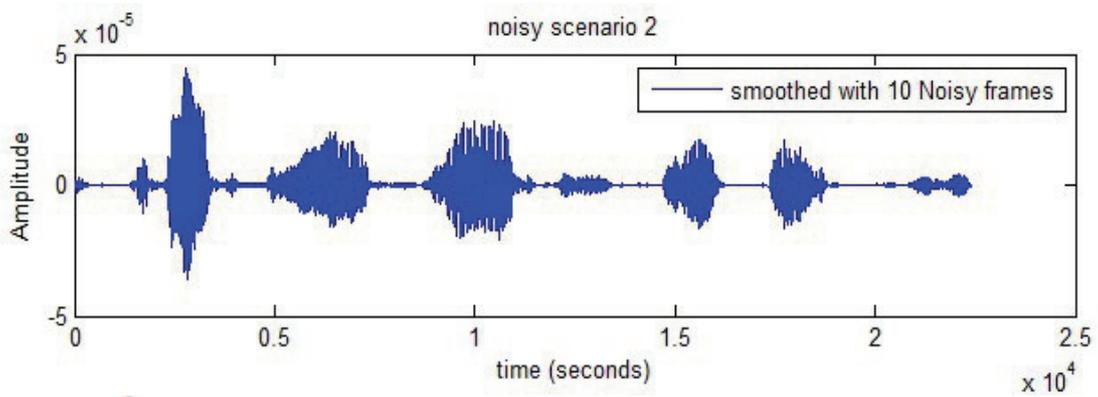
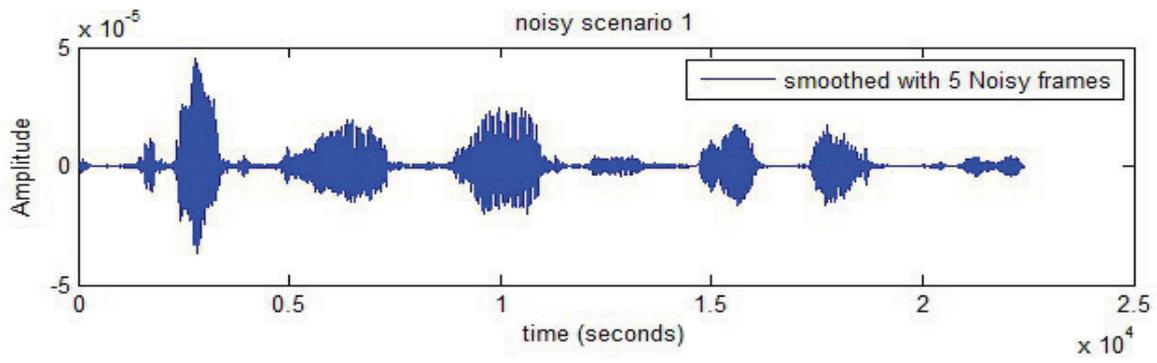


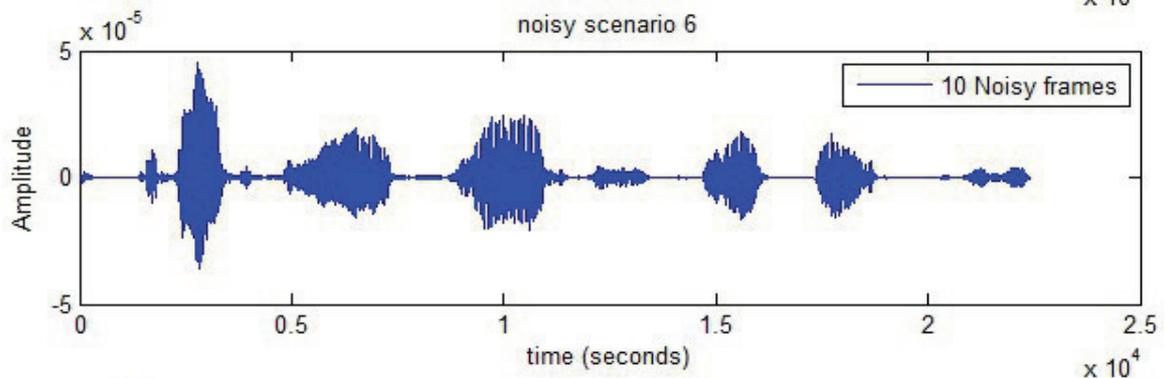
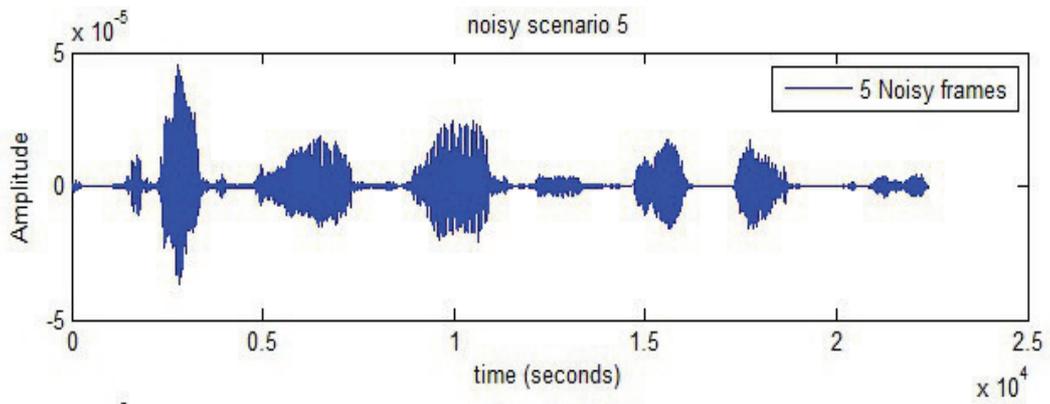
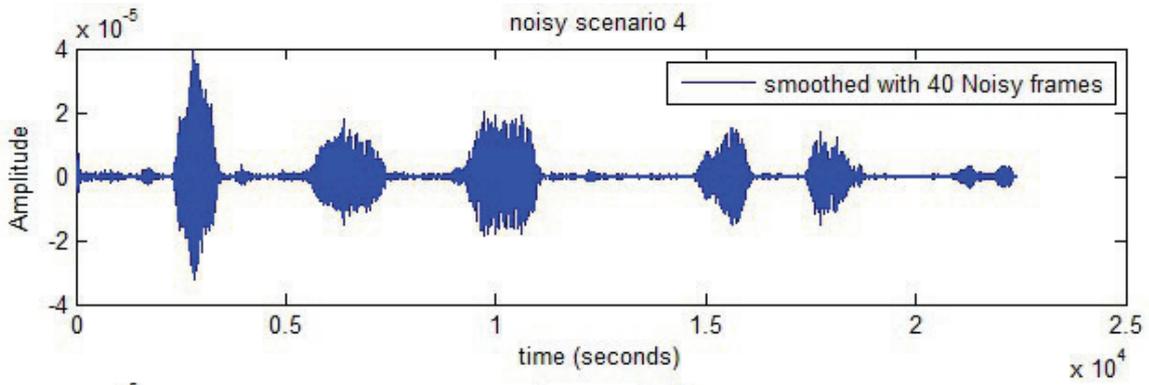


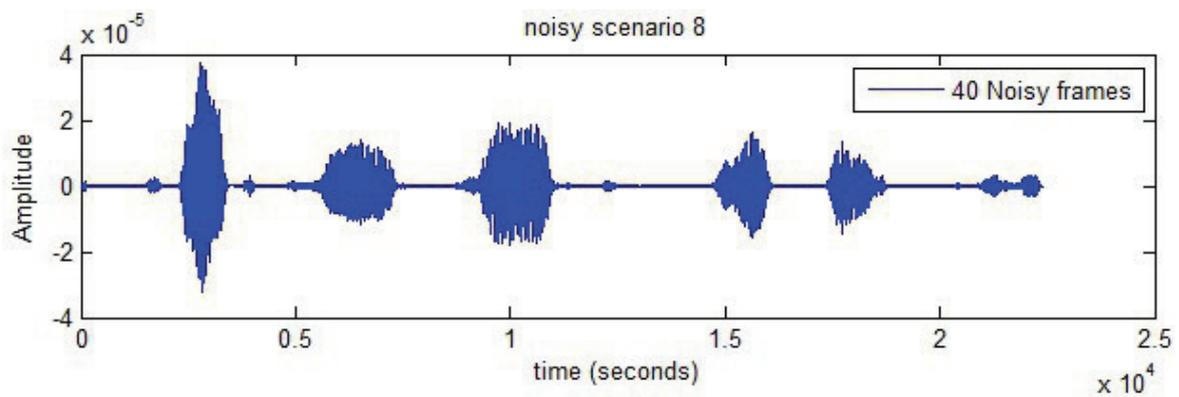
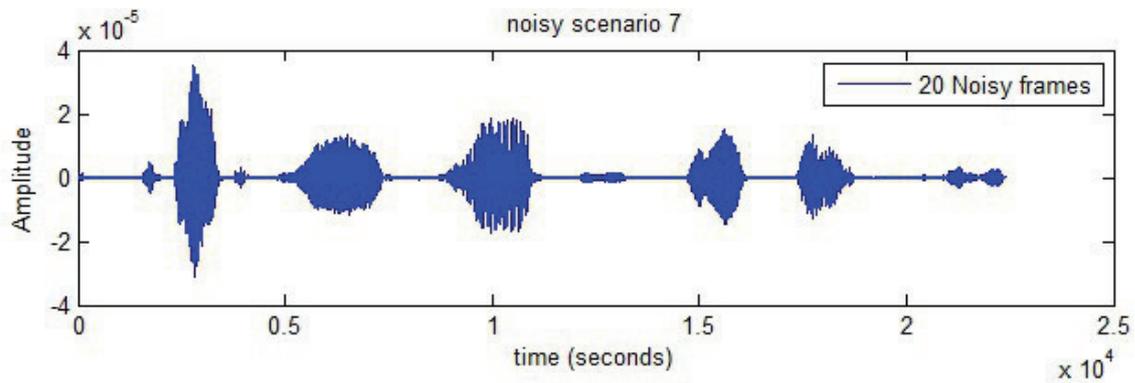












Παραπλήσια με αυτό που παρατηρήθηκε και στο προηγούμενο παράδειγμα, τα καλύτερα αποτελέσματα από πλευράς ποιότητας ήχου κατά την αποθρομβοποίηση αποκτήθηκαν για τις περιπτώσεις όπου είχαμε 5 και 10 σημεία για την εξομάλυνση του σήματος.

## Βιβλιογραφία

- [1] [http://nemertes.lis.upatras.gr/jspui/bitstream/10889/6001/3/Nimertis\\_Syros\(ele\).pdf](http://nemertes.lis.upatras.gr/jspui/bitstream/10889/6001/3/Nimertis_Syros(ele).pdf)
- [2] [http://users.ionio.gr/~floros/lessons/avarts\\_acoustics\\_LabNotes.pdf](http://users.ionio.gr/~floros/lessons/avarts_acoustics_LabNotes.pdf)
- [3] <http://users.ionio.gr/~floros/thesis/iSEE.pdf>
- [4] <http://repository.library.teimes.gr/xmlui/bitstream/handle/>
- [5] [http://nefeli.lib.teicrete.gr/browse/sefe/mta/2012/LeledakisIoannis/attached-document-1347358704-543252-30630/Leledakis\\_Ioannis\\_2012.pdf](http://nefeli.lib.teicrete.gr/browse/sefe/mta/2012/LeledakisIoannis/attached-document-1347358704-543252-30630/Leledakis_Ioannis_2012.pdf)
- [6] [http://artemis.cslab.ntua.gr/el\\_thesis/artemis.ntua.ece/DT20080158/DT2008-0158.pdf](http://artemis.cslab.ntua.gr/el_thesis/artemis.ntua.ece/DT20080158/DT2008-0158.pdf)
- [7] <http://thesis.ekt.gr/thesisBookReader/id/11050#page/26/mode/2up>
- [8] [http://nemertes.lis.upatras.gr/jspui/bitstream/10889/8310/3/Nimertis\\_PapadopoulosT\(ele\).pdf](http://nemertes.lis.upatras.gr/jspui/bitstream/10889/8310/3/Nimertis_PapadopoulosT(ele).pdf)
- [9] [http://artemis.cslab.ntua.gr/el\\_thesis/artemis.ntua.ece/DT2014-0170/DT2014-0170.pdf](http://artemis.cslab.ntua.gr/el_thesis/artemis.ntua.ece/DT2014-0170/DT2014-0170.pdf)
- [10] <http://nemertes.lis.upatras.gr/jspui/bitstream/10889/3833/1.pdf>
- [11] <http://docplayer.gr/7290267-Anaptyxi-efaruogis-epexergasias-ihoy-se-grafiko-perivallon-prograuuatisuoy.html>
- [12] ΤΣΟΥΚΑΛΑΣ-ΣΤΑΘΑΚΗΣ, ΔΙΟΝΥΣΙΟΣ (1997, *Πανεπιστήμιο Πατρών*), ΤΕΧΝΙΚΕΣ ΑΦΑΙΡΕΣΗΣ ΘΟΡΥΒΟΥ ΑΠΟ ΣΗΜΑΤΑ ΟΜΙΛΙΑΣ ΚΑΙ ΜΟΥΣΙΚΗΣ
- [13] “Perceptually Motivated Gain Filter Smoothing for Noise Suppression” 123<sup>ο</sup> Συνέδριο (5–8 Οκτώβρη 2007, Νέα Υόρκη) - “Audio Engineering Society”.
- [14] Audio Coding Standards, Chi-Min Liu, Wen-Whei Chang
- [15] Audio Signal Processing and Coding, Andreas Spanias, Ted Painter, Venkatraman Atti
- [16] Introduction to Digital Audio Coding and Standards, Marina Bosi, Richard E. Goldberg
- [17] Audio Coding. Theory and Applications, Y. You
- [18] Συμπύεση ηχητικών σημάτων: Από την ADPCM στο MPEG-4, Δρ. Μιχάλης Παρασκευάς, Επικ. Καθηγητής Γιάννης Μουρτζόπουλος
- [19] Ζαρούχας Θωμάς, «Μοντελοποίηση και Επεξεργασία Ηχητικών Δεδομένων για Αναπαραγωγή σε Χώρους με Αντήχηση», Διδακτορική Διατριβή, Πάτρα Ιούνιος 2010.
- [20] M. Berouti, R. Schwartz, και J. Makhoul, “Enhancement of speech corrupted by musical noise,”