

Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

2η Εργαστηριακή Άσκηση

Μάθημα: Ψηφιακή Επεξεργασία Σήματος

Εξάμηνο: 6^ο

Ονοματεπώνυμο: Αλεξοπούλου Γεωργία, Γκενάκου Ζωή

Θέμα: Κωδικοποίηση σημάτων Μουσικής βάσει του ψυχοακουστικού μοντέλου (Perceptual Audio Coding)

Μέρος 1ο - Ψυχοακουστικό Μοντέλο 1

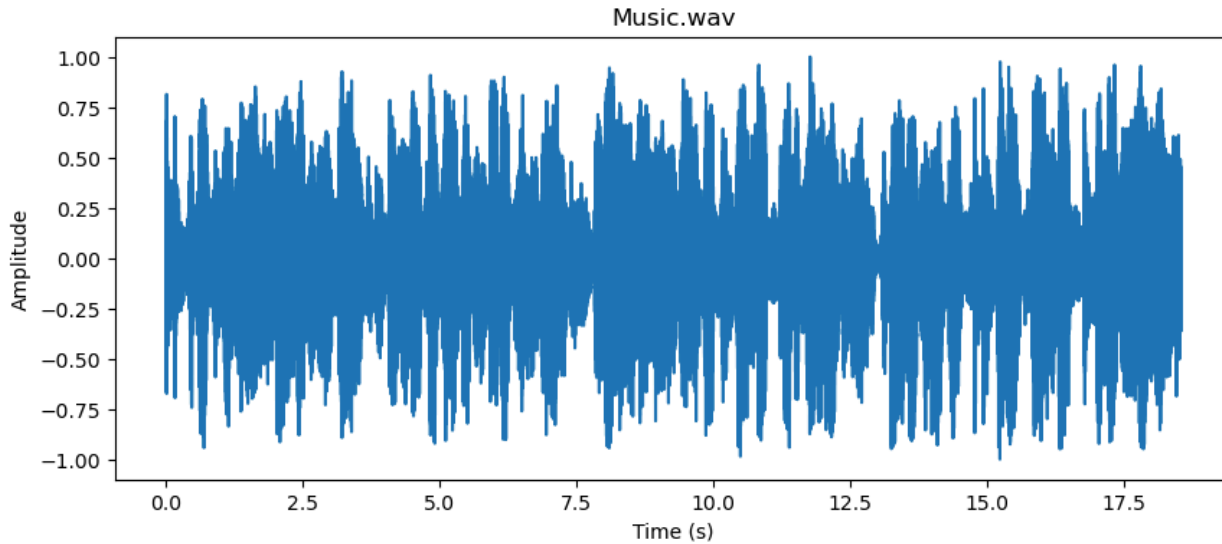
Βήμα 1.0: Προεπεξεργασία του σήματος

Αρχικά, διαβάζουμε το σήμα ήχου “music.wav”, αφού πρωτίστως το μετατρέψουμε από stereo σε mono format (συνάρτηση `convert_to_mono`), χρησιμοποιώντας τη συνάρτηση `read_wav_file`, η οποία επιστρέφει τον ρυθμό δειγματοληψίας (`sample_rate`) και τα κανονικοποιημένα δεδομένα (`normalized_data`). Για την κανονικοποίηση του σήματος, αξιοποιείται η συνάρτηση `normalize_signal`. Στη συνέχεια, παραθυροποιούμε το σήμα χρησιμοποιώντας 512 δείγματα και παράθυρα Hamming. Έτσι, στην έξοδο λαμβάνουμε τα εξής αποτελέσματα σχετικά με τις τιμές τόσο του ρυθμού δειγματοληψίας, όσο και των τιμών των κανονικοποιημένων δεδομένων (πράγματι κυμαίνονται μεταξύ $[-1,1]$):

Sample Rate: 44100

Normalized Data: [-0.02382895 0.02011556 0.06694087 0.10487403 0.12212546 0.12345761
0.12239189 0.13889398 0.17847568 0.21547633]

Έπειτα, απεικονίζουμε το κανονικοποιημένο σήμα με χρήση της εντολής `plt.plot()` και προκύπτει το παρακάτω γράφημα:



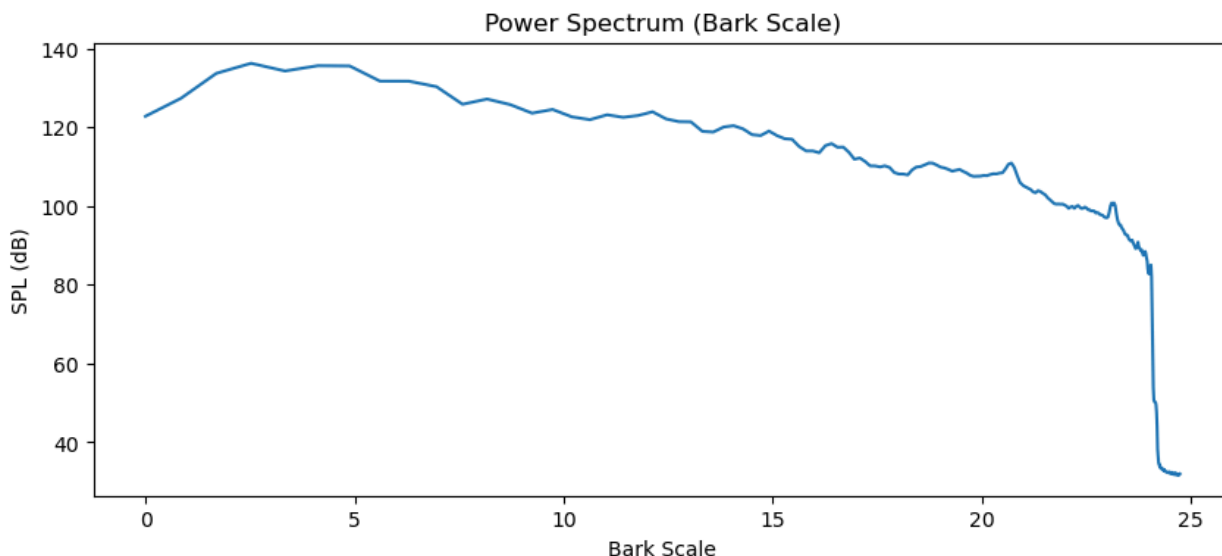
Βήμα 1.1: Φασματική Ανάλυση

Σε αυτό τον κώδικα, η συνάρτηση "bark_scale" υπολογίζει τις τιμές της κλίμακας Bark για μια συχνότητα που δίνεται, βάσει της εξίσωσης που παρέχεται. Απλώς, η συνάρτηση αυτή εφαρμόζει τη μαθηματική λογική της εξίσωσης για να υπολογίσει τις τιμές της κλίμακας Bark.

Από την άλλη πλευρά, η συνάρτηση "calculate_power_spectrum" λαμβάνει ως είσοδο το σήμα και το μέγεθος του παραθύρου. Στη συνέχεια, χρησιμοποιεί ένα παράθυρο Hanning και την εξίσωση για το μονόπλευρο φάσμα ισχύος που δίνεται, προκειμένου να υπολογίσει το φάσμα ισχύος του σήματος. Ουσιαστικά, η συνάρτηση αυτή εφαρμόζει τις αντίστοιχες μαθηματικές εξισώσεις και πράξεις για να παράξει το φάσμα ισχύος του σήματος, χρησιμοποιώντας το συγκεκριμένο παράθυρο και την εξίσωση που δίνεται.

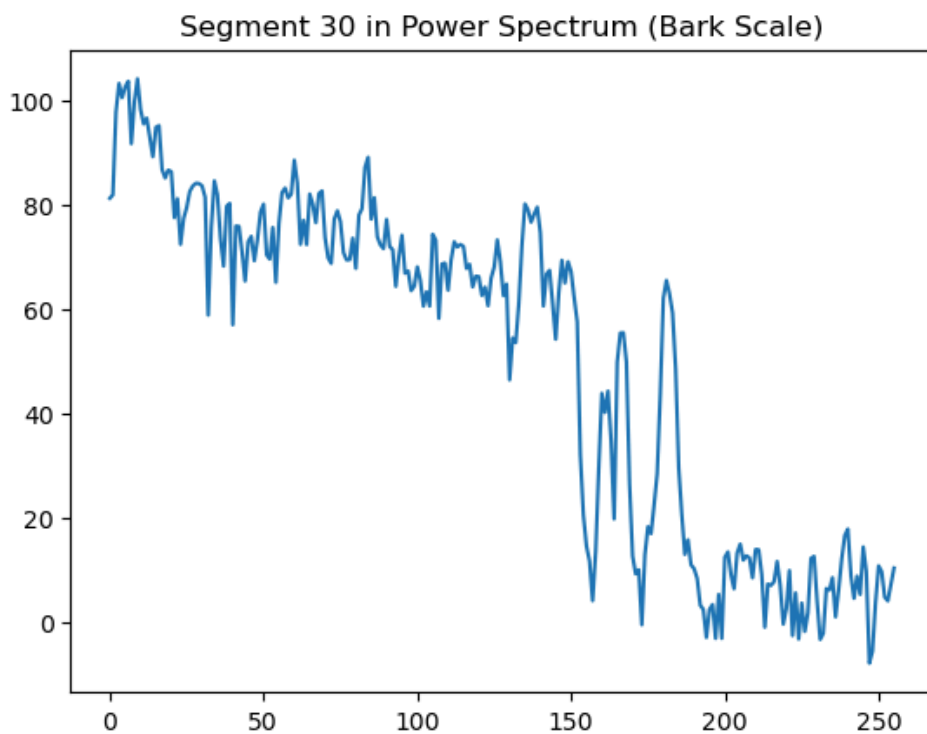
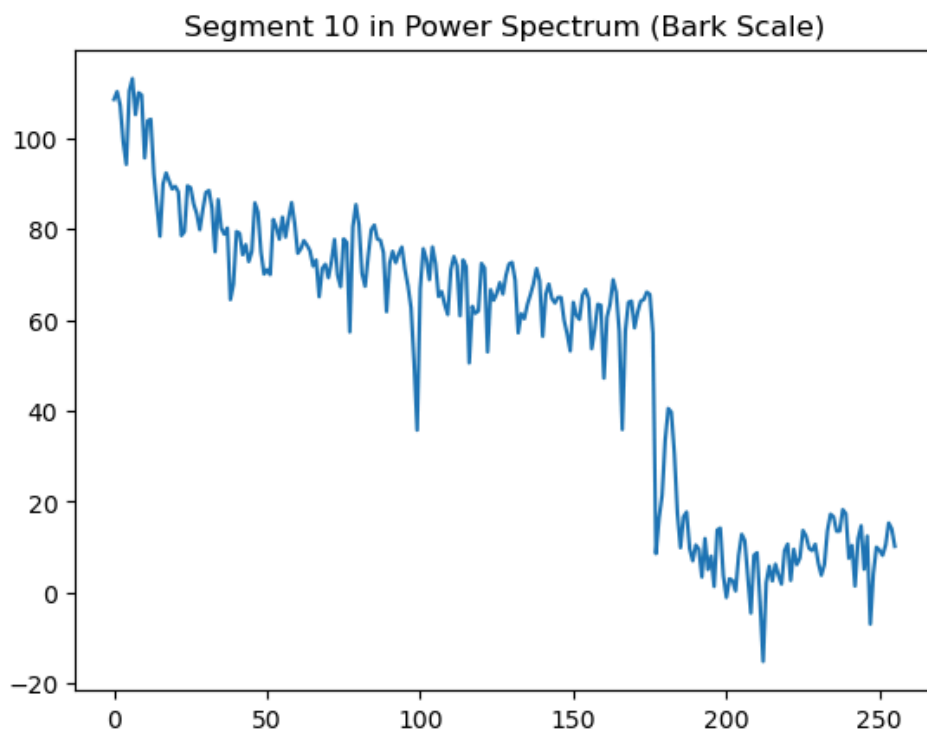
Ουσιαστικά, και οι δύο συναρτήσεις χρησιμοποιούν μαθηματικές εξισώσεις και λογική για να υπολογίσουν και να παράξουν τις απαιτούμενες τιμές και φάσματα στο πλαίσιο της επεξεργασίας του σήματος.

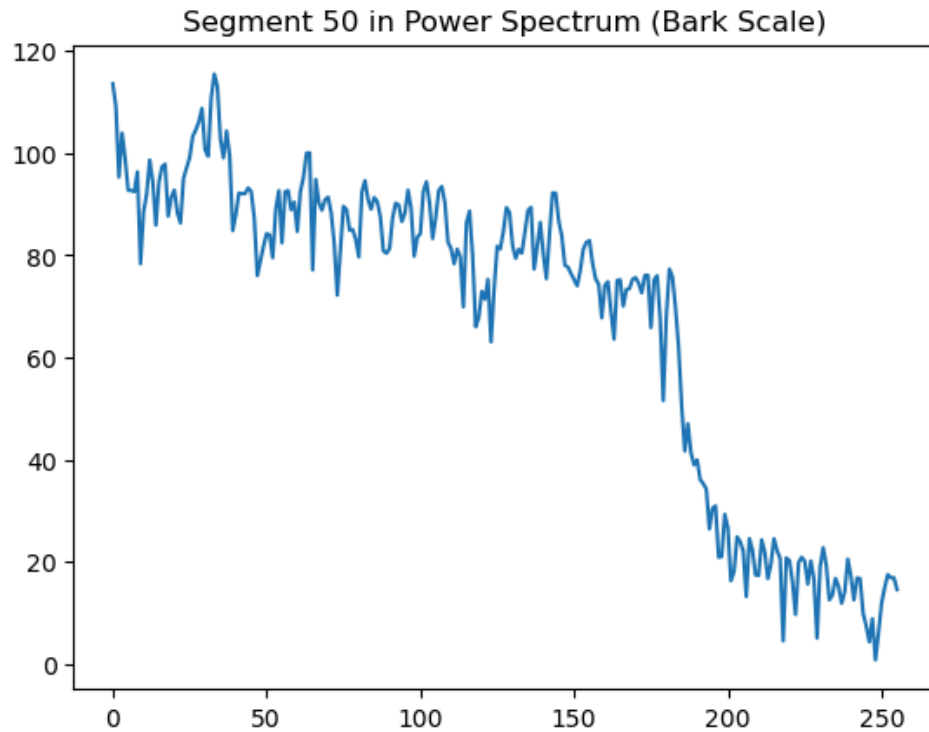
Έτσι, η απεικόνιση του φάσματος SPL (χωρίς παραθυροποίηση) φαίνεται παρακάτω:



Για να αποκτήσουμε το φάσμα ισχύος του σήματος σε κλίμακα Bark με περισσότερη ακρίβεια για συγκεκριμένα πλαίσια (π.χ. το 10ο, το 30ο και το 50ο), πραγματοποιούμε μερικές τροποποιήσεις στη συνάρτηση "calculate_power_spectrum" στον παραπάνω κώδικα και εκτυπώνουμε το φάσμα ισχύος για κάθε ένα από αυτά τα πλαίσια.

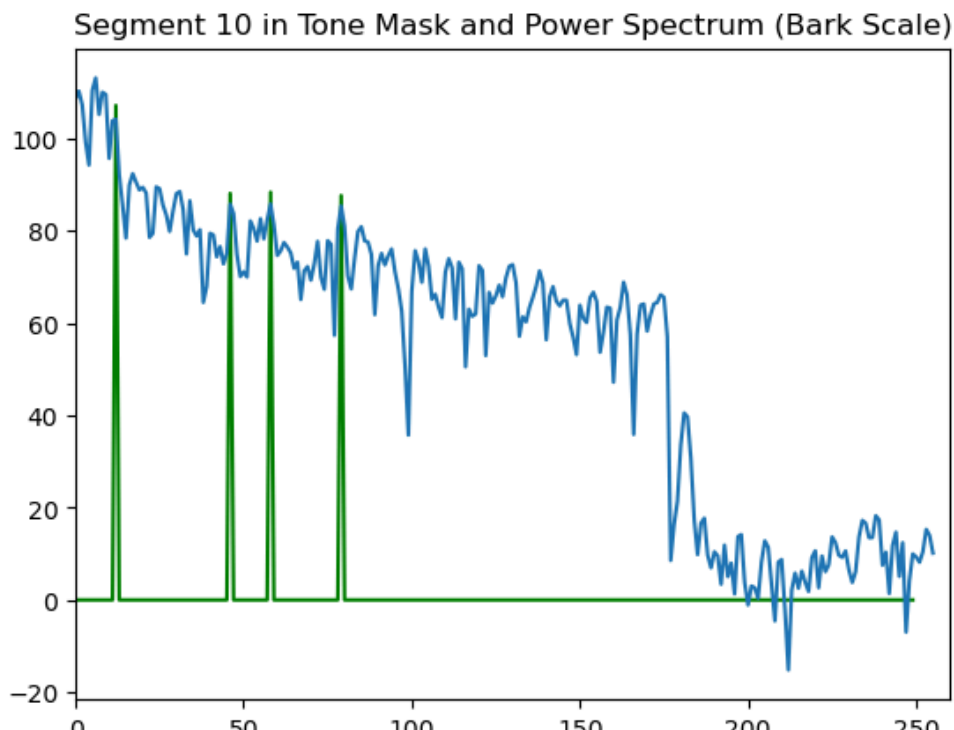
Αυτές οι αλλαγές περιλαμβάνουν την προσθήκη λειτουργιών για τη διαίρεση του σήματος σε παράθυρα και την εφαρμογή παραθύρου Hanning. Στη συνέχεια, υπολογίζουμε το φάσμα ισχύος για κάθε ένα από αυτά τα παράθυρα. Τέλος, εκτελούμε μερικούς υπολογισμούς και εκτυπώνουμε το φάσμα ισχύος για τα επιλεγμένα πλαίσια:

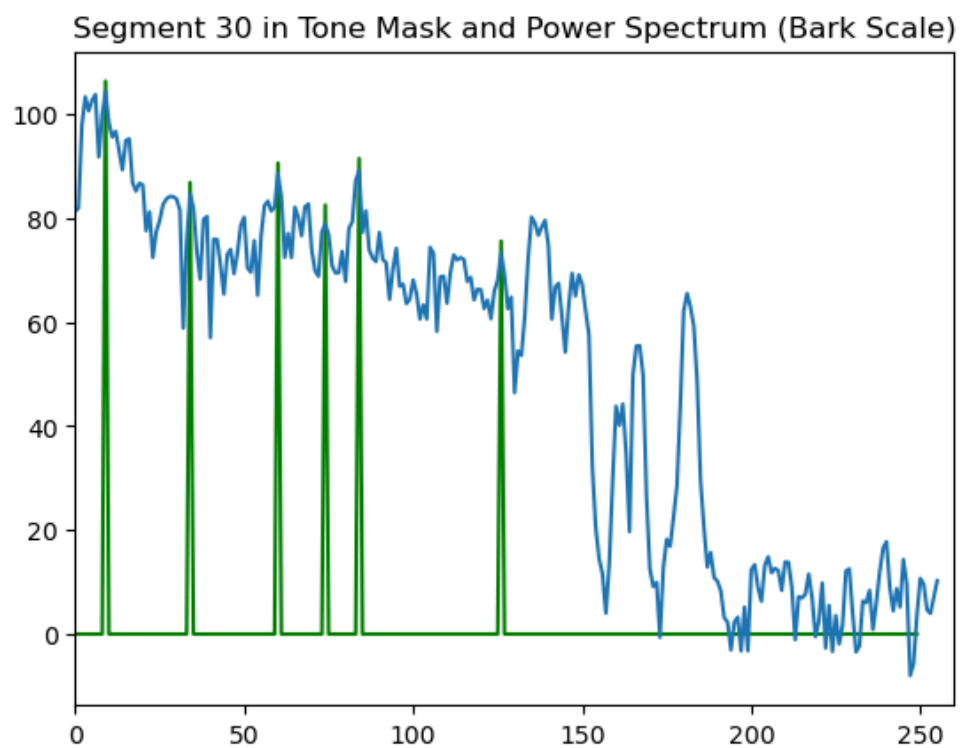




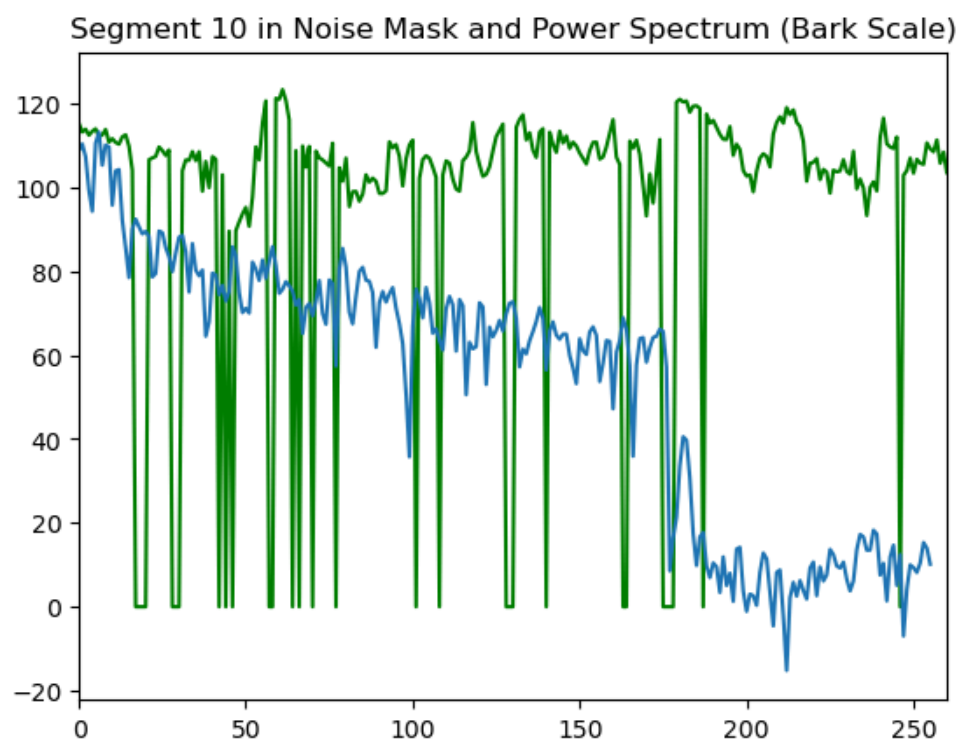
Βήμα 1.2: Εντοπισμός масκών τόνων και θορύβου (Maskers)

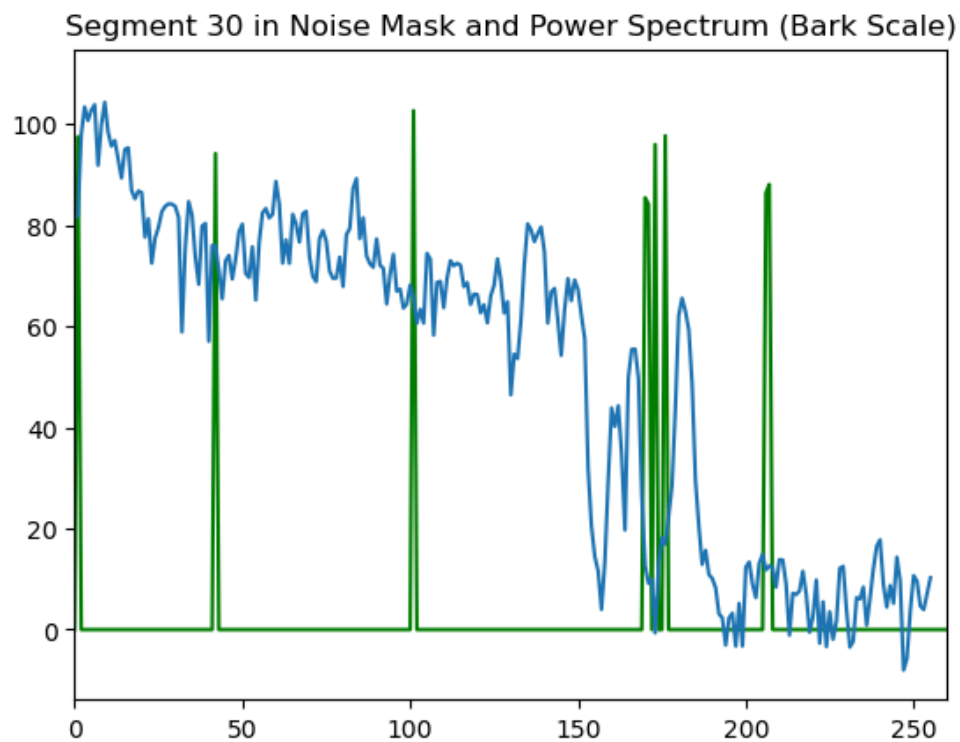
Σε αυτό το βήμα, πραγματοποιούμε τον υπολογισμό των τονικών масκών, που είναι τοπικά μέγιστα που υπερβαίνουν τις γειτονικές τους συχνότητες κατά τουλάχιστον 7dB. Η συνάρτηση $ST(k)$ επιστρέφει λογικές τιμές για το εάν η θέση k είναι τονική μάσκα. Στη συνέχεια, χρησιμοποιούμε τη συνάρτηση $PTM(k)$ για να υπολογίσουμε την ισχύ των τονικών масκών (εάν η θέση k δεν είναι τονική μάσκα, τότε η τιμή είναι 0). Για να είμαστε πλήρεις, κάνουμε plot το power spectrum του 10ου segment μαζί με το $P_TM[9]$, καθώς και για το 30ο segment αντίστοιχα.





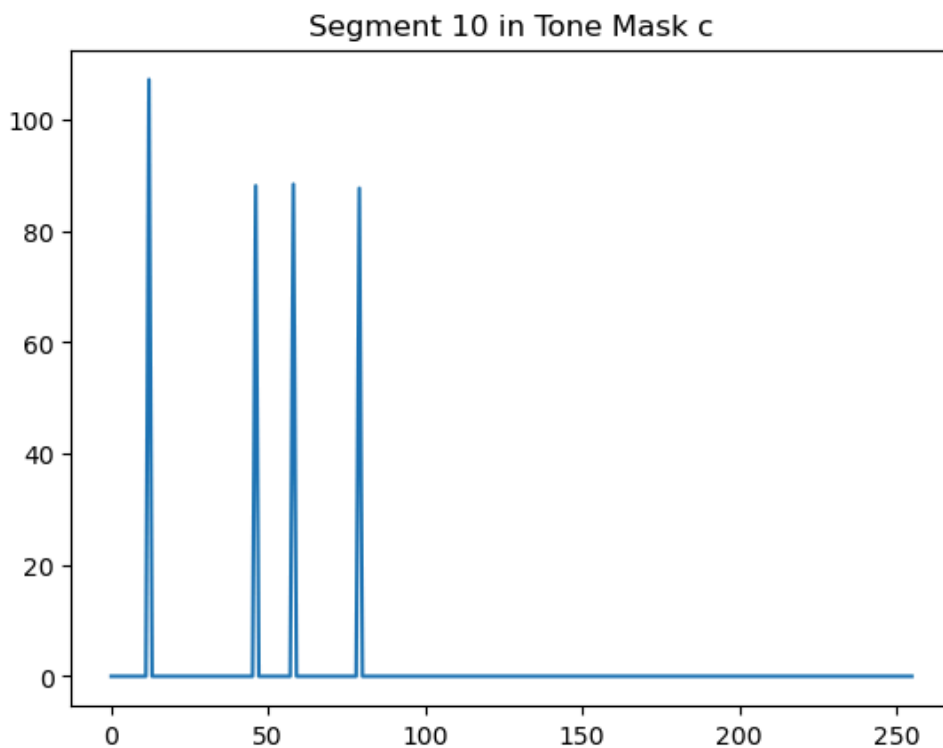
Την ίδια διαδικασία ακολουθούμε και για το P_{NM} :

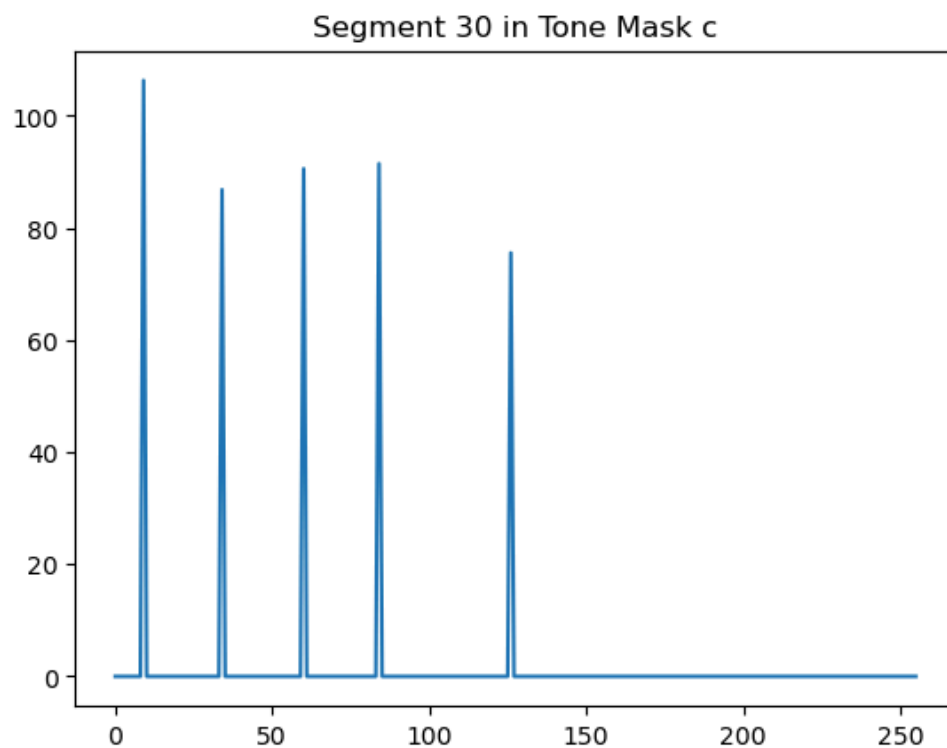




Βήμα 1.3: Μείωση και αναδιοργάνωση των масκών

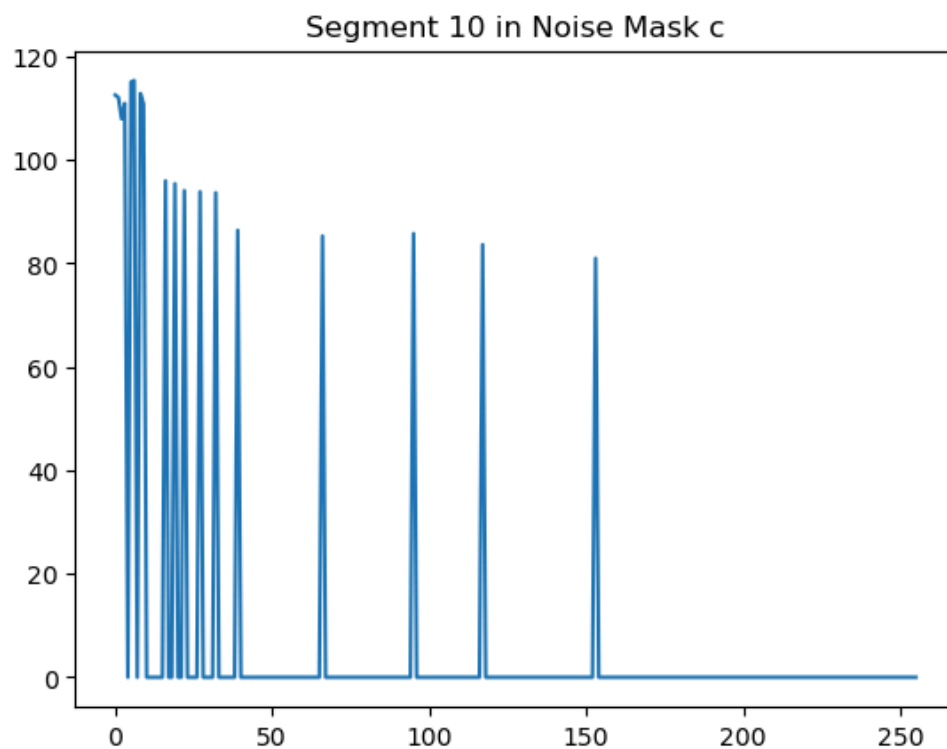
Για να εισάγουμε τα δεδομένα του αρχείου P_NM-2023, χρησιμοποιούμε τη συνάρτηση `np.load()`. Προκειμένου να διασφαλίσουμε τη σωστή λειτουργία του προγράμματος μας, εκτελούμε τη συνάρτηση "stem" για το P_TM (το οποίο αντιστοιχεί στο P_TM που υπολογίσαμε στο Βήμα 1.2) καθώς και για το P_TM_c.

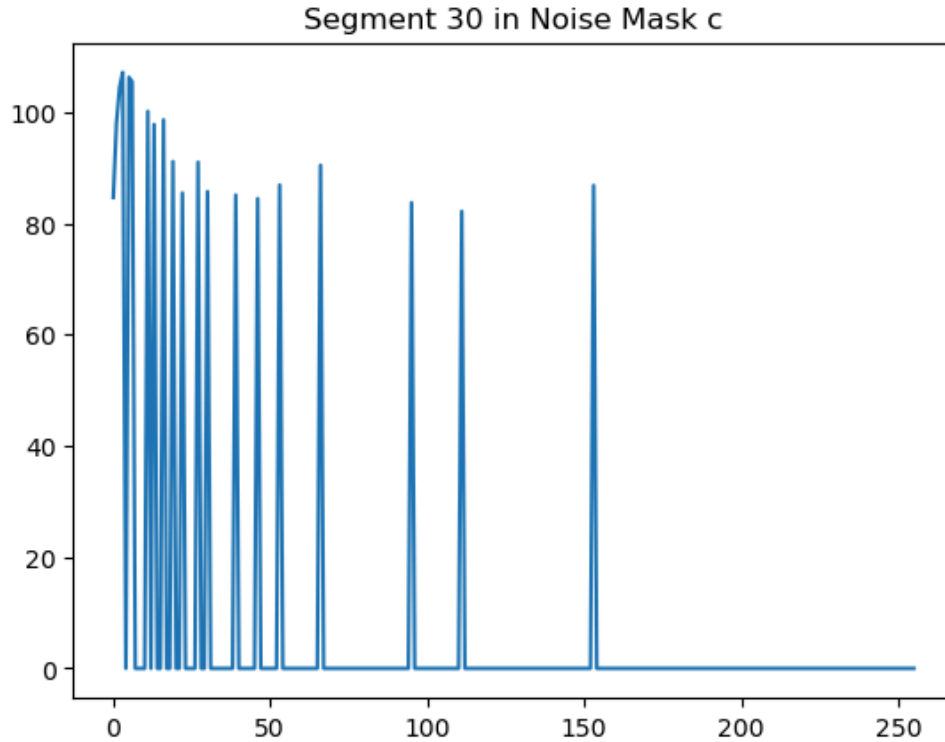




Παρατηρούμε ότι οι μουσικές συχνότητες που προκαλούν τονικές μάσκες έχουν μειωθεί, ενώ αυτές που παραμένουν φαίνονται στο P_{TM} .

Για να έχουμε μια πλήρη εικόνα, κάνουμε επίσης γραφικές παραστάσεις μερικών P_{Nmc} .





Βήμα 1.4: Υπολογισμός των δύο διαφορετικών κατωφλίων κάλυψης (Individual Masking Thresholds)

Η θεωρία του Global Masking Threshold (Καθολικού Ορίου Κάλυψης) αναφέρεται στον τρόπο με τον οποίο το ανθρώπινο αυτί αντιλαμβάνεται και μετατρέπει ηχητικά σήματα σε αντίληψη και αίσθηση του ήχου. Ο κώδικας που δίνεται χρησιμοποιεί αυτήν τη θεωρία για να υπολογίσει την αντίληψη του θορύβου και της έντασης του ήχου σε μια δεδομένη συνήθηση.

Η συνάρτηση $B(i)$ υπολογίζει την αξία του ορίου κάλυψης (masking threshold) για μια συχνότητα i . Η τιμή του ορίου κάλυψης εξαρτάται από τον ρυθμό δειγματοληψίας ($sample_rate$) και το μέγεθος του παραθύρου ($window_size$), τα οποία έχουμε υπολογίσει στο Βήμα 1.0. Η συνάρτηση $DB(i, j)$ υπολογίζει τη διαφορά του ορίου κάλυψης μεταξύ δύο συχνοτήτων i και j .

Η συνάρτηση $SF(P_Mask, i, j)$ χρησιμοποιείται για να υπολογίσει τον Individual Masking Threshold (Ατομικό Κατώφλι Κάλυψης) για δύο συχνότητες i και j . Ανάλογα με τη διαφορά $DB(i, j)$, επιλέγεται μια από τις τέσσερις περιπτώσεις για τον υπολογισμό του καθολικού ορίου κάλυψης, σύμφωνα με τη δοσμένη εξίσωση [10]. Οι συναρτήσεις $DB(i, j)$ και $SF(P_Mask, i, j)$ είναι υποσυναρτήσεις που χρησιμοποιούνται εντός της $SF(P_Mask, i, j)$.

Οι συναρτήσεις $T_TM(P_TM, i, j)$ και $T_NM(P_NM, i, j)$ υπολογίζουν τον Θόρυβο Καθολικού Ορίου (T_TM) και τον Μη Θόρυβο Καθολικού Ορίου (T_NM) αντίστοιχα. Αυτές οι συναρτήσεις χρησιμοποιούν τις

παραμέτρους P_{TM} και P_{NM} , που αντιστοιχούν στον θόρυβο και τον μη θόρυβο θερμαντήρα για μια συνήθηση.

Ο κώδικας επαναλαμβάνει τους υπολογισμούς για διάφορες τιμές των P_{TM} και P_{NM} σύμφωνα με τις εξισώσεις [8] και [9] και αποθηκεύει τα αποτελέσματα στις λίστες T_{TM} και T_{NM} αντίστοιχα. Τέλος, ο αριθμητής counter αυξάνεται κατά 1 όταν k είναι ίσο με 10.

Βήμα 1.5: Υπολογισμός του συνολικού κατωφλίου κάλυψης (Global Masking Threshold)

Σε αυτό το βήμα, αξιοποιώντας τους υπολογισμούς των κατωφλίων του *Βήματος 1.4*, υπολογίζουμε και απεικονίζουμε το συνολικό κατώφλι, όπως αυτό περιγράφεται από την εξίσωση [11].

Πιο συγκεκριμένα, σε αυτό το κομμάτι του κώδικα, συνάρτηση $T_q(i)$ υπολογίζει την αντίληψη της έντασης ήχου σε dB SPL για μια δεδομένη συχνότητα i . Η τιμή υπολογίζεται με βάση την εξίσωση που αναφέρεται στην συνάρτηση.

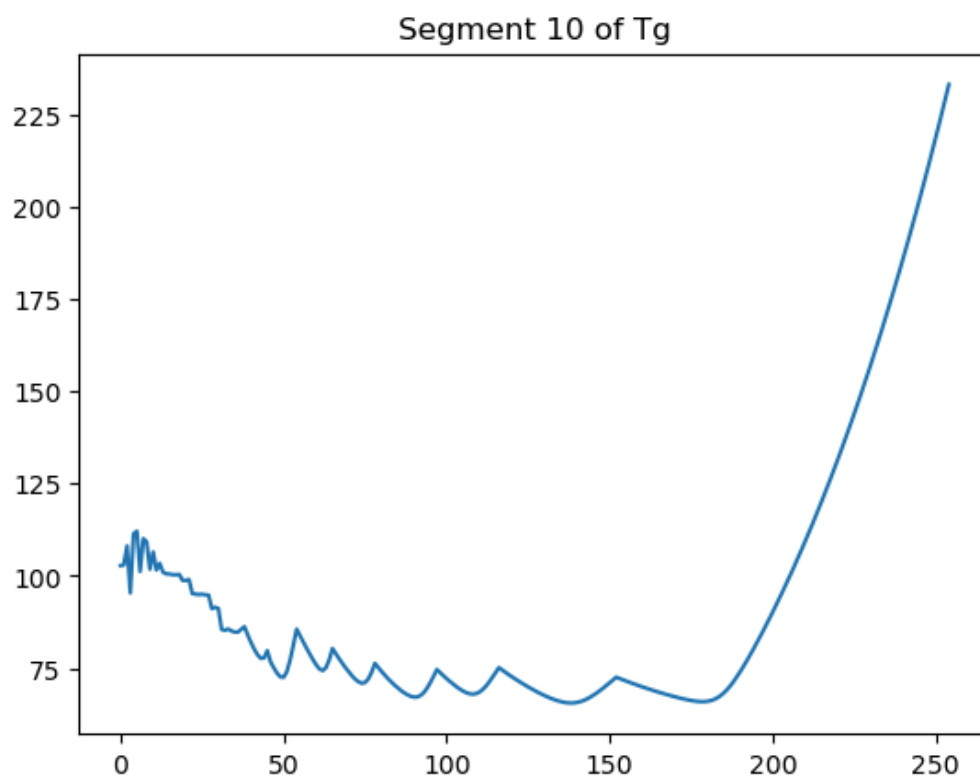
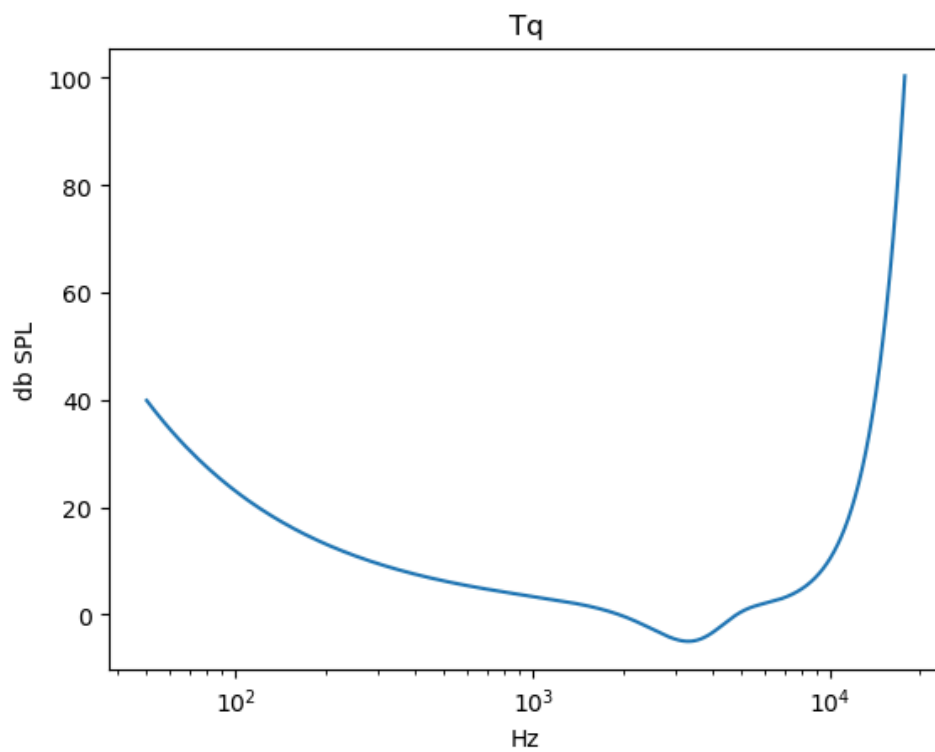
Η συνάρτηση $T_g(i, T_{TM}, T_{NM}, P_{TMC}, P_{NMC})$ υπολογίζει την αντίληψη του γενικού θορύβου (T_g) για μια δεδομένη συχνότητα i . Χρησιμοποιεί τις παραμέτρους T_{TM} και T_{NM} που υπολογίστηκαν στον προηγούμενο κώδικα, καθώς επίσης και τις παραμέτρους P_{TMC} και P_{NMC} που αντιστοιχούν στον θόρυβο και τον μη θόρυβο θερμαντήρα για μια συνήθηση. Οι τιμές T_{TM} και T_{NM} χρησιμοποιούνται για να υπολογιστεί το άθροισμα της ισχύος του θορύβου και του μη θορύβου για τις συχνότητες του T_g .

Η λίστα T_g αποθηκεύει τα αποτελέσματα της συνάρτησης T_g για διάφορες συχνότητες και συνήθησεις. Η συνάρτηση $T_g(i, T_{TM}[k], T_{NM}[k], P_{TMC}[k], P_{NMC}[k])$ καλείται για κάθε συχνότητα i και συνήθεια k και τα αποτελέσματα αποθηκεύονται στη λίστα T_g .

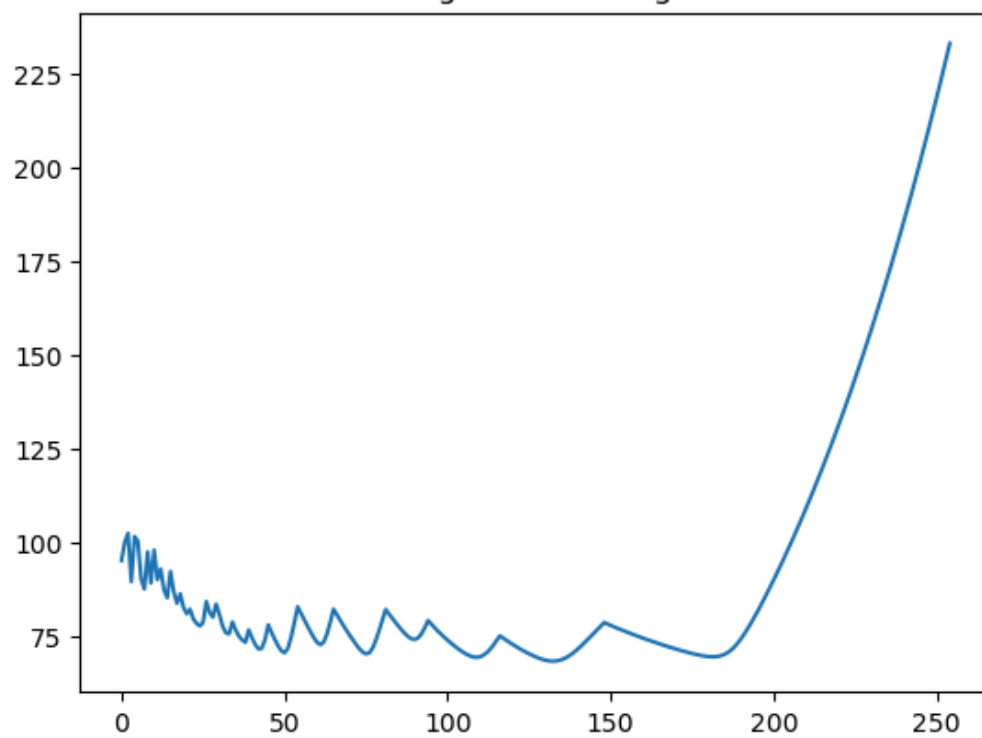
Στη συνέχεια, ο κώδικας προετοιμάζει για αποθήκευση τη λίστα T_g σε ένα αρχείο numpy με την εντολή `np.save()`, προκειμένου αυτό να χρησιμοποιηθεί ως δεδομένο εισόδου για το **Μέρος 2**. Ακολουθεί η παρουσίαση των αποτελεσμάτων χρησιμοποιώντας τη βιβλιοθήκη `matplotlib`. Πρώτα πραγματοποιείται γραφική παράσταση της συνάρτησης $T_q(i)$ για μια σειρά από συχνότητες `freqs`, ενώ στη συνέχεια παρουσιάζονται γραφικές παραστάσεις για τα 10th και 30th segments της λίστας T_g .

Όπως ήδη αναφέραμε, τα σήματα T_q και T_g αναπαριστούν την αντίληψη της έντασης ήχου σε διάφορες συχνότητες. Η συνάρτηση T_q υπολογίζει την αντίληψη του Global Masking Threshold, ενώ η συνάρτηση T_g υπολογίζει την αντίληψη του Global Masking Noise. Συνεπώς, η ομοιότητα μεταξύ των δύο σημάτων, η οποία γίνεται αντιληπτή παρακάτω, υποδεικνύει ότι η αντίληψη του γενικού θορύβου (T_g) μπορεί να επηρεάζεται από την αντίληψη του θορύβου καθολικού ορίου (T_q) σε διάφορες συχνότητες. Η ομοιότητα μεταξύ των δύο σημάτων μπορεί να υποδείξει κοινούς παράγοντες που επηρεάζουν την αντίληψη του θορύβου, καθώς και την σχέση μεταξύ της έντασης ήχου και της συχνότητας. Αυτό μπορεί να βοηθήσει στην κατανόηση των

μηχανισμών που εμπλέκονται στην αντίληψη του θορύβου από το ανθρώπινο αντί και στην ανάπτυξη αλγορίθμων επεξεργασίας ήχου που προσομοιώνουν την ανθρώπινη αντίληψη.

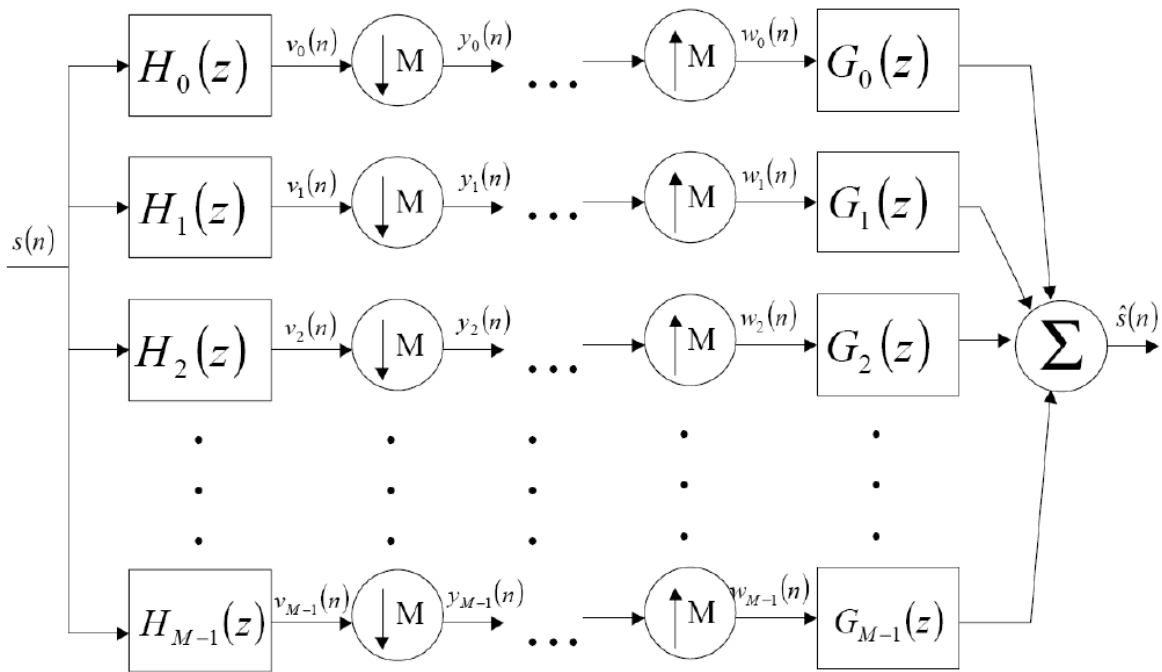


Segment 30 of Tg



Μέρος 2ο - Χρονο-Συχνотική Ανάλυση με Συστοιχία Ζωνοπερατών Φίλτρων

Στο Μέρος 2 φιλτράρουμε τα παράθυρα $x(n)$ του σήματος που εμπεριέχεται στο “music.wav” με τη συστοιχία φίλτρων ανάλυσης $h_k(n)$ και σύνθεσης $g_k(n)$. Στόχος είναι η δημιουργία μιας συνάρτησης που υλοποιεί τη διαδικασία του Σχ.3, η οποία παίρνει σαν είσοδο το κάθε πλαίσιο ανάλυσης $x(n)$, τη συστοιχία φίλτρων και το συνολικό κατώφλι κάλυψης που υπολογίστηκε στο Μέρος 1. Η έξοδος της συνάρτησης είναι το ανακατασκευασμένο σήμα $\hat{x}'(n)$ και ο αριθμός των bits που χρησιμοποιήθηκαν για τη δημιουργία του.



Σχήμα 3: Uniform M-Band Maximally Decimated Analysis-Synthesis Filterbank.

Βήμα 2.0: Συστοιχία Ζωνοπερατών Φίλτρων (Filterbank)

Στο πλαίσιο της ανάλυσης σήματος, η χρήση συστοιχιών φίλτρων ζώνης διέλευσης διατρέχει κρίσιμο ρόλο στην αποσύνθεση των σημάτων στα κρίσιμα συστατικά τους. Αυτές οι συστοιχίες φίλτρων έχουν σχεδιαστεί με βάση μια τροποποιημένη έκδοση του γνωστού διακριτού μετασχηματισμού που ονομάζεται Modified Discrete Cosine Transform (MDCT). Το MDCT έχει επιλεγεί για την αναστρεψιμότητα και τις ιδιότητες κωδικοποίησης χωρίς σφάλματα, καθιστώντας το ιδανική επιλογή για συστήματα συμπίεσης.

Σε αυτό το βήμα της υλοποίησης του συστήματος συμπίεσης, χρησιμοποιούμε filterbank βασισμένα σε MDCT που περιλαμβάνουν 32 φίλτρα ανάλυσης και 32 φίλτρα σύνθεσης. Τα φίλτρα ανάλυσης, που δηλώνονται ως « $h_k(n)$ », είναι υπεύθυνα για τη διάσπαση του σήματος εισόδου στα συστατικά στοιχεία του, ενώ τα φίλτρα σύνθεσης, που δηλώνονται ως « $g_k(n)$ », ανασυνθέτουν το αρχικό σήμα από αυτά τα στοιχεία.

Οι ειδικές συνθέσεις των φίλτρων ανάλυσης και σύνθεσης είναι οι εξής:

Φίλτρο ανάλυσης:

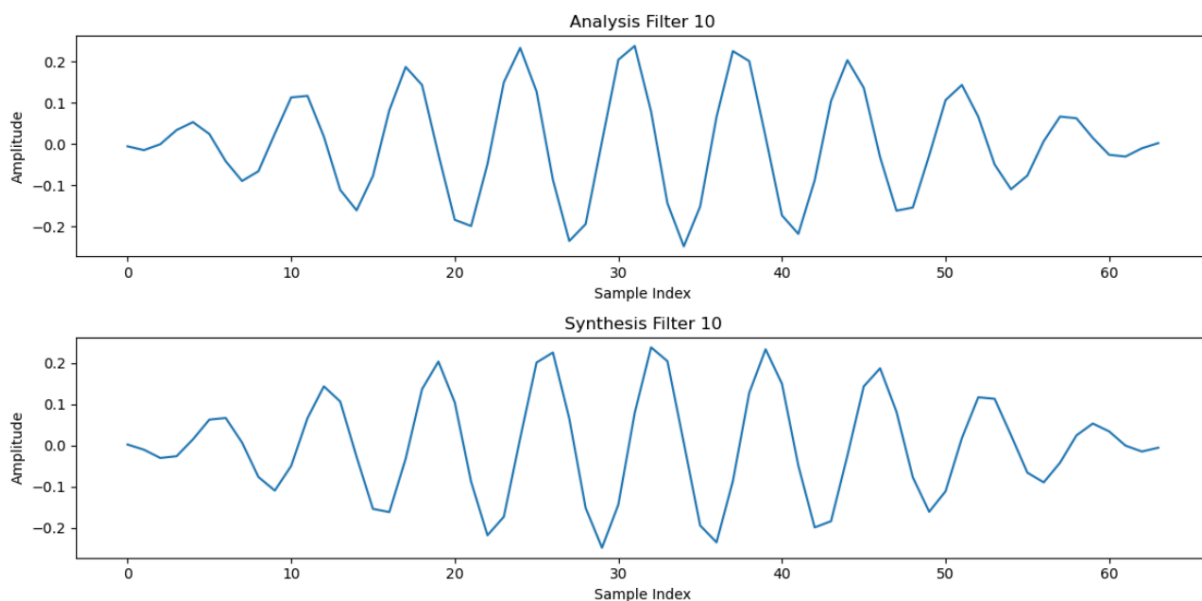
$$h_k(n) = \sin \left[\left(n + \frac{1}{2} \right) \frac{\pi}{2M} \right] \sqrt{\frac{2}{M}} \cos \left[\frac{(2n + M + 1)(2k + 1)\pi}{4M} \right]$$

Φίλτρο σύνθεσης:

$$g_k(n) = h_k(2M - 1 - n)$$

Και τα δύο φίλτρα ανάλυσης και σύνθεσης έχουν μήκος $L = 2M$. Ο όρος $\sin[(n + 1/2) * \pi / 2M]$ αντιπροσωπεύει ένα χαμηλοπερατό φίλτρο γνωστό ως "ημιτονοειδές παράθυρο" που χρησιμοποιείται για τη διαμόρφωση της απόκρισης συχνότητας του φίλτρου.

Στον κώδικα για το βήμα υλοποιούμε δύο συναρτήσεις που υπολογίζουν τα φίλτρα ανάλυσης και σύνθεσης και απεικονίζουμε γραφικά το 10ο φίλτρο ανάλυσης και το 10ο φίλτρο σύνθεσης από το σύνολο των σχεδιασμένων φίλτρων.



Βήμα 2.1: Ανάλυση με Συστοιχία Φίλτρων

Στο Βήμα 2.1 της υλοποίησης του συστήματος συμπίεσης, εστιάζουμε στην ανάλυση της συστοιχίας φίλτρων. Αυτό το βήμα περιλαμβάνει δύο κύριες λειτουργίες που εφαρμόζονται στο σήμα εισόδου « $x(n)$ » χρησιμοποιώντας τα φίλτρα σύνθεσης « $h_k(n)$ ». Ο στόχος είναι να διαιρεθεί το αρχικό σήμα στις συνιστώσες του χρόνου του, ελαχιστοποιώντας παράλληλα την παραποίηση του σήματος.

Η πρώτη λειτουργία στο Βήμα 2.1 είναι η συνέλιξη, όπου το σήμα « $x(n)$ » συνελίσσεται με κάθε φίλτρο σύνθεσης « $h_k(n)$ ». Το αποτέλεσμα της λειτουργίας συνέλιξης είναι το φιλτραρισμένο σήμα « $u_k(n)$ », το οποίο προκύπτει αθροίζοντας τα σταθμισμένα γινόμενα των δειγμάτων σήματος εισόδου και τους συντελεστές φίλτρου.

$$v_k(n) = h_k(n) * x(n) = \sum_{m=0}^{L-1} x(n-m)h_k(m), \quad k = 0, 1, \dots, M-1,$$

Η λειτουργία συνέλιξης μας επιτρέπει να εξαγάγουμε τα στοιχεία του σήματος εισόδου που σχετίζονται με κάθε φίλτρο σύνθεσης.

Η δεύτερη λειτουργία στο Βήμα 2.1 είναι η μείωση ή η υποδειγματοληψία. Το φιλτραρισμένο σήμα « $u_k(n)$ » υποβάλλεται σε δειγματοληψία με συντελεστή « M », με αποτέλεσμα τα χρονικά στοιχεία « $y_k(n) = v_k(Mn)$ ». Η μείωση δειγματοληψίας μειώνει τον αριθμό των δειγμάτων στο σήμα διατηρώντας παράλληλα τα σημαντικά στοιχεία συχνότητας. Αυτή η μείωση δειγματοληψίας εκτελείται στο μέγιστο δυνατό βαθμό χωρίς να εισάγονται εφέ παραφωνίας.

Στον κώδικα υπολογίζουμε και εκτυπώνουμε γραφικές παραστάσεις για την οπτικοποίηση διαφορετικών πτυχών του συστήματος συμπίεσης. Συγκεκριμένα:

1. Αρχικό σήμα:

- Αυτό το διάγραμμα δείχνει τις διακυμάνσεις πλάτους του αρχικού σήματος.
- -Παρέχει μια οπτική αναπαράσταση της ακουστικής κυματομορφής πριν από οποιαδήποτε επεξεργασία ή ανάλυση.

2. Σήμα με παράθυρο:

- Αυτό το διάγραμμα εμφανίζει τις παραλλαγές πλάτους ενός συγκεκριμένου πλαισίου παραθύρου του σήματος.
- Η λειτουργία παραθύρου (σε αυτή την περίπτωση, παράθυρο Hanning) εφαρμόζεται στο πλαίσιο, που τονίζει το κέντρο και μειώνει την πρόσκρουση των άκρων.
- Μας επιτρέπει να παρατηρούμε τα αποτελέσματα της λειτουργίας παραθύρου στο σήμα.

3. Φιλτραρισμένο σήμα:

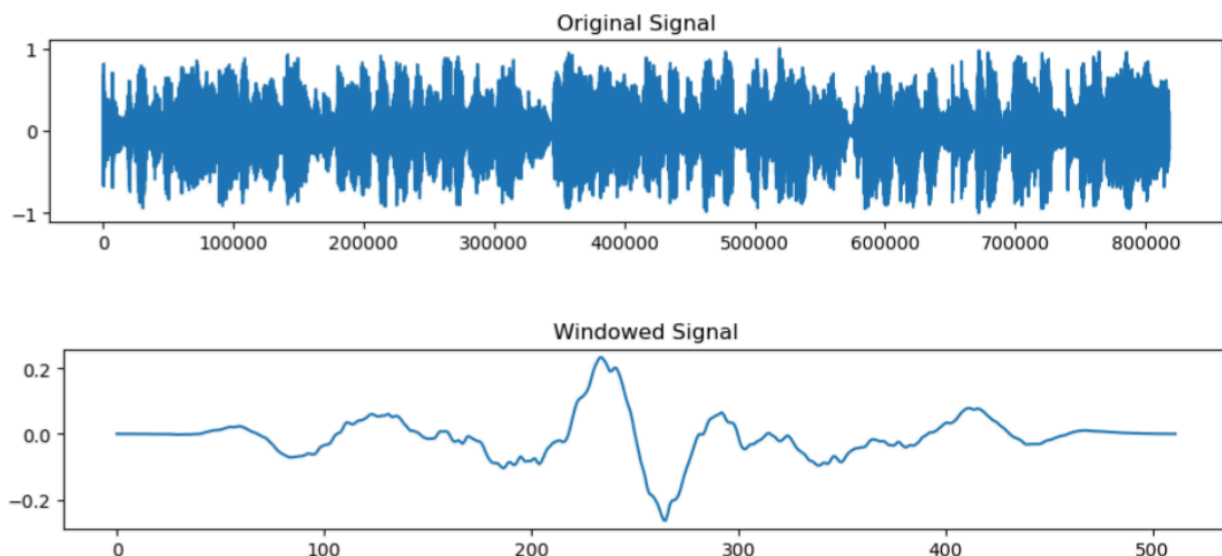
- Αυτή η γραφική παράσταση δείχνει τις διακυμάνσεις του πλάτους ενός συγκεκριμένου πλαισίου του σήματος αφού το περιέλιξε με τα φίλτρα ανάλυσης.
- Κάθε γραμμή αντιπροσωπεύει την έξοδο ενός συγκεκριμένου φίλτρου ανάλυσης που εφαρμόζεται στο πλαίσιο σήματος με παράθυρο.
- Οπτικοποιεί τα στοιχεία συχνότητας που εξάγονται από τα φίλτρα ανάλυσης.

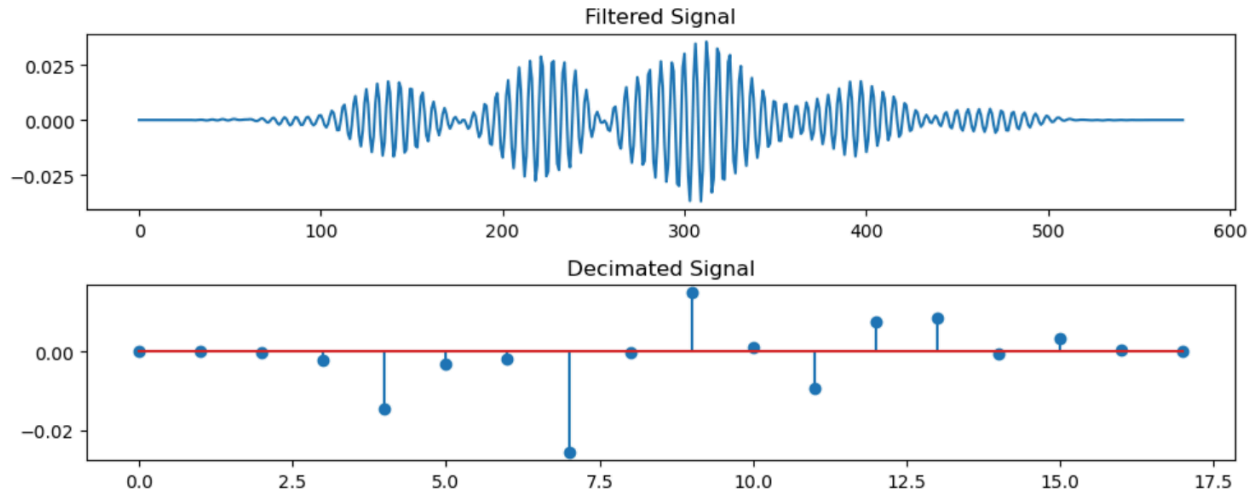
4. Αποδεκατισμένο σήμα:

- Αυτή η γραφική παράσταση εμφανίζει το αποδεκατισμένο σήμα που λαμβάνεται μετά τη μείωση της δειγματοληψίας του φιλτραρισμένου σήματος.
- Κάθε στέλεχος αντιπροσωπεύει ένα συγκεκριμένο στοιχείο του σήματος που λαμβάνεται από ένα διαφορετικό φίλτρο ανάλυσης.
- Απεικονίζει τον μειωμένο ρυθμό δειγματοληψίας και τις χρονικές συνιστώσες του σήματος.

Αυτές οι γραφικές παραστάσεις παρέχουν πληροφορίες για τα στάδια επεξεργασίας σήματος που εμπλέκονται στην ανάλυση της διάταξης φίλτρων, συμπεριλαμβανομένου του παραθύρου, του φιλτραρίσματος και της μείωσης δειγματοληψίας. Βοηθούν στην οπτικοποίηση των αποτελεσμάτων κάθε βήματος και της αποσύνθεσης του σήματος στα χρονικά του στοιχεία.

Όσον αφορά το aliasing, αν και τα μη ιδανικά φίλτρα διέλευσης ζώνης που χρησιμοποιούνται σε αυτό το βήμα μπορεί να προκαλέσουν κάποιες τύψεις ή τεχνουργήματα μεταξύ των στοιχείων του σήματος, θεωρούνται αμελητέα στο πλαίσιο της συνολικής διαδικασίας συμπίεσης.





Βήμα 2.2: Κβαντοποίηση

Το Βήμα 2.2 του συστήματος συμπύεσης εστιάζει στην κρίσιμη πτυχή της κβαντοποίησης. Η κβαντοποίηση περιλαμβάνει τη χαρτογράφηση των τιμών του δείγματος μιας διακριτής ακολουθίας σε έναν πεπερασμένο αριθμό επιπέδων κβαντισμού. Αυτό το βήμα παίζει σημαντικό ρόλο στη μείωση της ακρίβειας της αναπαράστασης του σήματος, ενώ ελαχιστοποιεί την αντιληπτική επίδραση στο ανακατασκευασμένο σήμα. Σε αυτήν την ενότητα της έκθεσης, θα εμβαθύνουμε στην εφαρμογή της διαδικασίας κβαντοποίησης, διερευνώντας τόσο έναν προσαρμοζόμενο όσο και έναν μη προσαρμοζόμενο κβαντιστή.

Ο προσαρμοζόμενος κβαντιστής που χρησιμοποιείται εδώ είναι ένας ομοιόμορφος κβαντιστής επιπέδου 2^{B_k} , όπου το B_k αντιπροσωπεύει τον αριθμό των bit κωδικοποίησης ανά δείγμα στο τρέχον πλαίσιο ανάλυσης. Ο αριθμός των bit κωδικοποίησης ανά δείγμα καθορίζεται από το όριο συνολικής κάλυψης του ψυχοακουστικού μοντέλου, το οποίο υπολογίστηκε στο Μέρος 1. Ο τύπος που χρησιμοποιείται για τον υπολογισμό του αριθμού των κωδικών bit, B_k , δίνεται από την επόμενη εξίσωση. Αυτή η προσαρμοστική προσέγγιση κβαντισμού επιτρέπει για δυναμική προσαρμογή των επιπέδων κβαντοποίησης με βάση το ψυχοακουστικό μοντέλο, βελτιστοποιώντας την αντιστάθμιση μεταξύ συμπίεσης και πιστότητας δεδομένων.

Επιπλέον, ένα πείραμα προσαρμοζόμενο κβαντιστή, αριθμό bits, $B_k = 8$ bit, για ανάλυσης. Το βήμα κβαντισμού, Δ , προσδιορίζεται με βάση το υποτιθέμενο σταθερό εύρος τιμών σήματος $[-1, 1]$ και τον αριθμό των bit ανά δείγμα, όπως περιγράφεται στην εξίσωση:

$$B_k = \text{int} \left(\log_2 \left(\frac{R}{\min(T_g(i))} \right) - 1 \right) \quad \text{διεξάγεται με έναν μη χρησιμοποιώντας έναν σταθερό κβαντισμό σε κάθε πλαίσιο}$$

$$\Delta = \frac{x_{max} - x_{min}}{2^{B_k}}$$

Αυτή η προσέγγιση μη προσαρμοστικής κβαντοποίησης παρέχει πληροφορίες για την απόδοση συμπίεσης όταν Ένας σταθερός αριθμός bit χρησιμοποιείται σε όλα τα καρέ, ανεξάρτητα από τα χαρακτηριστικά του ψυχοακουστικού μοντέλου.

Μέσω της εφαρμογής προσαρμοσμένων και μη προσαρμοζομένων κβαντιστών, στοχεύουμε να αξιολογήσουμε την επίδραση διαφορετικών στρατηγικών κβαντοποίησης στο συμπιεσμένο σήμα. Λαμβάνοντας υπόψη το συνολικό κατώφλι κάλυψης του ψυχοακουστικού μοντέλου, το δυναμικό εύρος του σήματος και τις σταθερές παραμέτρους κβαντοποίησης, μπορούμε να αξιολογήσουμε τις αντισταθμίσεις μεταξύ της απόδοσης συμπίεσης και της αντιληπτικής ποιότητας του ανακατασκευασμένου σήματος.

Όσον αφορά τον κώδικα, ξεκινάμε φορτώνοντας τις προηγουμένως υπολογισμένες τιμές κατωφλίου συνολικής κάλυψης, Tg, από το αρχείο "Tgi.npy". Αυτές οι τιμές κατωφλίου ελήφθησαν από το ψυχοακουστικό μοντέλο στο Μέρος 1.

Για να απεικονίσουμε τις τιμές Tg[0], τις σχεδιάζουμε στο Σχήμα 1. Αυτή η γραφική παράσταση παρέχει πληροφορίες για τις τιμές κατωφλίου που ορίζουν το ψυχοακουστικό μοντέλο.

Στη συνέχεια, ορίζουμε μια συνάρτηση f(k) που επιστρέφει το εύρος συχνοτήτων που σχετίζεται με κάθε φίλτρο ανάλυσης. Αυτές οι περιοχές συχνοτήτων χρησιμοποιούνται για τον υπολογισμό των προσαρμοστικών παραμέτρων κβαντοποίησης.

Χρησιμοποιώντας τις τιμές Tg και τα εύρη συχνοτήτων, υπολογίζουμε τον αριθμό των δυαδικών ψηφίων κωδικοποίησης ανά δείγμα, B_k, για κάθε πλαίσιο ανάλυσης στο αποδεκατισμένο φιλτραρισμένο σήμα. Οι τιμές B_k καθορίζονται από την ελάχιστη τιμή Tg εντός του αντίστοιχου εύρους συχνοτήτων, όπως περιγράφεται στην αναφορά. Αυτές οι παράμετροι κβαντοποίησης αντιπροσωπεύουν τα ψυχοακουστικά χαρακτηριστικά του σήματος.

Υπολογίζουμε το βήμα κβαντισμού, δέλτα, με βάση το εύρος τιμών σε κάθε αποδεκατισμένο φιλτραρισμένο πλαίσιο σήματος και τις αντίστοιχες τιμές B_k. Αυτό το μέγεθος βήματος καθορίζει την ευαισθησία της διαδικασίας κβαντοποίησης.

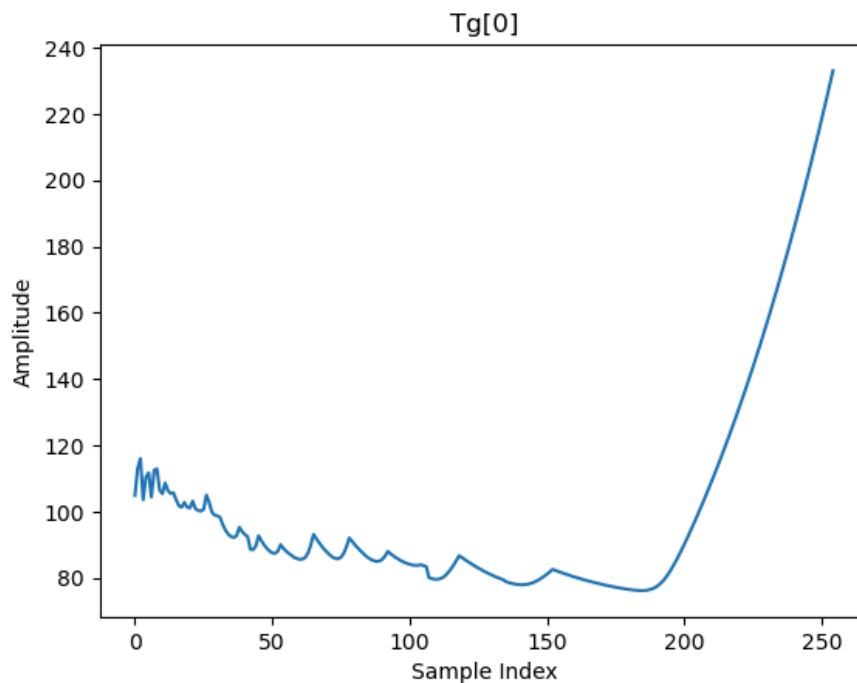
Έχοντας στη θέση τους τις παραμέτρους κβαντοποίησης, προχωράμε στον κβαντισμό του αποδεκατισμένου φιλτραρισμένου σήματος. Η συνάρτηση κβαντοποίησης εφαρμόζεται σε κάθε δείγμα του σήματος, χαρτογραφώντας το στο πλησιέστερο επίπεδο κβαντοποίησης με βάση τις τιμές δέλτα και B_k. Το κβαντισμένο σήμα αποθηκεύεται στη μεταβλητή "quantized_signal".

Επιπρόσθετα, διερευνούμε την απόδοση ενός μη προσαρμοσμένου κβαντιστή με σταθερό αριθμό bit κωδικοποίησης, $\text{test_Bk} = 8$. Το βήμα κβαντοποίησης, test_delta , προσδιορίζεται με βάση το εύρος τιμών σε κάθε αποδεκατισμένο φίλτραρισμένο πλαίσιο σήματος και τον σταθερό αριθμό bit. Το κβαντισμένο σήμα που χρησιμοποιεί τον μη προσαρμοστικό κβαντιστή αποθηκεύεται στη μεταβλητή "test_quantized".

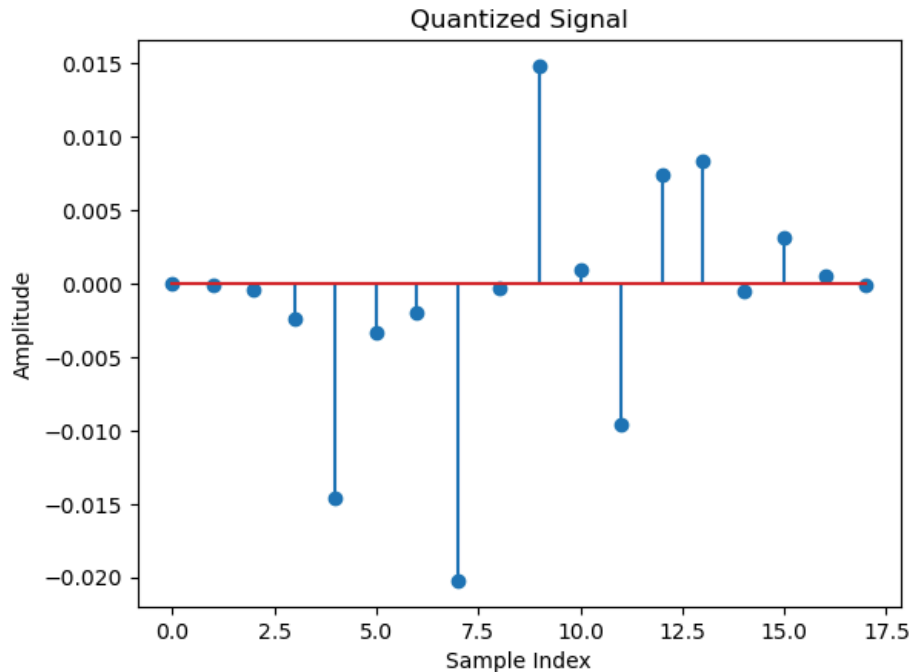
Για να διευκολυνθεί η επακόλουθη διαδικασία σύνθεσης, παρεμβάλλουμε το κβαντισμένο σήμα. Αυτό περιλαμβάνει την επέκταση κάθε κβαντισμένου πλαισίου με την εισαγωγή δειγμάτων μηδενικής τιμής $M - 1$ μεταξύ κάθε δείγματος. Το σήμα παρεμβολής αποθηκεύεται στη μεταβλητή "interpolated".

Για να απεικονίσουμε τα κβαντισμένα και παρεμβαλλόμενα σήματα, σχεδιάζουμε το κβαντισμένο σήμα στο Σχήμα 2 και το παρεμβαλλόμενο σήμα στο Σχήμα 3. Αυτά τα διαγράμματα παρέχουν μια οπτική αναπαράσταση των επιδράσεων κβαντοποίησης και παρεμβολής στο σήμα.

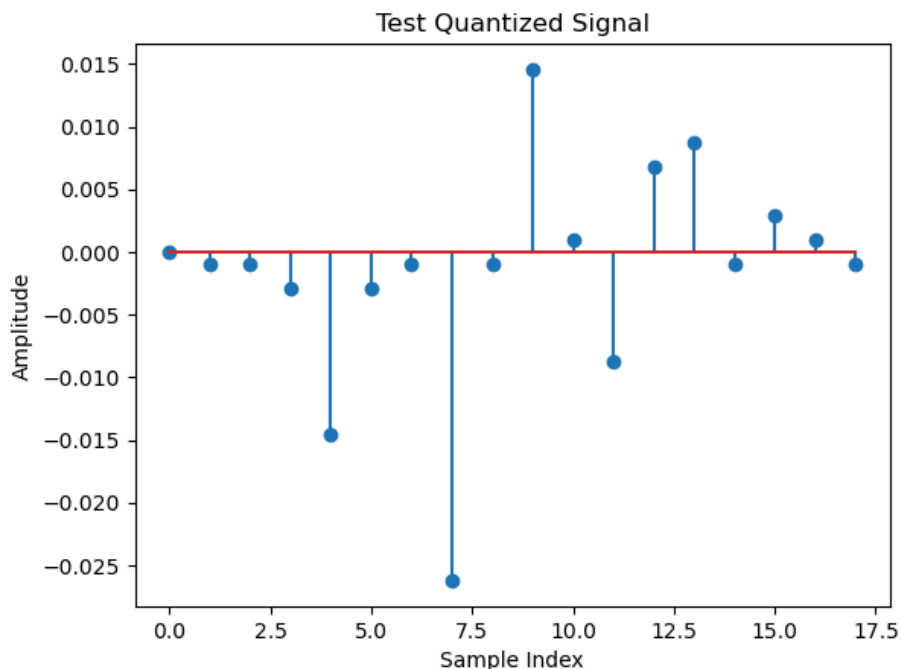
Άρα τελικά για το Βήμα 2.2 τυπώνουμε τα παρακάτω:



Αυτό το διάγραμμα δείχνει τις τιμές του $Tg[0]$, που αντιπροσωπεύει το συνολικό όριο κάλυψης του ψυχοακουστικού μοντέλου για το πρώτο πλαίσιο ανάλυσης. Το $Tg[0]$ υπολογίζεται με βάση το εύρος κεντρικής συχνότητας που ορίζεται από το φίλτρο ανάλυσης. Η γραφική παράσταση εμφανίζει τις τιμές πλάτους $Tg[0]$ πάνω από τον δείκτη του δείγματος.



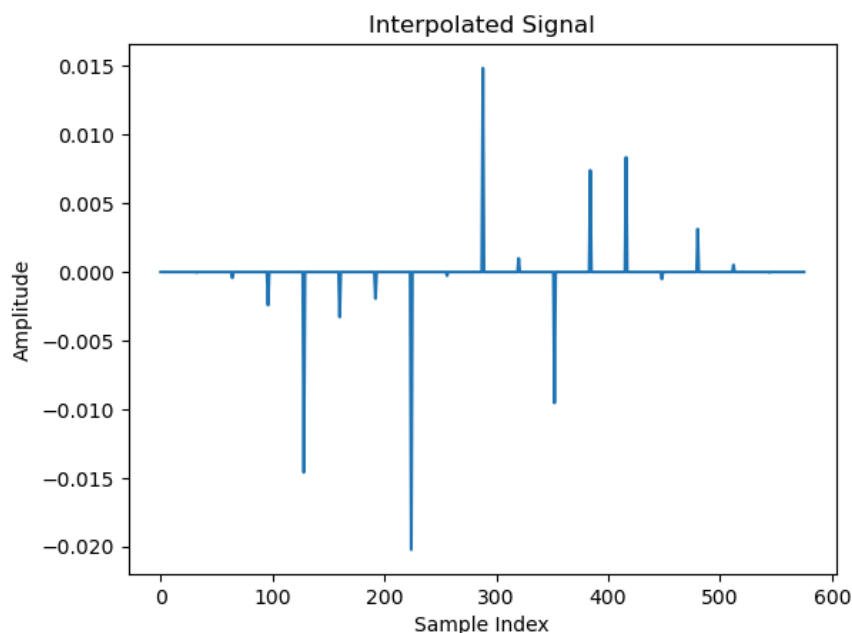
Αυτό το διάγραμμα εμφανίζει ένα αρχικό διάγραμμα του κβαντισμένου σήματος. Κάθε στέλεχος αντιπροσωπεύει ένα κβαντισμένο δείγμα από το αποδεκατισμένο φιλτραρισμένο σήμα. Η διαδικασία κβαντοποίησης περιλαμβάνει την αντιστοίχιση κάθε δείγματος σε ένα συγκεκριμένο επίπεδο κβαντοποίησης με βάση το μέγεθος του βήματος κβαντισμού (Δ_s) και τον αριθμό των bit (b) που καθορίζεται από τον προσαρμοσμένο κβαντιστή. Το διάγραμμα δείχνει τις τιμές πλάτους του κβαντισμένου σήματος πάνω από τον δείκτη του δείγματος.



Αυτό το
το κβαντισμένο
λαμβάνεται
έναν μη

διάγραμμα δείχνει
σήμα που
χρησιμοποιώντας
προσαρμοστικό

κβαντιστή με σταθερό αριθμό bit (8 bit) και ένα σταθερό εύρος τιμών σήματος $[-1, 1]$. Το αρχικό διάγραμμα εμφανίζει τις τιμές πλάτους του κβαντισμένου σήματος πάνω από τον δείκτη δείγματος.



Αυτό το διάγραμμα εμφανίζει το παρεμβαλλόμενο σήμα, το οποίο λαμβάνεται με την εισαγωγή δειγμάτων μηδενικής τιμής μεταξύ κάθε κβαντισμένου δείγματος προκειμένου να αποκατασταθεί ο αρχικός ρυθμός δειγματοληψίας. Το διάγραμμα δείχνει τις τιμές πλάτους του παρεμβαλλόμενου σήματος πάνω από τον δείκτη δείγματος.

Παρατηρούμε πως δεν είναι εμφανείς οι διαφορές των κβαντοποιημένων παραμέτρων με αυτά πριν από την κβάντιση και αυτό οφείλεται στο γεγονός ότι έχουμε πολύ μικρό βήμα κβαντισμού καθώς και πολλές στάθμες σε πολύ μικρό εύρος τιμών.

Όσον αφορά τον κβαντιστή 8bit παρατηρούμε πως τα αποτελέσματα είναι σχετικά παρόμοια με τα παραπάνω καθώς πάλι έχουμε πολλές στάθμες (2^8) σε ένα πολύ μικρό εύρος τιμών, μόνο που τώρα επειδή χρησιμοποιούμε κοινό Δ σε όλα τα διαφορετικά k κάθε segment παρατηρούμε πως είναι πιο κβαντοποιημένο.

Βήμα 2.3: Σύνθεση

Το Βήμα 2.3 της διαδικασίας κωδικοποίησης είναι κρίσιμο για τη σύνθεση και την ανακατασκευή των κβαντισμένων αλληλουχιών που λαμβάνονται στο Βήμα 2.2. Σε αυτό το βήμα, οι κβαντισμένες ακολουθίες αποστέλλονται στον αποκωδικοποιητή, όπου υποβάλλονται σε μια σειρά λειτουργιών για την ανακατασκευή του αρχικού μουσικού σήματος.

Ο αποκωδικοποιητής ξεκινά προγραμματίζοντας τις κβαντισμένες ακολουθίες με μηδενικά M και την υπερδειγματοληψία τους. Αυτή η υπερδειγματοληψία επιτυγχάνεται με τη δημιουργία μιας νέας ακολουθίας,

$wk(n)$, η οποία περιέχει τις κβαντισμένες τιμές σε συγκεκριμένα διαστήματα, ενώ τα υπόλοιπα δείγματα τίθενται στο μηδέν. Αυτή η υπερδειγματοληψία διασφαλίζει ότι ο αποκωδικοποιητής λειτουργεί με υψηλότερο ρυθμό δειγματοληψίας, επιτρέποντας μια πιο ακριβή ανακατασκευή.

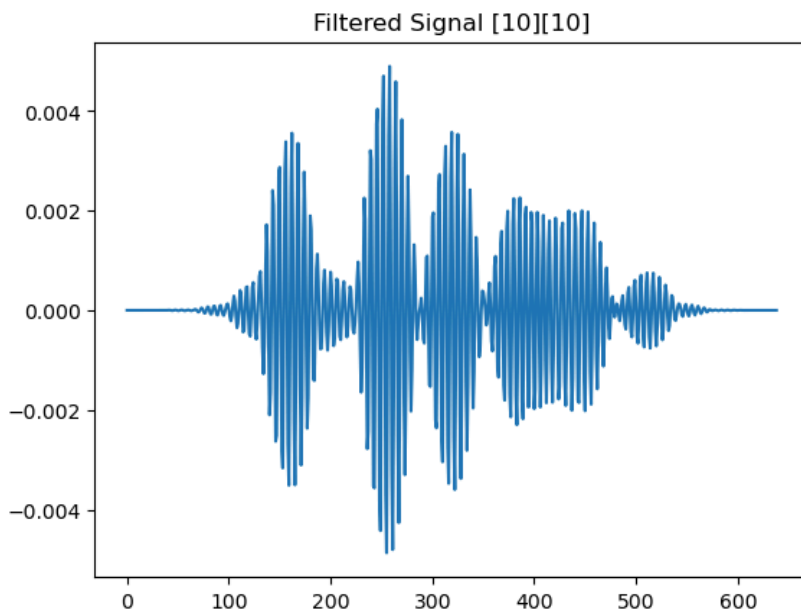
Η υπερδειγματοληψία ακολουθία, $wk(n)$, στη συνέχεια συνδυάζεται με τα φίλτρα σύνθεσης. Αυτά τα φίλτρα σύνθεσης έχουν σχεδιαστεί για να ικανοποιούν τον περιορισμό της γραμμικής φάσης και λειτουργούν ως το αντίστροφο των φίλτρων ανάλυσης που χρησιμοποιούνται στο Βήμα 2.1. Εφαρμόζοντας τα φίλτρα σύνθεσης, ο αποκωδικοποιητής στοχεύει στην ανακατασκευή του αρχικού σήματος αναίρωντας τα αποτελέσματα των φίλτρων ανάλυσης.

Η έξοδος της συστοιχίας φίλτρων στον αποκωδικοποιητή συμβολίζεται ως $\hat{x}(n)$ και αντιπροσωπεύει το ανακατασκευασμένο σήμα. Ωστόσο, λόγω σφαλμάτων κβαντισμού που εισήχθησαν στο Βήμα 2.2, το ανακατασκευασμένο σήμα μπορεί να μην είναι ακριβές αντίγραφο του αρχικού σήματος εισόδου, $x(n - n_0)$, όπου το n_0 αντιπροσωπεύει τη θέση του πρώτου επιπέδου.

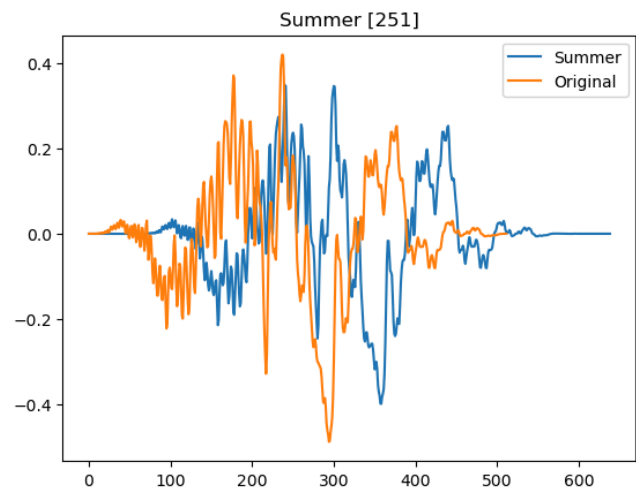
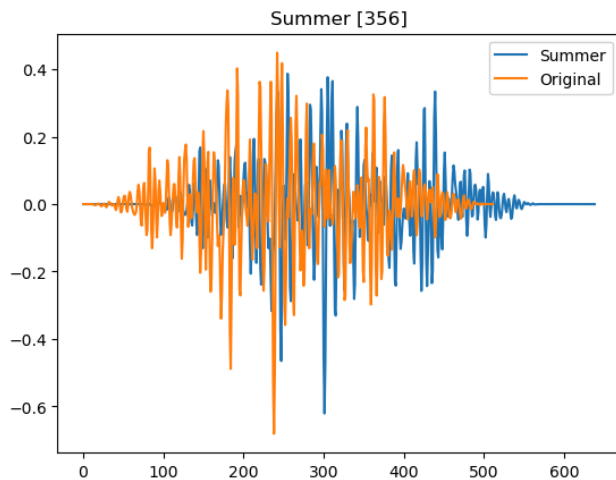
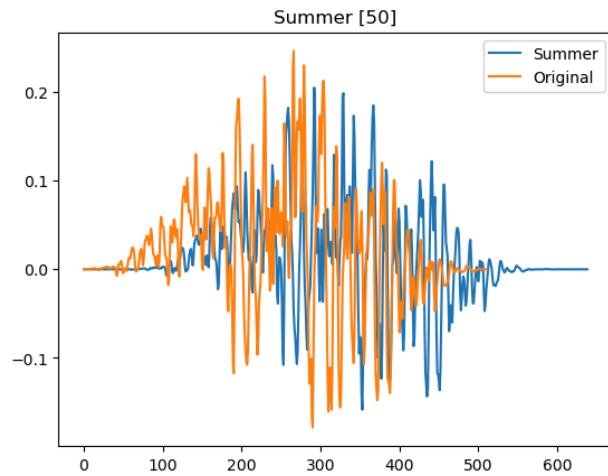
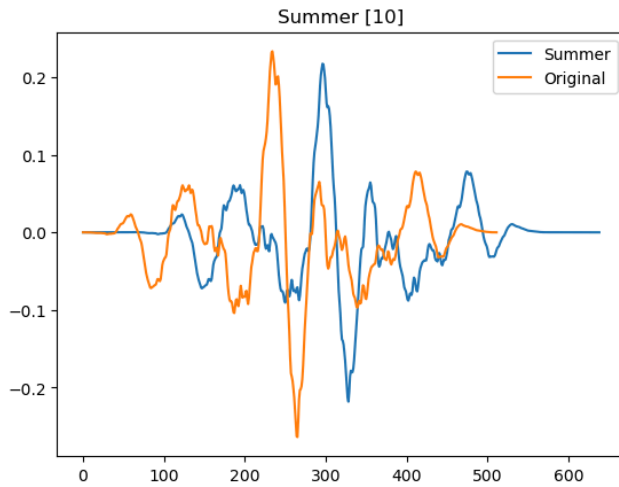
Για να διευκολυνθεί η διαδικασία ανακατασκευής, μεταδίδονται και κωδικοποιούνται πρόσθετες πληροφορίες όπως οι δείκτες σταθμού, η θέση του πρώτου επιπέδου και το μέγεθος του βήματος κβαντοποίησης (D) χρησιμοποιώντας 16 bit στον αποκωδικοποιητή.

Τέλος, η ανακατασκευή του μουσικού σήματος, που συμβολίζεται ως $\hat{s}(n)$, πραγματοποιείται με την εφαρμογή της τεχνικής OverLap-Add (OLA) χωρίς επικάλυψη μεταξύ διαδοχικών πλαισίων ανάλυσης. Αυτή η διαδικασία διασφαλίζει ότι το ανακατασκευασμένο σήμα ευθυγραμμίζεται και ενοποιείται, με αποτέλεσμα την πιστή αναπαράσταση της πρωτότυπης μουσικής.

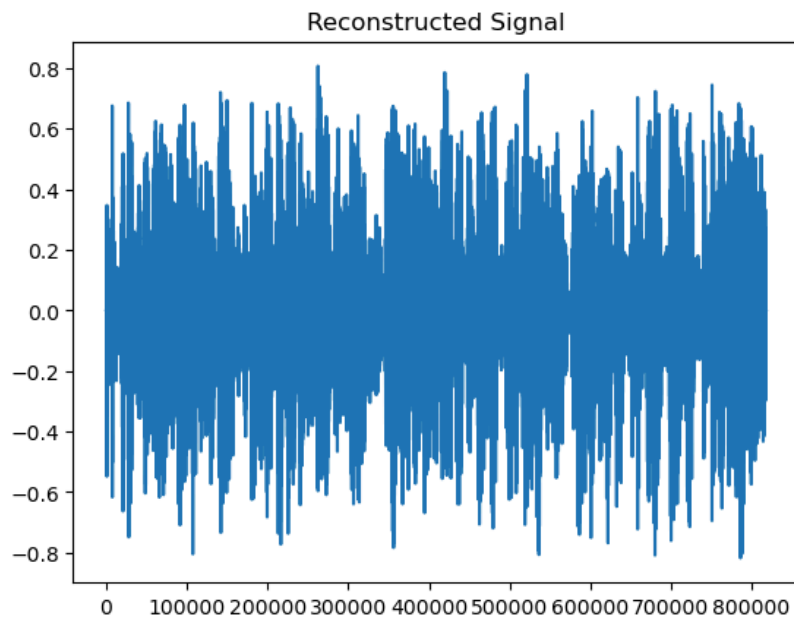
Συνοπτικά, το Βήμα 2.3 εστιάζει στη σύνθεση των κβαντισμένων ακολουθιών στον αποκωδικοποιητή, στην αποκωδικοποίησή τους χρησιμοποιώντας φίλτρα σύνθεσης και στην επακόλουθη ανακατασκευή του μουσικού σήματος μέσω της διαδικασίας OLA. Αυτές οι λειτουργίες παίζουν ζωτικό ρόλο στην επίτευξη μιας υψηλής πιστότητας ανακατασκευής του αρχικού μουσικού σήματος.



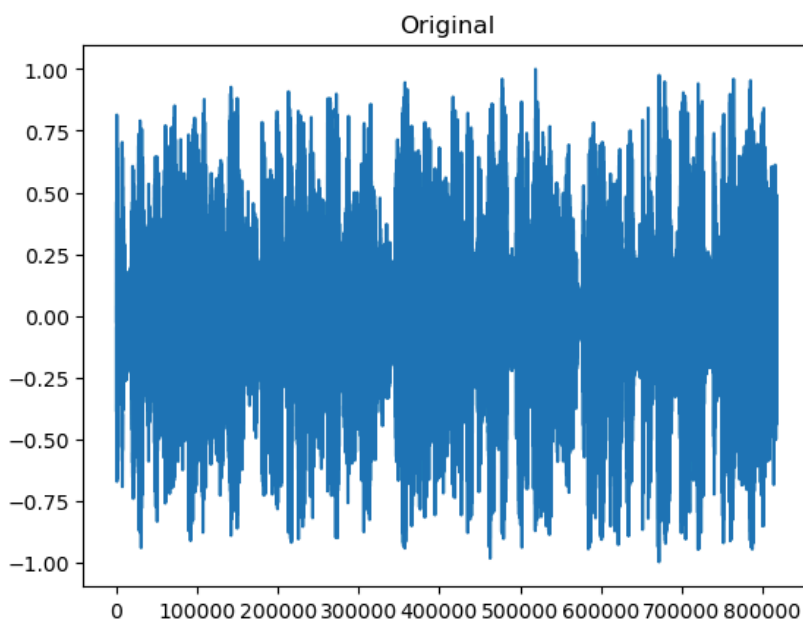
Αυτή η γραφική παράσταση εμφανίζει το φιλτραρισμένο σήμα που λαμβάνεται από την περιέλιξη του παρεμβalόμενου σήματος με τα φίλτρα σύνθεσης. Αντιπροσωπεύει ένα συγκεκριμένο τμήμα του ανακατασκευασμένου σήματος.



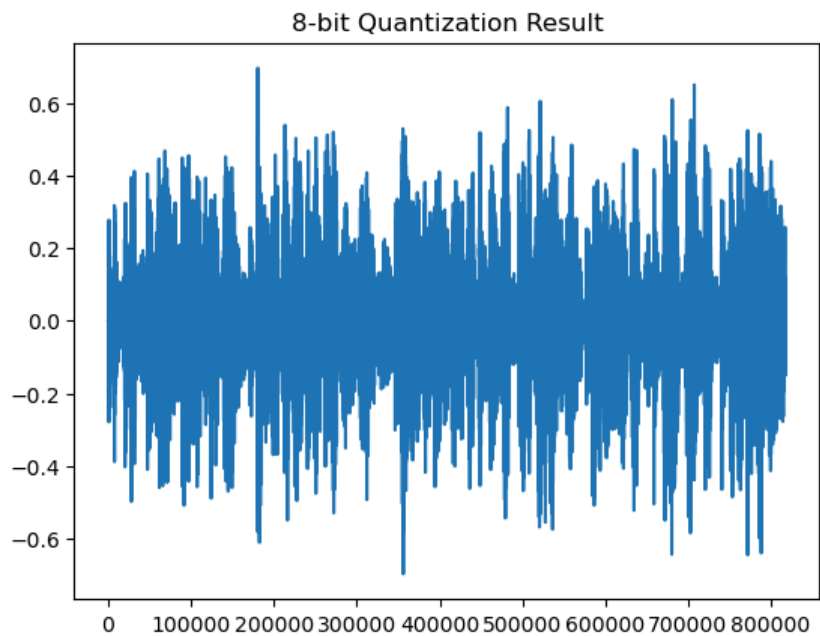
Αυτά τα διαγράμματα δείχνουν την έξοδο του αθροιστή, η οποία συνδυάζει τα φιλτραρισμένα σήματα από κάθε υποζώνη. Η συγκρίνεται με το αρχικό σήμα με παράθυρο για να εκτιμηθεί η ακρίβεια της ανακατασκευής.



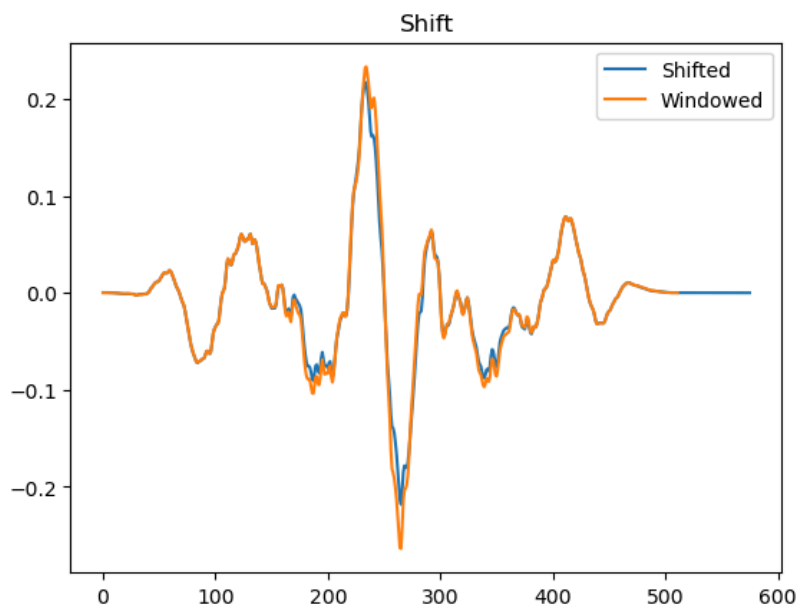
Αυτό το διάγραμμα παρουσιάζει το ανακατασκευασμένο μουσικό σήμα που λαμβάνεται μέσω της τεχνικής OverLap-Add (OLA). Δείχνει το τελικό αποτέλεσμα του συνδυασμού των επιμέρους τμημάτων του ανακατασκευασμένου σήματος.



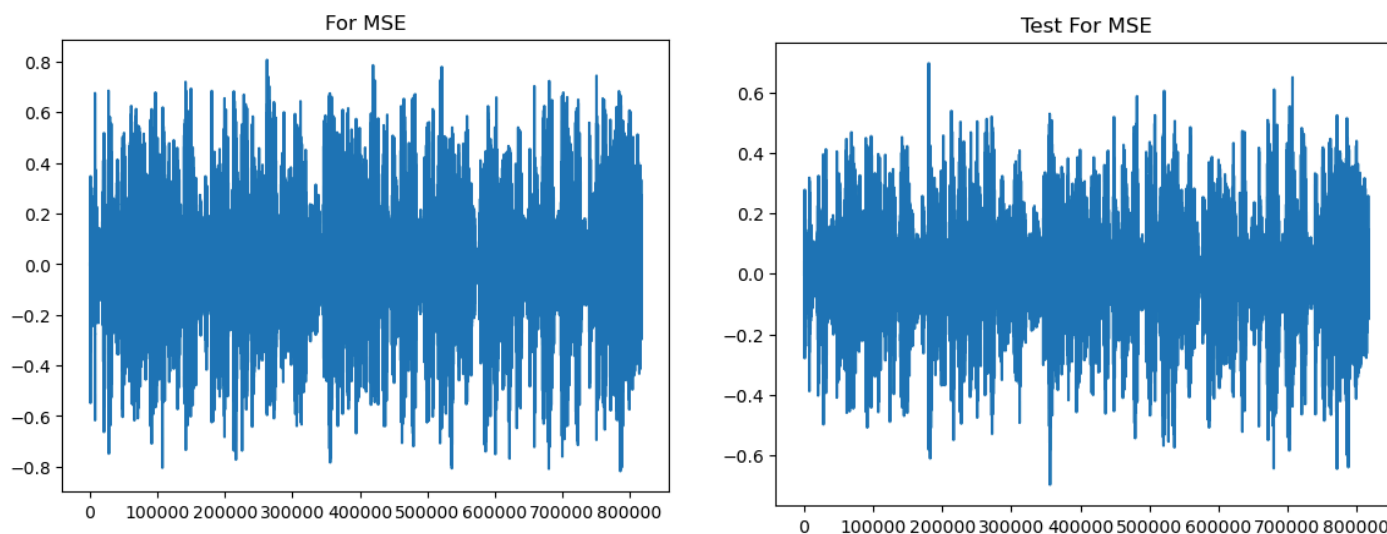
Αυτή η γραφική παράσταση εμφανίζει το αρχικό σήμα εισόδου μουσικής για σύγκριση με το ανακατασκευασμένο σήμα.



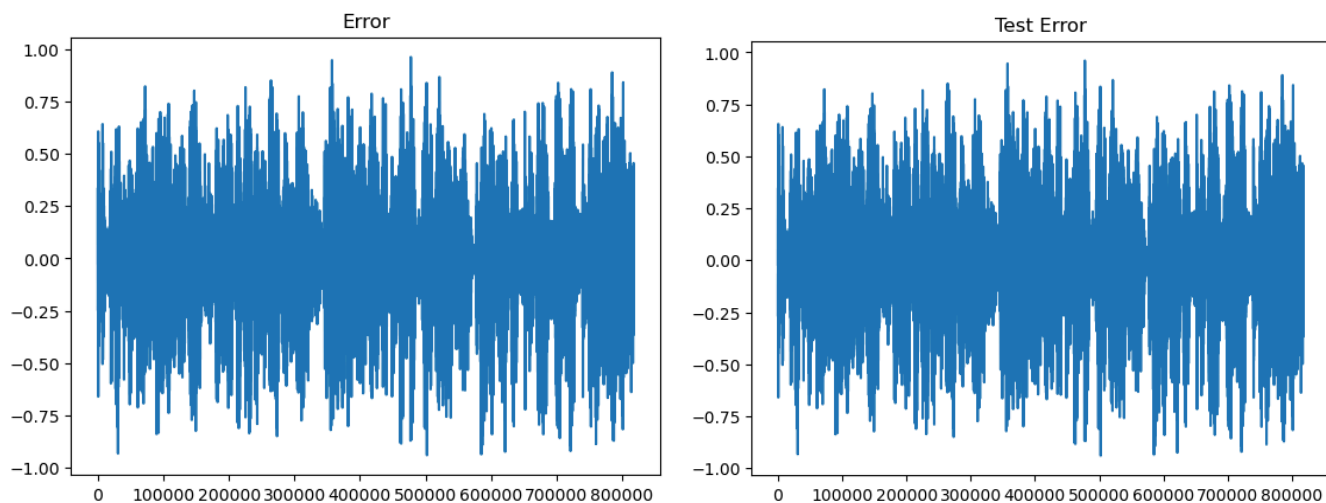
Αυτή η γραφική παράσταση αντιπροσωπεύει το ανακατασκευασμένο μουσικό σήμα μετά από κβαντισμό με 8 bit. Δείχνει την επίδραση της μείωσης του αριθμού των bit στην ποιότητα της ανακατασκευής.



Αυτή η γραφική παράσταση απεικονίζει τη λειτουργία μετατόπισης που εκτελείται στο ανακατασκευασμένο σήμα για την ευθυγράμμιση του με το αρχικό σήμα με παράθυρο. Βοηθά στην ευθυγράμμιση του ανακατασκευασμένου σήματος με το πρωτότυπο για περαιτέρω ανάλυση.



Αυτά τα διαγράμματα δείχνουν τα σήματα που λαμβάνονται μετά την εφαρμογή της τεχνικής OverLap-Add (OLA) και τη μετατόπιση για τον υπολογισμό του μέσου τετραγώνου σφάλματος (MSE). Αντιπροσωπεύουν τη διαφορά μεταξύ των ανακατασκευασμένων και των αρχικών σημάτων.



Αυτά τα σχέδια απεικονίζουν τα σήματα σφάλματος που λαμβάνονται αφαιρώντας τα ανακατασκευασμένα σήματα από τα αρχικά σήματα. Υπογραμμίζουν τις αποκλίσεις μεταξύ των ανακατασκευασμένων και των αρχικών σημάτων.

Επιπλέον, η αναφορά παρέχει διάφορες ποσοτικές αξιολογήσεις:

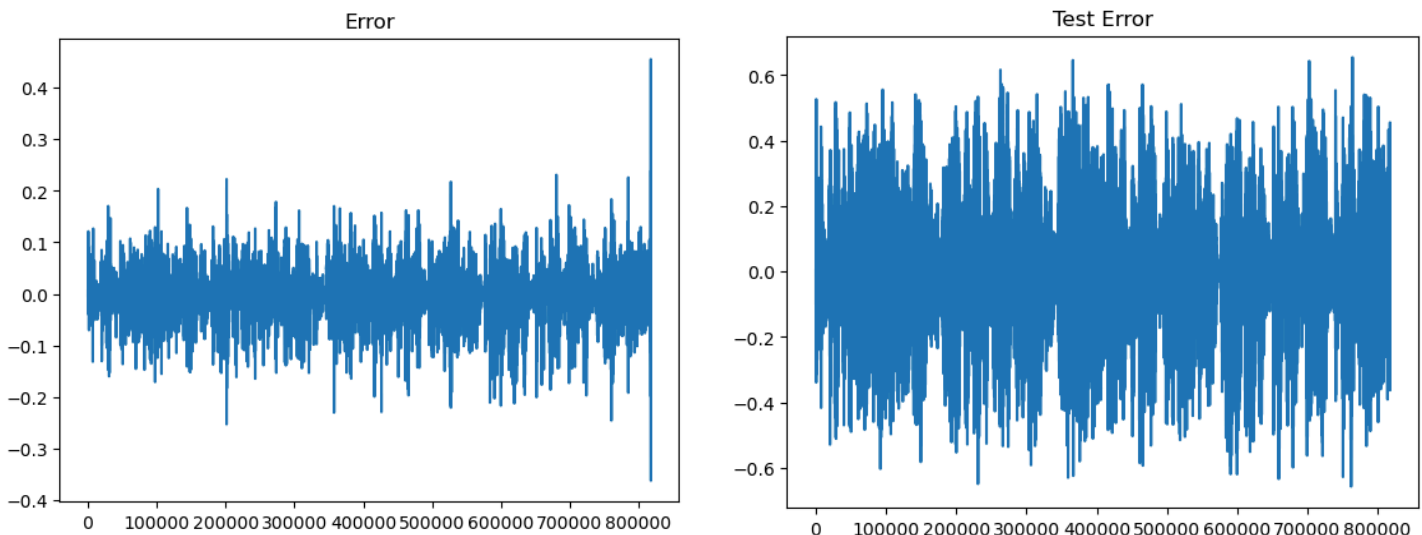
- MSE (Mean Squared Error): Οι τιμές MSE υπολογίζονται μεταξύ του αρχικού σήματος και των ανακατασκευασμένων σημάτων για την αξιολόγηση της ποιότητας της ανακατασκευής. Το MSE μετρά τη μέση τετραγωνική διαφορά μεταξύ του αρχικού και του ανακατασκευασμένου σήματος.
- Συμπίεση: Ο αρχικός αριθμός bit που απαιτείται για την αναπαράσταση του αρχικού σήματος συγκρίνεται με τον συμπίεσμένο αριθμό bit μετά την κβαντοποίηση. Το ποσοστό συμπίεσης που επιτυγχάνεται υπολογίζεται ως η διαφορά μεταξύ του αρχικού και του συμπίεσμένου bit διαιρούμενο με τα αρχικά bit.

```
MSE is 0.01870306604927925
Test MSE is 0.026787202717439118
Initial number of bits: 13086208
Compressed number of bits: 8181504
Compression (%): 37.47994835478696
```

Αναλύοντας αυτές τις πλοκές και ποσοτικές εκτιμήσεις, μπορούμε να αξιολογήσουμε την πιστότητα και την ποιότητα του ανακατασκευασμένου μουσικού σήματος και να αξιολογήσουμε την αποτελεσματικότητα της τεχνικής συμπίεσης που χρησιμοποιήθηκε.

Παρατήρηση:

Εάν αντί για παράθυρο Hanning χρησιμοποιήσουμε τετραγωνικό παράθυρο, τότε το τελικό ανακατασκευασμένο σήμα φαίνεται και ακούγεται να είναι πιο κοντά στο αρχικό.



Αρχικά και μόνο από τα παραπάνω διαγράμματα σφάλματος μπορούμε να παρατηρήσουμε ότι τα τελικά σήματα διαφέρουν σαφώς λιγότερο από ό,τι τα σήματα που προκύπτουν με το παράθυρο Hanning.

Το τετράγωνο παράθυρο έχει πιο απλό σχήμα με απότομες μεταβάσεις στις άκρες. Ενώ παρέχει εξαιρετική ακρίβεια πλάτους εντός του παραθύρου τμήματος, εισάγει μεγαλύτερη φασματική διαρροή σε σύγκριση με το παράθυρο Hanning. Οι αιχμηρές μεταβάσεις του τετράγωνου παραθύρου προκαλούν υψηλότερα επίπεδα πλευρικών λοβών, οι οποίοι μπορούν να διασπείρουν ενέργεια σε ένα ευρύτερο φάσμα συχνοτήτων. Ενώ αυτό μπορεί να οδηγήσει σε ευρύτερη εξάπλωση συχνότητας, μπορεί επίσης να έχει ως αποτέλεσμα τη μείωση του αντίκτυπου του θορύβου στενής ζώνης ή των παρεμβολών. Η φασματική διαρροή που προκαλείται από το τετράγωνο παράθυρο μπορεί να διασπείρει την ενέργεια του θορύβου σε ένα ευρύτερο φάσμα συχνοτήτων, μειώνοντας ενδεχομένως την επίδρασή της στο ανακατασκευασμένο σήμα.

Μαζί με την εργασία μας, επισυνάπτουμε τα αρχεία "output_square.wav" και "output_8bit_square.wav", τα αποτελέσματα των παρατηρήσεων μας, για να ολοκληρωθεί η σύγκριση.