

# Ψηφιακή Επεξεργασία Σημάτων



## *1η Εργαστηριακή Άσκηση*

Ψηφιακή Επεξεργασία Σημάτων

Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών ΕΜΠ

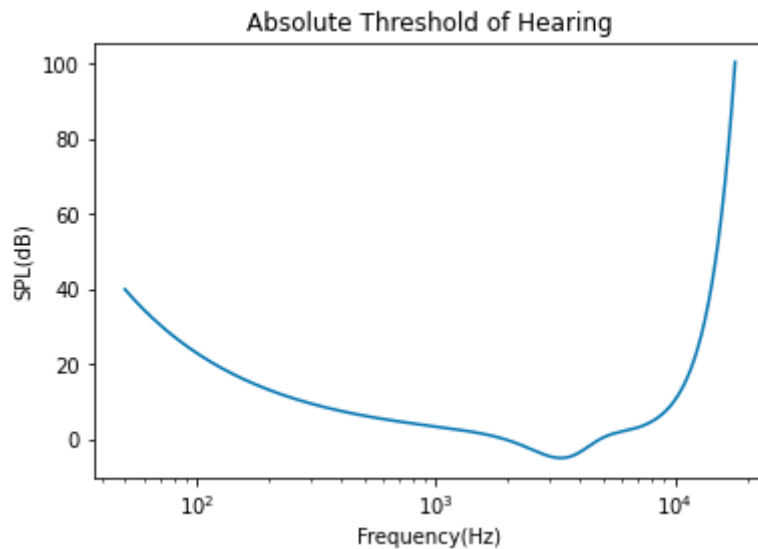
Εξάμηνο 6ο

Σερλής Εμμανουήλ Αναστάσιος, Α.Μ.:03118125, email: manosanastassis@hotmail.com

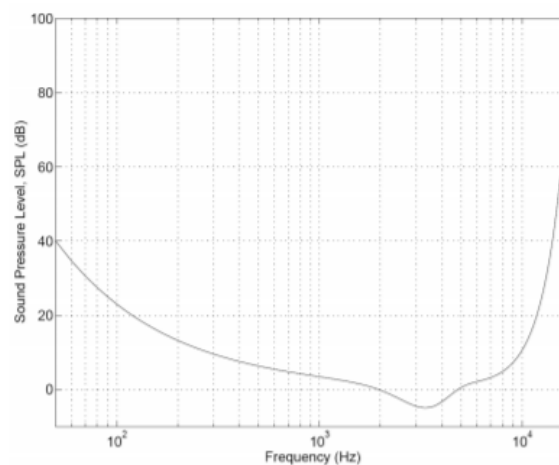
Δημήτρης Μητρόπουλος, Α.Μ.: 03118608, email: dimitris.7.mitropoulos@gmail.com

## *Absolute Threshold of Hearing*

Χρησιμοποιώντας την εξίσωση (1), πήραμε το κάτωθι αποτέλεσμα:



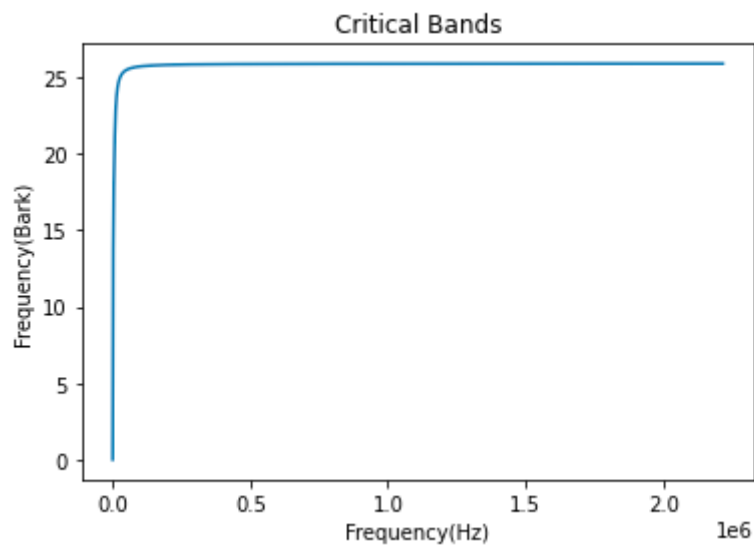
Παρατηρούμε ότι για υψηλές συχνότητες έχουμε μεγάλο threshold δηλαδή οι υψηλές συχνότητες απαιτούν υψηλή ένταση (σε dB) ώστε να γίνουν αντιληπτές από το ανθρώπινο αυτί. Μάλιστα, η παραπάνω γραφική παράσταση σχεδόν ταυτίζεται με το αποτέλεσμα της εκφώνησης:



Σχήμα 1: Απόλυτη τιμή κατωφλίου ακοής: Absolute Threshold of Hearing.

## *Critical Bands*

Μέσω της (2), έγινε η μετατροπή των συχνοτήτων από Hz σε Bark, παίρνοντας την αντίστοιχη γραφική:



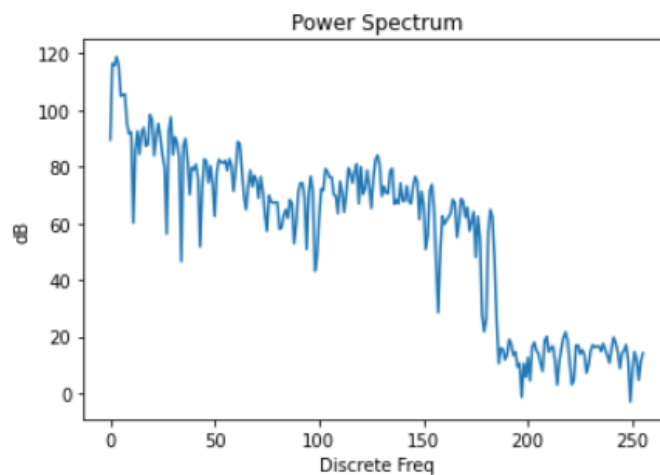
Πράγματι, επαληθεύεται ότι οι bark συχνότητες λαμβάνουν μέγιστη τιμή 25, μιας και σε αυτή την τιμή τείνει και η γραφική μας (ασυμπτωτικά)

# Μέρος 1. Ψυχοακουστικό Μοντέλο 1

Σημείωση: Για τα ακόλουθα ερωτήματα, θα δοθούν τα αποτελέσματα που αφορούν το πρώτο πλαίσιο

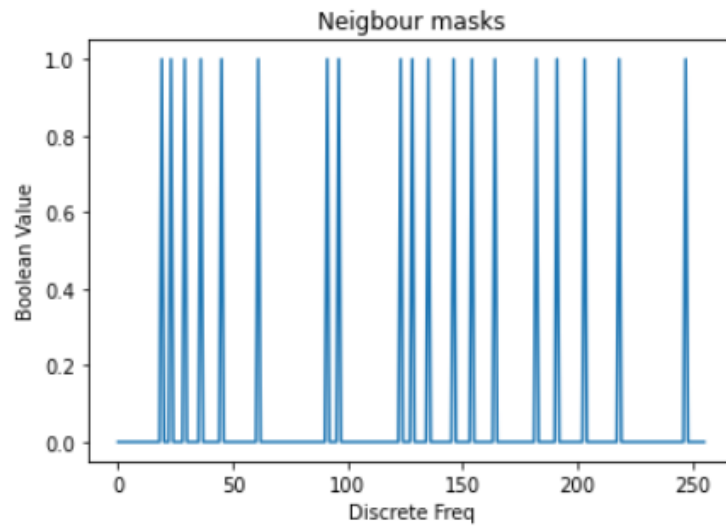
1.0) Πραγματοποιήσαμε ανάγνωση του σήματος εισόδου μέσω της librosa.load() και με παράμετρο mono=True ώστε να συγχωνευθούν τα 2 κανάλια σε ένα ενιαίο. Επιπλέον, το σήμα μας διέθετε ήδη τιμές στο πεδίο τιμών  $[-1,1]$  γεγονός που κατέστησε αχρείαστη την κανονικοποίησή του.

1.1) Χρησιμοποιούμε την εξ.(3) με βοήθεια των (2) και (4) ώστε να υπολογίσουμε το power spectrum για το πρώτο πλαίσιο ( $k=0$ ):

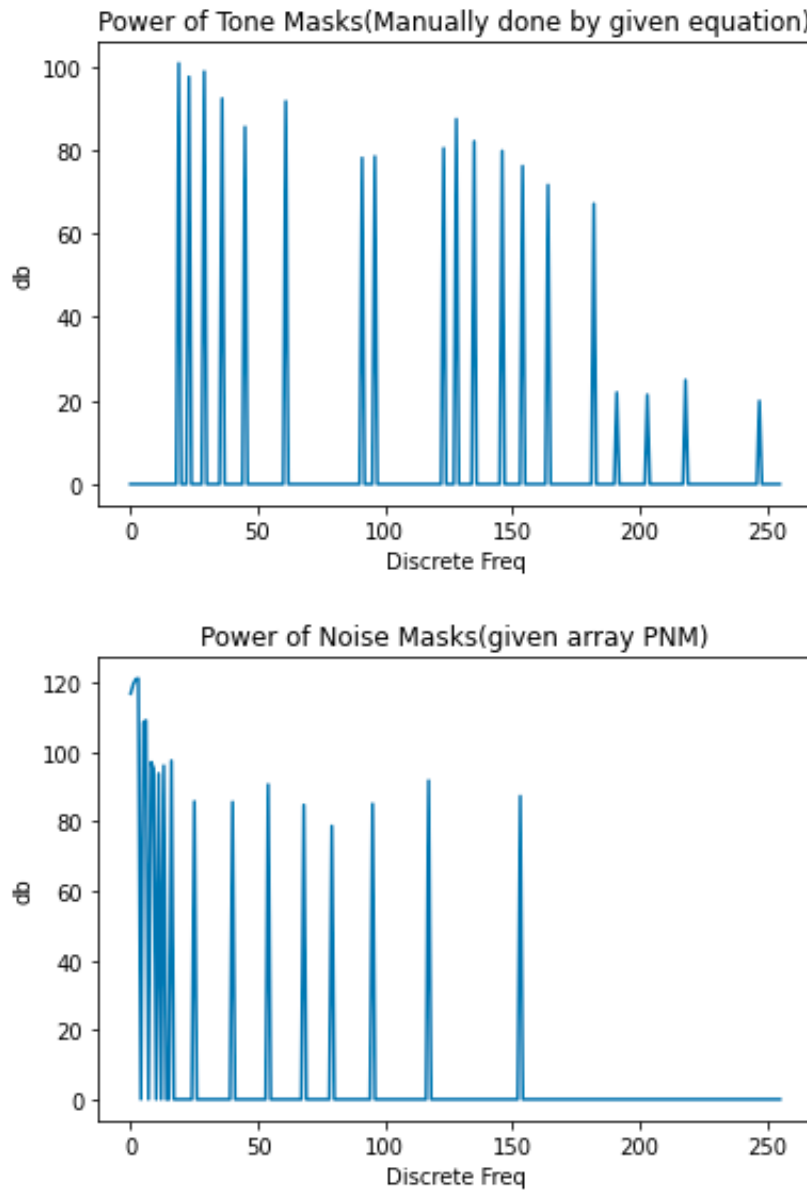


Αξίζει να σημειωθεί ότι το power spectrum καταλήγει με 256 δείγματα, παρά τα 512 που έχει το αρχικό σήμα (ανά πλαίσιο ανάλυσης). Αυτό συμβαίνει, λόγω της συμμετρίας του fft, ο οποίος πραγματοποιείται μεταξύ των σημάτων  $w(n)$  και  $x(n)$ , γεγονός που μας οδηγεί στην απόρριψη των αποτελεσμάτων  $P(k)$  για  $k > 255$ , αφού αυτά ταυτίζονται με τα αποτελέσματα για  $k$  μεταξύ 0 και 255.

1.2) Αρχικά βρίσκουμε την  $ST(k)$ , δηλαδή τις συχνότητες-ανά πλαίσιο- οι οποίες λειτουργούν ως μάσκες μέσω της (5) και με την βοήθεια των (4) και (6):



Στην συνέχεια, οι (4) και (5) μας οδηγούν στην ισχύ των τονικών масκών, δηλαδή στην υλοποίηση της εξ. (7), ενώ δείχνουμε και την ισχύ των масκών θορύβου για το ίδιο πλαίσιο:



Παρατηρούμε ότι οι μάσκες θορύβου είναι πολλές σε πλήθος και μεγάλες σε ισχύ σε χαμηλές discrete συχνότητες, πράγμα εύλογο αν αναλογισθούμε ότι ο θόρυβος αποτελεί ένα βαθυπερατό σήμα. Απεναντίας, οι τονικές μάσκες είναι πιο ομοιόμορφα κατανεμημένες στο συχνοτικό φάσμα.

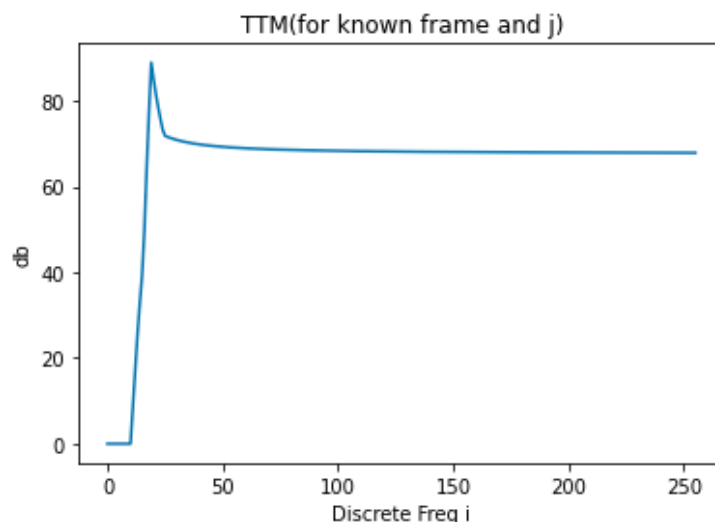
1.3) Για τα παρακάτω ερωτήματα, καλούμαστε να χρησιμοποιήσουμε τους πίνακες PNM\_c και PTM\_c που αποτελούν τους πίνακες ισχύος των масκών έπειτα από μείωση και αναδιοργάνωση:

Σημείωση: Οι εν λόγω πίνακες έπρεπε να αναστραφούν (μέσω της εντολής *np.transpose*) καθώς και να προστεθούν σε αυτούς μηδενικά για  $k < 3$  και  $k > 250$ , μιας και σε κάποια πλαίσια παρατηρήσαμε ότι οι εν λόγω θέσεις είχαν τιμή διάφορη του μηδενός.

1.4) Οι εξισώσεις (8) και (9) – και μέσω της (10) – θα μας δώσουν τα κατώφλια κάλυψης θορύβου και τόνων για κάθε πλαίσιο. Πιο αναλυτικά, για το 1ο πλαίσιο, καλούμαστε να αποθηκεύσουμε ποιες από τις μάσκες  $j$  του ερωτ 1.2 πληρούν την προϋπόθεση  $PTM(j) > 0$  και στην συνέχεια-για κάθε κατάλληλο  $j$ -επιλέγουμε τα  $i$  για τα οποία ισχύει  $b(i) \in [b(j) - 3, b(j) + 8]$ . Για τα εν λόγω ζεύγη  $(i, j)$  εφαρμόζουμε τις (8), (9) και (10). Τέλος αποθηκεύουμε τόσο τα  $j$ 's ανά πλαίσιο όσο και τα  $i$ 's που “καλύπτονται” από κάθε τέτοιο  $j$ . Για το πλαίσιο  $k=0$ , παρατηρούμε ότι υφίστανται οι κάτωθι μάσκες:

[19, 23, 29, 36, 45, 61]

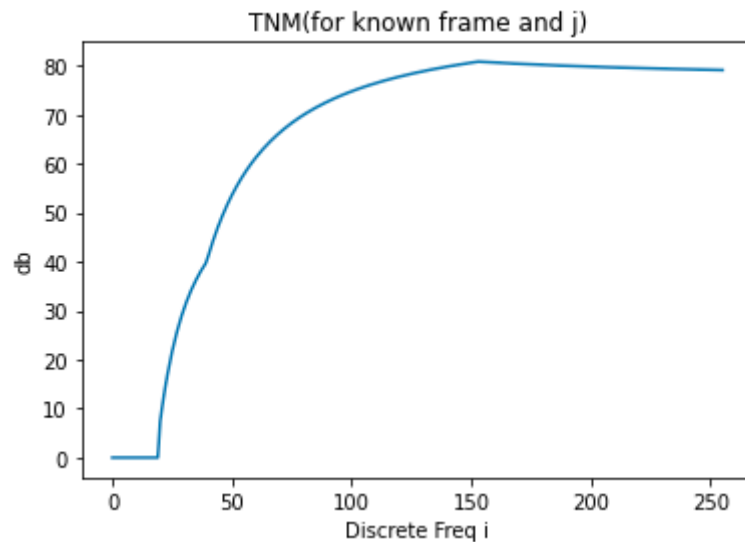
Επιλέγουμε  $j=19$  και πλοτάρουμε το TTM για όλα τα  $i$ 's (με τιμές  $> 0$  στα  $i$ 's που επικαλύπτει η εν λόγω μάσκα)



Ομοίως, λειτουργούμε και για το TNM:

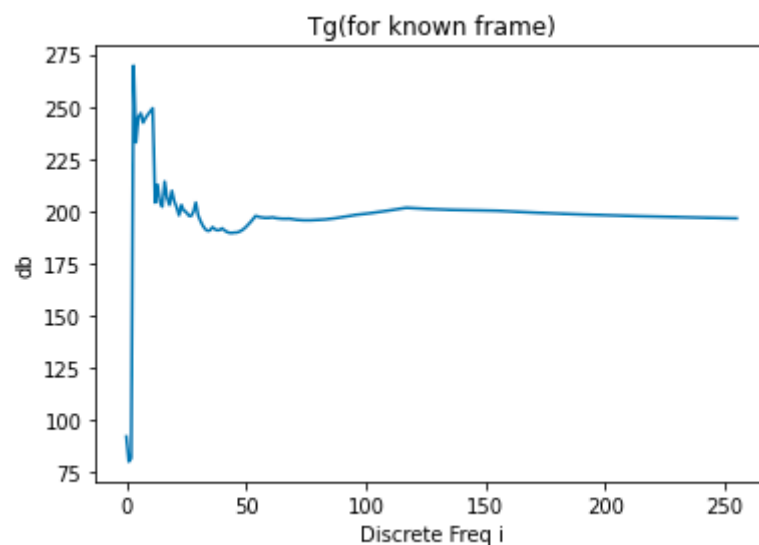
[3, 5, 6, 8, 9, 11, 13, 16, 25, 40, 54, 68, 79, 95, 117, 153]

Επιλέγουμε  $j=153$  και έχουμε:



Προφανώς, η παραπάνω διαδικασία οπτικοποίησης των αποτελεσμάτων λειτουργεί και ανάποδα, δηλαδή θα γινόταν να επιλέξουμε μία σταθερή συχνότητα  $i$  και να αναζητήσουμε τις  $j$  συχνότητες που λειτουργούν ως μάσκες της.

1.5) Με βάση τα αποτελέσματα της 1.4 υπολογίζουμε το συνολικό κατώφλι μέσω της (11). Για το 1ο πλαίσιο, έχουμε:





Επιπλέον, η γραφική  $Tg(i)$  δείχνει ότι η δυσκολία αυτή αφορά χαμηλές συχνότητες (που αντιστοιχούν σε χαμηλή τιμή του δείκτη  $i$ ), δηλαδή συχνότητες στις οποίες επικρατούν οι θορυβικές συνιστώσες. Απεναντίας, στην ενδιάμεση ζώνη συχνοτήτων έχουμε σταθερή μείωση του συνολικού κατωφλίου, άρα και διευκόλυνση στην ακοή της εκάστοτε πληροφορίας.

Με βάση τις συνθήκες (1) και (2), πραγματοποιήθηκε η μείωση και αναδιοργάνωση των μασκών, ώστε να βγάλουμε τους δικούς μας πίνακες PNMC και PTMC. Αποτελέσματα θα αναδειχθούν για το πρώτο πλαίσιο ανάλυσης:

[illegible]

Ο δοθέν PTMC πίνακας του βήματος 1.3:

[illegible]

Ο δικός μας PNMС πίνακας:

[illegible]

Ο δοθέν PNMС πίνακας από το βήμα 1.3:

[illegible]

Παρατηρούμε ότι-παρά την σχετική ομοιότητα μεταξύ δοθέντων και προκυπτόντων αποτελεσμάτων-δεν έχουμε πλήρη ταύτιση, γεγονός που πιθανότατα σχετίζεται με διαφορά στην λήψη των 0.5 Bark παραθύρων. Εξού και για τα βήματα 1.4 και 1.5 προτιμήθηκαν οι βοηθητικοί πίνακες PTMC και PNM, ώστε να είναι εγγυημένη η αξιοπιστία των επόμενων αποτελεσμάτων.

Σημείωση: Η ζητούμενη συνάρτηση *checkMasks* έσπασε σε 2 επιμέρους, τις *checkMasksPTM* και *checkMasksPNM* όπου η πρώτη οδηγεί στον PTMC πίνακα και η δεύτερη στον PNM πίνακα

## Μέρος 2

2.0) Σε αυτό το βήμα υπολογίζουμε τις συστοιχίες φίλτρων  $h_k[n]$  και  $g_k[n]$ , βάση των εξισώσεων (12) και (13), για  $0 \leq n \leq L-1$  και  $0 \leq k \leq M-1$ , όπου  $M=32$  (ο αριθμός των φίλτρων και  $L=2*M$ ).

2.1) Σε αυτό το βήμα κάνουμε συνέλιξη στο κάθε πλαίσιο του αρχικού σήματος με τα φίλτρα σύνθεσης ( $v_k(n)=h_k(n)*x(n)$ ) και έπειτα υποδειγματοληπτούμε με παράγοντα  $M=32$ , ώστε να πάρουμε το  $y_k[n]=v_k[n*M]$ .

2.2) Εδώ υλοποιούμε τον προσαρμοζόμενο κβαντιστή και έπειτα τον κβαντιστή 8-bit. Η διαδικασία και για τους 2 κβαντιστές είναι ίδια, με μόνες διαφορές τα levels του κβαντιστή και το βήμα, όπου στον προσαρμοσμένο αλλάζει ανάλογα με το σήμα. Η διαδικασία κβάντισης είναι η εξής:

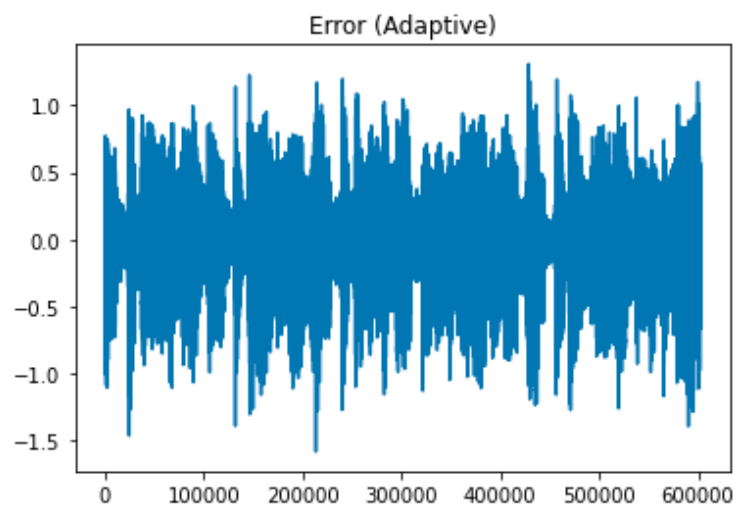
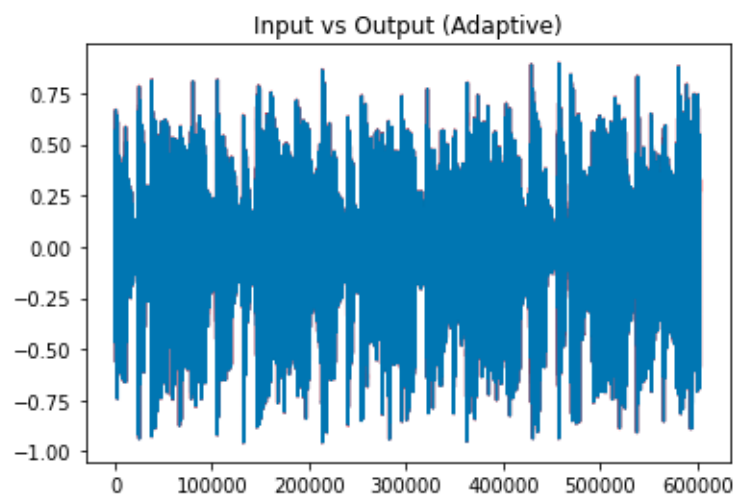
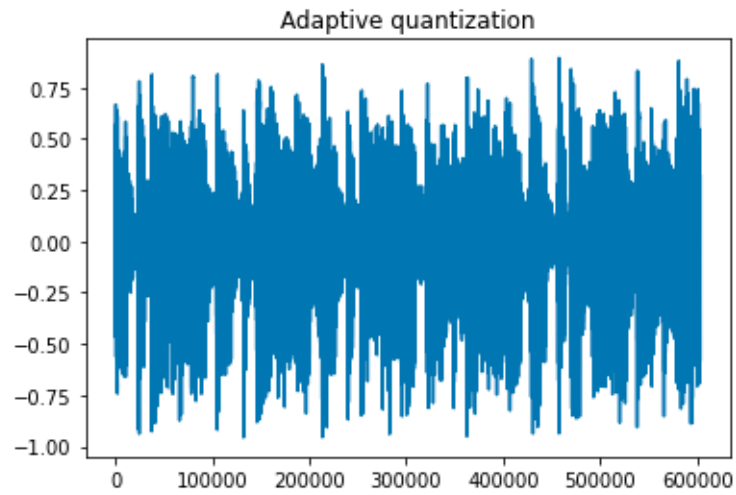
- υπολογίζουμε (για τον προσαρμοσμένο τα  $Tg(i)$  όπου  $Tg$  το συνολικό κατώφλι κάλυψης του ψυχοακουστικού μοντέλου όπως προέκυψε στο Μέρος 1
- υπολογισμός του  $b_k$  βάση της εξίσωσης (15) (στον 8bit κβαντιστή το ορίζουμε ως 8)
- υπολογισμός του βήματος  $\Delta$  του κβαντιστή ως  $(x_{\max}-x_{\min})/\text{levels}$ , όπου να επίπεδα είναι το τετράγωνο της αντίστοιχης τιμής του  $b_k$ .
- βάση των παραπάνω τιμών κβάντιση του κάθε πλαισίου

2.3) Μετά την κβάντιση το σήμα μας στέλνεται στον αποκωδικοποιητή, όπου υπερδειγματοληπτούμε κατά παράγοντα 32 το σήμα μας σύμφωνα με τη σχέση (16) και έπειτα συνελίσσουμε το σήμα μας με τα φίλτρα ανακατασκευής όπως τα ορίσαμε στο βήμα (2.0). Τέλος βάση της μεθόδου overlap-and-add ανακατασκευάζεται το σήμα μας.

## Αποτελέσματα

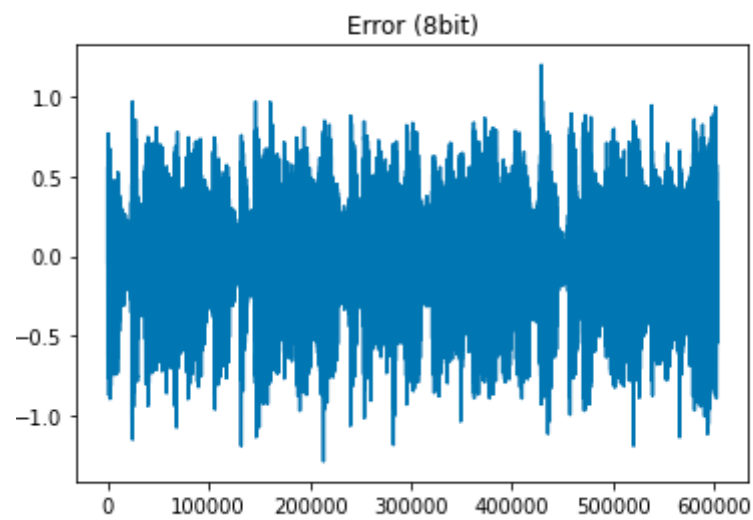
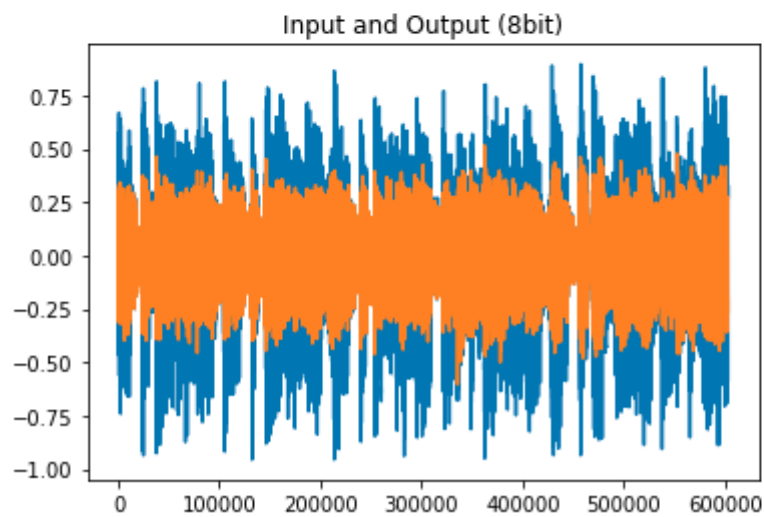
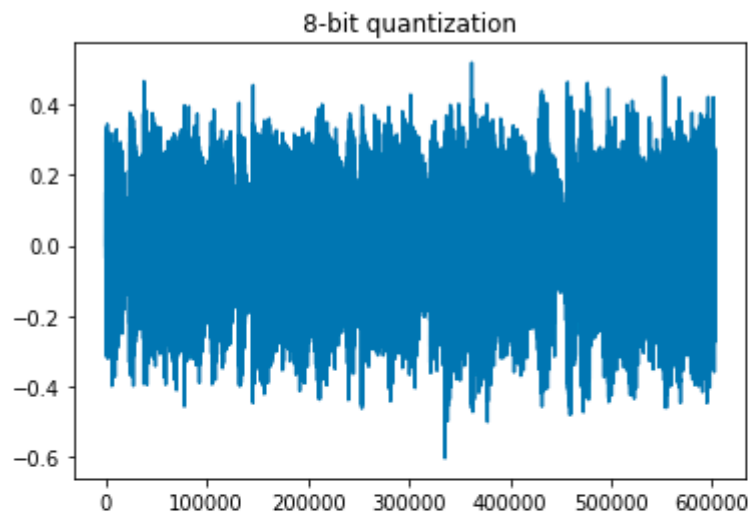
Mean Square error is: 0.09890486663729042

Compression rate for adaptive quantizer is: 52.4179768284805 %



Mean Square error is: 0.07192241012191379

Compression rate for adaptive quantizer is: 56.236107312464256 %



Παρατηρούμε από τα αποτελέσματα πως ο 8 bit decoder έχει μεγαλύτερο compression rate, ωστόσο το αποτέλεσμα είναι πολύ πιο “τεχνητό”. Αυτό συμβαίνει γιατί διώξαμε πολλά bits, ωστόσο κάποια από αυτά ήταν χρήσιμα και επηρεάζουν την ποιότητα του τελικού ήχου. Ο προσαρμοζόμενος κβαντιστής από την άλλη, μπορεί να έχει χαμηλότερο compression rate, ωστόσο τα bits που διώξαμε με τον κβαντιστή είναι bits που δεν χρησιμεύουν στο τελικό αποτέλεσμα και άρα η ποιότητα του ήχου είναι σαφώς καλύτερη.