ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΜΜΕ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΚΑΤΑΣΤΟΛΗ ΘΟΡΥΒΟΥ ΣΕ ΣΗΜΑ ΟΜΙΛΙΑΣ ΜΕ ΧΡΗΣΗ ΦΙΛΤΡΟΥ ΕΞΟΜΑΛΥΝΣΗΣ ΠΑΡΑΜΕΤΡΟΠΟΙΗΜΕΝΟ ΜΕ ΑΝΤΙΛΗΠΤΙΚΑ ΚΡΙΤΗΡΙΑ

ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΣΠΟΥΔΑΣΤΗ: ΣΕΤΣΙΚΑ ΑΓΓΕΛΙΚΗ
ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ: ΚΟΥΤΡΑΣ ΑΘΑΝΑΣΙΟΣ

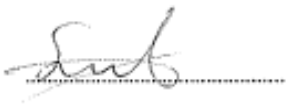
ΠΥΡΓΟΣ, ΑΥΓΟΥΣΤΟΣ 2016

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΜΗ ΛΟΓΟΚΛΟΠΗΣ

Βεβαιώνω/ουμε ότι είμαι/είμαστε ο/οι συγγραφέας/εις αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα/είχαμε για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία.

Επίσης, έχω/έχουμε αναφέρει τις οποίες πηγές από τις οποίες έκανα /κάναμε χρήση δεδομένων, ιδεών η λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες.

Ακόμη δηλώνω/ουμε ότι αυτή η γραπτή εργασία προετοιμάστηκε από εμένα/εμάς προσωπικά και αποκλειστικά και ειδικά για την συγκεκριμένη πτυχιακή εργασία ότι θα αναλάβω/ουμε πλήρως τις συνέπειες εάν η εργασία αυτή αποδειχτεί ότι δεν μου/μας ανήκει.

ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΣΠΟΥΔΑΣΤΗ 1	ΑΡΙΘ.ΜΗΤΡΩΟΥ	ΥΠΟΓΡΑΦΗ
ΑΓΓΕΛΙΚΗ ΣΕΤΣΙΚΑ	1489	
ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΣΠΟΥΔΑΣΤΗ 2	ΑΡΙΘ.ΜΗΤΡΩΟΥ	ΥΠΟΓΡΑΦΗ
.....	.....	.....
ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΣΠΟΥΔΑΣΤΗ 3	ΑΡΙΘ.ΜΗΤΡΩΟΥ	ΥΠΟΓΡΑΦΗ
.....	.....	.....

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΚΟΝΕΣ

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου για τη στήριξη, την υπομονή και την πολύτιμη συμβολή τους στην ολοκλήρωση αυτής της πτυχιακής εργασίας. Και τον κύριο Κούτρα για τη βοήθεια, καθοδήγηση και επίβλεψη κατά τη διάρκεια εκπόνησής της.

ΠΕΡΙΛΗΨΗ

Με την πάροδο των χρόνων οι αλγόριθμοι αποθορυβοποίησης εξελίσσονται και γίνονται ακόμα πιο αποτελεσματικοί. Η πτυχιακή εργασία παρουσιάζει κάποιες μεθόδους που βασίζονταν σε συγκεκριμένα στοιχεία του προς επεξεργασία σήματος και μελετά ένα νέο προτεινόμενο μοντέλο. Ξεκινώντας με τη μελέτη της βασικής δομής ενός σήματος, τον τρόπο που μετατρέπεται σε ψηφιακό, αλλά και πώς φιλτράρεται, η εργασία συνεχίζει με την κατανόηση της ανθρώπινης ομιλίας και του θορύβου. Παράλληλα αναλύεται το ψυχο-ακουστικό μοντέλο, το οποίο βασίζεται στο σύστημα ακοής του ανθρώπου. Αφού έχουν μελετηθεί οι παραπάνω παράμετροι, μπορεί να γίνει μελέτη και ανάλυση κάποιων αλγόριθμων αποθορυβοποίησης, καθώς και της Φασματικής Αφαίρεσης, πάνω στην οποία βασίζονται οι προς μελέτη αλγόριθμοι αλλά και η προτεινόμενη μέθοδος. Τέλος, πραγματοποιείται ένα πείραμα για το πόσο καλά ανταποκρίνονται οι αλγόριθμοι σε διαφορετικές συνθήκες θορύβου (ελάχιστο – μέγιστο SNR). Ακολουθεί μια υποκειμενική αξιολόγηση, βασισμένη στην κρίση ακροατών, για να διαπιστωθεί πώς γίνονται αντιληπτά σήματα που έχουν υποστεί αποθορυβοποίηση. Μετά το τέλος της διαδικασίας παρουσιάζονται αποτελέσματα και συμπεράσματα.

ABSTRACT

As the years passing by, noise suppression algorithms are evolving and becoming more effective. This thesis demonstrates some of these methods, which are based on specific characteristics of the signal, and also analysis a new suggested model. Starting from the base of a signal, how it is transformed to a digital signal and how it is filtered, the thesis continues with understanding the human's speech and noise characteristics. At the same time, a Psychoacoustic model is being analyzed, which is based on the human's auditory system. By having the above parametres studied, an analysis of noise suppression algorithms is performed, and alongside Spectral Subtraction (upon which are based the above algorithms and the proposed method). Finally, an experiment takes place on how the algorithms perform under different noise circumstances (minimum – maximum SNR values). A subjective assessment follows, based on human's perception, about how people perceive signals after noise reduction. Closing this thesis, are included results and conclusions about this experiment.

ΕΙΣΑΓΩΓΗ

Η χρήση της τεχνολογίας στη ζωή μας έχει πολλές εφαρμογές. Ένα χαρακτηριστικό κομμάτι είναι η χρήση ηχητικών σημάτων, είτε στην επικοινωνία ή σε άλλες εφαρμογές που κύριο ρόλο έχει η ανθρώπινη φωνή. Και επειδή σε πραγματικές συνθήκες τα ηχητικά αυτά σήματα προκύπτουν “φυσικά” και όχι με τεχνητούς τρόπους, είναι αναμενόμενο να εισέρχεται ανεπιθύμητος θόρυβος από το περιβάλλον.

Μέχρι σήμερα έχουν αναπτυχθεί αρκετοί αλγόριθμοι που στοχεύουν στην ενίσχυση της ομιλίας και τη συμπίεση του θορύβου. Μπορούν να ομαδοποιηθούν με βάση δύο διαφορετικά χαρακτηριστικά. Στο πρώτο εντάσσονται αυτοί που στοχεύουν στην βελτιστοποίηση της ομιλίας με χαμηλό *λόγο σήματος προς θόρυβο* (SNR), κυρίως για να διευκολυνθεί η επικοινωνία και η καταληπτότητα. Στο δεύτερο εντάσσονται αυτοί που στοχεύουν στην βελτιστοποίηση της ομιλίας με σχετικά υψηλό *λόγο σήματος προς θόρυβο* (SNR), κυρίως για ενίσχυση της ποιότητας αλλά και καταστολής του θορύβου.

Η παρούσα εργασία στοχεύει στην μελέτη και παρουσίαση ενός ακόμα αλγορίθμου αποθορυβοποίησης που έχει προταθεί για το πρόβλημα της εισαγωγής προσθετικού θορύβου σε ένα σήμα ομιλίας. Επίσης, παρουσιάζονται και συγκρίνονται ορισμένοι υπάρχοντες αλγόριθμοι, δίνοντάς μας την ευκαιρία να εκτιμήσουμε αν το αποτέλεσμα που προκύπτει μετά από την εφαρμογή τους είναι ικανοποιητικό ή όχι.

Οι αλγόριθμοι που παρουσιάζονται είναι εκείνοι που έχουν ως κύριο αντικείμενο μελέτης τη συμπίεση θορύβου ή την ενίσχυση της ομιλίας μέσω συμπίεσης θορύβου.

Πιο συγκεκριμένα, στο πρώτο κεφάλαιο παρουσιάζονται τα χαρακτηριστικά των σημάτων και πως λειτουργούν ως σύστημα. Πώς, από μια ηχογράφηση θα πάρουμε ένα ψηφιακό σήμα. Πώς επεξεργάζεται ένα σήμα ώστε να φιλτραριστεί.

Στο δεύτερο κεφάλαιο γίνεται μια προσπάθεια κατανόησης του τρόπου που παράγεται η ανθρώπινη ομιλία και ο θόρυβος. Πώς επηρεάζεται ένα σήμα ομιλίας και ποια χαρακτηριστικά είναι αυτά που μελετούνται στην επεξεργασία σήματος ομιλίας.

Στο τρίτο κεφάλαιο παρουσιάζονται αναλυτικά οι αλγόριθμοι αποθορυβοποίησης που θα χρησιμοποιηθούν αργότερα ως σύγκριση.

Στο τέταρτο κεφάλαιο αναλύεται η Φασματική Αφαίρεση. Ένας τρόπος μέσω του οποίου γίνεται η συμπίεση θορύβου. Επίσης, μελετάται και ένα νέο προτεινόμενο φίλτρο αποθορυβοποίησης.

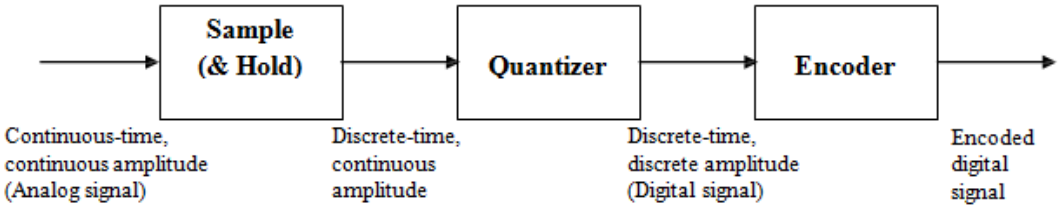
Στο πέμπτο κεφάλαιο εφαρμόζονται πρακτικά οι αλγόριθμοι σε σήματα ομιλίας. Γίνονται παρατηρήσεις και μια αξιολόγηση από ακροατές σχετικά με την αποτελεσματικότητά τους.

Στο έκτο και τελευταίο κεφάλαιο δίνονται κάποια συμπεράσματα μετά την εφαρμογή των αλγορίθμων.

1. Ψηφιακά Σήματα και Συστήματα

1.1 Δειγματοληψία – Κωδικοποίηση

Στο περιβάλλον, τα σήματα είναι αναλογικά, δηλαδή "συνεχή" και εκτείνονται στο χρόνο απεριόριστα. Στον τομέα της τεχνολογίας όμως, για να επεξεργαστούμε τέτοια σήματα θα πρέπει να έχουν μετατραπεί σε ψηφιακά. Η διαδικασία της μετατροπής ενός αναλογικού σήματος σε ψηφιακό αποτελείται από τρία βασικά μέρη: *δειγματοληψία (sampling), κβαντοποίηση (quantization), κωδικοποίηση (encoding)* (Εικόνα 1). Στη συνέχεια μελετάται αναλυτικά κάθε στάδιο με τα χαρακτηριστικά του



Εικόνα 1:Analog to Digital Convert ion

Δειγματοληψία

Αναφορικά με τη δειγματοληψία [1] μπορούμε να τη διακρίνουμε σε:

- Περιοδική: είναι η πιο συνήθης. Τα χρονικά δείγματα ισαπέχουν κατά την περίοδο Τ. Τέτοια δειγματοληψία κάνουν οι ADC.
- Τυχαία όταν γίνεται σε άτακτα χρονικά διαστήματα (κυρίως όταν η περίοδος των δειγμάτων είναι πολύ μεγάλη).
- Πολλαπλού ρυθμού.
- Πολλαπλής τάξης

Κατά τη διαδικασία της δειγματοληψίας πολλαπλασιάζουμε ένα σήμα συνεχούς χρόνου με μια περιοδική ακολουθία μοναδιαίων ώσεων. Η εξίσωση (1.1.1) περιγράφει τη γενική ιδέα υπό ιδανικές συνθήκες:

$$x_s(t) = x_a(t)s_a(t) = \sum_{n=-\infty}^{\infty} x_a(nT_s)\delta(t - nT_s)$$

όπου $x_a(t)$ το σήμα διακριτού χρόνου και $s_a(t)$ η περιοδική ακολουθία μοναδιαίων ώσεων. Επίσης, αν τα διαστήματα που μελετούνται είναι διαδοχικά τότε η διαδοχική απόσταση μεταξύ τους T_s είναι η περίοδος δειγματοληψίας και ορίζεται η συχνότητα δειγματοληψίας ως $f_s = {}^{1}/_{T_s}$. Από τα παραπάνω προκύπτει το διακριτό σήμα:

$$x(n) = x_a(nT_s) \qquad (1.1.2)$$

Θεώρημα Δειγματοληψίας –Nyquist

Στο θεώρημα δειγματοληψίας, που ανέπτυξε αρχικά ο Claude Shannon, [2] αναφέρεται ότι: *κάθε αναλογικό σήμα με περιορισμένο εύρος ζώνης* ***(band limited)*** *μπορεί να περιγραφεί ακριβώς από μια ακολουθία δειγμάτων της στάθμης του, αρκεί αυτά τα δείγματα να έχουν ληφθεί με συχνότητα τουλάχιστον διπλάσια μιας χαρακτηριστικής συχνότητας του αναλογικού σήματος που ονομάζεται* ***fc****.*
Η συχνότητα fc είναι η υψηλότερη συχνότητα (***fmax***) ενός αναλογικού σήματος περιορισμένου εύρους ζώνης.
Από την παραπάνω προϋπόθεση του θεωρήματος προκύπτει ότι το αναλογικό σήμα μπορεί να ανακτηθεί με ακρίβεια και χωρίς καμία παραμόρφωση. Έτσι, η συχνότητα δειγματοληψίας ***fs*** *ή Συχνότητα* ***Nyquist*** πρέπει να είναι τουλάχιστον ίση ή μεγαλύτερη από το διπλάσιο της ***fmax****.*

$$f_s \geq 2f_{max} \tag{1.1.3}$$

Κβαντισμός

Η ακολουθία συνεχούς πλάτους του σήματος εισόδου που προκύπτει από τη δειγματοληψία μετατρέπεται σε ακολουθία διακριτού πλάτους. Κατά τη διαδικασία, τα δείγματα του σήματος εισόδου αντιστοιχίζονται στην κοντινότερη διακριτή τιμή και προκύπτει από την σχέση:

$$\hat{x}(n) = Q[x(n)] \tag{1.1.4}$$

Ειδικότερα σύμφωνα με [3], μετά τη δειγματοληψία προκύπτει μια συνεχής ακολουθία παλμών, οι οποίοι χωρίζονται κατά ένα ποσό του χρόνου δειγματοληψίας Ts. Ένα σετ παλμών μπορεί να οριστεί στο εύρος του χρόνου από έναν πεπερασμένο αριθμό πιθανών στοιχείων, κάτι το οποίο δεν μπορεί να γίνει στο εύρος του πλάτους. Ας υποθέσουμε ότι το πλάτος κυμαίνεται μεταξύ των τιμών A_{max} και A_{min}. Αν ένα σήμα έχει τιμές φορτίου από -5 έως +5 Volt, για μια χρονική στιγμή δειγματοληψίας το πλάτος μπορεί να είναι 3.0002V. Σε τέτοια περίπτωση το σήμα χαρακτηρίζεται ως *σήμα συνεχούς τιμής διακριτού χρόνου* ***(DTCV)****.* Η διαδικασία με την οποία το σετ πεπερασμένων τιμών πλάτους μπορεί να οριστεί ακριβώς ονομάζεται *Κβαντισμός*.
Ας υποθέσουμε ότι ένα σετ ορίζεται με τιμές Volt 1, 2, 3, 4, 5 με τα αντίστοιχα αρνητικά τμήματα. Κάθε τιμή πλάτους μετά τη δειγματοληψία θα υπολογιστεί προσεγγιστικά στην πιο κοντινή τιμή ορισμού. Στο γράφημα (x.x.x) φαίνεται η παραπάνω διαδικασία. Οι μέγιστες τιμές πλάτους ορίζονται ως A_{max} και A_{min}. Όλο το εύρος (A_{max} και A_{min}) χωρίζεται ισότροπα σε τέσσερεις τομείς μήκους S οι οποίοι ορίζονται $\Delta_0 - \Delta_3$. Η μέση κάθε Δ-τομέα είναι το πιθανό επίπεδο πλάτους q.
Σύμφωνα με τον αλγόριθμο κβαντισμού «αν το πλάτος του σήματος δειγματοληψίας είναι εντός ενός εύρους Δ_i, το πλάτος θα κβαντιστεί στο πλάτος q_i».

Κωδικοποίηση

Στη συνέχεια, για κάθε επίπεδο q γίνεται κωδικοποίηση 2-bit ώστε να πάρει την κατάλληλη ψηφιακή μορφή του αντίστοιχου αναλογικού σήματος.
Στο στάδιο της κωδικοποίησης το αναλογικό σήμα μετατρέπεται σε μια ακολουθία αριθμών (bits) όπου κάθε μοναδικός δυαδικός αριθμός αντιστοιχεί σε ένα επίπεδο κβαντισμού.
Υπάρχουν πολλοί τρόποι κωδικοποίησης [4], με πιο βασικές τις PCM, DPCM, LPC.

1.2 Φίλτρα FIR

Είναι φίλτρα *πεπερασμένης χρονικής διάρκειας απόκρισης*. Σε πρακτικές εφαρμογές συνήθως προτιμώνται λόγω της απλότητας κατά την υλοποίηση τους (μπορούν να σχεδιαστούν και χωρίς μεγάλο κόστος για συστήματα με χαμηλές απαιτήσεις). Ταυτόχρονα, πλεονεκτούν έναντι των IIR λόγω της ευστάθειας που παρουσιάζουν (ακόμα και μετά από τον κβαντισμό των συντελεστών του φίλτρου) και της γραμμικής φάσης.

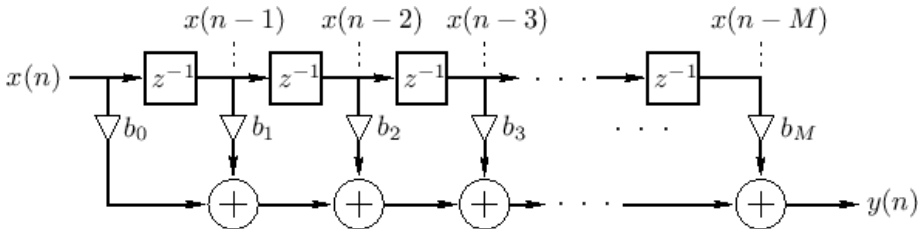
Ένα αιτιατό φίλτρο FIR για μια είσοδο x(n) θα δώσει μια έξοδο

$$y(n) = \sum_{k=0}^{N} h(k)x(n-k) \tag{1.2.1}$$

και έχοντας μια συνάρτηση μεταφοράς H(z) η οποία είναι πολυώνυμο του z^{-1}.

$$H(z) = \sum_{n=0}^{N} h(n)z^{-n} \tag{1.2.2}$$

Στην Εικόνα 2 φαίνεται η γενική δομή των FIR φίλτρων με μια γραμμή καθυστέρησης.



Εικόνα 2: Γενική δομή FIR φίλτρων

Στη συνέχεια γίνεται αναφορά στους τρόπους υλοποίησης αυτού του συστήματος.

Δομή

Στη βιβλιογραφία *[5]*, για τη δομή αυτών των φίλτρων μπορούμε να ξεχωρίσουμε τρεις βασικούς τύπους:

- Ευθεία μορφή
- Μορφή καταρράκτη
- Γραμμικής φάσης

Ευθεία μορφή

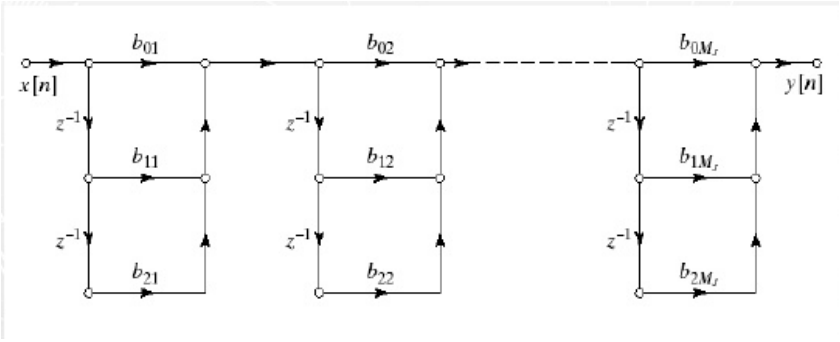
Αυτός ο τύπος χρησιμοποιείται αρκετά συχνά και ουσιαστικά είναι η γενική μορφή που αναφέρεται στο σχήμα (εικόνα 2). Για ένα φίλτρο τάξης N, έχουμε N+1 συντελεστές, N+1 πολλαπλασιασμούς, N προσθέσεις και μια *διακλαδιζόμενη γραμμή καθυστέρησης*.

Μορφή καταρράκτη

Η συνάρτηση μεταφοράς H(z) [6] ενός high-pass FIR φίλτρου μπορεί να γραφεί ως γινόμενο πρώτου ή δευτέρου βαθμού:

$$H(z) = A\prod_{k=1}^{Ns}[1 + b_k(1)z^{-1} + b_k(2)z^{-2}] \tag{1.2.3}$$

Η παραπάνω υλοποίησή της H(z) είναι μια σύνδεση σε σειρά φίλτρων $2^{ης}$ τάξης (καταρράκτης).



Εικόνα 3:Φίλτρο FIR 2ης τάξης σε σύνδεση καταρράκτη

Γραμμικής φάσης

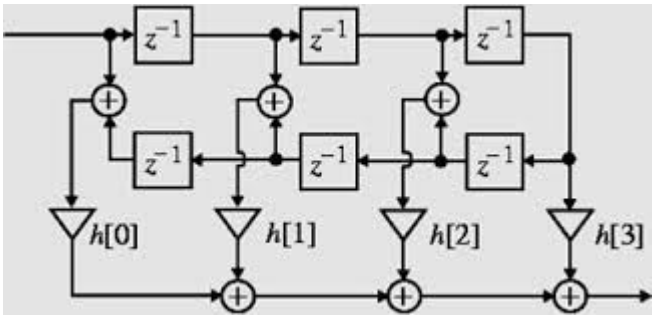
Για τα φίλτρα αυτού του τύπου, η κρουστική απόκριση μπορεί να είναι είτε συμμετρική

$$h(n) = h(N-n)$$

είτε αντισυμμετρική

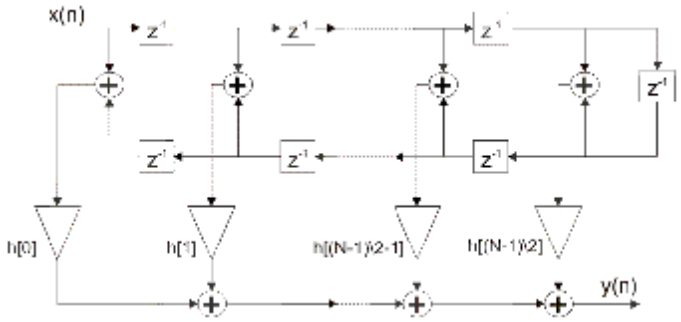
$$h(n) = -h(N-n)$$

Στην περίπτωση που η h(n) είναι συμμετρική μπορούμε να απλοποιήσουμε τη δομή. Ένας τύπος φίλτρου, για παράδειγμα, είναι αν το N είναι άρτιο και η h(n) συμμετρική έχουμε το σχήμα της Εικόνας 4



Εικόνα 4:Linear phase τύπου I

Ενώ αν η h(n) είναι συμμετρική και το N περιττό, έχω το σχήμα της Εικόνας 5



Εικόνα 5:Linear phase τύπου II

Και στις δύο περιπτώσεις μπορούμε να δούμε ότι χρειαζόμαστε μόνο 4 πολλαπλασιαστές ενώ σε μια ευθεία μορφή θα ήθελα 7 για τον τύπο (I) και 8 για τον τύπο (II).

Σχεδίαση

Για τη σχεδίαση των FIR ισχύει ότι σχεδιάζονται ώστε να έχουν γραμμική φάση. Έτσι, σε αυτή την ενότητα έχουμε σαν δεδομένο ότι τα φίλτρα FIR είναι γραμμικής φάσης.

Η διαδικασία της σχεδίασης περιλαμβάνει κάποια βασικά στάδια ώστε να είναι ολοκληρωμένη και αυτά είναι: καθορισμός των προδιαγραφών του φίλτρου, υπολογισμός των συντελεστών του φίλτρου, καθορισμός δομής υλοποίησης, ανάλυση σφαλμάτων. Ο υπολογισμός των συντελεστών h(n) είναι απαραίτητος ώστε το φίλτρο να πληροί τις προδιαγραφές του μέτρου και της φάσης στη συχνότητα. Όταν καθοριστούν αυτές οι προδιαγραφές τότε μπορεί να σχεδιαστεί το φίλτρο.

Η απόκριση συχνότητας ενός φίλτρου FIR είναι

$$H(e^{j\omega}) = \sum_{n=0}^{N} h(n) e^{-jn\omega} \qquad (1.2.4)$$

Και εδώ διακρίνουμε τρεις βασικούς τρόπους σχεδίασης:

- με χρήση παραθύρων
- με δειγματοληψία συχνότητας
- βέλτιστο φίλτρο

Σχεδίαση με χρήση παραθύρων

Σχεδιάζοντας FIR φίλτρα με τη μέθοδο των παραθύρων είναι σχετικά εύκολη διαδικασία. Έστω $h_D(n)$ η κρουστική απόκριση για ένα ιδανικό φίλτρο. Οι συντελεστές της $h_D(n)$ υπολογίζονται μέσω του αντίστροφου μετασχηματισμού Fourier της απόκρισης συχνότητας $H_D(e^{j\omega})$.

$$h_D(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_D(e^{j\omega}) e^{j\omega n} d\omega \qquad (1.2.5)$$

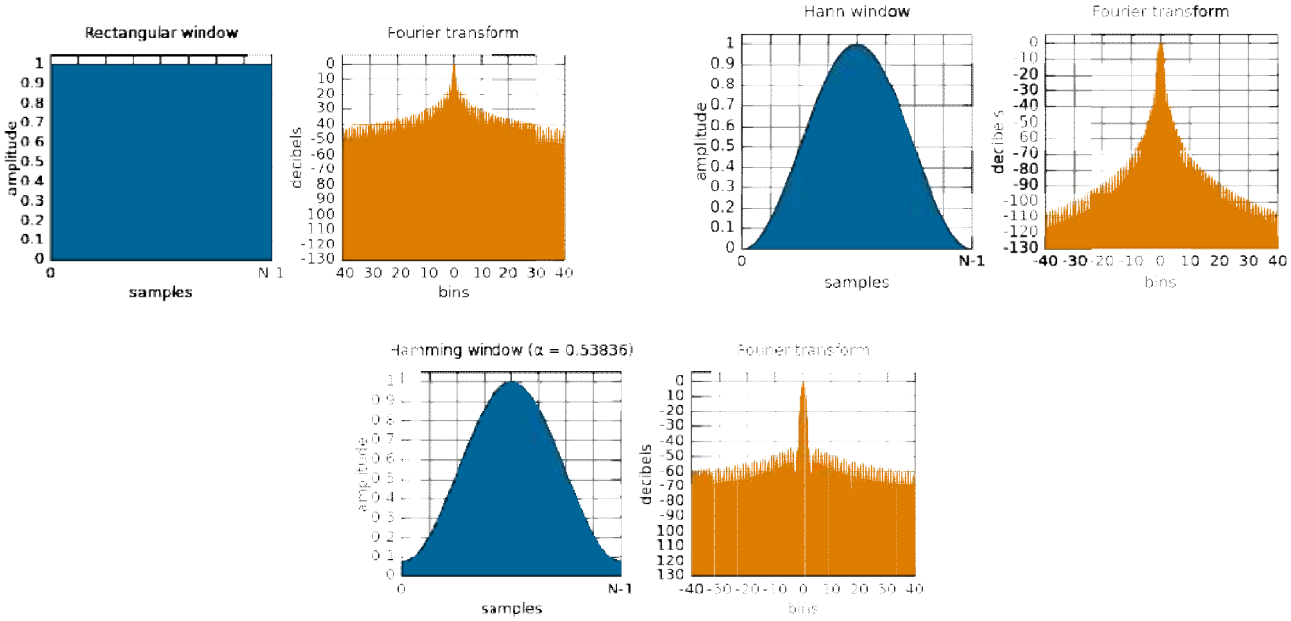
Όμως η $h_D(n)$ έχει άπειρο μήκος και κατά τη διαδικασία αποκοπής θα έχουμε ως αποτέλεσμα την αλλοίωση της ιδανικής απόκρισης στη συχνότητα (φαινόμενο Gibbs). Έτσι, προκειμένου να αποφύγουμε αυτού του είδους το πρόβλημα και ταυτόχρονα να είναι η κρουστική απόκριση $h_D(n)$ πεπερασμένου μήκους και συμμετρική, χρησιμοποιούμε μια συνάρτηση «παραθύρου».

$$h(n) = h_D(n)w(n) \tag{1.2.6}$$

όπου w(n) η συνάρτηση παραθύρου πεπερασμένου μήκους, η οποία για τιμές έξω από το διάστημα $0 \leq n \leq N$ είναι μηδενική. Έτσι, προκύπτει η τελική απόκριση συχνότητας:

$$H(e^{j\omega}) = \frac{1}{2\pi} H_D(e^{j\omega}) * W(e^{j\omega}) \tag{1.2.7}$$

Μπορούμε να χρησιμοποιήσουμε πολλούς τύπους συναρτήσεων παραθύρου. Οι πιο γνωστοί είναι: ορθογώνιο, Hann, Hamming.



Είναι αυτονόητο ότι αφού η διεργασία ολοκληρωθεί σε ένα block, τότε, το παράθυρο ανάλυσης μεταφέρεται στο αμέσως επόμενο block δεδομένων.
Επιπλέον, στη βιβλιογραφία, ορίζεται το *μέγεθος ολίσθησης παραθύρου* (window slide). Είναι η απόσταση μεταξύ των κέντρων δύο συνεχόμενων παραθύρων. *Αν το μέγεθος αυτό είναι μικρότερο από το μέγεθος του παραθύρου, τότε θα υπάρξει κάποιο ποσοστό επικάλυψης, που σημαίνει ότι κάποια δείγματα θα χρησιμοποιηθούν παραπάνω από μια φορά στη διαδικασία ανάλυσης.*
Για την $W(e^{j\omega})$ [6] ισχύει ότι, σε ιδανική περίπτωση το εύρος του κύριου λοβού θα πρέπει να είναι όσο δυνατό πιο στενό και το πλάτος πλευρικών λοβών όσο δυνατό πιο περιορισμένο.
Κάποιες ιδιότητες [6] των συναρτήσεων παραθύρου είναι α) αύξηση του μήκους N της συνάρτησης οδηγεί σε μείωση του εύρους του κύριου λοβού β) το πλάτος των πλευρικών λοβών δεν εξαρτάται από το μήκος της συνάρτησης παραθύρου γ) αν θέλουμε σκόπιμα να μειώσουμε το πλάτος των πλευρικών λοβών, τότε θα αυξηθεί το εύρος του κύριου λοβού.

1.3 Μετασχηματισμός στο πεδίο της συχνότητας

1.3.1 Μετασχηματισμός Fourier

Όπως είναι κοινώς αποδεκτό, για την επεξεργασία ενός σήματος από ένα υπολογιστικό μέσο, θα πρέπει το σήμα να είναι διακριτό και μεταφερμένο στο πεδίο της συχνότητας. Η μεταφορά

στο πεδίο της συχνότητας γίνεται με τον Μετασχηματισμό Fourier. Και για σήματα διακριτού χρόνου, αντίστοιχα, Διακριτός Μετασχηματισμός Fourie.
Το πρόβλημα που προκύπτει εδώ είναι το μεγάλο πλήθος των πράξεων, και κατά συνέπεια και του χρόνου, που απαιτείται για τον υπολογισμό του Μετασχηματισμού για όλες τις τιμές του σήματος. Πιο συγκεκριμένα, για ένα παράθυρο N δειγμάτων, ο ΔMF απαιτεί N2 πολλαπλασιασμούς και N(N-1) επιπλέον προσθέσεις.

1.3.2 Fast-Fourier-Transform (FFT)

Τη λύση στο πρόβλημα του μεγάλου πλήθους πράξεων που απαιτεί ο DFT έρχεται να δώσει ο Fast-Fourier-Transform (FFT). Μετά από έρευνες χρόνων πάνω σε αυτό τον τομέα, έχουν βρεθεί αλγόριθμοι οι οποίοι βασίζονται στον υπολογισμό πολλών DFT (με μήκος μικρότερο από N σημεία) και ο συνδυασμός τους τελικά παράγει τον DFT N σημείων. Η κύρια θεωρία ορίζει το N ως μια ακέραια δύναμη 2 με σκοπό η πολυπλοκότητα να γίνει ανάλογη του:

$$Nlog_2N \qquad (1.3.2.1)$$

Ωστόσο, κάτι τέτοιο δεν ισχύει αν λάβουμε υπόψη άλλες παραλλαγές του FFT.

1.3.3 Short-time Fourier Transform (STFT) & Spectrogram

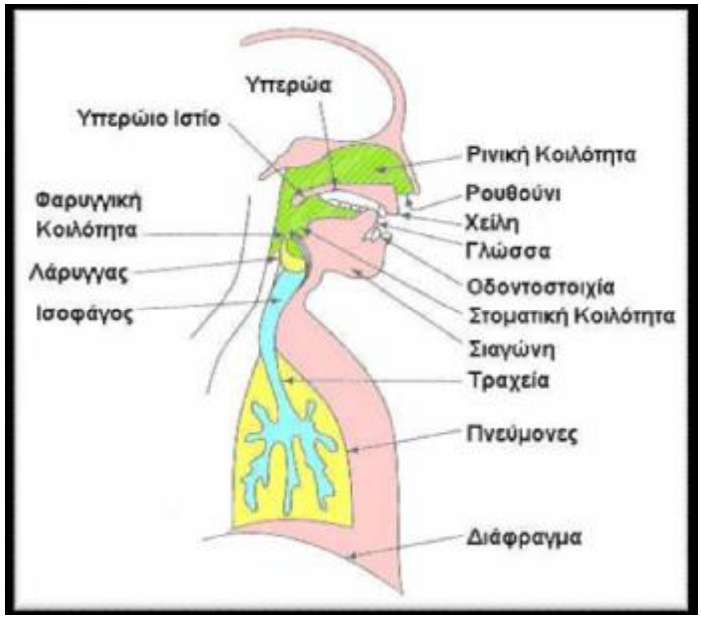
Πρακτικά, η επεξεργασία σήματος πραγματοποιείται μόνο σε ένα μέρος του όλου σήματος.
Σε συνθήκες πραγματικού χρόνου, επειδή δεν ξέρουμε πότε τελειώνει το σήμα, θα πρέπει να βρούμε τρόπο ώστε να επεξεργαστούμε το σήμα γρήγορα, να δοθεί μία έξοδος σε real-time και να αποφευχθεί η καθυστέρηση (delay). Είναι σημαντικό καθώς το ανθρώπινο αυτί μπορεί να αντιληφθεί μια λανθάνουσα κατάσταση των 50ms τουλάχιστον. Όσο μεγαλώνει αυτό το μέγεθος, η κατάσταση αυτή (delay) είναι πολύ πιο αισθητή και ενοχλητική.
Σε αυτή τη λογική βασίζεται ο μετασχηματισμός Fourier-σύντομου χρόνου (Short-time Fourier Transform). Στην μέθοδο του STFT χρησιμοποιούμε την ανάλυση Fourier χωρίς να είναι γνωστό το μήκος του σήματος.
Εδώ επιλέγεται ένα τμήμα του σήματος διάρκειας M δειγμάτων. Καθώς περνάει ο χρόνος, μόλις συμπληρωθεί αυτό το “παράθυρο” (frame) M-δειγμάτων, ξεκινάει η επεξεργασία με τον STFT.
Αυτή η διαδικασία εξασφαλίζει ικανοποιητικά χαμηλή καθυστέρηση και ταυτόχρονα επεξεργασία σε real-time.

2. Ομιλία και Θόρυβος

2.1 Ομιλία

2.1.1 Σύστημα ομιλίας

Κατά την παραγωγή ομιλίας συμμετέχουν πολλά όργανα του ανθρώπινου σώματος. Συγκεκριμένα, εκείνα που ανήκουν στο αναπνευστικό σύστημα και εκείνα που ανήκουν στο ανώτερο πεπτικό σύστημα. Το ανώτερο τμήμα του πεπτικού συστήματος, δηλαδή το στόμα και ο φάρυγγας, παίζει σημαντικό ρόλο στο σχηματισμό των φωνηέντων και συμφώνων, καθώς και στη συγκεκριμένη φωνητική ποιότητα που αυτά αποκτούν. Το σχήμα και το μέγεθος αυτών των κοιλοτήτων μπορεί να αλλάζει με τη βοήθεια των μυών της γλώσσας, της μαλακής υπερώας και του φάρυγγα [7]. Η «συνεργασία» αυτών των δύο συστημάτων έχει ως αποτέλεσμα τη δημιουργία ήχων οι οποίοι σχηματίζουν την ομιλία.



Εικόνα 6: Παραγωγή ομιλίας

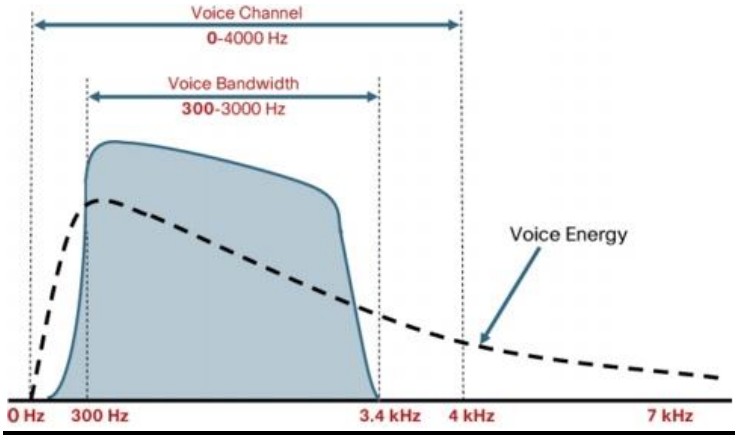
2.1.2 Παραγωγή ομιλίας

Ένα σήμα ομιλίας δημιουργείται από το μηχανισμό παραγωγής ομιλίας του ανθρώπου [7], προκαλώντας διακυμάνσεις στη πίεση του αέρα. Η ομιλία, κατά κύριο λόγο παράγεται στο χωρικό διάστημα μεταξύ φωνητικών χορδών και χειλιών. Ταυτόχρονα, η ρινική κοιλότητα συμμετέχει για να παραχθούν τα ένρινα φωνήματα.
Πιο ειδικά, η ομιλία ξεκινάει από τη θωρακική κοιλότητα η οποία συστέλλεται ή διαστέλλεται. Με τον τρόπο αυτό παράγεται συμπιεσμένος αέρας από τους πνεύμονες και έπειτα από τον λάρυγγα (μέσω της γλωττίδας) κατευθύνεται προς τη γλώσσα και τη στοματική κοιλότητα. Τέλος, αυτή η ροή αέρα απελευθερώνεται από το στόμα, τη μύτη ή και τα δύο, ώστε να γίνει αντιληπτή ως ομιλία.

Οι ήχοι ομιλίας παράγονται με δύο τρόπους με κύρια πηγή την ταλάντωση των φωνητικών χορδών. Αν οι φωνητικές χορδές είναι τεντωμένες, η ροή το αέρα θα προκαλέσει τη δόνησή τους σε μια σταθερή συχνότητα. Αν οι φωνητικές χορδές είναι χαλαρές και η ροή αέρα συναντήσει στένωση κατά μήκος του φαρυγγικού σωλήνα, δημιουργείτε μια ταραχώδης ροή αέρα η οποία έχει τα χαρακτηριστικά θορύβου.
Η ροή αέρα που εκπέμπεται από τη γλωττίδα ορίζεται ως το σήμα διέγερσης. Ο ρυθμός ταλάντωσης των φωνητικών χορδών καθορίζει τη θεμελιώδη συχνότητα F_0 του ακουστικού σήματος.

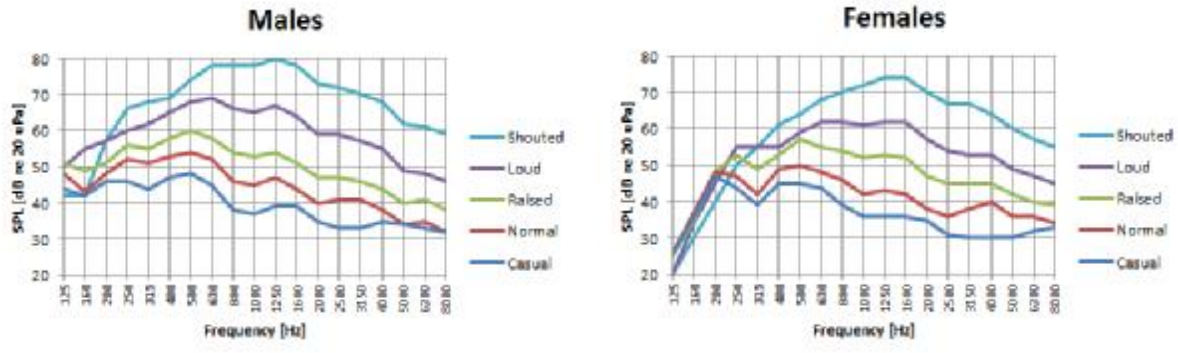
2.1.3 Φασματικά χαρακτηριστικά

Γενικά, ένα σήμα ανθρώπινης ομιλίας έχει εύρος συχνότητας που κυμαίνεται από 100 έως 7000Hz. Μέσω κάποιων μετατροπών στην ποιότητα, το εύρος αυτό μειώνεται στα 300-3400Hz. Θεωρείτε κατάλληλο για σήματα που μεταφέρονται μέσω τηλεφώνου, καθώς σε αυτό το εύρος η επικοινωνία είναι βέλτιστη σε ακουστότητα.



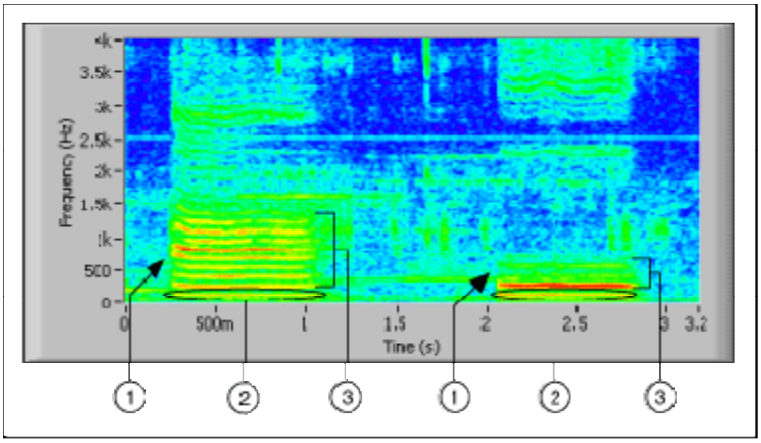
Εικόνα 7:Εύρος ενός σήματος ομιλίας

Στο παραπάνω σχήμα της Εικόνας 7, παρατηρούμε ότι, φυσιολογικά, το εύρος ενός σήματος ομιλίας είναι εξασθενημένο στις υψηλές συχνότητες, κάτι το οποίο θα πρέπει να αντισταθμίζεται.



Για να μπορέσουμε να εντοπίσουμε τα φασματικά χαρακτηριστικά του σήματος εφαρμόζουμε τον Μετασχηματισμό Fourier όπως αναφέρεται και σε προηγούμενο κεφάλαιο. Έχοντας το φάσμα δύναμης ενός σήματος μπορούμε να δούμε, πιθανή κρυμμένη πληροφορία στο

σήμα ή την κατανομή της ενέργειας(τη σχετική ένταση ενέργειας για κάθε συχνότητα του σήματος).
Επιπλέον, παίρνοντας τον μετασχηματισμό Fourier βραχέως-χρόνου και εξάγοντας το φασματογράφημά του (spectrogram), μπορούμε να εντοπίσουμε τις περιόδους απουσίας ομιλίας, τη θεμελιώδη συχνότητα και τις αρμονικές.



Εικόνα 8:STFT Spectrogram

Στο παραπάνω σχήμα της Εικόνας 8, τα χρώματα από το μπλε ως το κόκκινο μας δείχνουν τη συγκέντρωση της ενέργειας για ένα σήμα ομιλίας από πιο ασθενής σε πιο δυνατή [8].
Υπάρχουν τρεις κύριοι τομείς που μπορούμε να «χωρίσουμε» ένα σήμα: α) διαστήματα ομιλίας, β) διαστήματα με απουσία ομιλίας, γ) κλειστά ή διαστήματα κρότου.
Στην πρώτη περίπτωση έχουμε μια αρμονική δομή που καθορίζεται αρκετά καθαρά και αποτελείται από έναν τόνο και τις αρμονικές του.
Στη δεύτερη περίπτωση έχουμε ένα σχεδόν επίπεδο φάσμα το οποίο παραπέμπει σε τυχαίο σήμα θορύβου.
Στην τρίτη περίπτωση έχουμε ένα ευρύ φάσμα για πολύ σύντομο χρονικό διάστημα.

2.1.4 Προσομοίωση Ομιλίας

Για να περιγράψουμε και να αναλύσουμε το φαινόμενο της ομιλίας, θα πρέπει να αναπτυχθεί ένα μοντέλο το οποίο θα περιγράφει με απλό τρόπο το σύστημα ομιλίας. Τα περισσότερα μοντέλα ανάλυσης και σύνθεσης ομιλίας είναι γραμμικά συστήματα.
Το μοντέλο πηγής-φίλτρου είναι ένα από αυτά [7]. Βασίζεται στη θεωρία για την παραγωγή ομιλίας η οποία ονομάζεται *θεωρία πηγής-φίλτρου* και προέκυψε μέσω των πειραμάτων του Johannes Müller. Σύμφωνα με τη θεωρία αυτή, το σήμα ομιλίας δημιουργείται διεγείροντας το σύστημα του φωνητικού καναλιού με μια διέγερση ευρείας ζώνης. Το φωνητικό κανάλι αλλάζει σχήμα σχετικά αργά στο χρόνο και επομένως μπορεί να προτυποποιηθεί σαν ένα χρονικά-μεταβαλλόμενο φίλτρο το οποίο επιβάλλει τις ιδιότητες της απόκρισης συχνότητάς του στο φάσμα της διέγερσης. Το φωνητικό κανάλι χαρακτηρίζεται από τις φυσικές του συχνότητες, οι οποίες ονομάζονται formants και αντιστοιχούν σε συντονισμούς (resonances) στην απόκριση της συχνότητάς του.
Ως μαθηματική σχέση μπορεί να προσδιοριστεί:

$$s(t) = e(t) * v(t) = [p(t) + a(t)] * v(t) \qquad (2.1.4.1)$$

όπου e(t) η πηγή διέγερσης, v(t) το φίλτρο, p(t) η συνιστώσα για το κανονικό πρότυπο δονήσεων των φωνητικών χορδών και a(t) η συνιστώσα για τις ανωμαλίες στους εύφωνους και άφωνους ήχους.

2.2 Θόρυβος

Το ενθόρυβο σήμα προς επεξεργασία αποτελείται από ένα σήμα ομιλίας και ένα σήμα θορύβου.
Σε πραγματικές συνθήκες δεν μπορούμε απλά να “διώξουμε” τον θόρυβο, καθώς έχουμε πρόσβαση μόνο στο τελικό σήμα. Σε αυτές τις περιπτώσεις προσπαθούμε να μειώσουμε τον θόρυβο σε τέτοιο επίπεδο ώστε, από τη μία η μείωση να είναι ικανοποιητική (σχεδόν ανύπαρκτος) και από την άλλη η ποιότητα του καθαρού, πλέον, σήματος ομιλίας να είναι όσο το δυνατόν καλύτερη. Βέβαια, κατά την επεξεργασία πρέπει να ληφθεί υπ' όψη το γεγονός ότι στο σήμα εξόδου θα δημιουργηθούν artifacts όπως θα δούμε και σε επόμενη ενότητα τα οποία, στον ακροατή, δεν θα πρέπει να γίνονται αντιληπτά.

2.2.1 Γενικά Χαρακτηριστικά Θορύβων

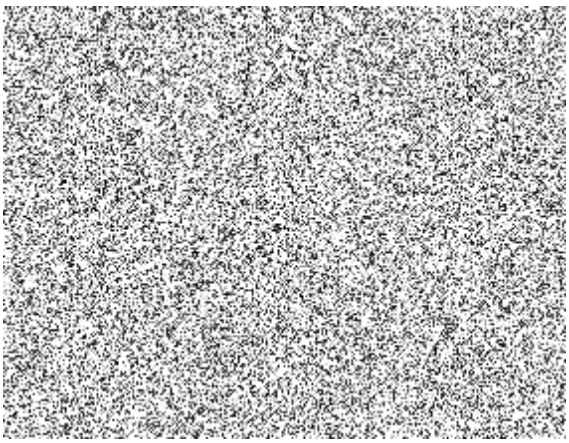
Με μια γενική προσέγγιση, ο θόρυβος γίνεται αντιληπτός ως ένα τυχαίο σήμα. Δημιουργείται από εξωτερικούς παράγοντες ή γεννάται μέσα στο ίδιο το σήμα ή το μέσο μετάδοσης της πληροφορίας. Ωστόσο, μπορούν να οριστούν συγκεκριμένες φασματικές, χρονικές και χωρικές ιδιότητες.

2.2.2 Φασματικά Χαρακτηριστικά και είδη θορύβου

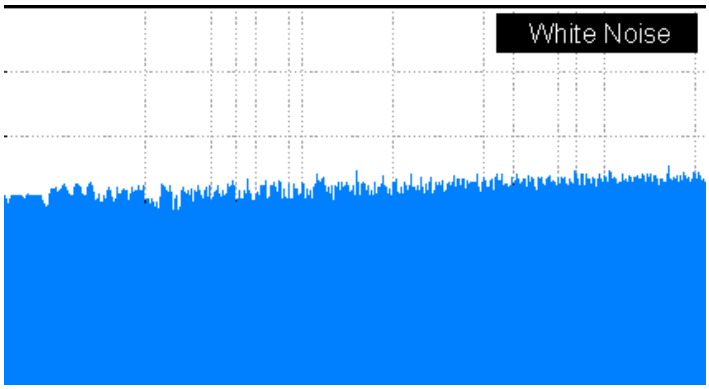
Στις φασματικές ιδιότητες ανήκει η πυκνότητα φάσματος (***spectral density***), η μέση τιμή (***mean***) και η διακύμανση (***variation***).

Λευκός θόρυβος

Είναι ο θόρυβος όπου στο φάσμα παρατηρείται ίση κατανομή ενέργειας σε όλες τις συχνότητες ενός σήματος, καθώς και τυχαία φάση. Η ονομασία «λευκός» προκύπτει από την αναλογία φωτός και παρουσία συχνοτήτων. Λόγω της παρουσίας όλων των συχνοτήτων, συνεπώς και όλων των χρωμάτων, προκύπτει το λευκό.
Στην παρούσα εργασία μελετάται ο συγκεκριμένος τύπος θορύβου, καθώς μπορεί να μοντελοποιηθεί και να εκτιμηθεί πιο εύκολα.
Ωστόσο, δεν ανταποκρίνεται σε πραγματικές συνθήκες, αλλά είναι μια αποδεδειγμένα καλή λύση για την ανάπτυξη των αλγορίθμων.



Εικόνα 9: Λευκός θόρυβος



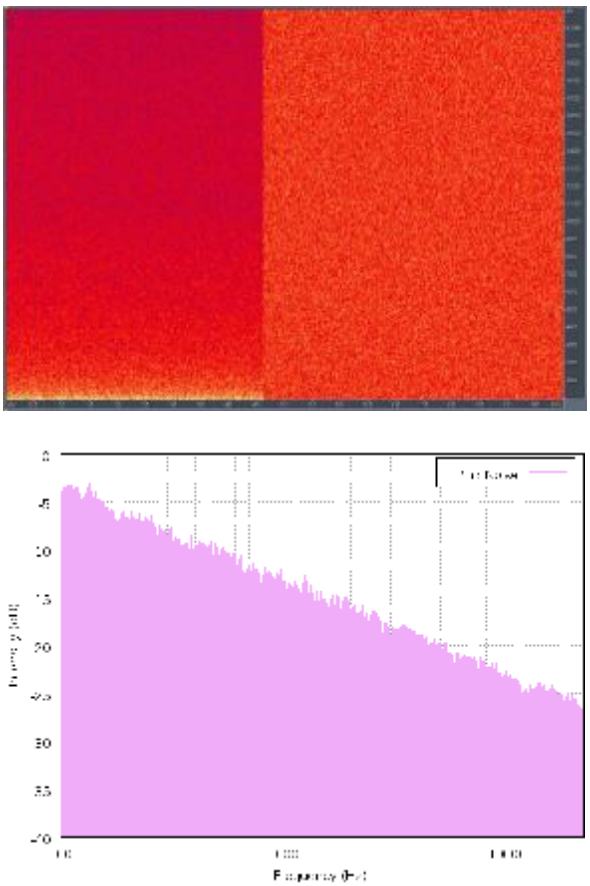
Εικόνα 10: Φάσμα λευκού θορύβου

Πολύχρωμος θόρυβος

Εδώ έχουμε ένα σήμα θορύβου με φάσμα που αντιπροσωπεύει ένα χρώμα και με τυχαία φάση. Χρησιμοποιείται συνήθως για αντικατάσταση σημάτων στο υπόβαθρο.
Σε αυτή την κατηγορία ανήκουν δύο χαρακτηριστικά παραδείγματα, ο ροζ και καφέ θόρυβος.

Ροζ θόρυβος

Αυτός ο τύπος θορύβου πλησιάζει περισσότερο στο φάσμα που παρατηρείται να έχει ένας θόρυβος σε εφαρμογές ομιλίας. Παρατηρείται ότι στις υψηλές συχνότητες έχει κάθοδο 6dB ανά οκτάβα.
Αντίστοιχα με τον λευκό θόρυβο, το χρώμα προκύπτει με τον ίδιο τρόπο. Το κόκκινο χρώμα έχει χαμηλή συχνότητα, εκεί δηλαδή όπου υπάρχει η μεγαλύτερη συγκέντρωση συχνοτήτων.



2.2.3 Παραγωγή θορύβου

Γενικά, τα σήματα ομιλίας θεωρούνται μη-στάσιμα αλλά, σε μικρού-μήκους frame 10-40ms εκλαμβάνονται ως στάσιμα.

Μια από τις πιο χρησιμοποιούμενες παραδοχές για ένα σήμα θορύβου είναι η υπόθεση ότι ο θόρυβος είναι ένα στάσιμο σήμα. Θεωρείται, δηλαδή, ότι οι στατιστικές ιδιότητες τόσο στο πεδίο του χρόνου όσο και στης συχνότητα δεν αλλάζουν όσο περνάει ο χρόνος. Στη βιβλιογραφία, ένα σήμα θορύβου μπορεί να αναφέρεται ως “quasi-stationary” (ημι-στάσιμο). Με τον όρο αυτό σημειώνεται ότι ναι μεν οι ιδιότητες του σήματος θορύβου δεν αλλάζουν, αλλά προκύπτει σε σύγκριση με το σήμα που αντιπαρατίθεται (συνήθως σήμα ομιλίας). Έτσι, οι αλλαγές των ιδιοτήτων μπορεί να συμβαίνουν αλλά είναι πολύ πιο αργές από αυτές του δεύτερου παρόντος σήματος.

Από τα προαναφερθέντα είδη θορύβου, ο λευκός θόρυβος Gauss *(ακολουθία αμοιβαία ανεξάρτητων μεταβλητών με την ίδια κατανομή)* είναι ένα στατιστικό μοντέλο που μπορεί να περιγράψει αρκετά καλά τα περισσότερα σήματα θορύβου σε επεξεργασία ομιλίας.

Για τη μοντελοποίηση του χρησιμοποιείται διακύμανση σ^2_N, μέση τιμή $\mu_N=0$ και κανονική συνάρτηση πυκνότητας πιθανότητας

$$p(x|\mu_N,\sigma_N) = \frac{1}{\sigma_N\sqrt{2\pi}}\exp\left(-\frac{(x-\mu_N)^2}{2\sigma_N^2}\right) = \frac{1}{\sigma_N\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma_N^2}\right) \qquad (2.2.3.1)$$

Σε αυτή την περίπτωση θορύβου η μέση τιμή και η αυτοδιασπορά είναι ανεξάρτητες από το χρόνο και η διεργασία είναι δευτέρας τάξεως στάσιμη. Επίσης είναι και αυστηρώς στάσιμη.
Το φάσμα ισχύος του λευκού θορύβου είναι σταθερό και δεν εξαρτάται από τις τιμές της συχνότητας

2.3 Ψυχο-ακουστικό μοντέλο

Μηχανισμός ακοής – Φαινόμενο επικάλυψης (masking)

Το ανθρώπινο σύστημα ακοής μπορεί να διαιρεθεί σε δύο υποσυστήματα, το περιφερειακό και το νευρικό [9].
Το περιφερειακό αποτελείται από το εξωτερικό, μέσω και εσωτερικό αυτί. Από τα τρία αυτά μέρη κύριο ρόλο έχει το εσωτερικό αυτί. Αποτελείται από δύο μέρη, τον κοχλία και τον λαβύρινθο. Στον κοχλία υπάρχουν δύο οργανικά υγρά μέσω των οποίων οι ταλαντώσεις που έρχονται από το μέσο αυτί μεταδίδονται στη *Βασική Μεμβράνη*. Εκεί, λαμβάνει χώρα ένας μετασχηματισμός χωρικός και φαίνεται ότι αυτός είναι ο τρόπος που το αυτί πραγματοποιεί κάποια μορφή ανάλυσης φάσματος.

Επικάλυψη(Masking)

Αποτελεί φαινόμενο του εσωτερικού αυτιού (κοχλίας) και συγκεκριμένα πραγματοποιείται πάνω στη Βασική μεμβράνη. Εκεί λαμβάνει χώρα ένας μετασχηματισμός χωρικός και φαίνεται ότι αυτός είναι ο τρόπος που το αυτί πραγματοποιεί κάποια μορφή ανάλυσης φάσματος.
Πιο συγκεκριμένα, η επικάλυψη οφείλεται σε ισχυρές ταλαντώσεις από σήματα υψηλής στάθμης σε κάποιο σημείο της μεμβράνης, με αποτέλεσμα άλλα σήματα να δημιουργούν αμελητέα ή καθόλου ταλάντωση. Σε μεγάλο βαθμό τα φαινόμενα επικάλυψης οφείλονται σε ανώτερα επίπεδα του νευρικού συστήματος.
Το φαινόμενο της επικάλυψης διακρίνεται κυρίως σε *Ταυτόχρονη* και *Παροδική* [10].

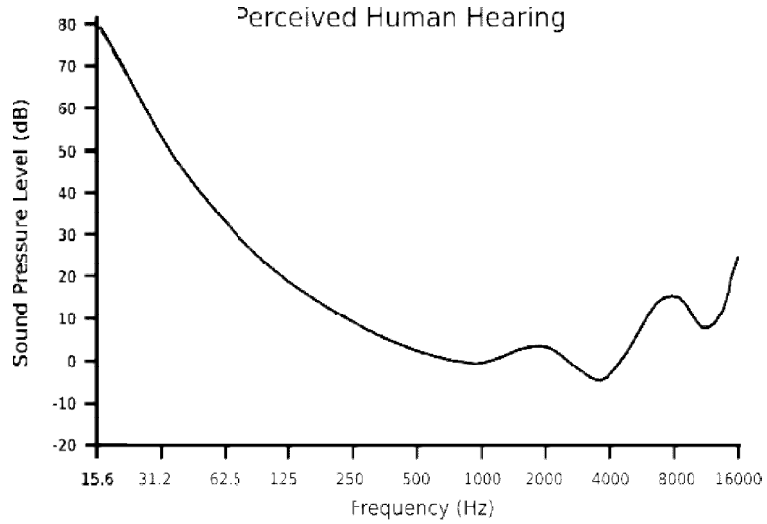
1. Στην *Ταυτόχρονη Επικάλυψη* ένα σήμα (το οποίο προκαλεί φαινόμενα επικάλυψης) χρησιμοποιείται για την μέτρηση ενός κατωφλίου, το οποίο ονομάζεται *Ακουστό Κατώφλι Επικάλυψης* κάτω από τη στάθμη του οποίου ένα σήμα δεν γίνεται αντιληπτό. Η μεθοδολογία που χρησιμοποιείται: υπολογισμός της ενέργειας του σήματος ανά κρίσιμη περιοχή, συνέλιξη με κάποια συνάρτηση διασποράς που προσομοιάζει τη μορφή μιας τυπικής καμπύλης επικάλυψης από ημιτονικό σήμα και στη συνέχεια βελτίωση ανάλογα με το φασματικό περιεχόμενο του σήματος.
2. Η *Παροδική Επικάλυψη* είναι ένα φαινόμενο που κατά κύριο λόγο οφείλεται σε πιο κεντρικές νευρικές επεξεργασίες και λιγότερο στον κοχλία. Έχει δειχθεί ότι σήματα πεπερασμένης διάρκειας παράγουν επικάλυψη στο πεδίο του χρόνου που χωρίζεται σε τρεις περιοχές: προ-επικάλυψη, ταυτόχρονη, μετεπικάλυψη. Η μετεπικάλυψη παίζει τον κυριότερο ρόλο στα παροδικά φαινόμενα. Μπορεί να διαρκέσει μέχρι 200msec και επιπλέον η κλίση της είναι ανάλογη με τη στάθμη του επικαλυπτή σήματος και με τη διάρκειά του. Προσέγγιση μοντελοποίησης: ένα μοντέλο προσαρμοσμένο σε πειραματικά δεδομένα, θέτει το ποσό επικάλυψης ανάλογο της στάθμης του επικαλυπτή σήματος και ανάλογο του λογαρίθμου της καθυστέρησης από τη στιγμή που ξεκινάει το τεστ σήμα.

Απόλυτο Κατώφλι Ακουστότητας

Η ανθρώπινη ακοή έχει ένα συγκεκριμένο εύρος. Το κατώτερο όριο αυτής της περιοχής είναι το *"Απόλυτο Κατώφλι Ακουστότητας"*. Σε συνάρτηση με τη συχνότητα υποδηλώνει εκείνη την ηχητική στάθμη που μόλις γίνεται αντιληπτή.
Τυπικά, ορίζεται ως απόλυτο κατώφλι ακουστότητας η ποσότητα της ενέργειας που απαιτείται για έναν καθαρό ήχο, ώστε να γίνει αντιληπτός από έναν ακροατή σε περιβάλλον χωρίς θόρυβο. Εκφράζεται σε dB SPL και προσδιορίζεται από την εξίσωση(όπου f σε kHZ):

$$T_q(f) = 3.64f^{-0.8} - 6.5e - 0.6^{(f-3.3)^2} + 10^{-3}f^4 \quad (2.3.1)$$

Στο σχήμα φαίνεται το κατώφλι ακουστότητας με Tq σε dB και f σε HZ.
Σε εφαρμογές συμπίεσης σημάτων, μπορεί να ερμηνευθεί ως το μέγιστο επιτρεπτό όριο ενέργειας, ώστε να μην γίνονται αντιληπτές κατά την κωδικοποίηση οι παραμορφώσεις που εισάγονται στο πεδίο της συχνότητας.



Εικόνα 11:Απόλυτο κατώφλι ακουστότητας

Κρίσιμες Περιοχές Συχνότητας

Η θεωρία των "Κρίσιμων Περιοχών Συχνότητας" εισήχθη αρχικά από τον Fletcher, υποθέτοντας ότι *το τμήμα του φάσματος του θορύβου που επικαλύπτει ένα τονικό σήμα, είναι μέρος του φάσματος που βρίσκεται κοντά στο τονικό σήμα.*

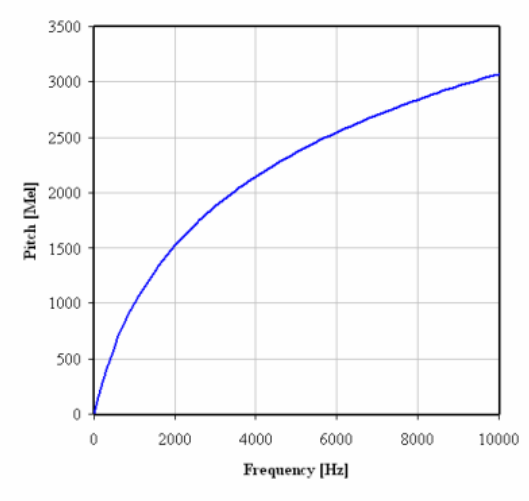
Στον παρακάτω πίνακα παρουσιάζονται οι κρίσιμες συχνότητες (εύρος/κεντρική συχνότητα) σύμφωνα με τον Zwicker:

Band No.	Range (Hz)	Center Freq. (Hz)
1	0–100	50
2	100–200	150
3	200–300	250
4	300–400	350
5	400–510	450
6	510–630	570
7	630–770	700
8	770–920	840
9	920–1080	1000
10	1080–1270	1175
11	1270–1480	1370
12	1480–1720	1600
13	1720–2000	1850
14	2000–2320	2150
15	2320–2700	2500
16	2700–3150	2900
17	3150–3700	3400
18	3700–4400	4000

Εικόνα 12:Κρίσιμες συχνότητες

Κλίμακα MEL

Κατά την ακοή ο άνθρωπος δεν αντιλαμβάνεται το ύψος του τόνου με γραμμικό τρόπο(ως γραμμική διεργασία). Η ερμηνεία του ύψους του τόνου αυξάνεται αντίστοιχα με τη συχνότητα, κάτι το οποίο μπορεί να είναι μη επιθυμητό σε ορισμένες εφαρμογές. Έτσι, σαν αντιστάθμιση αναπτύχθηκε η κλίμακα Mel. Βασίζεται στην ανθρώπινη ακοή και αναπτύχθηκε μέσω πειραματισμών πάνω σε αυτή. Σκοπός αυτής της έρευνας ήταν η γραμμική προσέγγιση της ακοής.



Εικόνα 13:Κλίμακα Mel

Όπως φαίνεται στο διάγραμμα, το ύψος του τόνου γίνεται αντιληπτό ως γραμμικό στις συχνότητες 0 – 1000Hz. Πάνω από τα 1000Hz γίνεται λογαριθμικό.
Ο υπολογισμός της κλίμακας δίνεται από τον εξής τύπο:

$$F_{mel} = \frac{1000}{\log 2}\left[1 + \frac{F_{Hz}}{1000}\right] \qquad (2.3.2)$$

όπου Fmel το αποτέλεσμα σε κλίμακα mel σε μέτρηση mel και Fhz, και Fhz η κανονική συχνότητα σε Hz.

3.Φίλτρα Αποθορυβοποίησης

Στα πλαίσια της εργασίας γίνεται αναφορά στους αλγόριθμους αποθορυβοποίησης Boll, Beroutti, Scalart, Ephraim – Malah, Kamath, καθώς και στον αλγόριθμο Τσουκαλά ο οποίος βασίζεται στην "Ψυχό-ακουστική". Οι παραπάνω αλγόριθμοι αποτελούν τη βάση μελέτης και ανάπτυξης του προτεινόμενου από την εργασία αλγόριθμου.

Οι πέντε πρώτοι αλγόριθμοι βασίζονται καθαρά σε στατιστικά χαρακτηριστικά του σήματος προκειμένου να αναπτυχθεί μια διαδικασία κατάλληλη για την αφαίρεση του ανεπιθύμητου θορύβου. Τέτοια χαρακτηριστικά είναι η μέση τιμή, η διακύμανση, ο λόγος σήματος προς θόρυβο SNR, η συνάρτηση πυκνότητας πιθανότητας, κ.τ.λ..

Ο αλγόριθμος του Τσουκαλά εισάγει ένα νέο τρόπο για την εκτίμηση και αφαίρεση του θορύβου, ο οποίος βασίζεται στο Ψυχό-ακουστικό μοντέλο. Λαμβάνει υπόψη δηλαδή, χαρακτηριστικά της ανθρώπινης ακοής και αντίληψης.

3.1 Αλγόριθμος Boll

Η προσέγγιση του Boll [11] για τη συμπίεση του θορύβου, στο γενικό πλαίσιο, είναι να προκύψει μια εκτίμηση για το φάσμα συχνότητας του καθαρού σήματος ομιλίας, αφαιρώντας το φάσμα του θορύβου από το ενθόρυβο φάσμα ομιλίας. Για τη διαδικασία αυτή απαιτείται μια εκτίμηση του τρέχοντος φάσματος θορύβου. Το μέγεθος αυτό υπολογίζεται προσεγγιστικά χρησιμοποιώντας το μέσο όρο του θορύβου που μετρήθηκε κατά τη διάρκεια απουσίας ομιλίας.

Ωστόσο, μετά την ολοκλήρωση της διαδικασίας συμπίεσης θορύβου μπορεί να απομείνει θόρυβος γνωστός ως "residual noise".

Σύμφωνα με τον αλγόριθμο το ηχητικό αποτέλεσμα του παραπάνω θορύβου, αφαιρείται παίρνοντας μια ελάχιστη τιμή φάσματος για κάθε στοιχείο και όχι θέτοντάς το στο μηδέν. Πιο συγκεκριμένα, για κάθε πλαίσιο συχνοτήτων ο θόρυβος θα συμπιεστεί αντικαθιστώντας κάθε τρέχουσα τιμή του με την ελάχιστή του η οποία επιλέγεται από τα εκάστοτε παράθυρα ανάλυσης. Η ελάχιστη αυτή τιμή επιλέγεται μόνο όταν το μέτρο του $\hat{S}(e^{j\omega})$ είναι μικρότερο από τη μέγιστη τιμή του θορύβου που απομένει και έχει υπολογιστεί κατά τα διαστήματα απουσίας ομιλίας.

$$\left|\widehat{S_i}(e^{j\omega})\right| = \begin{cases} \left|\widehat{S_i}(e^{j\omega})\right| \;\; \gamma\iota\alpha \left|\widehat{S_i}(e^{j\omega})\right| \geq max\left|N_R(e^{j\omega})\right| \\ \min_{j=i-1,i,i+1}\left\{\left|\widehat{S_i}(e^{j\omega})\right|\right\} \;\; \gamma\iota\alpha \left|\widehat{S_i}(e^{j\omega})\right| < max\left|N_R(e^{j\omega})\right| \end{cases} \qquad (3.1.1)$$

Από την παραπάνω εξίσωση προκύπτει ότι στις περιπτώσεις που πιθανώς έχουμε θόρυβο ή είναι ένα κομμάτι χαμηλής ενέργειας (πιθανότατα ομιλίας), θα γίνει προσπάθεια διατήρησης της πληροφορίας παίρνοντας μια ελάχιστη φασματική τιμή.

3.2 Αλγόριθμος Berouti

Πραγματοποιείται αφαίρεση του φάσματος θορύβου κατά έναν παράγοντα α, όπου α ένας αριθμός μεγαλύτερος από την μονάδα και ποικίλει από πλαίσιο σε πλαίσιο. Έπειτα δίνεται προσοχή ώστε να αποφευχθούν φασματικά στοιχεία του προς επεξεργασία σήματος από το να πάνε κάτω από ένα συγκεκριμένο κατώφλι το οποίο ονομάζεται φασματικό κατώφλι (spectral floor). Κατόπιν εκφράζουμε το φασματικό κατώφλι σαν ένα μέρος β του αυθεντικού φάσματος ισχύος Pn(w) [12].

3.3 Αλγόριθμος Multiband

Ο θόρυβος στον πραγματικό κόσμο είναι κυρίως έγχρωμος (colored noise), δηλαδή δεν επηρεάζει το σήμα ομιλίας ομοιόμορφα σε όλο το μήκος του φάσματός του. Πάνω σε αυτό, αναπτύχθηκε η *Πολυζωνική Αφαίρεση* [13], που λαμβάνει υπόψη το γεγονός ότι αυτός ο τύπος θορύβου επηρεάζει διαφορετικά το φάσμα ομιλίας στις διαφορετικές συχνότητες. Στην προσέγγιση αυτή, επιτυγχάνεται μείωση των παραμορφώσεων σε μεγάλη έκταση ενώ διατηρείται υψηλού επιπέδου ποιότητα ομιλίας.
Το φάσμα ομιλίας διαχωρίζεται σε N μη-επικαλυπτόμενες ζώνες και εφαρμόζεται φασματική αφαίρεση ανεξάρτητα σε κάθε ζώνη. Σε επόμενη ενότητα δίνεται η λεπτομερής ανάλυση της *Πολυζωνικής Αφαίρεσης*.

Μία παραλλαγή της Φασματικής Αφαίρεσης είναι η *Πολυζωνική Φασματική Αφαίρεση* [13]. Σε σχετική αναφορά στην ενότητα του θορύβου, ο θόρυβος ο οποίος ανταποκρίνεται σε πραγματικές συνθήκες είναι ο colored και δεν επηρεάζει το σήμα ομιλίας ομοιόμορφα σε όλο το φάσμα.
Η μέθοδος αυτή υπερτερεί της τυπικής μεθόδου Φασματικής Αφαίρεσης, αποσκοπώντας σε ανώτερη ποιότητα ομιλίας και σε μεγάλο βαθμό μειωμένη από μουσικό θόρυβο.
Αν και το αντικείμενο αυτής της εργασίας δεν επεκτείνεται σε αυτό τον τομέα, είναι πολύ σημαντικό να αναφερθεί καθώς όπως είπαμε ανταποκρίνεται σε ρεαλιστικές συνθήκες.
Η διαδικασία είναι κοινή μέχρι το σημείο της εκτίμησης του φάσματος ισχύος.

$$|Y(k)|^2 \approx |X(k)|^2 + |D(x)|^2 \tag{3.3.1}$$

Στη συνέχεια, *το φάσμα ομιλίας διαιρείται σε N μη-επικαλυπτόμενες ζώνες και η φασματική αφαίρεση εκτελείται ανεξάρτητα σε κάθε ζώνη. Η διαδικασία της διάσπασης του σήματος ομιλίας σε διαφορετικές ζώνες μπορεί να πραγματοποιηθεί είτε στο πεδίο του χρόνου χρησιμοποιώντας ζωνοπερατά φίλτρα ή στο πεδίο της συχνότητας χρησιμοποιώντας κατάλληλα παράθυρα.*
Η εκτίμηση του φάσματος του καθαρού σήματος στην ι-οστή ζώνη προκύπτει από:

$$|\hat{X}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i \delta_i |\hat{D}_i(k)|^2 \quad b_i \le k \le e_i \tag{3.3.2}$$

όπου τα b_i και e_i είναι η αρχή και το τέλος των τομέων συχνότητας της ι-οστής μπάντας συχνότητας, α_i είναι ο παράγοντας υπεραφαίρεσης της ι-οστής ζώνης και δ_i είναι ένας επιπλέον παράγοντας που μπορεί να οριστεί ξεχωριστά για κάθε μπάντα συχνοτήτων για να προσαρμόσει τις ιδιότητες της αφαίρεσης θορύβου.
Αρνητικές τιμές που μπορεί να προκύψουν στο ενισχυμένο φάσμα της παραπάνω εξίσωσης, ελαχιστοποιούνται στο φάσμα θορύβου:

$$|\hat{X}_i(k)|^2 = \begin{cases} |\hat{X}_i(k)|^2 & |\hat{X}_i(k)|^2 > 0 \\ \beta |Y_i(k)|^2 & αλλού \end{cases} \tag{3.3.3}$$

όπου η παράμετρος ελαχιστοποίησης β ορίζεται στην τιμή 0.002.
Ο παράγοντας υπεραφαίρεσης α_i είναι μια συνάρτηση του τμηματικού SNR για την ι-οστή μπάντα συχνοτήτων και δίνεται από τον τύπο:

$$SNR_i(dB) = 10\log_{10}\left(\frac{\sum_{k=b_i}^{e_i}|Y_i|^2}{\sum_{k=b_i}^{e_i}\left|\widehat{D}_i\right|^2}\right) \tag{3.3.4}$$

Έχοντας προσδιορίσει την τιμή SNR από την παραπάνω εξίσωση, υπολογίζουμε το α_i ως:

$$\alpha_i = \begin{cases} 5 & SNR_i < -5 \\ 4 - \frac{3}{20}(SNR_i) & -5 \le SNR_i \le 20 \\ 1 & SNR_i > 20 \end{cases} \tag{3.3.5}$$

Οι τιμές του παράγοντα δ_i της εξίσωσης (3.3.2) προσδιορίζονται εμπειρικά από:

$$\delta_i = \begin{cases} 1 & f_i \le 1kHz \\ 2.5 & 1kHz < f_i \le \frac{Fs}{2} - 2kHz \\ 1.5 & f_i > \frac{Fs}{2} - 2kHz \end{cases} \tag{3.3.6}$$

όπου fi είναι η άνω συχνότητα της ι-οστής μπάντας, και Fs είναι η συχνότητα δειγματοληψίας.
Κατά την εφαρμογή εφαρμόζεται στο σήμα |Y(n)| προ-επεξεργασία όπου, το φάσμα ισχύος ομαλοποιείται ώστε να μειωθεί ο υπολειμματικός θόρυβος κατά τη διαδικασία αφαίρεσης. *Ένας ζυγισμένος φασματικός μέσος όρος επιβάλλεται στα πλαίσια ομιλίας που προηγούνται και ακολουθούν ως εξής:*

$$\bar{Y}_j(k) = \sum_{i=-M}^{M} W_i Y_{j-i}(k) \tag{3.3.7}$$

όπου $\bar{Y}_j(k)$ είναι το μέτρο του φάσματος του θορυβώδους σήματος του ι-οστού πλαισίου, j είναι ο δείκτης πλαισίου. Ο αριθμός των πλαισίων M περιορίζεται σε 2 για να αποφευχθεί η αλλοίωση της φασματικής πληροφορίας. Τα βάρη του φίλτρου Wi ορίζονται W=[0.09, 0.25, 0.32, 0.25, 0.09].

3.4 Αλγόριθμος Scalart

Σε αυτή τη μέθοδο [14] έχουμε το σενάριο εκτίμησης ενός a-priori SNR που αποδεδειγμένα οδηγεί σε καλύτερα υποκειμενικά αποτελέσματα. Μέσω αυτού επιτυγχάνεται μια εξισορρόπηση μεταξύ αποτελεσματικής μείωσης θορύβου και χαμηλής υπολογιστικής ισχύος για real-time διεργασίες.
Θεωρούμε s(t) και b(t) σήμα ομιλίας και θορύβου αντίστοιχα. Το υπό εξέταση σήμα δίνεται από τη συνάρτηση $x(t) = s(t) + b(t)$. Έστω $S_k = A_k e^{ja_k}$, B_k, $X_k = R_k e^{jv_k}$, δηλώνουν το k-οστο φασματικό στοιχείο του s(t), τον θόρυβο b(t) και τον θόρυβο που εξετάζουμε x(t) στο διάστημα ανάλυσης [0,T] όπου η ημι-στατικότητα του σήματος ομιλίας εξασφαλίζεται για την περίοδο T. Είναι χρήσιμο να λάβουμε υπόψη ότι η εκτίμηση πλάτους $\hat{A}_k$ προκύπτει από το X_k μέσω μιας πολλαπλασιαστικής μη-γραμμικής συνάρτησης κέρδους ορισμένη ως $G(f_k) \triangleq \hat{A}_k / X_k$.
Μπορούμε να εκφράσουμε τη βέλτιστη συνάρτηση κέρδους ως προϊόν της γενικής συνάρτησης κέρδους με έναν όρο ο οποίος συνεισφέρει στο κομμάτι της «χαλαρής-απόφασης» (soft-decision)

της εκτίμησης όπως δίνεται από τον τύπο:

$$G(f_k) = \frac{\Lambda(X_k, q_k)}{1 + \Lambda(X_k, q_k)} \quad G_\circ(f_k) \tag{3.4.1}$$

Όπου $\Lambda(X_k, q_k)$ είναι ο γενικευμένος λόγος πιθανότητας λαμβάνοντας υπόψη την αβεβαιότητα παρουσίας ομιλίας στο υπό εξέταση ενθόρυβο σήμα και ορίζεται ως:

$$\Lambda(X_k, q_k) = \mu_k \frac{p(X_k/H_k^1)}{p(X_k/H_k^0)} \tag{3.4.2}$$

με $\mu_k \triangleq (1 - q_k)/q_k$, όπου q_k είναι η πιθανότητα απουσίας σήματος στο k-στο φασματικό στοιχείο και p(.) δηλώνει τη συνάρτηση πυκνότητας πιθανότητας. Τα H_k^0, H_k^1 δηλώνουν τις δύο υποθέσεις απουσίας και παρουσίας σήματος, αντίστοιχα, στο k-στο φασματικό στοιχείο. Να σημειωθεί ότι αν $q_k = 0$, τότε $\Lambda/(1 + \Lambda)$ ισούται με τη μονάδα, και $G(f_k)$ ισούται με τη γενική συνάρτηση $G_\circ$ όταν είναι πάντα παρόν στο υπό εξέταση ενθόρυβο σήμα.
Ορίζουμε ένα τοπικό a-posteriori και a-priori SNR:

$$SNR_{post}(f_k) \triangleq \frac{|X_k|^2}{E\{|B_k|^2\}} \qquad SNR_{prio}(f_k) \triangleq \frac{E\{|S_k|^2\}}{E\{|B_k|^2\}} \tag{3.4.3}$$

Αυτές οι εκτιμήσεις πλάτους (3.4.3) προέκυψαν από την υπόθεση ότι είναι γνωστά το a-priori SNR και συνάρτηση πυκνότητας πιθανότητας. Ωστόσο, σε πρακτικές εφαρμογές συστημάτων βελτίωσης ομιλίας, αυτές οι παράμετροι είναι ήδη γνωστές καθώς υπάρχει πρόσβαση στο ενθόρυβο σήμα. Επιπλέον, έχει αναφερθεί ότι το a-priori SNR λειτουργεί σαν παράμετρος κλειδί (παρά σαν διασπορά του θορύβου) στη ελάττωση των διαστρεβλώσεων ομιλίας και του μουσικού θορύβου. Προκειμένου λοιπόν να έχουμε αλγορίθμους που λειτουργούν σε πραγματικό χρόνο, υπάρχουν δύο περιπτώσεις: στην πρώτη περίπτωση επιχειρείται ο ορισμός, μέσω ανεπίσημης ακρόασης, μιας βέλτιστης τιμής a-priori SNR (5 έως 7 dB). Στη δεύτερη περίπτωση αντιστοιχίζονται οι άγνωστοι παράμετροι από τις ακόλουθες εκτιμήσεις για την πυκνότητα φάσματος ισχύς του θορύβου, το a-posteriori και a-priori SNR αντίστοιχα:

$$\hat{P}_B^t(f_k) = \lambda . \hat{P}_B^{t-1}(f_k) + (1 - \lambda) . |B^t(f_k)|^2 \tag{3.4.4}$$

$$\widehat{SNR}_{post}^t(f_k) = \frac{|X_k|^2}{\hat{P}_B^t(f_k)} \tag{3.4.5}$$

$$\widehat{SNR}_{priot}^t(f_k) = (1 - \beta) . P\left[\widehat{SNR}_{post}^t(f_k) - 1\right] + \beta . \frac{|\hat{S}^{t-1}(f_k)|^2}{\hat{P}_B(f_k)} \tag{3.4.6}$$

όπου P[.] δηλώνει μια ημι-κυματική διόρθωση (half-wave rectification) και $(.)^t$ το παρόν χρονικό διάστημα.
Μετά από μετρήσεις καταλήγουμε στην πρόταση ενός a-priori SNR σε κλασικές μεθόδους βελτίωσης ομιλίας, όπως Wiener, φασματική αφαίρεση κ.α. Αυτό γίνεται θεωρώντας $E\{SNR_{post}(f_k)\} = 1 + SNR_{prio}(f_k)$, το οποίο οδηγεί:

$$G_{\diamond}^{PE} = \sqrt{\frac{SNR_{prio}(f_k)}{1 + SNR_{prio}(f_k)}} \quad (3.4.7)$$

$$G_{\diamond}^{W} = \frac{SNR_{prio}(f_k)}{1 + SNR_{prio}(f_k)} \quad (3.4.8)$$

$$G_{\diamond}^{ML} = \frac{1}{2}\left[1 + \sqrt{\frac{SNR_{prio}(f_k)}{1 + SNR_{prio}(f_k)}}\right] \quad (3.4.9)$$

όπου η εκτίμηση του a-priori SNR εκτιμάται με την προσέγγιση «καθοδηγούμενης-απόφασης» (3.4.6). Σημειώνεται επίσης ότι οι προτεινόμενοι εκτιμητές δίνουν χαμηλό χρόνο υπολογισμού για εφαρμογές πραγματικού χρόνου.

3.4.1 Φίλτρα Wiener

Η γενική θεωρία στα φίλτρα Wiener [10] είναι ότι εφαρμόζεται ένα βέλτιστο φίλτρο με υπολογισμό του *Μέσου Τετραγωνικού Σφάλματος* και είναι το καλύτερο σε αυτή την κατηγορία. Τα φίλτρα αυτού του τύπου είναι γραμμικά. Ωστόσο, δεν μπορούν να εφαρμοστούν *"άμεσα στην αποθορυβοποίηση ομιλίας αφού το σήμα ομιλίας δεν είναι στάσιμο και η πυκνότητα του φάσματος ισχύος δεν είναι γνωστή"*. Υπάρχουν δύο τύποι αυτού του φίλτρου, τα πεπερασμένης κρουστικής απόκρισης (FIR) και άπειρης κρουστικής απόκρισης (IIR). Χρησιμοποιούνται συνήθως τα φίλτρα διάταξης FIR επειδή είναι σταθερά και η προκύπτουσα έξοδος είναι γραμμική, οπότε και υπολογιστικά εύκολη να υλοποιηθεί.
Το φίλτρο παίρνει σαν είσοδο ένα σήμα y(n), συνήθως μια "παραμορφωμένη" εκδοχή του επιθυμητού σήματος x(n), και παράγει μια έξοδο x^(n) η οποία είναι μια εκτίμηση του ελάχιστου μέσου τετραγωνικού σφάλματος του x(n). Θεωρώντας σύστημα FIR έχουμε τη σχέση εισόδου – εξόδου του σήματος:

$$\hat{x}(n) = \sum_{k=0}^{M-1} h_y y(n-k) \quad n = 0,1,2,\ldots \quad (3.6.1)$$

όπου n είναι η τιμή διακριτού χρόνου, y() σήμα εισόδου, x^(n) η έξοδος του φίλτρου και h_k ο συντελεστής του φίλτρου. Τέλος, το λάθος εκτίμησης e(n) ορίζεται ως η διαφορά μεταξύ του επιθυμητού σήματος x(n) και της εξόδου του φίλτρου x^(n):

$$e(n) = x(n) - \hat{x}(n) \quad (3.6.2)$$

Στη θεωρία των φίλτρων Wiener, το Μ.Ε.Τ.Σ χρησιμοποιείται ως κριτήριο ελαχιστοποίησης και θεωρείται βέλτιστο για σήματα κατανομής Gauss. Σε φίλτρα FIR, το Μ.Ε.Τ.Σ, οδηγεί σε γραμμικές και κλειστού τύπου λύσεις.

<u>Φίλτρα Wiener με εφαρμογή στη συμπίεση θορύβου</u>

Το σήμα εξόδου $\hat{x}(n)$ προκύπτει από τη συνέλιξη της απόκρισης του σήματος h_k με το σήμα y(n):

$$\hat{x}(n) = h(n) * y(n) \tag{3.6.3}$$

Μεταφέροντας την εξίσωση (3.6.3) στο πεδίο της συχνότητας έχουμε:

$$\hat{X}(n) = H(n)Y(n) \tag{3.6.4}$$

Συνεπώς, το λάθος εκτίμησης στη συχνότητα ορίζεται ως:

$$E(n) = X(n) - \hat{X}(n) = X(n) - H(n)Y(n) \tag{3.6.5}$$

Από τους παραπάνω παράγοντες, θα πρέπει να υπολογίσουμε το H(n), το οποίο θα μας δώσει μια βέλτιστη εξίσωση για το ελαχιστοποιημένο σφάλμα. Το σφάλμα $E\{|e(n)|^2\}$ δίνεται από την εξίσωση: (γίνεται πολλαπλασιασμός με συζυγή)

$$E\{|e(n)|^2\} = E\{e(n)e*(n)\} \dots \tag{3.6.6}$$

Μετά από πράξεις και θέτοντας μεταβλητές όπως Pyy και Pyd και χρησιμοποιώντας πρώτη παράγωγο, προκύπτει η γενική εξίσωση H(n) ως:

$$H(n) = \frac{P_{dy}(n)}{P_{yy}(n)} \tag{3.6.7}$$

Και μετά από άλλους υπολογισμούς καταλήγουμε στο βέλτιστο φίλτρο Wiener για συμπίεση:

$$H(n) = \frac{P_{xx}(n)}{P_{xx}(n) + P_{mm}(n)} \tag{3.6.8}$$

Έπειτα ορίζουμε τη μεταβλητή ξ_k:

$$\xi_k = \frac{P_{xx}(n)}{P_{nn}(n)} \tag{3.6.9}$$

Ο τελικός παράγοντας συμπίεσης προκύπτει μέσω του παραπάνω εκ των προτέρων SNR ως:

$$H(k) = \frac{\xi_k}{1 + \xi_\kappa} \tag{3.6.10}$$

Το συμπέρασμα που προκύπτει για το φίλτρο Wiener είναι ότι *τονίζει μέρη του φάσματος όπου το SNR είναι υψηλό και μειώνει μέρη όπου το SNR είναι χαμηλό.*

3.5 Αλγόριθμος Ephraim – Malah

Σε αυτό τον αλγόριθμο [15] μελετάται ένας βέλτιστος εκτιμητής του STSA, ο οποίος προκύπτει άμεσα από το ενθόρυβο προς μελέτη σήμα. Γίνεται εστίαση ώστε να δημιουργηθεί ένας εκτιμητής MMSE STSA (minimum mean-square error short-time spectral amplitude), και πώς εφαρμόζεται στην ενίσχυση ενός σήματος ομιλίας.
Για να πάρουμε αυτό τον εκτιμητή, πρέπει να είναι γνωστά: η εκ των προτέρων (a priori) πιθανότητα κατανομής των συντελεστών του σήματος και της έκτασης Fourier του θορύβου. Εφόσον είναι άγνωστα, μπορεί να τεθεί σαν υπόθεση ένα λογικό στατιστικό μοντέλο. Στο υπό συζήτηση πρόβλημα, το σήμα ομιλίας και ίσως και του θορύβου δεν είναι ούτε σταθερά ούτε εργοδικά. Από αυτό προκύπτει ότι εύκολα μπορεί να προκύψουν οι παραπάνω κατανομές πιθανότητας τη μακροχρόνια συμπεριφορά κάθε διεργασίας.
Έτσι, χρησιμοποιείται ένα στατιστικό μοντέλο. Το μοντέλο αυτό κάνει χρήση στατιστικών ασυμπτωτικών ιδιοτήτων των συντελεστών έκτασης Fourier.
Ειδικά, θεωρείται ότι οι συντελεστές έκτασης Fourier κάθε διεργασίας μπορούν να μοντελοποιηθούν ως στατιστικά ανεξάρτητες τυχαίες Gaussian μεταβλητές. Η μέση τιμή κάθε συντελεστή θεωρείται μηδενική, αφού οι διεργασίες που λαμβάνουν χώρα θεωρούνται κι αυτές μηδενικής μέσης τιμής. Λόγω της μη σταθερότητας του σήματος ομιλίας, η διακύμανση κάθε συντελεστή ομιλίας της έκτασης Fourier είναι χρονικά μεταβαλλόμενη.

Αυτό το Gaussian στατιστικό μοντέλο προκύπτει από το κεντρικό οριακό θεώρημα, αφού κάθε συντελεστής έκτασης Fourier, είναι ένα «ζυγισμένο» άθροισμα τυχαίων μεταβλητών που προκύπτουν από τα δείγματα της διαδικασίας.

Εδώ, όσον αφορά την περίοδο ανάλυσης, δεν πρέπει να είναι πολύ μεγάλη λόγω της ημι-στατικότητας του σήματος ομιλίας. Όπως και να ‘χει όμως διατηρείται αυτή η υπόθεση ώστε να είναι απλός ο προκύπτον αλγόριθμος.
Πρακτικά, εφαρμόζεται ένα κατάλληλο παράθυρο στο ενθόρυβο σήμα, το οποίο μειώνει τη συσχέτιση μεταξύ φασματικών συντελεστών που έχουν μεγάλη απόσταση, σε βάρος της αύξησης της συσχέτισης μεταξύ γειτονικών φασματικών συντελεστών.

Μετά από μελέτες σημειώνεται ότι, αφού το πραγματικό στατιστικό μοντέλο είναι μη-προσβάσιμο, η εγκυρότητα αυτού του μοντέλου που προτείνεται μπορεί να κριθεί εκ των υστέρων (a posteriori) με βάση τα αποτελέσματα που προκύπτουν εδώ. Επιπρόσθετα, το βέλτιστο των εκτιμητών που προκύπτουν εδώ συνδέεται με το στατιστικό μοντέλο που τέθηκε σαν υπόθεση.

Θέτονται σήματα x(t), d(t) καθαρό και ενθόρυβο. Το υπό μελέτη σήμα ορίζεται ως:

$$y(t) = x(t) + d(t) \qquad 0 \le t \le T \qquad (3.5.1)$$

Έχουμε $X_k \triangleq A_k \exp(j\alpha_k)$, D_k, $Y_k \triangleq R_k \exp(j\theta_k)$ να δηλώνουν τον k-οστό φασματικό συντελεστή του σήματος x(t), του θορύβου d(t) και του ενθόρυβου σήματος y(t) αντίστοιχα για την περίοδο ανάλυσης [0, T].
Y_k και παρόμοια X_k, D_k δίνονται από

$$Y_k = \frac{1}{T}\int_0^T y(t) \exp\left(-j\frac{2\pi}{T}kt\right) dt \qquad (3.5.2)$$

Σκοπός είναι η εκτίμηση του A_k του σήματος με τον θόρυβο. Εκφράζοντας το σήμα με συντελεστές φάσματος Y_k έχουμε

$$y(t) = \lim_{k=-k} Y_k \exp\left(j\frac{2\pi}{T}kt\right) \qquad 0 \le t \le T \tag{3.5.3}$$

Όπου lim είναι όριο μέσης τιμής. Επιπλέον η (3.5.3) συγκλίνει σχεδόν απόλυτα στο y(t) για κάθε t∈ $[0, T]$. Έτσι προκύπτει ότι {y(t), $0 \le t \le T$} και $\{Y_0, Y_1, \ldots\}$ φέρουν την ίδια πληροφορία. Δηλαδή χρειάζεται να υπολογιστεί μόνο το A_k και όχι όλη η σειρά $\{Y_0, Y_1, \ldots\}$. Το A_k προκύπτει ως εξής:

$$\begin{aligned}\hat{A}_k &= E\{A_k | y(t), \quad 0 \le t \le T\} \\ &= E\{A_k | Y_0, Y_1 \ldots\} \\ &= E\{A_k | Y_k\}\end{aligned}$$

$$= \frac{\int_0^\infty \int_0^{2\pi} a_k p(Y_k|a_k, a_k) p(a_k, a_k) da_k da_k}{\int_0^\infty \int_0^{2\pi} p(Y_k|a_k, a_k) p(a_k, a_k) da_k da_k} \tag{3.5.4}$$

όπου Ε{.} είναι ο παράγοντας αναμονής, και p{.} μια συνάρτηση πυκνότητας πιθανότητας.
Μετά από αντικαταστάσεις προκύπτει ότι:

$$\hat{A}_k = \Gamma(1{,}5)\frac{\sqrt{v_k}}{\gamma_k} M(-0.5; 1; -v_k) R_k$$

$$= \Gamma(1{,}5)\frac{\sqrt{v_k}}{\gamma_k} \exp\left(-\frac{v_k}{2}\right)\left[(1 + v_k) I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right)\right] R_k \tag{3.5.5}$$

όπου Γ(.) δηλώνει τη συνάρτηση gamma, I_0 και I_1 δηλώνουν τροποποιημένη συνάρτηση Bessel και v_k ορίζεται ως

$$v_k \triangleq \frac{\xi_k}{1 + \xi_k}\gamma_\kappa \tag{3.5.6}$$

όπου ξ_k και γ_κ τα a priori και a posteriori SNR αντίστοιχα.
Ενδιαφέρον παρουσιάζει να εξεταστεί ο $\hat{A}_k$ σε υψηλό SBR δηλαδή $\xi_k \gg 1$, όπου προκύπτει τελικά:

$$\hat{A}_k \cong \frac{\xi_k}{1 + \xi_k} R_k \qquad high\ SNR$$

$$\triangleq A_k^\omega \tag{3.5.7}$$

Αφού τελικά υπολογίσαμε την εκτίμηση για τον φασματικό συντελεστή $X_k = A_k \exp(j\alpha_k)$ από $\hat{X}_k = \hat{A}_k \exp(j\theta_k)$, όπου $\exp(j\theta_k)$ είναι ο μιγαδικός εκθέτης της φάσης θορύβου, παίρνουμε τελικά μια προσέγγιση για τον k-οστό εκτιμητή για τον συντελεστή φάσματος του σήματος:

$$\hat{X}_k \cong \frac{\xi_k}{1 + \xi_k} Y_k \qquad high\ SNR$$

$$\triangleq X_k^\omega \tag{3.5.8}$$

Ουσιαστικά αυτός ο εκτιμητής είναι ο MMSE εκτιμητής του k-οστού συντελεστή του φάσματος σήματος. Γι' αυτό αναφέρεται και ως εκτιμητής πλάτους Wiener.
Είναι χρήσιμο να πούμε ότι ο εκτιμητής πλάτους της (3.5.5) προκύπτει από το R_k μέσω μιας

συνάρτησης κέρδους:

$$G_{MMSE}(\xi_k, \gamma_\kappa) \triangleq \frac{\hat{A}_k}{R_k} \tag{3.5.9}$$

Από την (3.5.5) βλέπουμε ότι εξαρτάται μόνο από τα a priori και a posteriori SNR ξ_k, γ_κ αντίστοιχα.
Επίσης, γίνεται η παραδοχή ότι ο εκτιμητής έχει μια συγκεκριμένη απόδοση που βασίζεται στον συμβιβασμό μεταξύ του ποια πληροφορία είναι γνωστή (a priori) και τι πληροφορία προκύπτει από τα ενθόρυβα δεδομένα.

3.5.1 MMSE log-spectral amplitude estimator

Για τον αλγόριθμο [16] αυτό χρησιμοποιείται το ίδιο στατιστικό μοντέλο με την παραπάνω μέθοδο.
Θεωρείται $X_k = A_k e^{j\alpha k}$, D_k, $Y_k = R_k e^{j\theta k}$ να δηλώνουν τον k-οστό συντελεστή της έκτασης Fourier για το φάσμα ομιλίας, τον θόρυβο, και την ενθόρυβη ομιλία για την περίοδο ανάλυσης [0,T].
Σύμφωνα με το πρόβλημα που εξετάζεται εδώ, αναζητείται ο εκτιμητής $\hat{A}_k$ ο οποίος ελαχιστοποιεί την ακόλουθη παραμόρφωση:

$$E\{(logA_k - log\hat{A}_k)^2\} \tag{3.5.1.1}$$

με γνωστή την ενθόρυβη ομιλία {y(t), 0≤t≤T}. Για τον εκτιμητή αποδεικνύεται ότι:

$$\hat{A}_k = \exp\{E[lnA_k|y(t), \quad 0 \le t \le T\} \tag{3.5.1.2}$$

και είναι ανεξάρτητο της βάσης που έχει επιλεγεί για την (3.5.1.1).
Όπως αποδείχτηκε παραπάνω, η αναμενόμενη τιμή του A_k με δοσμένο {y(t), 0≤t≤T } ισούται με την αναμενόμενη τιμή A_kμε δοσμένο μόνο το Y_k. Αφού αυτή η συνθήκη παραμένει αληθής όταν το A_k αντικατασταθεί με lnA_k η (3.5.1.2) ισούται με:

$$\hat{A}_k = \exp\{E[lnA_k|Y_k, \quad 0 \le t \le T\} \tag{3.5.1.3}$$

Η παραπάνω εξίσωση μπορεί να εφαρμοστεί και για ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος του λογαριθμικού φάσματος δύναμης το οποίο δίνεται από:

$$E\{(logA^2{}_k - log\hat{A}^2{}_k)^2\} \tag{3.5.1.4}$$

όπου $\tilde{A}^2{}_k$ η εκτίμηση του $A^2{}_k$ ώστε $\hat{A}_k = \sqrt{\tilde{A}^2{}_k}$.

Η αξιολόγηση του $E[lnA_k|Y_k]$ για το μοντέλο Gauss που τίθεται σαν υπόθεση εδώ, γίνεται χρησιμοποιώντας τη στιγμιαία συνάρτηση δημιουργίας του $\ln A_k$ με δοσμένο το Y_k.

Επίσης, δίνεται $Z_k = \ln A_k$, οπότε η παραπάνω συνάρτηση δημιουργίας προκύπτει από:

$$\Phi_{Z_k}|Y_k(\mu) = E\{\exp(\mu Z_k)\,|Y_k\}$$

$$= E\{A^{\mu}{}_k|Y_k\} \quad (3.5.1.5)$$

Το αποτέλεσμα του $E\{lnA_k|Y_k\}$ προκύπτει μέσω του $\Phi_{Z_k}|Y_k(\mu)$ από τη σχέση:

$$E\{lnA_k|Y_k\} = \frac{d}{d\mu}\Phi_{Z_k|Y_k}(\mu)|_{\mu=0} \quad (3.5.1.6)$$

Έτσι, σκοπός είναι να υπολογίσουμε το $\Phi_{Z_k|Y_k}(\mu)$ και μετά να πάρουμε το $E\{lnA_k|Y_k\}$ χρησιμοποιώντας την εξίσωση (3.5.1.6).
Έπειτα από υπολογισμούς και αντικαταστάσεις σε εξισώσεις με βάση το μοντέλο Gauss, το $\Phi_{Z_k|Y_k}(\mu)$ προκύπτει από:

$$\Phi_{Z_k|Y_k}(\mu) = \lambda_k^{\mu/2}\Gamma(\mu/2 + 1)M(-\mu/2\,; 1; -v_k) \quad (3.5.1.7)$$

όπου Γ(.) είναι η συνάρτηση gamma και M(a;c;x) είναι η συνάρτηση υπεργεωμετρικής συνάθροισης.
Το M(a;c;x) προκύπτει:

$$\mathrm{M(a;c;x)} = \sum_{r=0}^{\infty} \frac{\mathrm{(a)r\,x^r}}{\mathrm{(c)r\ r!}} \quad (3.5.1.8)$$

Και έπειτα προκύπτουν και τα παράγωγα του $M(-\mu/2\,; 1; -v_k)$.
Τελικά παίρνουμε τον εκτιμητή πλάτους από τη συνάρτηση

$$\hat{A}_k = \frac{\xi_k}{1 + \xi_k}\exp\left\{\frac{1}{2}\int_{v_k}^{\infty}\frac{e^{-t}}{t}dt\right\}R_k \quad (3.5.1.8)$$

Είναι χρήσιμο να θεωρηθεί ότι το $\hat{A}_k$ προκύπτει μέσω του R_k με μια μη-γραμμική πολλαπλασιαστική συνάρτηση κέρδους η οποία εξαρτάται μόνο από τα a priori και a posteriori SNR ξ_k και γ_k και ορίζεται ως:

$$G(\xi_k, \gamma_\kappa) \triangleq \frac{\hat{A}_k}{R_k} \quad (3.5.1.9)$$

Ενδιαφέρον παρουσιάζει το γεγονός ότι η παραπάνω συνάρτηση κέρδους πάντα δίνει χαμηλότερο κέρδος από αυτό που προκύπτει στην παραπάνω μέθοδο. Αυτό αποδεικνύεται μέσω της ισότητας Jensen:

$$\hat{A}_k = \exp\{E[\ln A_k|Y_k]\} \le exp\{\ln E[A_k|Y_k]\} = E[A_k|Y_k] \quad (3.5.1.10)$$

Συγκρίνοντας τις μεθόδους που προκύπτουν από τις εξισώσεις (3.5.1.8) και (3.5.5) παρατηρείται ότι για την πρώτη, στο τελικό αποθορυβοποιημένο σήμα μένει λιγότερος μουσικός θόρυβος ενώ δεν παρατηρείται καμία αλλαγή στην πληροφορία της ομιλίας. Ωστόσο, σε σχέση με την (3.5.5), ο θόρυβος αυτός δεν ακούγεται τόσο ομοιόμορφος. Παρόλα αυτά, λόγω του χαμηλού επιπέδου του μουσικού θορύβου, το παραπάνω πρόβλημα ομοιομορφίας θεωρείται αμελητέο. Το χαμηλό αυτό επίπεδο μουσικού θορύβου πιθανώς προκύπτει λόγω του χαμηλού κέρδους που προκύπτει από τη συνάρτηση (3.5.1.9).

3.6 Αλγόριθμος Τσουκαλά (Ψυχό-ακουστικό μοντέλο)

Ο τελευταίος αλγόριθμος που εξετάζεται [10], βασίζεται στη χρήση ενός πολύ γνωστού μηχανισμού του συστήματος ακοής, την επικάλυψη θορύβου. Επιπροσθέτως, εισάγονται διεργασίες εκτίμησης που μπορούν να τροποποιήσουν βέλτιστα ή υπό-όρους οι μεταβλητές που προκύπτουν από ψυχο-ακουστικά κριτήρια της συνάρτησης STSA.

Η προσέγγιση για την ενίσχυση που υιοθετήθηκε εδώ βασίζεται στον ορισμό ενός συντελεστή ακουστού ήχου του STSA, ο οποίος εκτείνεται και χρησιμοποιείται στην εξαγωγή ενός βέλτιστου τροποποιητή που πετυχαίνει συμπίεση του ακουστού θορύβου. Ακόμη, αυτή η τροποποίηση επιλεκτικά επηρεάζει τις -αντιληπτικά σημαντικές- φασματικές τιμές και γι' αυτό είναι πιο ισχυρή από μεθόδους οι οποίες επηρεάζουν το συνολικό STSA και λιγότερο επιρρεπής στην εισαγωγή μη επιθυμητών παραμορφώσεων.

Θεωρούμε τα φάσματα ισχύος του καθαρού σήματος ομιλίας και του ενθόρυβου:

$$Y_p(k,i) = |Y_w(k,i)|^2, \qquad 0 \le k \le K-1 \tag{3.5.1}$$

$$X_p(k,i) = |X_w(k,i)|^2, \qquad 0 \le k \le K-1 \tag{3.5.2}$$

Με βάση το απόλυτο κατώφλι επικάλυψης υπολογίζονται οι συντελεστές που βρίσκονται πάνω από αυτό και είναι αυτοί που πρέπει να αφαιρεθούν:

$$A_v(k,i) = max\{Y_p(k,i), T(k,i)\} = \begin{cases} Y_p(k,i), & \alpha\nu\ Y_p(k,i) \ge T(k,i) \\ T(k,i), & \alpha\nu\ Y_p(k,i) < T(k,i) \end{cases} \quad 0 \le k \le K-1 \tag{3.5.3}$$

$$A_\chi(k,i) = max\{X_p(k,i), T(k,i)\} = \begin{cases} X_p(k,i), \alpha\nu\ X_p(k,i) \ge T(k,i) \\ T(k,i), \alpha\nu\ X_p(k,i) < T(k,i) \end{cases} \quad 0 \le k \le K-1 \tag{3.5.4}$$

Συνεπώς, μπορούμε να ορίσουμε το Ακουστό Φάσμα προσθετικού θορύβου $A_d(k,i)$ μέσω των φασματικών συνιστωσών που εκλαμβάνονται ως θόρυβος από τον μηχανισμό ακοής μπορούν προσδιοριζόμενο από τη σχέση:

$$A_d(k,i) = A_v(k,i) - A_\chi(k,i), \quad 0 \le k \le N-1 \tag{3.5.5}$$

δηλαδή τη διαφορά μεταξύ των φασμάτων καθαρής ομιλίας και ενθόρυβης.

Πιο αναλυτικά ο Ακουστός θόρυβος προκύπτει από αντικατάσταση των (x.x) και (x.x) στην (3.5.5):

$$A_d(k,i) = \begin{cases} Y_p(k,i) - X_p(k,i), \alpha\nu\ Y_p(k,i) \ge T(k,i) \kappa\alpha\iota X_p(k,i) \ge T(k,i)\ I \\ Y_p(k,i) - T(k,i), \alpha\nu\ Y_p(k,i) \ge T(k,i) \kappa\alpha\iota X_p(k,i) < T(k,i)\ II \\ T(k,i) - X_p(k,i), \alpha\nu\ Y_p(k,i) < T(k,i) \kappa\alpha\iota X_p(k,i) \ge T(k,i)\ III \\ 0, \qquad \alpha\nu\ Y_p(k,i) < T(k,i) \kappa\alpha\iota X_p(k,i) < T(k,i)\ IV \end{cases}, 0 \le k \le K-1$$

Από τις παραπάνω τέσσερις περιπτώσεις μόνο οι δύο πρώτες πρέπει να επεξεργαστούν ώστε να καταστήσουν τον θόρυβο μη-ακουστό. Οι υπόλοιπες δύο δεν θα επηρεαστούν από τη φασματική τροποποίηση, αφού δεν συνεισφέρουν στα στοιχεία του Ακουστού Θορύβου.

Κύριος στόχος του αλγορίθμου είναι να κάνει το φάσμα του Ακουστού Θορύβου μικρότερο του

μηδενός ή ίσο με το μηδέν. Αυτό θα επιτευχθεί με κατάλληλη τροποποίηση του φάσματος του ενθόρυβου σήματος Yp(k,i). Έτσι, ο υπολειμματικός θόρυβος θα γίνει μη-ακουστός.
Συνεπώς, αν το φάσμα ισχύος του σήματος θορύβου Yp(k,i) τροποποιηθεί ώστε να παράγει το ενισχυμένο φάσμα ομιλίας, αναφερόμενο ως $\hat{X}_p(k,i)$, τότε το φάσμα Ακουστού Θορύβου, αναφερόμενο ως $\hat{A}_d(k,i)$, θα πρέπει να ικανοποιεί τη συνθήκη:

$$\hat{A}_d(k,i) \leq 0 \quad 0 \leq k \leq K-1 \tag{3.5.6}$$

Σε αυτό τον αλγόριθμο χρησιμοποιείται μια συνάρτηση κέρδους τύπου Wiener. Πιο συγκεκριμένα, έχουμε μια ευέλικτη συνάρτηση κέρδους για τις μεταβλητές α(k,i) και ν(k,i):

$$\hat{X}_p(k,i) = \frac{Y_p^{\nu(k,i)}(k,i)}{\alpha^{\nu(k,i)}(k,i) + Y_p^{\nu(k,i)}(k,i)} Y_p(k,i) \tag{3.5.7}$$

όπου α(k,i) και ν(k,i) παράμετροι που αλλάζουν στον χρόνο και στην συχνότητα. Θεωρούνται πάντα θετικές.
Για την εκτίμηση της παραμέτρου α(k,i) ορίζονται δύο λύσεις:

i. βασισμένη στην εκτίμηση του φασματικού ελαχίστου μέσα σε κάθε κρίσιμη ζώνη
ii. βασισμένη στην εκτίμηση του κατωφλίου ακουστικής επικάλυψης

3.7 VAD (voice activity detector)

Στις περισσότερες εφαρμογές συμπίεσης θορύβου χρησιμοποιείται ένας “ανιχνευτής δραστηριότητας ομιλίας”. Είναι αλγόριθμος απόφασης για τον εντοπισμό παρουσίας ομιλίας σε ένα σύνθετο σήμα που αποτελείται από θόρυβο και ομιλία. Η σύνθεσή του μπορεί να είναι από απλή έως αρκετά περίπλοκη.

4. Φασματική Αφαίρεση

4.1 Βασικές Αρχές

Στη βιβλιογραφία [17] [18] [19] [20] [21] αναφέρονται πολλοί μέθοδοι αφαίρεσης προσθετικού θορύβου. Η πιο διαδεδομένη και μελετημένη είναι αυτή της Φασματικής Αφαίρεσης. (οι αλγόριθμοι που αναφέρθηκαν στην αρχή βασίζονται κατά ένα κομμάτι στη φασματική αφαίρεση). Βασίζεται σε μία γενική θεωρία – αρχή – και από εκεί και πέρα υπάρχουν διάφορες παραλλαγές της.
Δουλεύοντας στο πεδίο της συχνότητας και εφαρμόζοντας τον STFT (έχει μελετηθεί), θα πάρουμε μία εκτίμηση του φάσματος του θορύβου από αυτό του ενθόρυβου σήματος κατά τα διαστήματα απουσίας ομιλίας. Στη συνέχεια αφαιρούνται τα δύο φάσματα (του θορύβου από το ενθόρυβο σήμα), και τέλος με αντίστροφο Fourier παίρνουμε το τελικό βελτιωμένο σήμα.
Εδώ να σημειώσουμε ότι έχει γίνει η εξής υπόθεση: ο προσθετικός θόρυβος που μελετάται θεωρείται στατικός.

Παράλληλα, θα πρέπει να δοθεί προσοχή στην ποσότητα του θορύβου που αφαιρείται. Σε περίπτωση πολύ μεγάλης αφαίρεσης (υπέρ-αφαίρεση), μπορεί να έχουμε απώλεια χρήσιμης πληροφορίας/ομιλίας. Ενώ, σε περίπτωση μικρής αφαίρεσης (υπό-αφαίρεσης), μπορεί να μείνει παρεμβατικός θόρυβος.

Το ενθόρυβο σήμα y(n) αποτελείται από το καθαρό σήμα x(n) και το σήμα θορύβου d(n)

$$y(n) = x(n) + d(n) \tag{4.1.1.1}$$

Εφαρμόζοντας STFT στις δύο πλευρές της εξίσωσης (x.x.x) θα έχουμε

$$Y[n] = X[n] + D[n] \tag{4.1.1.2}$$

Η παραπάνω εξίσωση μας δίνει το πεδίο της συχνότητας. Από την παραπάνω εξίσωση θα προκύψουν και τα μέτρα των μεγεθών |Y[n]|, |X[n]| και |D[n]| αντίστοιχα. Βέβαια, το μέτρο του θορύβου είναι άγνωστο, αλλά μπορεί να προσδιοριστεί παίρνοντας τη μέση τιμή του, η οποία από τα διαστήματα απουσίας ομιλίας.
Επίσης, μπορούμε να προσδιορίσουμε και το φάσμα δύναμης για την παραπάνω εξίσωση. Αυτό προκύπτει προσεγγιστικά ως:

$$|Y[n]|^2 \approx |X[n]|^2 + |D[n]|^2 \tag{4.1.1.3}$$

Σύμφωνα με την γενική αρχή της φασματικής αφαίρεσης, προκύπτει η εξίσωση για την εκτίμηση του τελικού φάσματος του σήματος

$$\left|\hat{X}[n]\right| = |Y[n]| - \left|\hat{D}[n]\right| \tag{4.1.1.4}$$

Επίσης, έχουμε και το φάσμα ισχύος που προκύπτει ως:

$$|\hat{X}[n]|^2 = |\hat{Y}[n]|^2 - |\hat{D}[n]|^2 \quad (4.1.1.5)$$

Όπου $\hat{D}[\ ]$ είναι η εκτίμηση του φάσματος του θορύβου.
Τέλος με τον αντίστροφο Fourier του $\hat{X}$[n] θα πάρω το ζητούμενο βελτιωμένο, και μερικώς απαλλαγμένο από θόρυβο σήμα.
Στη διάρκεια της διαδικασίας, εφαρμόζεται μια εξίσωση φιλτραρίσματος που μας δίνει τον παράγοντα συμπίεσης.
Η εξίσωση του φάσματος ισχύος μπορεί να γραφεί ως

$$|\hat{X}[n]|^2 = H^2[n]|Y(n)|^2 \quad (4.1.1.6)$$

όπου

$$H(n) = \sqrt{1 - \frac{|\hat{D}[n]|^2}{|Y[n]|^2}} \quad (4.1.1.7)$$

Το H(n) ως είθισται ορίζει τη συνάρτηση μεταφοράς. Στη φασματική αφαίρεση όμως ορίζει τη συνάρτηση καταστολής. Είναι πραγματικός αριθμός και πρέπει να είναι θετικός.
Έτσι, για κάθε αλγόριθμο συμπίεσης, ορίζεται και μία τέτοια συνάρτηση. Επίσης, κατά τη σύγκριση των μεθόδων συμπίεσης, είναι ο κύριος παράγοντας σύγκρισης.

4.1.1 Εκτίμηση Θορύβου

Ένα κρίσιμο σημείο κατά τη διαδικασία αποθορυβοποίησης είναι η εκτίμηση του στατικού θορύβου. Αυτό επιτυγχάνεται με επανεκτίμηση του short-time φάσματος θορύβου καθώς περνάει ο χρόνος. Αποτελείται από δύο διαδοχικά στάδια:
α)στα πρώτο προς ανάλυση frames υποθέτουμε ότι υπάρχει μόνο θόρυβος, οπότε υπολογίζουμε "τυφλά" μία εκτίμηση του θορύβου εντός αυτών των frames
β)για κάθε frame ελέγχουμε αν περιλαμβάνει μόνο θόρυβο (μέσω VAD) και μόνο τότε γίνεται επανεκτίμηση του θορύβου.

4.1.2 Artifacts & Distortions

Σε ιδανικές συνθήκες η συμπίεση θορύβου και γενικά η επεξεργασία σημάτων είναι εύκολη υπόθεση.
Στην πραγματικότητα, στους αλγόριθμους συμπίεσης, και κυρίως στην εφαρμογή της μεθόδου *Φασματική Αφαίρεση*, μπορεί να προκύψουν παραμορφώσεις και αλλοιώσεις κατά τη διαδικασία, και το τελικό σήμα να χαρακτηρίζεται από ηχητικά προβλήματα.
Ένα πρόβλημα που μπορεί να προκύψει είναι οι αρνητικές τιμές για την εξίσωση $H_k^{(n)}$ που ορίζει τη συμπίεση, αν έχουμε τιμές του $|Y_k^{(n)}|^2$ μικρότερες από τη διακύμανση του θορύβου. Αυτό συνεπώς οδηγεί και σε αρνητικές τιμές του τελικού φάσματος. Κάτι τέτοιο συνήθως αντιμετωπίζεται με το να θέσουμε τιμές της εξίσωσης συμπίεσης που δίνουν αρνητικό αποτέλεσμα ίσες με το μηδέν (διόρθωση half-wave). Με τη σειρά του όμως αυτό δημιουργεί άλλη μια ανεπιθύμητη κατάσταση. Απομονωμένες περιοχές στο φάσμα που κάπως μπλέκονται με τα frames και στον ακροατή ακούγονται έντονα και ενοχλητικά. Τα artifacts προκύπτουν σε κομμάτια του σήματος όπου δεν υπάρχει ομιλία δημιουργώντας τον *μουσικό θόρυβο* (musical noise).

Μουσικός Θόρυβος

Όπως αναφέρθηκε στην ενότητα Artifacts & Distortions, ένα από τα μειονεκτήματα της φασματικής αφαίρεσης είναι ο *μουσικός θόρυβο*. Αποτελεί κρίσιμο σημείο και η καταπολέμησή του είναι υψίστης σημασίας. Έχουν αναπτυχθεί αρκετές μέθοδοι ελάττωσης του. Εξίσου σημαντικό θεωρείται να δούμε και τους παράγοντες που συμβάλουν στην εμφάνισή του. (...)
Ο μουσικός θόρυβος δεν είναι απλός στην αφαίρεσή του. Είναι μια σημαντική λεπτομέρεια και σίγουρα υπάρχουν trade off μεταξύ της μείωσής του και της ποιότητας της ομιλίας στην τελική μορφή της.

Μη – γραμμική Φασματική Αφαίρεση

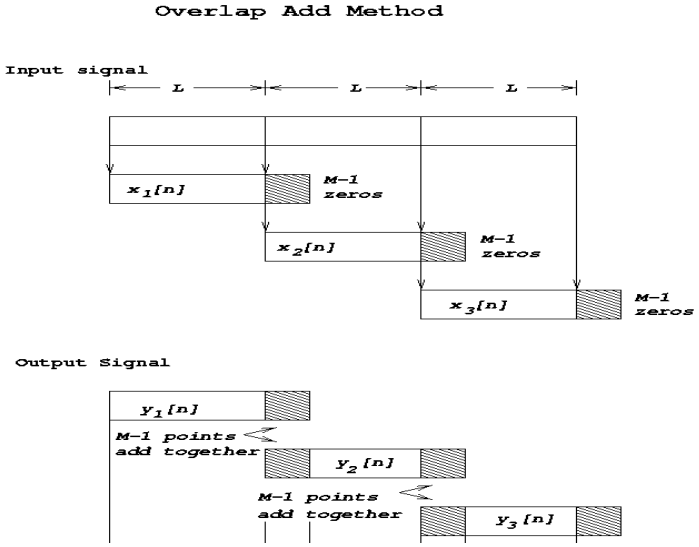
Για κάποιους θορύβους του πραγματικού κόσμου παρατηρούμε ότι επηρεάζουν το φάσμα θορύβου διαφορετικά σε χαμηλές και υψηλές συχνότητες. Γι' αυτό, εισάγεται ένας παράγοντας αφαίρεσης κατάλληλα μελετημένος, ώστε να διαχειρίζεται σωστά διαφορετικούς τύπους θορύβου.
Η Μη-γραμμική φασματική αφαίρεση ορίζεται ως μια *τροποποίηση της μεθόδου του* **Berouti**, *κάνοντας τον παράγοντα υπεραφαίρεσης εξαρτώμενο από την συχνότητα και την αφαιρετική διαδικασία μη-γραμμική. Μεγαλύτερες τιμές αφαιρούνται στις συχνότητες με χαμηλά* ***SNR*** *και μικρότερες στις συχνότητες με υψηλά* ***SNR.***

Αλγόριθμος Επικάλυψης – Overlapadd

Κατά τη διάρκεια της επεξεργασίας σημάτων σε real-time, αν δεν απαιτείται να επαναφέρουμε το σήμα ομιλίας στον τομέα του χρόνου (reconstruct), τότε η διαδικασία θα θεωρούνταν ολοκληρωμένη. Επειδή όμως δεν έχει νόημα να παραληφθεί το βήμα αυτό, απαιτείται ακόμη μία ενέργεια, η “επικάλυψη”.
Υπάρχουν δύο μέθοδοι. Η πιο διαδεδομένη όμως είναι η “Επικάλυψης – Πρόσθεσης” (Overlap-add).
Το πρόβλημα που προκύπτει και πρέπει να γίνει αυτή η δουλειά είναι το εξής: κατά την επαναφορά του επεξεργασμένου σήματος στον τομέα του χρόνου προκύπτουν μη-σωστές συνδέσεις (ή ασυνέχειες) στο “τέλος” των γειτονικών frames (δηλαδή έχουν ίδιες ή παρόμοιες τιμές). Έτσι δημιουργείται κάποιος ενοχλητικός ήχος και φυσικά γίνεται αντιληπτός στο ανθρώπινο αυτί.
Η διαδικασία είναι:
Έχουμε το μήκος των παραθύρων και κατόπιν το καθένα συμπληρώνεται με M-1 μηδενικά.
Έχουμε τους DFT φίλτρου και μπλοκ δεδομένων.
Πολλαπλασιάζονται.
Εφαρμόζεται IDFT κατά τον οποίο τα μπλοκ δεδομένων είναι απελευθερωμένα από aliasing.
“Μιας και κάθε μπλοκ δεδομένων συμπληρώνεται με M-1 μηδενικά, τα τελευταία M-1 δείγματα από κάθε μπλοκ πρέπει να επικαλυφθούν και να προστεθούν στα M-1 αρχικά δείγματα του επόμενου παραθύρου. Γι' αυτό και η μέθοδος πήρε το όνομα Επικάλυψη-Πρόσθεσης, και έχει σαν αποτέλεσμα μια ακολουθία της μορφής:

y(n) = {y1(0), y1(1),..., y1(L-1), y1(L) + y2(0), y1(L+1) + y2(1),..., y1(N-1) + y2(M-1), y2(M),...}”

Στο σχήμα της Εικόνας 14 παρουσιάζεται ο αλγόριθμος overlap-add με δεδομένα εισόδου-εξόδου:



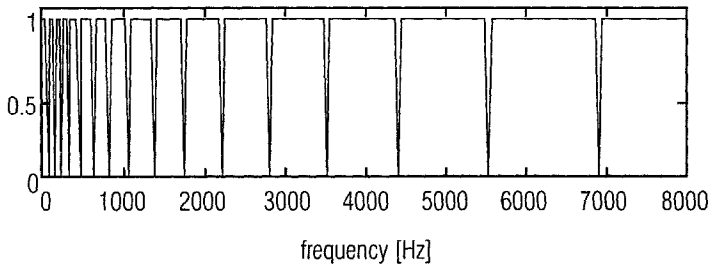
Εικόνα 14:Αλγόριθμος Επικάλυψης – Overlapadd

4.2 Φίλτρο Εξομάλυνσης Παραμετροποιημένο με Αντιληπτικά Κριτήρια

Στην παρούσα προτεινόμενη τεχνική συμπίεσης προσθετικού θορύβου [21], αναφερόμαστε στον μουσικό θόρυβο και θεωρούμε την εξομάλυνση παραμετροποιημένη σύμφωνα με τις ιδιότητες της ανθρώπινης αντίληψης. Μέσα από μελέτη προέκυψε ότι ο μουσικός θόρυβος οφείλεται, κυρίως, σε χρονικές και συχνοτικές διακυμάνσεις κατά την εκτίμησή του, κάτι το οποίο έχει άμεσο αντίκτυπο στο φίλτρο κέρδους. Κατά συνέπεια, εφαρμόζουμε σε αυτό το φίλτρο κέρδους εξομάλυνση τόσο στο πεδίο του χρόνου όσο και στης συχνότητας.
Συγκεκριμένα, εδώ εστιάζουμε σε ομαδοποιημένα μη-επικαλυπτόμενα κομμάτια με εύρος-ζώνης τέτοιο ώστε να μιμούνται τη συχνοτική ανάλυση του ανθρώπινου ακουστικού συστήματος.
Επιπλέον, η μελέτη γίνεται μόνο για την κεντρική συχνότητα των κομματιών.

-Εξομάλυνση Συχνότητας

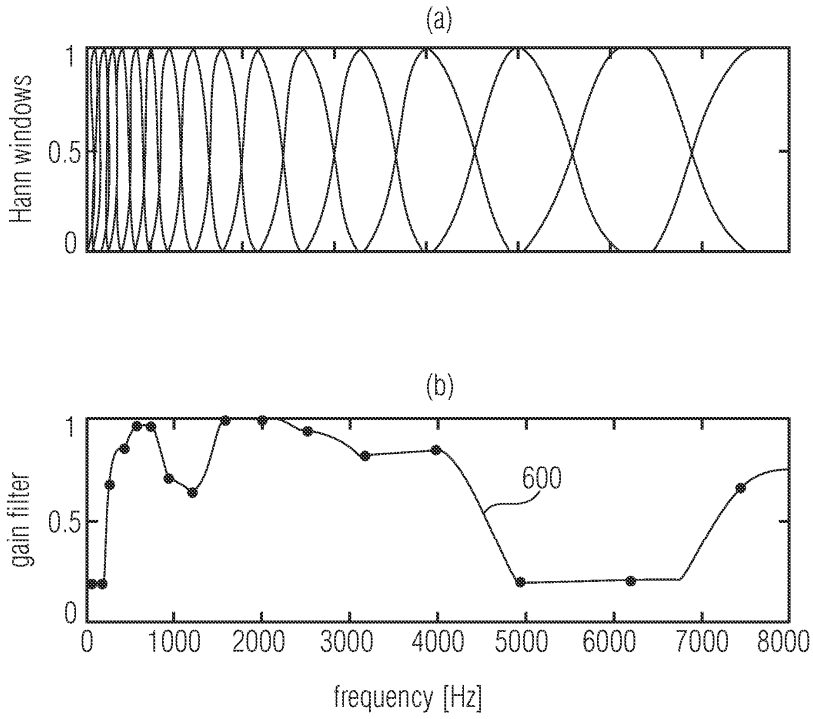
Όπως αναφέρθηκε παραπάνω, οι συντελεστές του φάσματος ενθόρυβου σήματος Y[k,m] ομαδοποιούνται σε μη-επικαλυπτόμενα κομμάτια με εύρος-ζώνης τέτοιο ώστε μιμούνται τη συχνοτική ανάλυση του ανθρώπινου ακουστικού συστήματος.
Ενδεικτικά για ένα δείγμα 16kHz, χρησιμοποιείται ένα μήκος DFT 512 και 15 κομμάτια με εύρος-ζώνης το καθένα, προσεγγιστικά, όσο δύο φορές το ισοδύναμο τετράγωνο εύρος-ζώνης (ERB). Στο παρακάτω γράφημα φαίνονται οι σχετικές συχνότητες.



Για το φίλτρο κέρδους με εξίσωση:

$$G_{PL}[k,m] = \max\left(10^{-\frac{L}{20}}, G_P\right) \tag{4.2.1}$$

υπολογίζεται μόνο για τις κεντρικές συχνότητες κάθε κομματιού, κάτι που οδηγεί σε μικρότερη υπολογιστική πολυπλοκότητα. Προγενέστερα αυτής της εφαρμογής στο φάσμα ομοιόμορφου σήματος STFT, γίνεται παρεμβολή με τη χρήση φίλτρων παρεμβολής Hann.



Η τεχνική της εξομάλυνσης συχνότητας οδηγεί σε πιο ήπιες διακυμάνσεις του φάσματος που προκύπτει σαν συνάρτηση της συχνότητας και γι' αυτό μειώνει τόσο τον μουσικό θόρυβο όσο και άλλες παραμορφώσεις.

-Εξομάλυνση Χρόνου

Η χρονική εξομάλυνση ελέγχεται σε συσχέτιση με το κατώφλι επικάλυψης. Όσο χαμηλό το κατώφλι επικάλυψης, τόσο μεγαλύτερη είναι η σταθερά-χρόνου εξομάλυνσης που επιλέγεται για τη

μείωση των παραμορφώσεων. Λειτουργεί ικανοποιητικά επειδή, τυπικά, το κατώφλι επικάλυψης είναι χαμηλό για σταθερά και τονικά σήματα. Αντιστρόφως, όσο μεγαλύτερο είναι αυτό το κατώφλι, τόσο μικρότερη είναι η σταθερά-χρόνου εξομάλυνσης, έτσι ώστε το φίλτρο κέρδους (gain) να προσαρμόζεται αρκετά γρήγορα ώστε να ακολουθεί το σήμα.
Τεχνικές που έχουν αναπτυχθεί σε υπάρχοντες αλγορίθμους (π.χ. Τσουκαλάς), χρησιμοποιούν το κατώφλι επικάλυψης για να προσαρμόσουν την ισχύ της συμπίεσης του ήχου έτσι ώστε *αν το κατώφλι επικάλυψης είναι υψηλό, ο θόρυβος που απομένει θα επικαλυφθεί φυσικά και θα γίνει μη-ακουστός. Ως εκ τούτου δεν χρειάζεται να μειωθεί προκειμένου να μείνουν οι παραμορφώσεις όσο δυνατόν χαμηλότερα. Ωστόσο, αν το κατώφλι επικάλυψης είναι χαμηλό, ο θόρυβος που μένει θα είναι ενοχλητικός στο ανθρώπινο αυτί, οπότε πρέπει να μειωθεί*.

Έτσι, με το να ελέγξουμε τη χρονική εξομάλυνση έναντι του ποσού του θορύβου που αφαιρείται σαν συνάρτηση του κατωφλίου επικάλυψης, το επιθυμητό αποτέλεσμα εισαγωγής ποσού παραμόρφωσης ακολουθώντας το κατώφλι, επιτυγχάνεται.

Φίλτρο κέρδους με συνάρτηση:

$$G_P[k,m] = \left(\frac{|Y[k,m]|^\alpha - \beta|\hat{V}[k,m]|^\alpha}{|Y[k,m]|^\alpha}\right)^{1/\alpha} \tag{4.2.2}$$

εξομαλύνεται χρονικά χρησιμοποιώντας μεσοποίηση μονού-πόλου:

$$\bar{G}_P[k,b] = \mu G_P[k,b] + (1-\mu)\bar{G}_P[k-1,b] \tag{4.2.3}$$

όπου ο θετικός αριθμός μ καθορίζει τον βαθμό εξομάλυνσης στο χρόνο. Η χρονική σταθερά της εκθετικής μείωσης σε sec είναι περίπου:

$$T_\mu = \frac{1}{\mu f_s} \tag{4.2.4}$$

Η σταθερά χρόνου και ισοδύναμα το μ, για κάθε κομμάτι ελέγχεται σε σχέση με το κατώφλι επικάλυψης.
Όσο πιο σταθερό και τονικό είναι ένα σήμα, τόσο λιγότερη επικάλυψη προκύπτει. Ενώ, όσο πιο in-stationary ή μεταβατικό (transient), τόσο περισσότερη επικάλυψη προκύπτει. Γι' αυτό, προκειμένου να πάρουμε μια προσεγγιστική εκτίμηση για το ποσοστό επικάλυψης, υπολογίζεται ένα *μέτρο διακύμανσης* για κάθε κομμάτι:

$$\Phi[k,b] = \frac{\sigma_y^2[k,b] - \sigma_y^2[k-1,b]}{\max\left(\sigma_y^2[k,b] - \sigma_y^2[k-1,b]\right)} \tag{4.2.5}$$

όπου

$$\sigma_y^2[k,b] = E\{Y[k,m]Y*[k,m]\} \tag{4.2.6}$$

και E{…} αντιπροσωπεύει τον βραχύχρονο μέσο όρο. Το μέτρο διακύμανσης Φ[k,b], έχει εύρος από [-1, 1]. Για διαφορετικούς τύπους συνιστωσών του σήματος, επιλέγονται διαφορετικές χρονικές σταθερές, που προκύπτουν από την επικάλυψη και την ανάγκη προσαρμογής του φίλτρου τόσο γρήγορα ώστε να ακολουθεί το σήμα.

5. Πειράματα

5.1 Πειραματική διαδικασία

Στο κεφάλαιο αυτό πραγματοποιείται υλοποίησης των πειραμάτων και έπειτα γίνεται σύγκριση των αλγορίθμων που αναλύθηκαν στο κεφάλαιο 3.
Η υλοποίηση έλαβε χώρα σε προγραμματιστικό περιβάλλον Matlab [Mathworks (2015)] για την παρούσα πτυχιακή εργασία.
Η διαδικασία αξιολόγησης πραγματοποιείται μέσω του MUSHRA test (windows standalone). Εισάγεται το σήμα αναφοράς καθώς και τα αποθορυβοποιημένα σήματα που προκύπτουν μετά από εφαρμογή του κάθε αλγόριθμου αποθορυβοποίησης στο ενθόρυβο σήμα.
Έλαβαν μέρος 10 άτομα ηλικίας 22-52. Η πειραματική διαδικασία διεξήχθη σε περιβάλλον με την ελάχιστη επιρροή εξωτερικού θορύβου.
Για τα ηχητικά σήματα χρησιμοποιήθηκε ρυθμός δειγματοληψίας 48kHz και για σήματα ήχου 16kHz πραγματοποιήθηκε επανα-δειγματοληψία στα 48kHz μέσω υψηλής ποιότητας εργαλείου επαναδειγματοληψίας.

Ο ακροατής / αξιολογητής καλείται να ακούσει όλα τα ηχητικά σήματα και να δώσει μια αξιολόγηση σύμφωνα με τα προσωπικά του κριτήρια. Η αξιολόγηση γίνεται σε σχέση με το σήμα αναφοράς (χωρίς θόρυβο). Η κλίμακα αξιολόγησης κυμαίνεται από 0 (κακό) έως 100 (πολύ καλό). Ανάλογα δηλαδή πόσο καλή αποθορυβοποίηση επιτεύχθηκε από τον εκάστοτε αλγόριθμο. Στα προς μελέτη ενθόρυβα σήματα συμπεριλαμβάνεται και το σήμα αναφοράς για να εξασφαλίσει την εγκυρότητα κάθε τεστ.

5.1.1 Προετοιμασία

Στα πλαίσια της εργασίας χρησιμοποιήθηκαν αρχεία καθαρής ομιλίας (wav), δημιουργώντας μια βάση δεδομένων με 25 άνδρες και 25 γυναίκες. Τα αρχεία ομιλίας που χρησιμοποιήθηκαν είναι από τη συλλογή διαθέσιμων αρχείων EUSTACE της ηλεκτρονικής σελίδας «The Center for Speech Technology Research» του Πανεπιστημίου του Εδιμβούργου.
Τροποποιώντας τα κατάλληλα, υπάρχει περίοδος ησυχίας (απουσία ομιλίας) στην αρχή του σήματος, για να αναγνωρισθεί το μοτίβο του θορύβου. Ο θόρυβος προκύπτει τεχνητά από ειδική συνάρτηση(SNR_set) προσαρμόζοντας ανάλογα τις μεταβλητές.
Για τη σύνθεση των αρχείων ομιλίας και θορύβου έγινε κατάλληλη προσαρμογή του μήκους τους.
Για την καλύτερη και αποτελεσματικότερη αξιολόγηση των αλγορίθμων συγκρίνονται σήματα ανδρικής και γυναικείας ομιλίας με υψηλό και χαμηλό SNR (λόγος σήματος προς θόρυβο). Αυτό γίνεται ώστε να δούμε πώς λειτουργούν οι αλγόριθμοι σε γυναικείες και ανδρικές συχνότητες, τόσο σε χαμηλό SNR όσο και σε υψηλό.

Στην παρούσα εργασία οι αλγόριθμοι που χρησιμοποιήθηκαν ως σύγκριση είναι οι: Boll, Berouti, Multiband, Wiener, Ephraim.
Επίσης, παρουσιάζονται τα διαγράμματα του σήματος αναφοράς, του ενθόρυβου σήματος αλλά και του σήματος αποθορυβοποίησης σε μια προσπάθεια αντικειμενικής αξιολόγησης και παρατήρησης.
Όπως έχει προαναφερθεί, η υποκειμενική αξιολόγηση γίνεται μέσω του MUSHRA test, το οποίο παρουσιάζεται στο κομμάτι της αξιολόγησης.

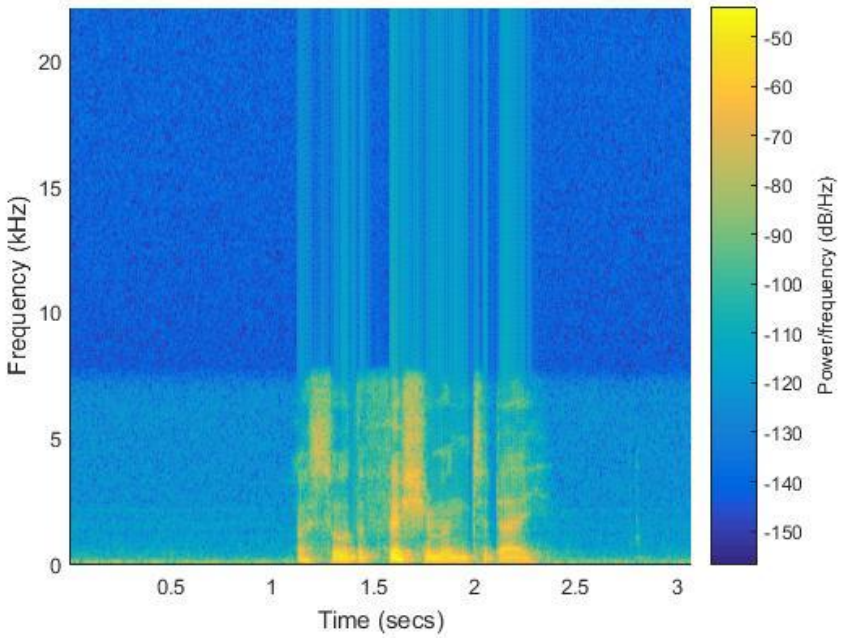
5.1.2 Εφαρμογή αλγορίθμων

Σε προγραμματιστικό περιβάλλον Matlab χρησιμοποιήθηκε το αρχείο experiments_batch.m για τη δημιουργία και την αποθήκευση των σημάτων που θα χρησιμοποιηθούν. Αρχικά, διαβάζεται το σήμα αναφοράς μέσα από τον φάκελο “results” και στη συνέχεια δημιουργείται το ενθόρυβο (μέσω της συνάρτησης SNR_set) αποθηκεύοντάς το επίσης στον ίδιο φάκελο. Έπειτα, στα ενθόρυβα σήματα, εφαρμόζονται οι αλγόριθμοι αποθορυβοποίησης και τέλος αποθηκεύονται τα νέα αρχεία που προκύπτουν.
Ως προετοιμασία για την αξιολόγηση δημιουργούνται φάκελοι για κάθε άνδρα και γυναίκα, στους οποίους περιέχονται υποφάκελοι με τους αλγόριθμους αποθορυβοποίησης για τις τιμές των SNR 0, 5, 10 ,15, 20 dB.

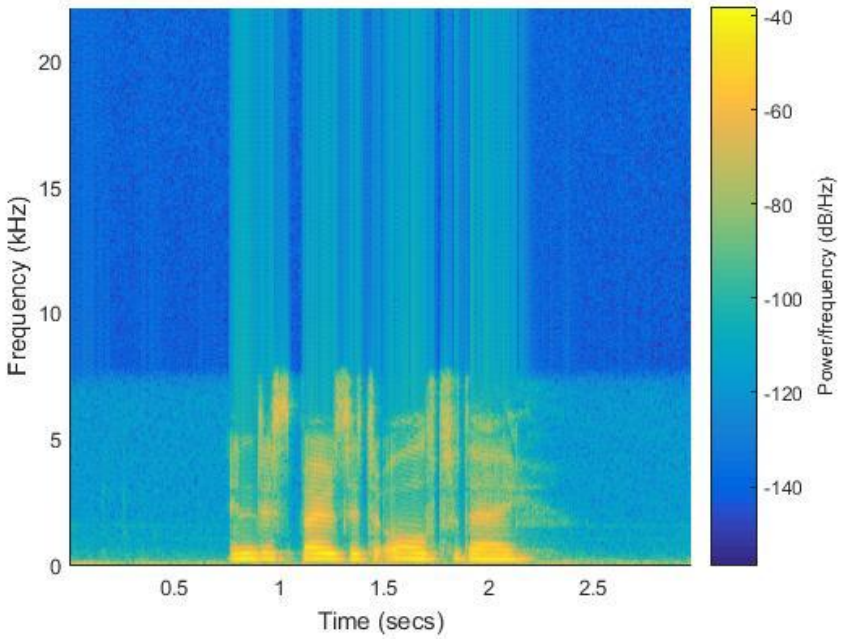
Κατά την προσθήκη θορύβου στο καθαρό σήμα με τις διαφορετικές τιμές SNR κάθε φορά, παρατηρούμε τις αλλαγές στο φάσμα τόσο στο ανδρικό δείγμα ομιλίας όσο και στο γυναικείο.
Στο αποτέλεσμα της αποθορυβοποίησης παρατηρούμε ότι ο θόρυβος σε ορισμένες συχνότητες συμπιέστηκε σε πολύ μεγάλο βαθμό, ενώ σε κάποιες άλλες όχι. Παράλληλα παρατηρήθηκε απώλεια χρήσιμης πληροφορίας σε ορισμένες περιπτώσεις, κάτι που είναι ανεπιθύμητο.

5.1.3 Αποτελέσματα – Αντικειμενικές παρατηρήσεις

Σε αυτό το κομμάτι παρουσιάζονται, ενδεικτικά, αποτελέσματα για ένα σήματα ανδρικής ομιλίας και ένα γυναικείας. Γίνεται χρήση γραφημάτων με απεικόνιση του φάσματος των σημάτων. Αρχικά έχουμε το σήμα αναφοράς το οποίο είναι η καθαρή πληροφορία του σήματος.

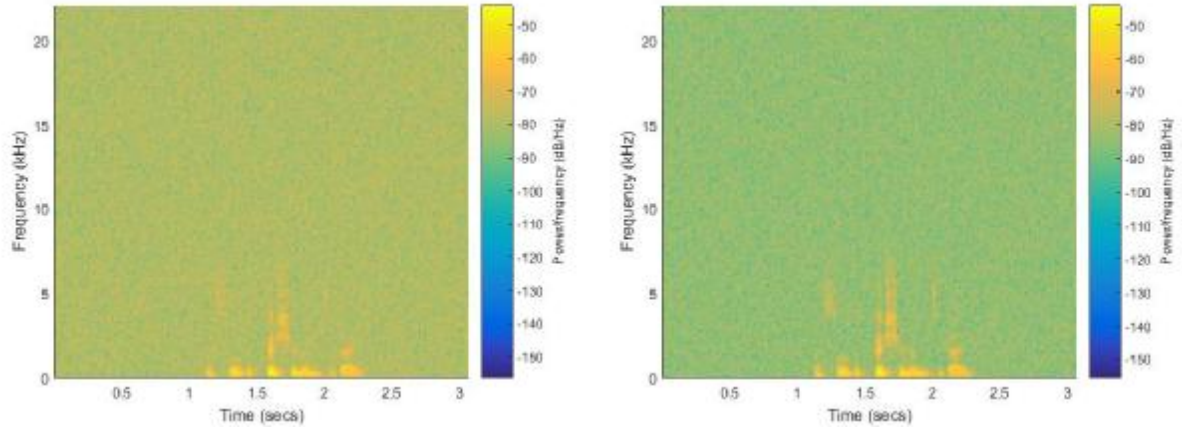


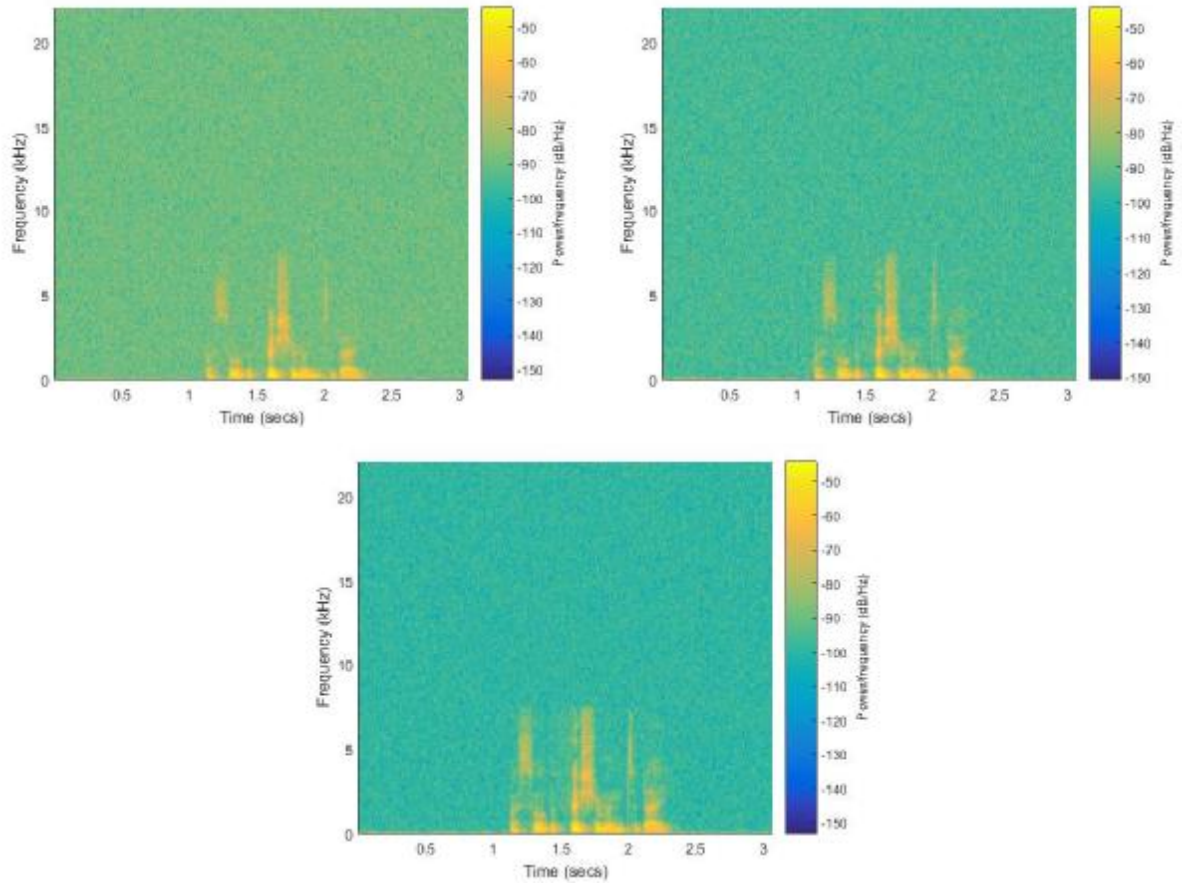
Εικόνα 15: man01-reference signal



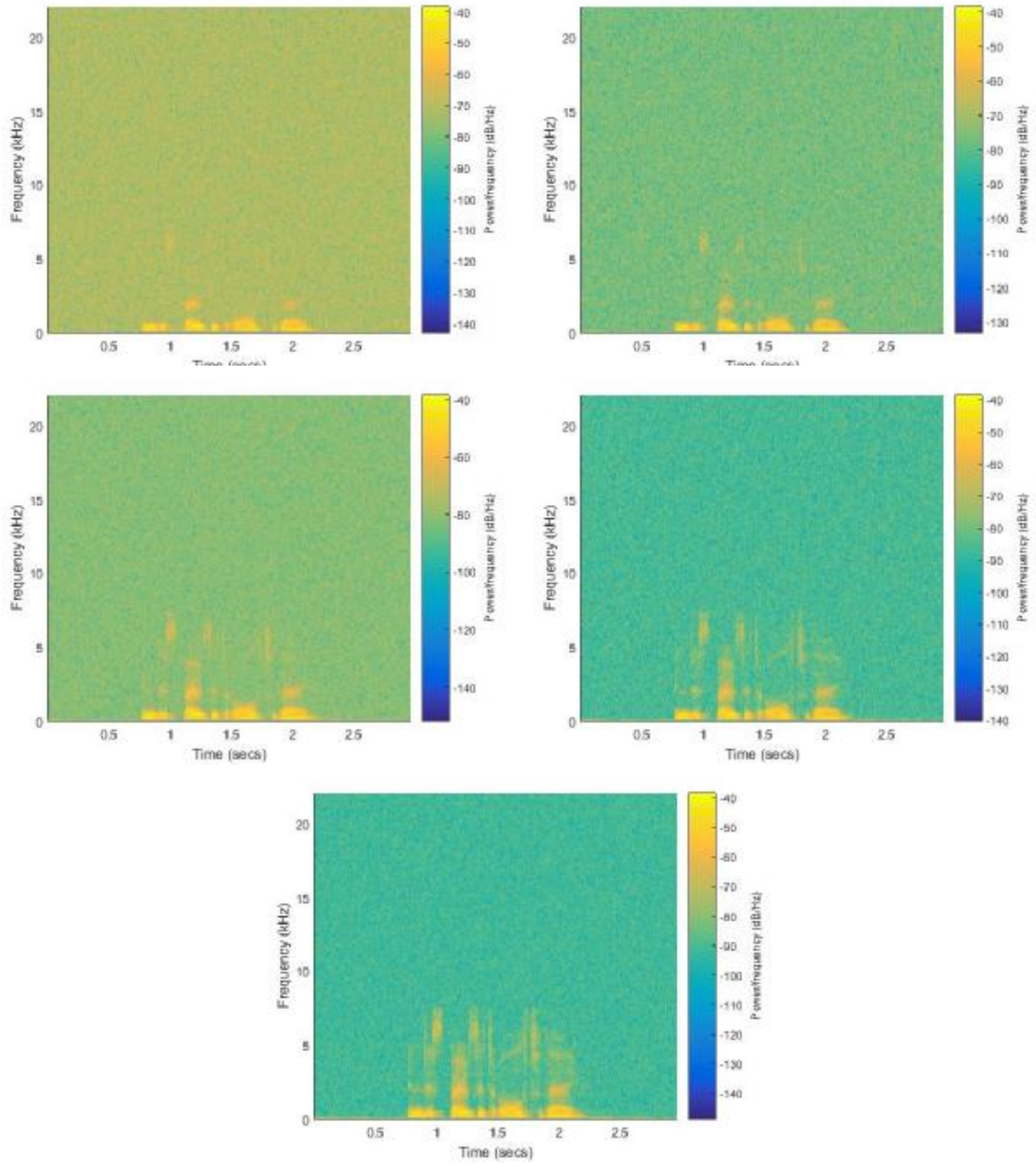
Εικόνα 16: woman02 - reference signal

Από τα παραπάνω γραφήματα μπορούμε να διακρίνουμε πού υπάρχει δραστηριότητα ομιλίας (χρονικά) και σε ποιες συχνότητες.
Στη συνέχεια παρουσιάζονται γραφήματα σχετικά με το πόσο επηρεάστηκαν τα αρχικά σήματα με την είσοδο διαφορετικών ποσοστών θορύβου.





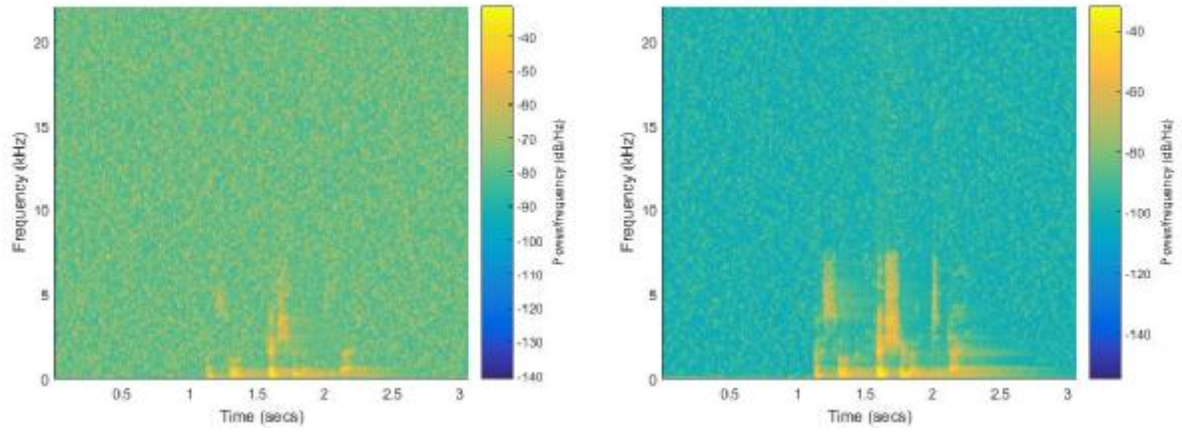
Εικόνα 17:man01 - multiple SNR



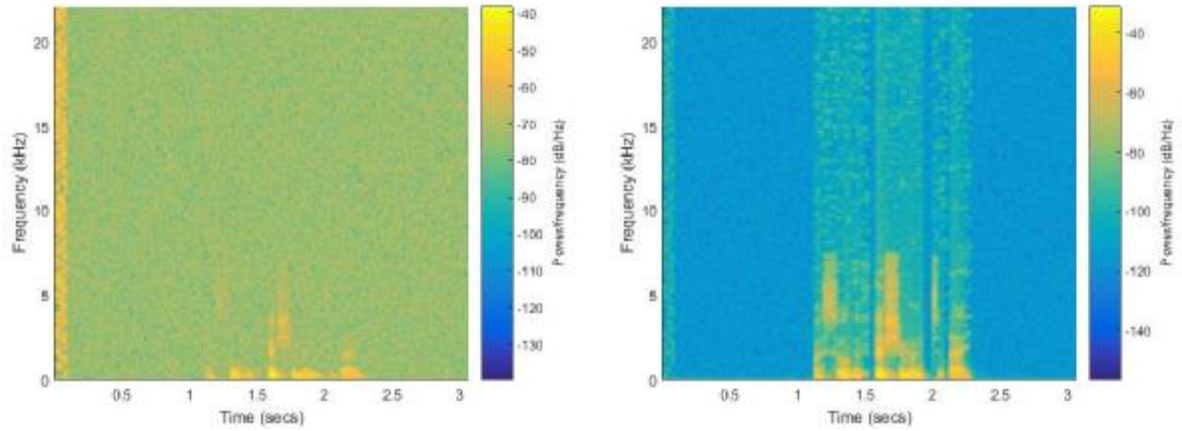
Εικόνα 18: woman02 - multiple SNR

Τα παραπάνω block εικόνων αντιστοιχούν σε τιμές SNR 0, 5, 10, 15, 20dB για κάθε σήμα. Γίνεται λοιπόν εύκολα αντιληπτό πόσο έχει επηρεάσει ο θόρυβος το σήμα. Όσο μεγαλύτερο είναι το SNR, τόσο περισσότερη πληροφορία του καθαρού σήματος διατηρείται. Δηλαδή, στην περίπτωση των 0dB έχει χαθεί μέρος πληροφορίας του καθαρού σήματος. Ενώ αντίθετα στην περίπτωση των 20dB η πληροφορία του καθαρού σήματος διατηρείται σε μεγαλύτερο ποσοστό. Είναι προφανές πως κάθε σήμα θα δώσει διαφορετικό αποτέλεσμα, καθώς έχει μοναδικά χαρακτηριστικά (ποιότητα, ένταση κ.ά.)
Έτσι, μετά από παρατηρήσεις διαπιστώθηκε ότι ιδιαίτερο ενδιαφέρον παρουσιάζουν οι περιπτώσεις για τις τιμές SNR=0dB και SNR=20dB.

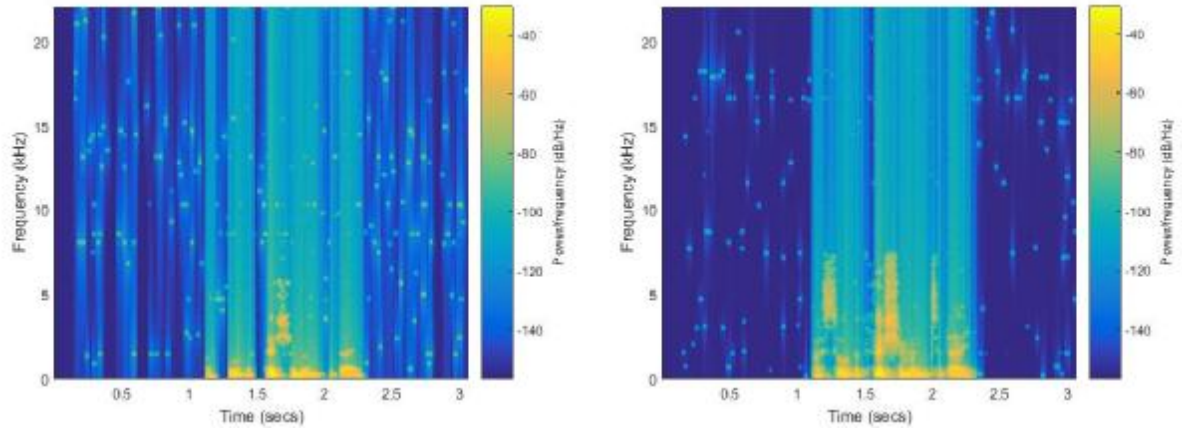
Όσον αφορά τη θεωρητική προσέγγιση των αποτελεσμάτων, παρουσιάζονται γραφήματα γι' αυτές τις δύο τιμές. Σκοπός είναι να αναδειχθεί η καλύτερη μέθοδος αποθορυβοποίησης τόσο για την ακραία χειρότερη, όσο και την ακραία καλύτερη τιμή SNR.



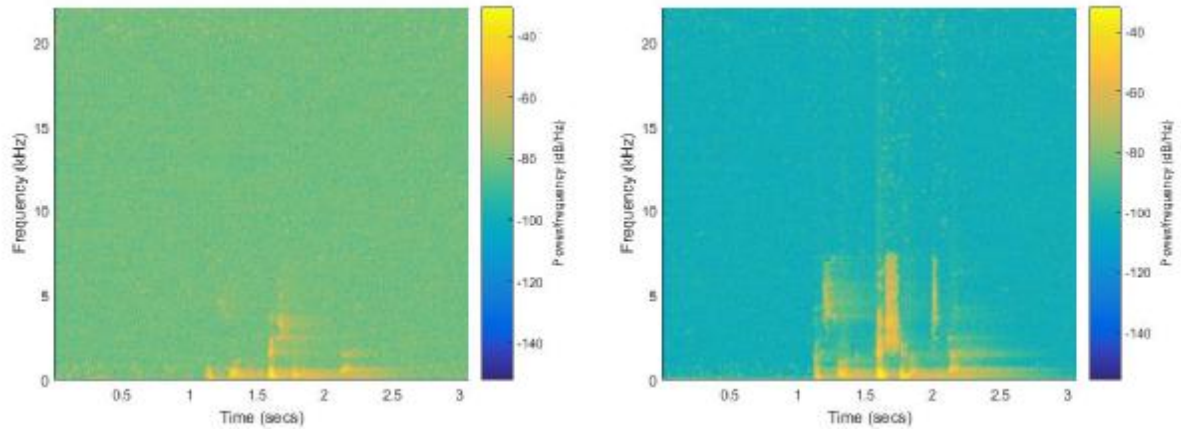
Εικόνα 19: man01 - Berouti



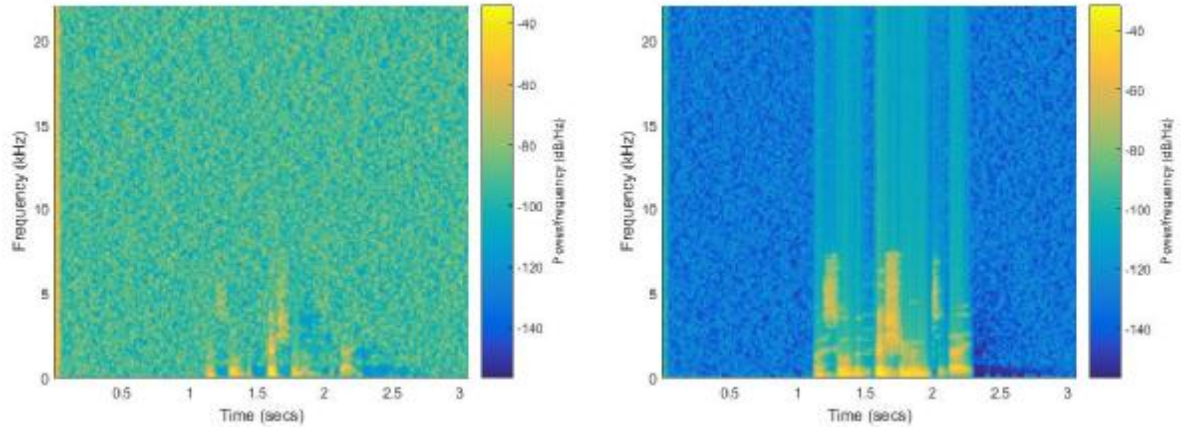
Εικόνα 20: man01 - Boll



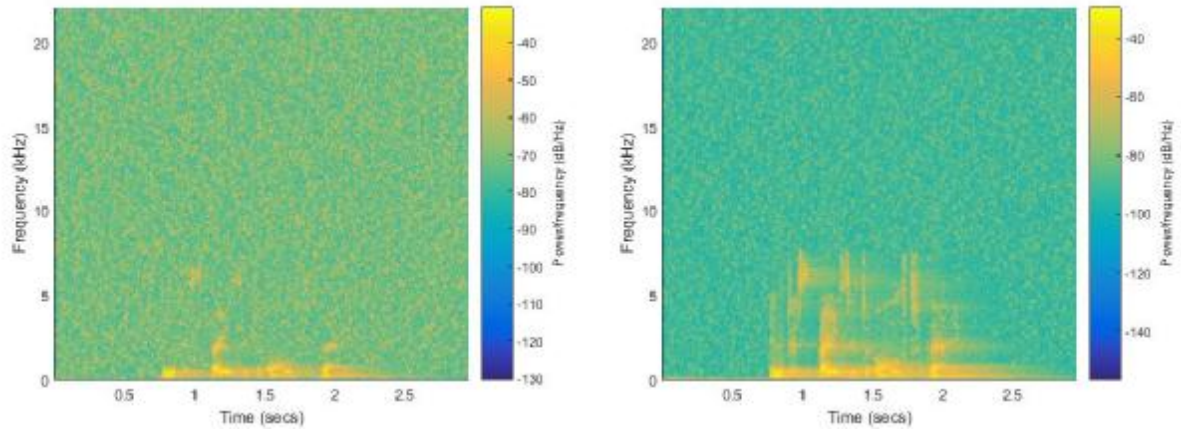
Εικόνα 21: man01 - Ephraim



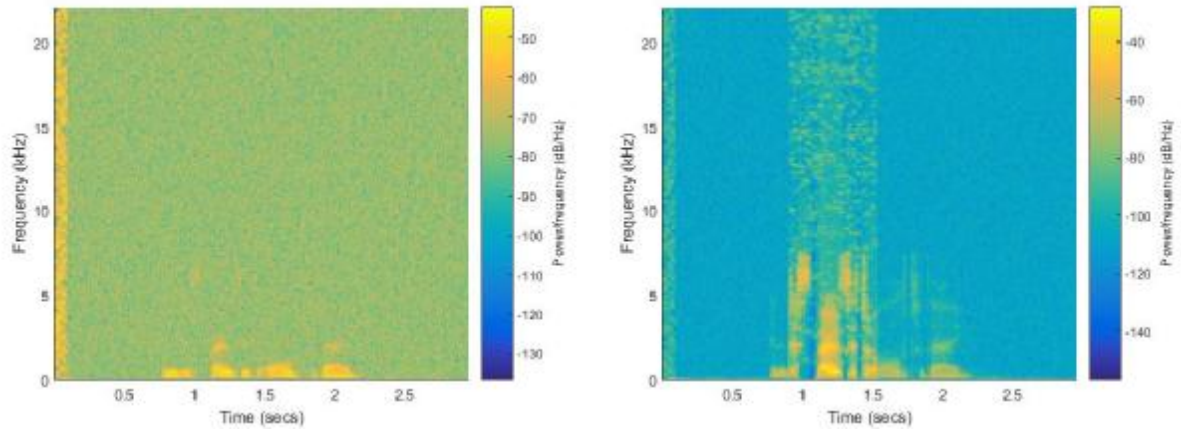
Εικόνα 22: man01 - Multiband



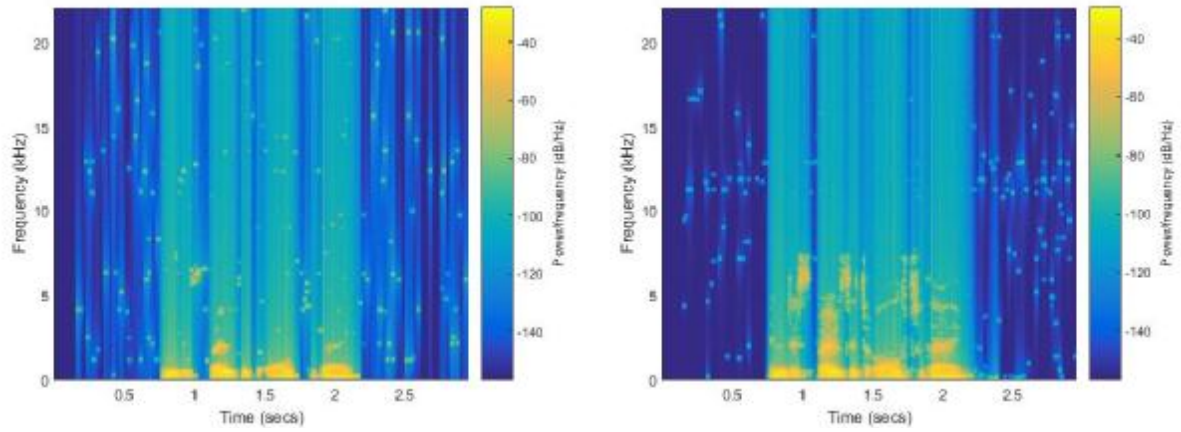
Εικόνα 23: man01 – Wiener



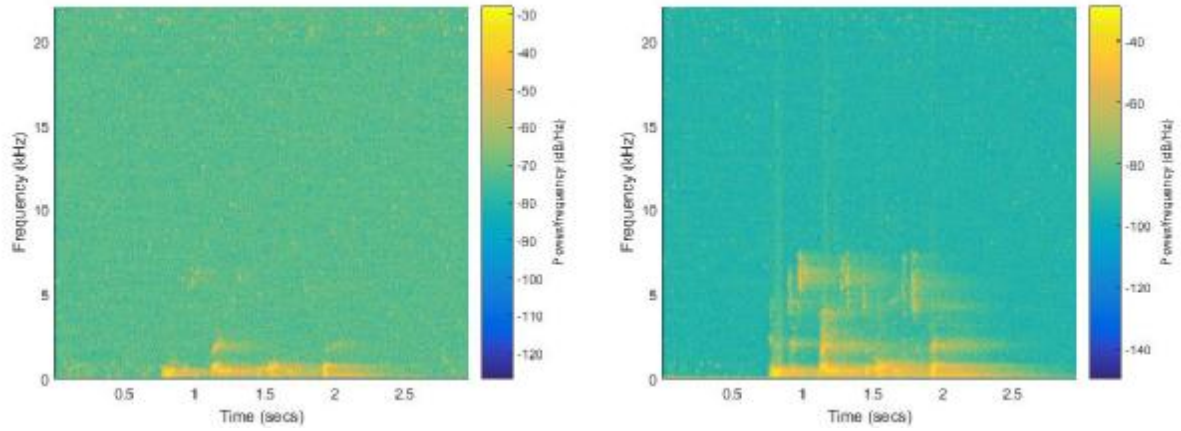
Εικόνα 24: woman02 - Berouti



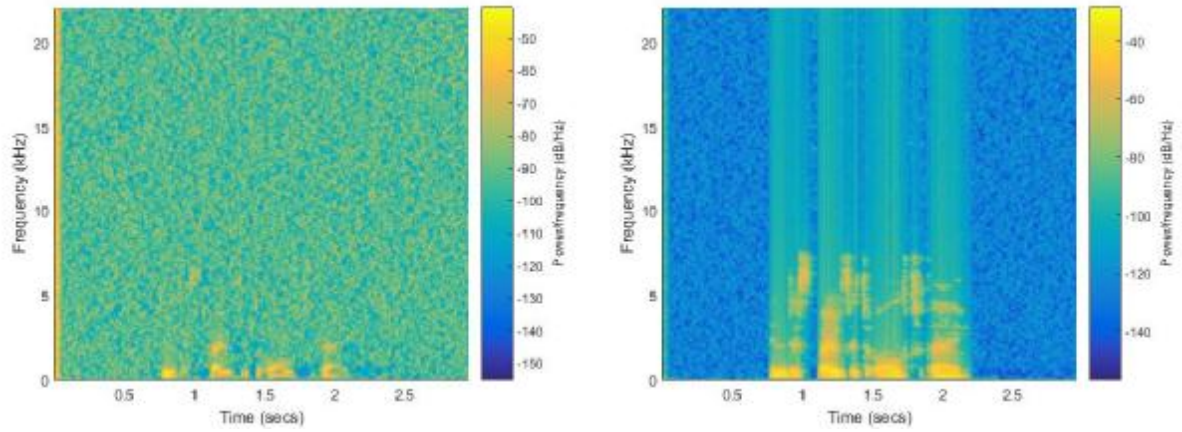
Εικόνα 25: woman02 - Boll



Εικόνα 26: woman02 - Ephraim



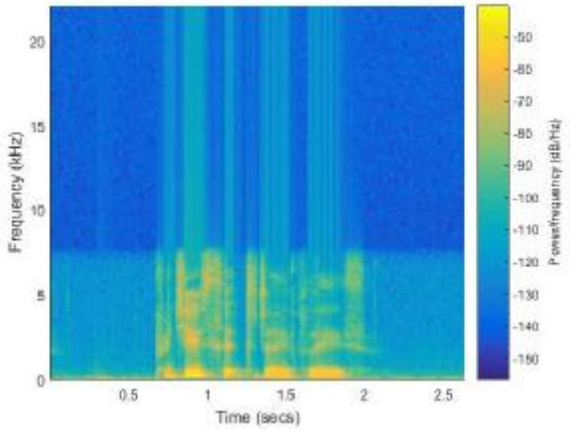
Εικόνα 27: woman02 - Multiband



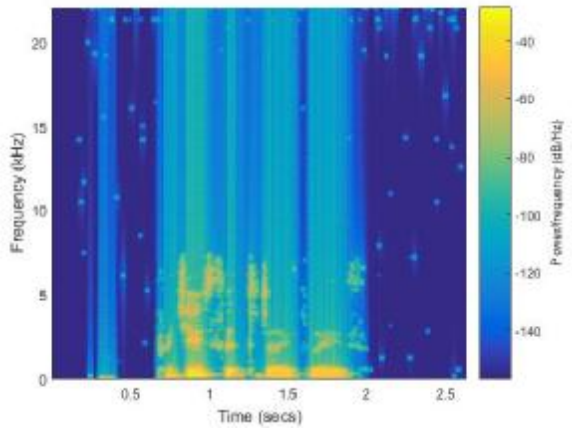
Εικόνα 28: woman02 – Wiener

Συγκρίνοντας τα παραπάνω αποτελέσματα παρατηρείται ότι ο αλγόριθμος Boll κατά την αποθορυβοποίηση χάνει το μεγαλύτερο ποσοστό της πληροφορίας είτε για χαμηλό SNR είτε για υψηλό και παράλληλα παραμένει αρκετός θόρυβος. Ακολουθεί ο αλγόριθμος Berouti όπου ενώ φαίνεται να διατηρεί περισσότερη πληροφορία, έχουμε αλλοίωση της ποιότητας του σήματος. Πολύ καλύτερα αποτελέσματα προκύπτουν από τους αλγόριθμους Multiband και Wiener. Τέλος, ο αλγόριθμος όπου διατηρείται αρκετά καλό ποσοστό της ομιλίας αλλά και συμπιέζεται ο θόρυβος (ακόμα και για SNR=0), είναι ο αλγόριθμος Ephraim. Ο συγκεκριμένος αλγόριθμος όχι μόνο συμπιέζει τον θόρυβο αλλά ενισχύει και το σήμα ομιλίας βελτιώνοντας έτσι την ποιότητα.

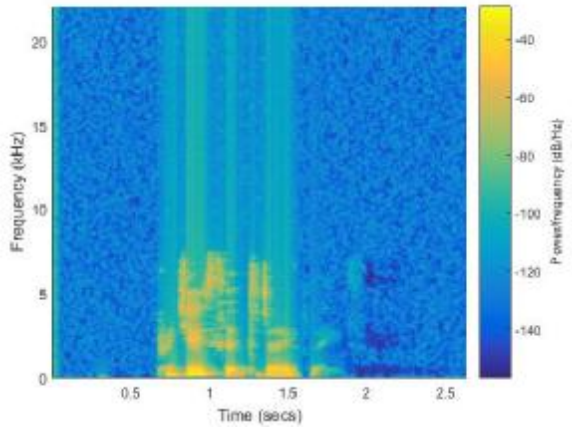
Στα παραπάνω γραφήματα παρατηρούμε ότι για SNR=20dB, οι αλγόριθμοι Wiener και Ephraim παρουσιάζουν τα πιο «καθαρά» γραφήματα. Συγκρίνοντάς τα και με το σήμα αναφοράς παρατηρείται ότι το αποτέλεσμα της αποθορυβοποίησης είναι αρκετά ικανοποιητικό για τον Wiener, ενώ για τον Ephraim είναι πολύ περισσότερο. Δηλαδή, στα σημεία με τη μεγαλύτερη δραστηριότητα (ομιλία) έχουμε μεγάλο ποσοστό αποκατάστασης.



Εικόνα 29: Σήμα αναφοράς



Εικόνα 30: ephraim SNR 20db



Εικόνα 31: wiener SNR 20dB

Ωστόσο, σε ορισμένα σήματα, συνεχίζει να υπάρχει θόρυβος αλλά είναι σε χαμηλό επίπεδο. Οπότε θεωρούμε ότι η αποθορυβοποίηση ήταν επιτυχείς.

Για την περίπτωση του αλγόριθμου Wiener μπορεί να παρατηρηθεί σε κάποια σήματα απώλεια πληροφορίας. Τότε δεν θεωρούμε ότι η αποθορυβοποίηση είναι επιτυχής ακόμα κι αν ο θόρυβος έχει συμπιεστεί πλήρως. Σκοπός είναι και να κατασταλεί ο θόρυβος αλλά και να διατηρήσουμε την ομιλία.

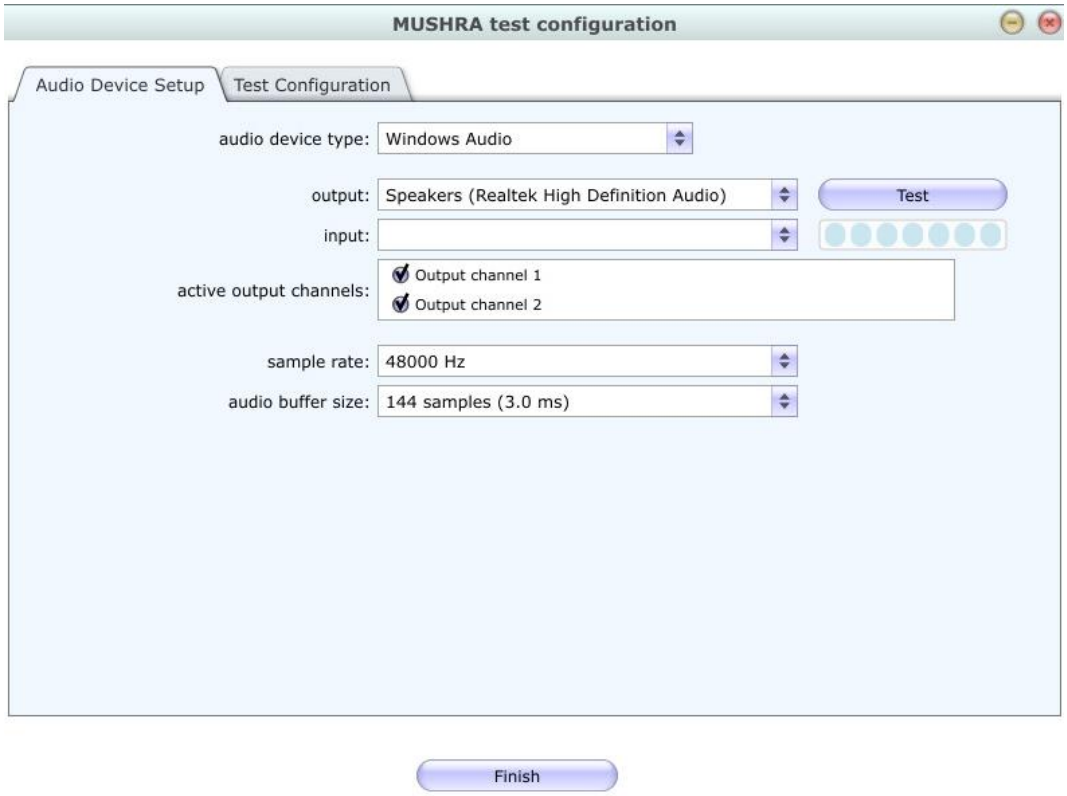
Για τις περιπτώσεις των αλγόριθμων Boll, Berouti, Multiband ο θόρυβος που απομένει είναι πολύ περισσότερος. Εκεί εξετάζουμε πόσο έχει βελτιωθεί έστω και στο ελάχιστο.

5.2 Αξιολόγηση

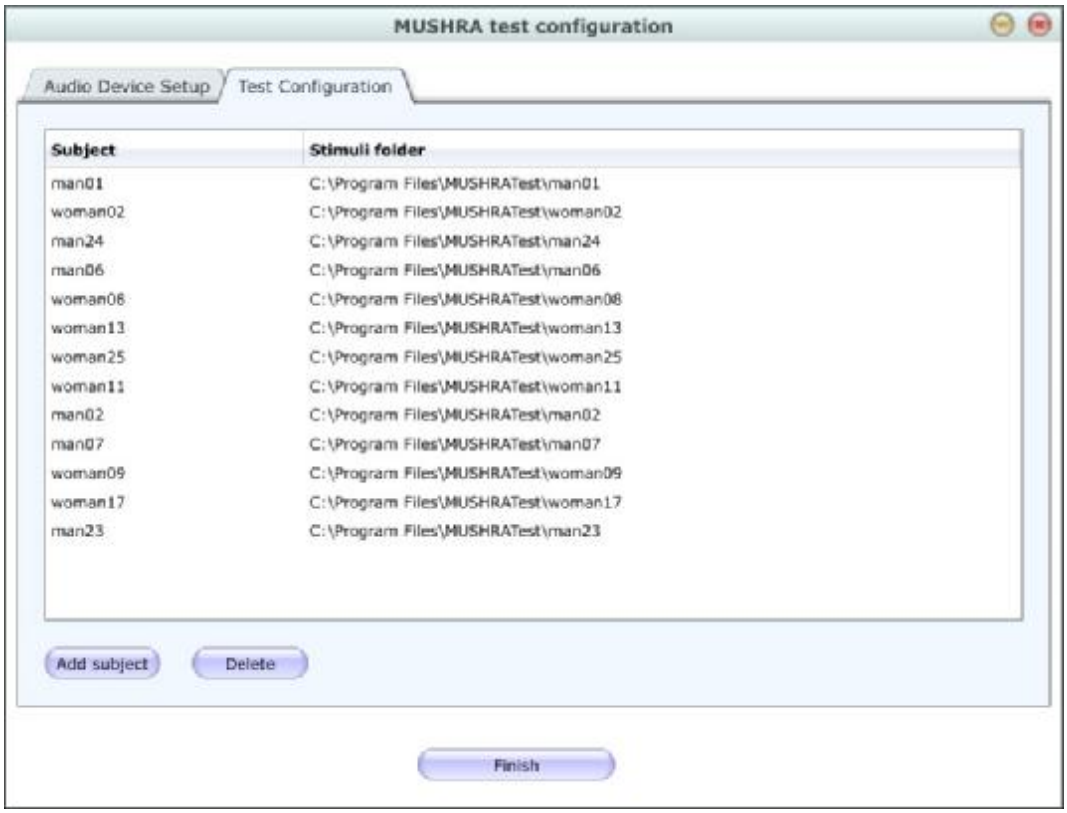
Υπάρχουν πολλοί τρόποι να αξιολογήσουμε το αποτέλεσμα της εξόδου της επεξεργασίας ενός σήματος ομιλίας που έχει υποστεί αποθορυβοποίηση. Κάποιοι βασίζονται σε αντικειμενικά και άλλοι σε υποκειμενικά κριτήρια. Οι συνήθως χρησιμοποιούμενες μέθοδοι είναι αυτές που βασίζονται σε αντικειμενικά κριτήρια. Μέχρι πρόσφατα ήταν αρκετά εύκολο, γρήγορο και χωρίς μεγάλο κόστος να υλοποιηθούν. Βέβαια, δεν είναι αρκετά αξιόπιστες, καθώς μας πληροφορούν από μαθηματικής άποψης αν η αποθορυβοποίηση είναι ικανοποιητική ή όχι. Μετράνε απλά τη διαφορά μεταξύ των δύο σημάτων – του σήματος αναφοράς και του σήματος που εξετάζεται κατά

περίπτωση.
Στον αντίποδα είναι οι μέθοδοι που βασίζονται σε υποκειμενικά κριτήρια. Εδώ, η αξιολόγηση γίνεται βασιζόμενη στην ανθρώπινη ακοή.

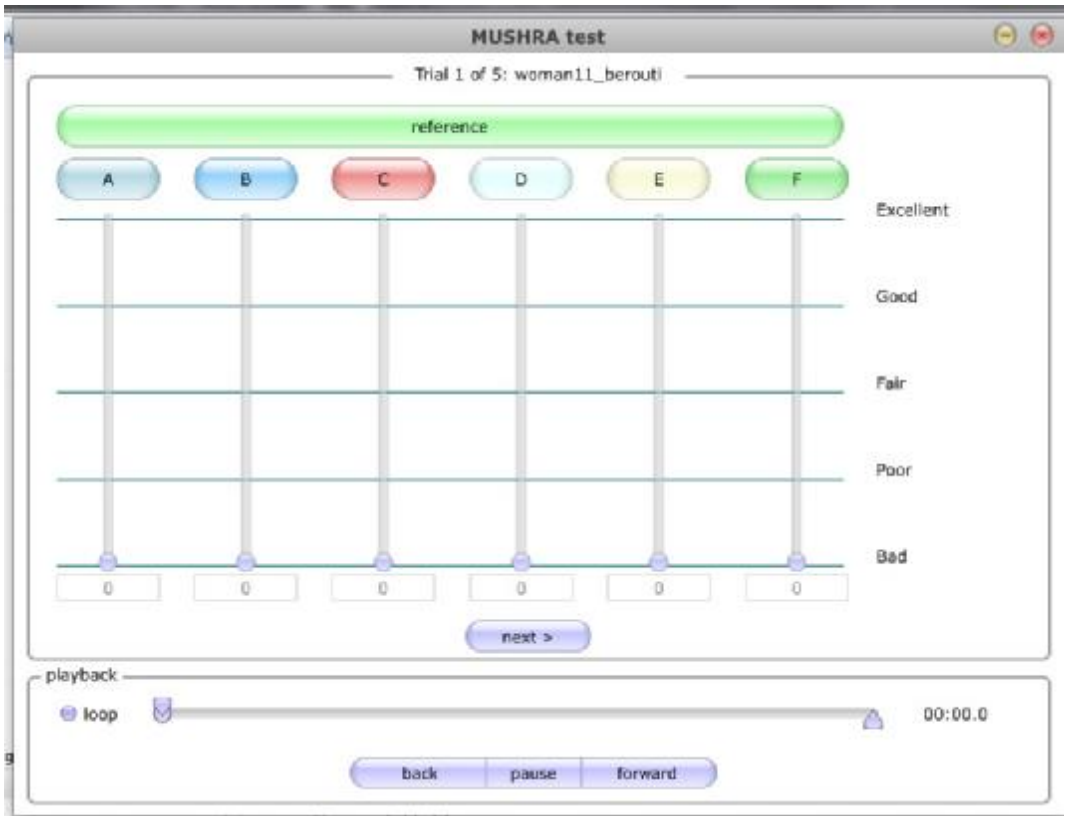
5.2.1 MUSHRA Test



Εικόνα 32: Ρυθμίσεις MUSHRA Test



Εικόνα 33: Επιλογή σημάτων προς αξιολόγηση



Εικόνα 34: Διαδικασία αξιολόγησης

5.2.1 Signal to Noise Ratio (SNR) - SSNR

Ένας από τους πιο κοινούς και αρκετά αξιόπιστους τρόπους αξιολόγησης της ποιότητας ενός σήματος, είναι ο *λόγος σήματος προς θόρυβο*. Ορίζεται ως ο λόγος των ενεργειών σήματος και θορύβου, με μονάδα μέτρησης τα dB(decibels). Είναι μια προσέγγιση η οποία βασίζεται σε αντικειμενικά κριτήρια και δεν μπορεί να μας δώσει "πραγματικά" αποτελέσματα, καθώς δεν σχετίζεται με το ανθρώπινο σύστημα ακοής.
Η εξίσωση που προσδιορίζει το SNR, *μετράει το ποσοστό του θορύβου στο ενθόρυβο σήμα ή της παραμόρφωσης σε σχέση με το σήμα αναφοράς*.

$$SNR_{dB} = 10\log_{10}\frac{\sigma^2_{signal}}{\sigma^2_{noise}} \tag{5.2.1.1}$$

Με το SNR μπορούμε να προσδιορίσουμε και τον συντελεστή θορύβου λ.

$$\lambda = \frac{\sum_{n=0}^{N-1} x^2(n)}{10^{\frac{SNR}{10}} \sum_{n=0}^{N-1} d^2(n)} \tag{5.2.1.2}$$

5.2.2 Υποκειμενικά κριτήρια

Όταν αναφερόμαστε σε υποκειμενικά κριτήρια αξιολόγησης ενός σήματος (ήχος/ομιλία), εννοούμε την αξιολόγησή του από φυσικούς ακροατές.
Η γενική προσέγγιση, λοιπόν, για να πάρουμε υποκειμενικά αποτελέσματα είναι: αναπαραγωγή και ακρόαση του ζητούμενου σήματος από πολλαπλούς χρήστες, αξιολόγηση με βάση την υποκειμενική αντίληψη του καθενός και αποτύπωση του αποτελέσματος σε μία κλίμακα (πχ. 0-100).

Στην παρούσα πτυχιακή εργασία παρουσιάζονται αποτελέσματα για τα σήματα με τη μεγαλύτερη συχνότητα αξιολόγησης. Έτσι, μπορεί να γίνει και σύγκριση μεταξύ χρηστών, αναδεικνύοντας τη μοναδικότητα της υποκειμενικής αξιολόγησης. {Ωστόσο, δεν θα ασχοληθούμε με τον αν είναι άντρας ή γυναίκα ο ακροατής καθώς δεν είναι παράγοντας που μελετάται στο πείραμα.}

Τα αποτελέσματα κρίνονται συνολικά με βάση την ανθρώπινη αντίληψη γενικά. Δεν γίνεται εστίαση ξεχωριστά για το αν οι αξιολογήσεις έγιναν από άνδρες ή γυναίκες. Θέλουμε απλά να δούμε πόσο καλά λειτουργούν οι αλγόριθμοι με βάση την υποκειμενική αξιολόγηση.

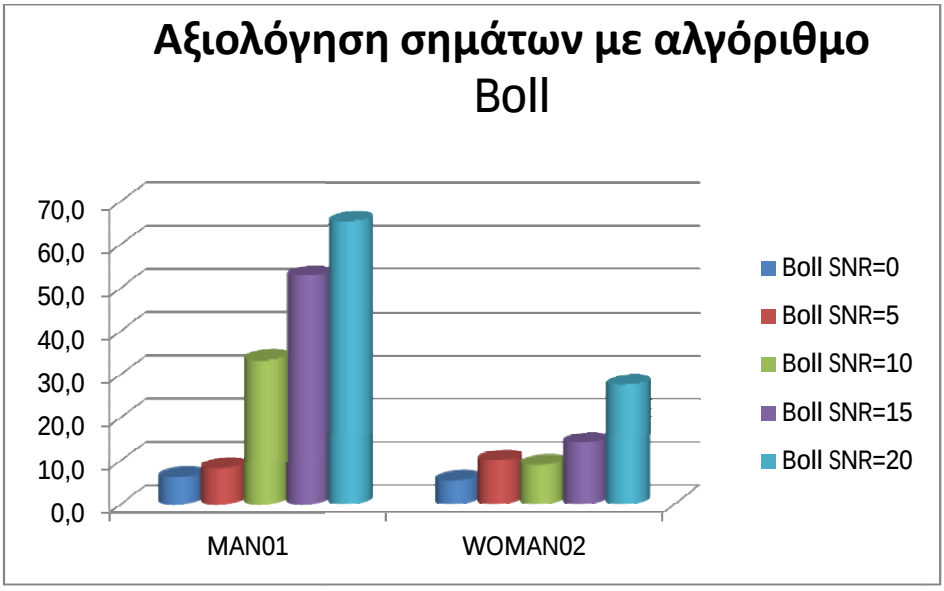
Ωστόσο, θα γίνει αναφορά σχετικά με το πώς γίνονται αντιληπτά τα γυναικεία και ανδρικά σήματα.

Λόγω του ότι η αξιολόγηση είναι υποκειμενική μπορούμε να καταλάβουμε ότι τα αποτελέσματα ίσως είναι διαφορετικά από αυτά της αντικειμενικής αξιολόγησης, καθώς ο χρήστης αξιολογεί ο ίδιος ποιο σήμα του ακούγεται καλύτερο κα ποιο χειρότερο. Προκύπτει λοιπόν σε πολλές περιπτώσεις σημάτων ασυμφωνία για τους δύο τρόπους αξιολόγησης. Φυσικά, δεν μπορούμε να πούμε ότι κάποιος χρήστης έχει κάνει λάθος ούτε ότι έπεσε έξω στην αξιολόγηση.

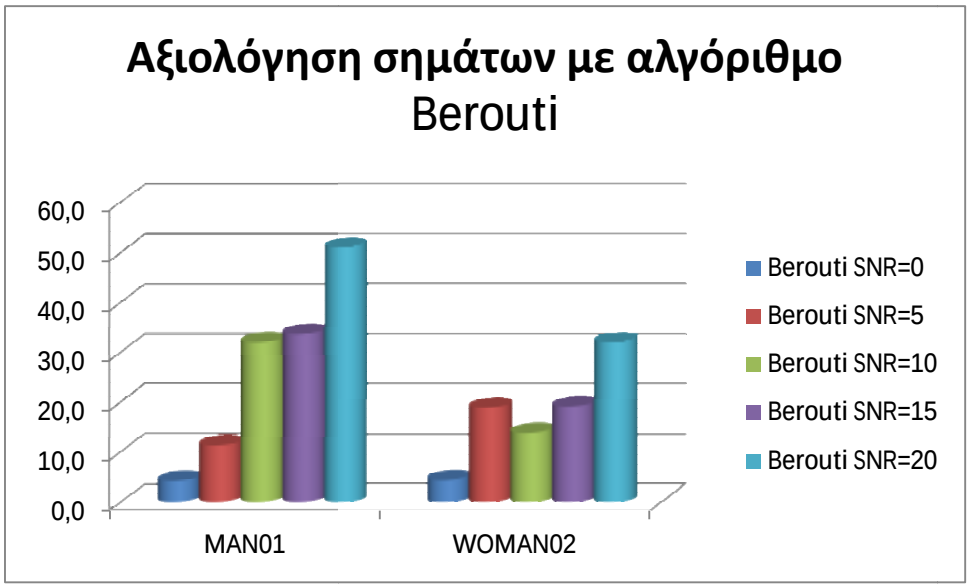
Η εργασία εστιάζει στον ακροατή και αναλύονται τα αποτελέσματα συγκριτικά για το πώς επηρεάζονται τα ανδρικά και γυναικεία σήματα.

Στο κομμάτι αυτό παραθέτονται τα αποτελέσματα του τεστ σύμφωνα με το πώς αξιολόγησαν οι ακροατές τα σήματα. Ουσιαστικά είναι η υποκειμενική αξιολόγηση των σημάτων.

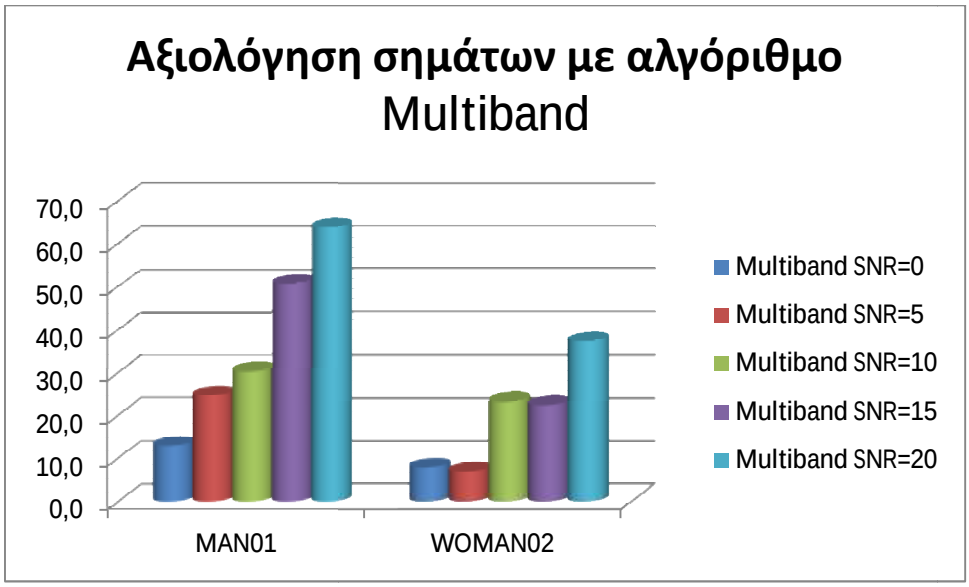
Σε σήματα με τη χρήση του αλγόριθμου Ephraim παρατηρήθηκε ότι για την τιμή SNR=20dB η αξιολόγηση ήταν μεγαλύτερη από το σήμα αναφοράς ή και ίση. Από αυτό το γεγονός προκύπτει το συμπέρασμα ότι όχι μόνο συμπιέστηκε απόλυτα ο θόρυβος αλλά και ότι το σήμα βελτιώθηκε αισθητά σε σύγκριση με το αρχικό (ακουστικό παράδειγμα για man01).



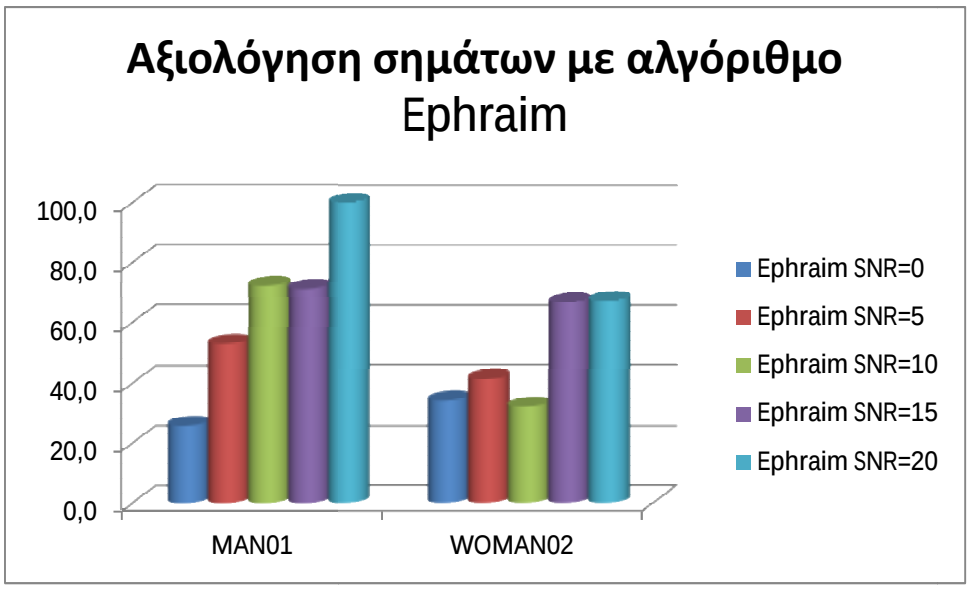
Εικόνα 35: Αξιολόγηση Boll



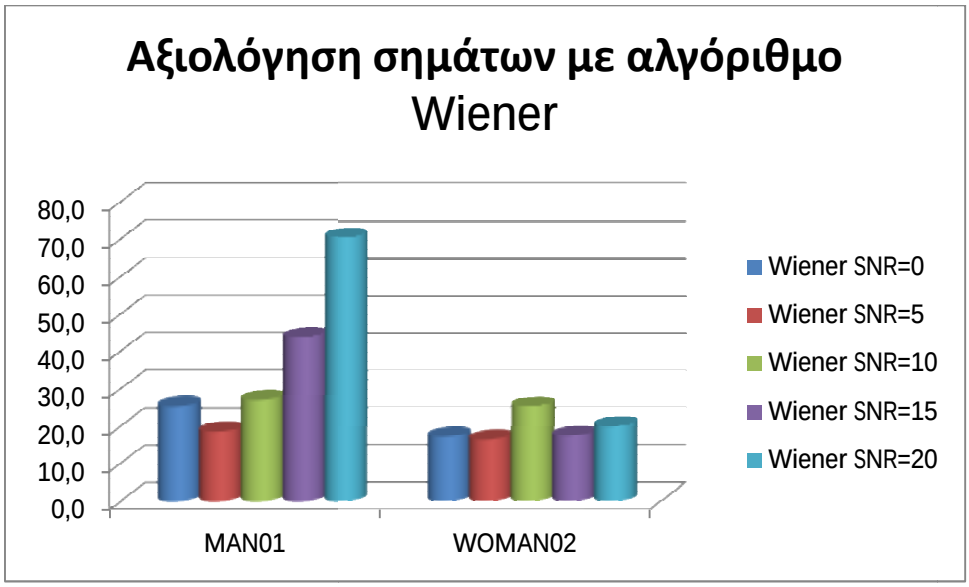
Εικόνα 36: Αξιολόγηση Berouti



Εικόνα 37: Αξιολόγηση Multiband



Εικόνα 38: Αξιολόγηση Ephraim



Εικόνα 39: Αξιολόγηση Wiener

Τα παραπάνω γραφήματα περιέχουν πληροφορίες σχετικά με τον μέσο όρο βαθμολογίας των σημάτων σύμφωνα με τα αποτελέσματα του τεστ.

Στη συνέχεια παρουσιάζεται ένας πίνακας με τα αποτελέσματα της αξιολόγησης σύμφωνα με το φύλο του ακροατή. Αφορά στον μέσο όρο βαθμολογίας των δύο σημάτων που παρουσιάζονται ως παράδειγμα.

man01	Boll SNR=0	Boll SNR=5	Boll SNR=10	Boll SNR=15	Boll SNR=20
άνδρες	0,0	0,0	24,0	47,0	74,0
γυναίκες	9,5	12,5	38,0	55,5	61,0
	Berouti SNR=0	Berouti SNR=5	Berouti SNR=10	Berouti SNR=15	Berouti SNR=20
άνδρες	0,0	0,0	24,0	50,0	74,0
γυναίκες	6,5	17,0	36,0	25,5	39,5
	Multiband SNR=0	Multiband SNR=5	Multiband SNR=10	Multiband SNR=15	Multiband SNR=20
άνδρες	0,0	12,0	40,0	46,0	38,0
γυναίκες	19,5	31,0	25,5	53,0	76,5
	Ephraim SNR=0	Ephraim SNR=5	Ephraim SNR=10	Ephraim SNR=15	Ephraim SNR=20
άνδρες	0,0	48,0	68,0	74,0	100,0
γυναίκες	39,0	55,5	74,5	70,0	100,0
	Wiener SNR=0	Wiener SNR=5	Wiener SNR=10	Wiener SNR=15	Wiener SNR=20
άνδρες	0,0	8,0	25,0	10,0	63,0
γυναίκες	38,0	24,0	28,0	60,5	74,0

woman02	Boll SNR=0	Boll SNR=5	Boll SNR=10	Boll SNR=15	Boll SNR=20
άνδρες	3	8	7,5	11	23
γυναίκες	8	12,5	11	17,5	32
	Berouti SNR=0	Berouti SNR=5	Berouti SNR=10	Berouti SNR=15	Berouti SNR=20
άνδρες	9,0	22,5	15,5	26,5	22,5
γυναίκες	0	15,5	12,5	12	41,5
	Multiband SNR=0	Multiband SNR=5	Multiband SNR=10	Multiband SNR=15	Multiband SNR=20
άνδρες	5,5	5,5	18,5	22,5	31,5
γυναίκες	10,5	8,5	28	22,5	43
	Ephraim SNR=0	Ephraim SNR=5	Ephraim SNR=10	Ephraim SNR=15	Ephraim SNR=20
άνδρες	40	47,5	20	58,5	60
γυναίκες	28,5	35,5	44,5	75	74,5
	Wiener SNR=0	Wiener SNR=5	Wiener SNR=10	Wiener SNR=15	Wiener SNR=20
άνδρες	20,5	20	33,5	22	27
γυναίκες	14	13	17	13	13

6. Συμπεράσματα

Τα τελευταία χρόνια, η εκτίμηση της ποιότητας ομιλίας κερδίζει όλο και περισσότερο έδαφος, αφού αποτελεί σημαντικό κομμάτι στον τομέα της επεξεργασίας σημάτων ομιλίας. Το ενδιαφέρον αυτής της διεργασίας είναι η χρησιμότητά της ως προς το τελικό αποτέλεσμα της επεξεργασίας σημάτων. Μέσω των αποτελεσμάτων από μετρήσεις και εκτιμήσεις μπορούμε να δώσουμε ακριβείς αριθμούς και να εξάγουμε επίσημα συμπεράσματα σχετικά με το προϊόν – σήμα που δουλεύουμε.

Μπορούμε να προσδιορίσουμε την ποιότητα ομιλίας ως «*το αποτέλεσμα της αξιολόγησης όλων των αναγνωρίσιμων χαρακτηριστικών, αυτών που μπορούν να προσδιοριστούν αλλά και των τιμών των χαρακτηριστικών ενός δείγματος ομιλίας που είναι υπό εξέταση, υπό όρους καταλληλότητας ώστε να πληροί τις προσδοκίες όλων των αναγνωρίσιμων χαρακτηριστικών, αυτών που μπορούν να προσδιοριστούν αλλά και των τιμών των χαρακτηριστικών των ατομικών προσδοκιών και/ή κοινωνικών απαιτήσεων και/ή απαιτήσεων*».

Λόγω της υποκειμενικότητας στην αξιολόγηση της ομιλίας, αντιλαμβανόμαστε ότι δεν είναι καθαρά ένα «γεγονός» αιτιοκρατικό και ελεγχόμενο από το σήμα. Επίσης, δεν θα οδηγήσει απαραίτητα στο ίδιο συμπέρασμα κάθε φορά που θα είναι παρούσες οι ίδιες ακουστικές και περιβαλλοντικές συνθήκες. Αντίθετα, μπορεί να προκύψουν εντελώς διαφορετικά δεδομένα ακόμα κι αν επαναληφθεί η διαδικασία της αξιολόγησης που σχετίζεται με τα ακουστικά σήματα ομιλίας.

Πολλά άτομα μπορεί να περιγράψουν με τον ίδιο τρόπο σήματα ομιλίας για τα οποία κλήθηκαν να αξιολογήσουν παρά το γεγονός ότι καθένας το αντιλαμβάνεται με δικά του υποκειμενικά κριτήρια.

Από μαθηματικής άποψης, η ποιότητα ομιλίας είναι μια συνάρτηση που σχετίζεται με πολλές μεταβλητές όπως αντίληψη, προσήλωση, υποκειμενικότητα, μνήμα, γνώση, νευρικό σύστημα κ.ά.

Η αντίληψη είναι μια ενεργή, ατομική διαδικασία.

Από τις παραπάνω μεταβλητές, η γνώση και η εμπειρία ελέγχουν πώς επιλέγουμε τα εισερχόμενα σήματα και έχουν κύριο ρόλο στο πώς επεξεργαζόμαστε και ερμηνεύουμε τα σήματα. Έτσι, ο ακροατής θα κρατήσει και θα επεξεργαστεί σαν εισερχόμενο σήμα μόνο αυτό που μπορεί να αξιολογήσει ως πηγή πληροφορίας σε μια δεδομένη στιγμή. Το σήμα αυτό φαίνεται ότι είναι αυτό που περιέχει το μεγαλύτερο ποσοστό χρήσιμης πληροφορίας για τον ακροατή.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Δ. Σκαρλάτος, Εφαρμοσμένη Ακουστική, Πάτρα: GOTSIS, 2008.

[2] Λ. Χαδέλης, Τεχνολογία Ήχου, Πανεπιστημιακές εκδόσεις "Αράκυνθος", 2010.

[3] A. Das, «Waveform Encoding,» σε *Digital Communication Principles and System Modeling*, Springer, 2010.

[4] A. Das, Digital Communication: Principles and System Modelling, Springer, 2010.

[5] S. K. Mitra, Digital Signal Processing: A Computer-Based Approach, Higher Education επιμ., McGraw-Hill, 4th Edition 2011.

[6] M. H. Hayes, Ψηφιακή Επεξεργασία Σήματος, Θεσσαλονίκη: Τζιόλα, 2000.

[7] Ε. Μιχαλέτου, *Διπλωματική Εργασία - Παραμετροποίηση Σήματος Ομιλίας Για Αναγνώριση Συναισθήματος Ομιλητή*, Πάτρα: Πανεπιστήμιο Πατρών, 2008.

[8] I. t. T. F. Analysis, «National Instruments,» 2014. [Ηλεκτρονικό]. Available: http://zone.ni.com/reference/en-XX/help/372656C-01/lvasptconcepts/tfa_intro/. [Πρόσβαση 4/9/2015 Σεπτέμβριος 2015].

[9] Θ. Σ. Ζαρούχας, *Διδακτορική Διατριβή - Μοντελοποίηση και Επεξεργασία Ηχητικών Δεδομένων για Αναπαραγωγή σε Χώρους με Αντήχηση*, Πάτρα: Πανεπιστήμιο Πατρών, Ιούνιος 2010.

[10] Δ. Ε. Τσουκαλάς - Σταθάκης, *Διδακτορική Διατριβή - Τεχνικές Αφαίρεσης Θορύβου από Σήματα Ομιλίας και Μουσικής*, Πάτρα: Πανεπιστήμιο Πατρών, Νοέμβριος 1997.

[11] S. F. Boll, ***Suppression of acoustic noise in speech using spectral subtraction,*** IEEE Transactions on acoustics, speech, and signal processing, vol. ASSP-27, No. 2, April 1979.

[12] M. Berouti, R. Schwartz, J. Makhoul, ***Enhancement of speech corrupted by musical noise,*** Proc. IEEE ICASSP, pp. 208-211, April 1979.

[13] S. Kamath, P. Loizou, ***A multi-band spectral subtraction method for enhancing speech corrupted by colored noise,*** in Proc. ICASSP, 2002.

[14] Pascal Scalart, Jozue Vieira Filho, ***Speech enhancement based on a priori signal to noise estimation,*** Proc. ICASSP, May 1996.

[15] Yariv Ephraim, David Malah, ***Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,*** IEEE Transactions on acoustics, speech, and signal processing, vol. ASSP-32, No 6, December 1984.

[16] Yariv Ephraim, David Malah, ***Speech enhancement using a minimum mean-square error log-spectral amplitude estimator,*** IEEE Transactions on acoustics, speech, and signal processing, vol. ASSP-33, No. 2, April 1985.

[17] Andreas Spanias, Ted Painter, Venkatraman Atti, Audio Signal Processing and Coding, New Jersey: A John Wiley & Sons, Inc., Publication, 2007.

[18] David Havelock, Sonoko Kuwano, Michael Vorlander, Handbook of Signal Processing in Acoustics vol.1, New York: Springer Science+Business Media, LLC, 2008.

[19] G. M. Davis, Noise Reduction in Speech Applications, CRC Press LLC, 2002.

[20] I. J. Tashev, Sound Capture and Processing, United Kingdom: John Wiley & Sons Ltd., 2009.

[21] Alexis Favrot, Christof Faller, ***Perceptually motivated gain filter smoothing for noise suppression,*** New York: Audio Engineering Society, Convention Paper 7169, October 2007.

[22] Dionysis E. Tsoukalas, John N. Mourjopoulos, George Kokkinakis, ***Speech enhancement based***

on audible noise suppression, IEEE Transactions on speech, and audio processing, vol. 5, No. 6, November 1997.

[23] S. V. Vaseghi, Advanced Digital Signal Processing and Noise Reduction, UK: John Wiley & Sons Ltd, 2006.

[24] «Department of Electronic Systems,» 2004. [Ηλεκτρονικό]. Available: http://kom.aau.dk/group/04gr742/pdf/MFCC_worksheet.pdf.

[25] «Εθνικό και Καποδιστριακο Πανεπιστήμιο Αθηνών,» Ιούνιος 2013. [Ηλεκτρονικό]. Available: http://moag.phys.uoa.gr/moag_gr/sites/default/files/moag_files/Telecom_Chapter_7.pdf.

[26] Α. Κούτρας, *Εργαστηριακές Σημειώσεις - Ψηφιακή Επεξεργασία Ήχου,* ΤΕΙ Πάτρας - Παράρτημα Πύργου, 2011-2012.

[27] White, L.S. and King, S., «The Center for Speech Technology Research,» University of Edinburgh, 2003. [Ηλεκτρονικό]. Available: http://www.cstr.ed.ac.uk/projects/eustace.

[28] U. Jekosch, Voice and Speech Quality Perception - Assessment and Evaluation, Springer, 2005.

[29] Ι. Χαραλάμπου, *Διπλωματική Εργασία - Μελέτη Αλγόριθμων Αποθορυβοποίησης Σήματος Ομιλίας,* Πάτρα: Πανεπιστήμιο Πατρών, Ιούνιος 2016.

[30] Α. Ρούσου, *Διπλωματική Εργασία - Ανάπτυξη Μηχανισμών Διεπικοινωνίας (user interfaces) Μέσω Διάδρασης με Συνθετική Ψηφιακή Φωνή,* Θεσσαλινίκη: Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 2011.

ΠΕΡΙΕΧΟΜΕΝΑ ΣΥΝΟΔΕΥΤΙΚΟΥ CD

Στο συνοδευτικό CD περιέχονται:

1. Η παρούσα πτυχιακή σε μορφή Word και PDF.
2. Ο φάκελος working code με τα αρχεία matlab που χρησιμοποιήθηκαν για την εφαρμογή των αλγόριθμων.
3. Η βάση δεδομένων για τα αρχεία ομιλίας speech_db.
4. Ο φάκελος results στον οποίο υπάρχουν αποτελέσματα για τα σήματα ομιλίας με εισαγωγή SNR 0dB, 5dB, 10dB, 15dB, 20dB. Επίσης περιέχονται και τα αποτελέσματα της αποθορυβοποίησης για όλους τους αλγόριθμους για τις παραπάνω τιμές SNR.
5. Ο φάκελος figures με τις αναπαραστάσεις φάσματος για όλα τα παραπάνω σήματα (ενθόρυβα – αποθορυβοποιημένα).
6. Ο φάκελος Excel με το αρχείο αποτελεσμάτων της αξιολόγησης για τα επιλεγμένα σήματα.