



ΑΡΙΣΤΟΤΕΛΕΙΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΕΣΣΑΛΟΝΙΚΗΣ



9<sup>ο</sup> Εξάμηνο : **«Συστήματα Πολυμέσων»**

Ακαδ. έτος 2022-23 | Ημ.Παράδοσης 19/2/2023

**Εργασία : «Simplified MP3 codec»**

του Σταύρου Βασίλειου Μπουλιόπουλου 9671

και του Βασίλειου Ευαγγελόπουλου 9767

Διδάσκων θεωρίας : Αναστάσιος Ντελόπουλος

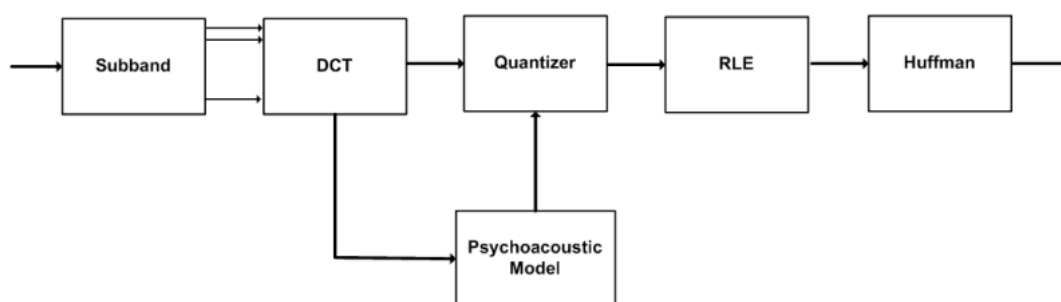
Επιβλέπων εργασίας : Αντώνιος Καρακώπτας

## Περιεχόμενα

1.Εισαγωγή και ζητούμενα.....	2
2. Εκτέλεση των αρχείων κώδικα .....	3
3. Επεξήγηση των προβλημάτων που αντιμετωπίσαμε, παραδοχές και η δομή του κώδικα...	3
4. Παρουσίαση αποτελεσμάτων.....	5

## 1.Εισαγωγή και ζητούμενα

Στην προκειμένη εργασία του μαθήματος υλοποιήσαμε μια απλοποιημένη εκδοχή της κωδικοποίησης-αποκωδικοποίησης ηχητικών σημάτων βάσει πρωτοκόλλου MPEG-1 LAYER III.



Εικόνα 1: Ροή της διαδικασίας

Ζητήθηκαν συναρτήσεις με τις οποίες υπολογίζουμε το κάθε στάδιο της διαδικασίας και υλοποιήθηκαν όλα τα στάδια της εργασίας (Subband-Filter bank, DCT για αποσυσχέτιση, psychoacoustic model για εκμετάλλευση του μηχανισμού της ανθρώπινης ακοής στην συμπίεση, ελεγχόμενος Quantizer βάσει του προηγούμενου μοντέλου, Run Length Encoding και Huffman για κωδικοποίηση ικανοποιητικά κοντά στην εντροπία).

## 2. Εκτέλεση των αρχείων κώδικα

Η απεικόνιση των αποτελεσμάτων(απεικόνιση των συναρτήσεων μεταφοράς των φίλτρων και  $x$ ,  $x_{hat}$ ) γίνεται χρησιμοποιώντας για γλώσσα προγραμματισμού Python(version=3.7) στο αρχείο εκτέλεσης **main.py** ,το οποίο περιέχει τις απαιτούμενες βιβλιοθήκες και τις ζητούμενες συναρτήσεις(η εκτέλεση των συναρτήσεων γίνεται στο τέλος του αρχείου).

## 3. Επεξήγηση των προβλημάτων που αντιμετωπίσαμε,παραδοχές και η δομή του κώδικα

Τα προβλήματα που ήρθαμε αντιμέτωποι ήταν:

- οι μικροαλλαγές στο κατώφλι ακουστότητας(όπως προτάθηκε στην εκφώνηση) κατά μείωση 15 dB προκειμένου να είναι ικανοποιητική η ποιότητα του ήχου(χωρίς την μείωση αυτή, ο ήχος ήταν αρκετά παραμορφωμένος σαν ένα ηχείο σε θάλασσα).
- ο υπολογισμός του μέγιστου συντελεστή ετεροσυσχέτισης για τον υπολογισμό της ολίσθησης ενός εκ των δύο σημάτων, έτσι ώστε τα δύο σήματα να έχουν κοινή αρχή με αποτέλεσμα να μπορούμε να υπολογίσουμε ορθά το SNR.
- παρατηρήσαμε(με εσωτερικά plots πριν το τελικό set-up) ότι οι συντελεστές DCT που αντιστοιχούν σε υψηλές συχνότητες στο βήμα κβάντισης(λαμβάνοντας υπόψιν και το ψυχοακουστικό μοντέλο) αντιστοιχούσανε στην πλειοψηφία των frames στην νεκρή ζώνη.
- στην ζητούμενη συνάρτηση `Dksparse(Kmax)` αρχικά μορφοποιήσαμε τις δικτυακές σχέσεις σε αραιό πίνακα μορφής COO. Ύστερα, από εκτέλεση των προγραμμάτων και profiling διαπιστώσαμε ότι ο χρόνος εύρεσης μίας σειράς του πίνακα ήταν αρκετά μεγάλος. Για αυτόν τον λόγο, επιλέξαμε

καταλληλότερη μορφή αραιού πίνακα CSR με αποτέλεσμα την μείωση του υπολογιστικού χρόνου κατά 7 φορές.

Τηρήσαμε τα ορίσματα εισόδου και εξόδου σε κάθε ζητούμενη συνάρτηση και προσθήσαμε μια κατηγορία από τις εξής βοηθητικές συναρτήσεις(UTILITY FUNCTIONS):

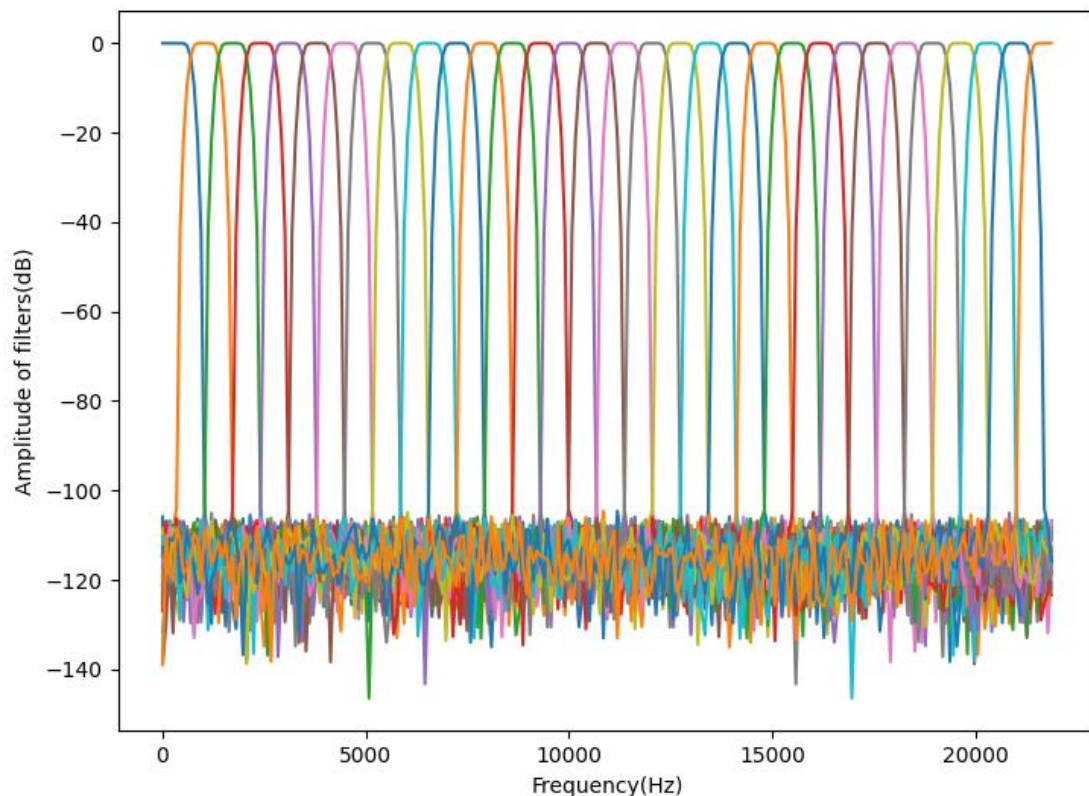
**-make\_mp3\_synthesisfb(h, M):** Δέχεται ως όρισμα τον πίνακα των  $h(n)\cos((2i + 1)\pi * \frac{n}{2M} + \frac{(2i+1)\pi}{4})$  όπου  $i$  στο εύρος 0 έως M (όσο και οι μπάντες) και επιστρέφει στην έξοδο τον πίνακα φίλτων σύνθεσης του filterbank  $G$  διαστάσεων  $L \times M$  που υπακούει στον γνωστό κανόνα ορθογονικού filter bank by construction , δηλαδή στην χρονική-αντιστροφή των φίλτρων ανάλυσης .

**-plotterV1(wavin, xhat, start,end):** Δέχεται ως όρισμα το αρχικό ηχητικό σήμα, το αποκωδικοποιημένο σήμα και την περιοχή μελέτης για ακριβέστερη οπτική ανάλυση (zoom-in μελέτη) προκειμένου να απεικονίσει τα σήματα αυτά, να υπολογίσει και να απεικονίσει το σφάλμα μεταξύ αυτών αλλά και τον συντελεστή **SNR**

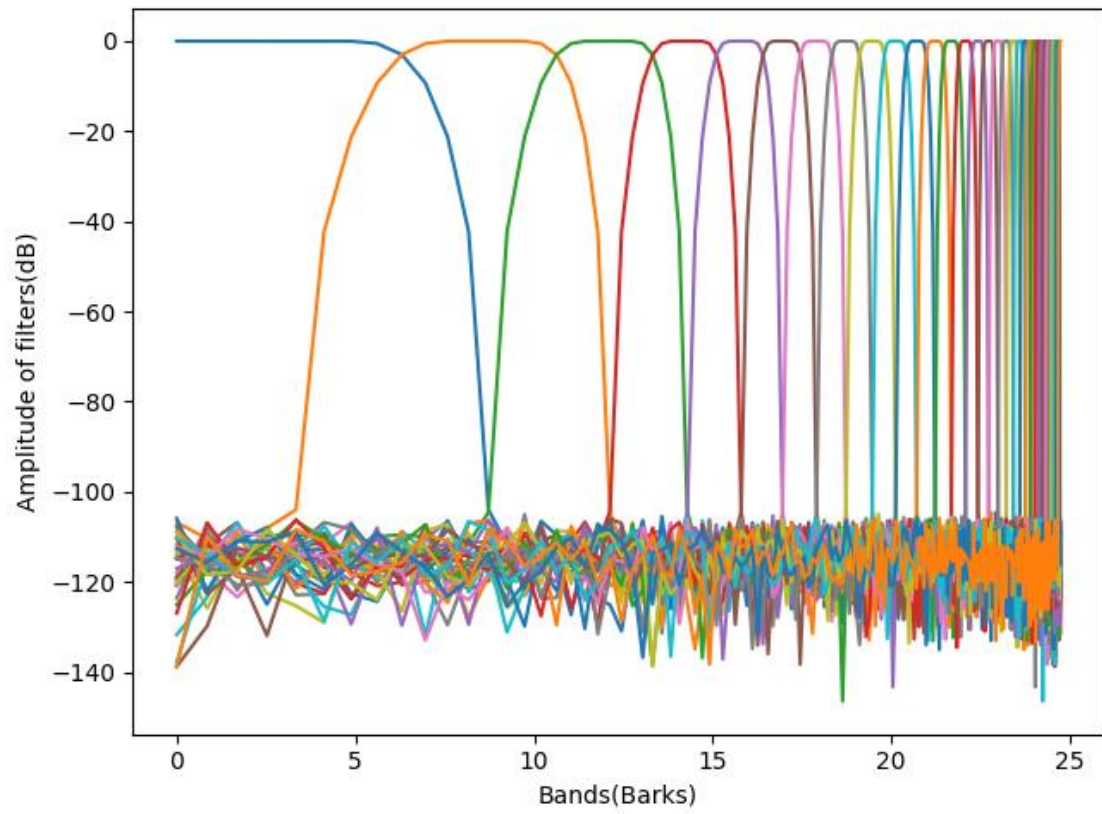
**-writerV1(xhat, wavin, start, end):** Δέχεται ως όρισμα τα ίδια με την παραπάνω συνάρτηση και αποθηκεύει το αποκωδικοποιημένο σήμα στο testDec2.wav πριν κάποια αλλαγή και στο testDec3.wav λαμβάνοντας υπόψιν την καθυστέρηση που εισάγει η χρήση της συνέλιξης και των buffers. Συγκεκριμένα, υπολογίζουμε την ολίσθηση αυτή βάσει του συντελεστή ετεροσυσχέτισης μεταξύ του αρχικού σήματος και του αποκωδικοποιημένου σήματος (η ολίσθηση που επιλέγεται είναι αυτή που δίνει τον μέγιστο συντελεστή ετεροσυσχέτισης όπως και περιμέναμε αφού  $L=512, M=32$   $L-M=480$ ).

**-writerV2(xhat, wavin, start, end):** Η συνάρτηση αυτή είναι μορφικά και υπολογιστικά παρόμοια με την **writerV1**, αλλά διαφέρει στα ονόματα αρχείων αποθήκευσης του ηχητικού σήματος μετά την αποκωδικοποίηση MP3. Αντιστοίχως: (πριν την ολίσθηση)testMP3Dec2.wav, (μετά τον υπολογισμό της ολίσθησης)testMP3Dec3.wav

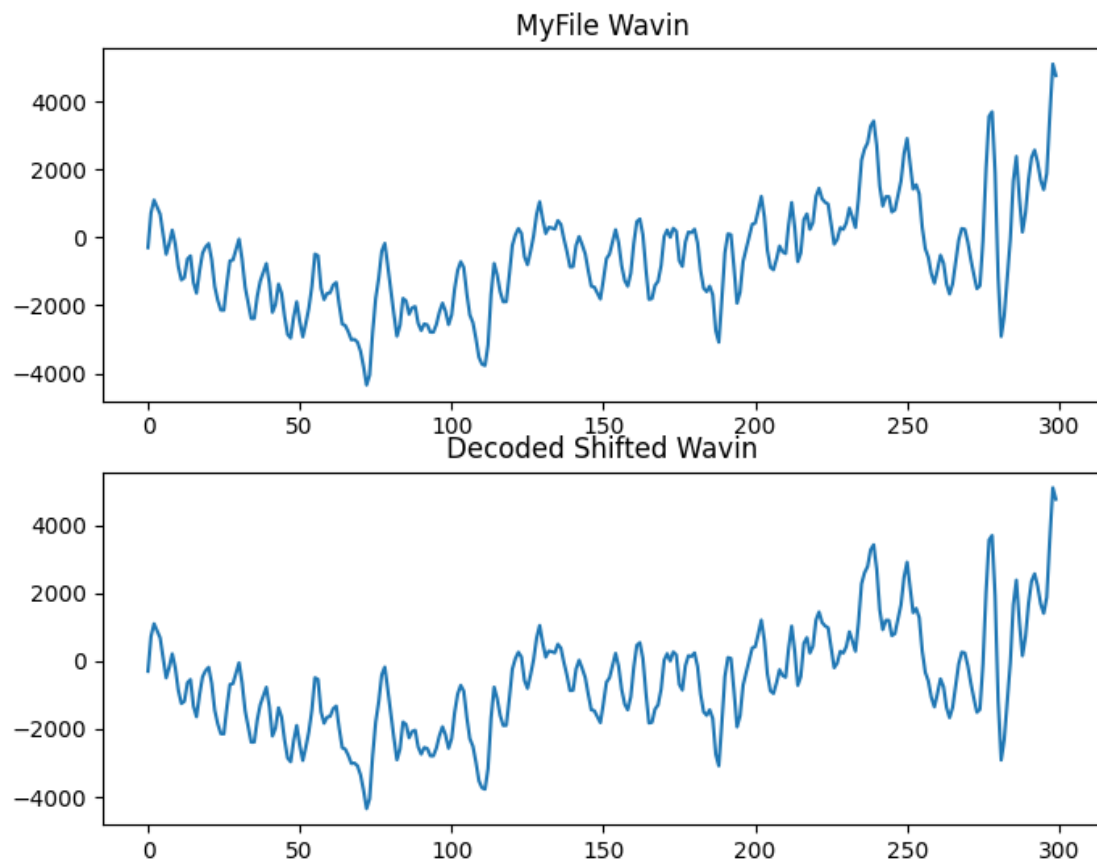
#### 4. Παρουσίαση αποτελεσμάτων



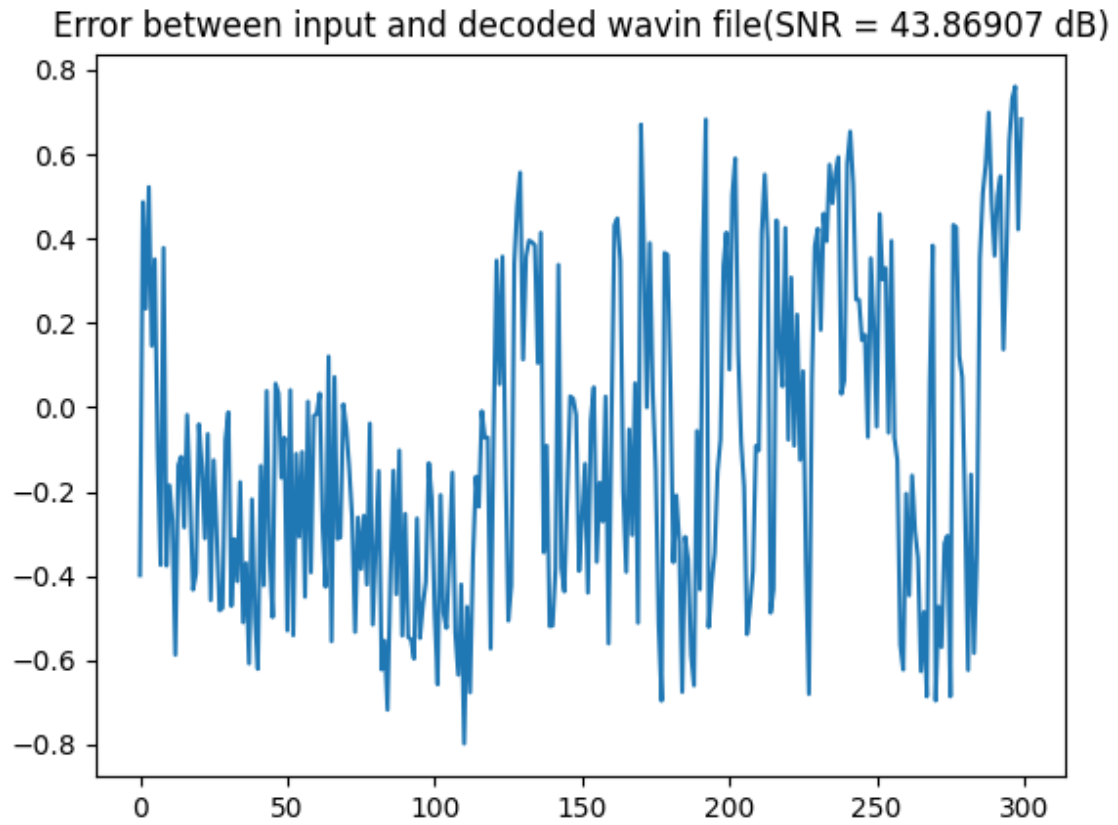
Εικόνα 2: Πλάτος συναρτήσεων μεταφοράς φίλτρων ανάλυσης ανά Hz



Εικόνα 3: Πλάτος συναρτήσεων μεταφοράς φίλτρων ανάλυσης ανά Barks

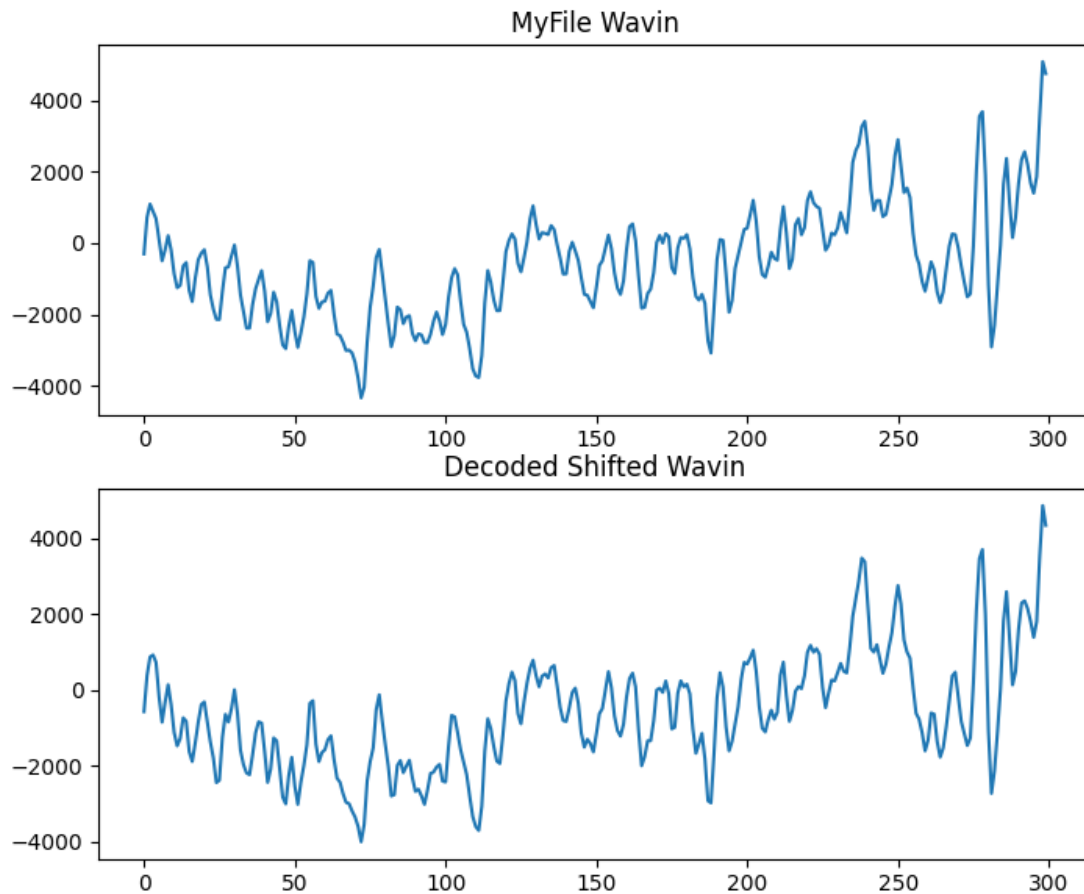


Εικόνα 4: Απεικόνιση του αρχικού σήματος και του σήματος μετά τα φίλτρα ανάλυσης και σύνθεσης (και κατάλληλης ολίσθησης)

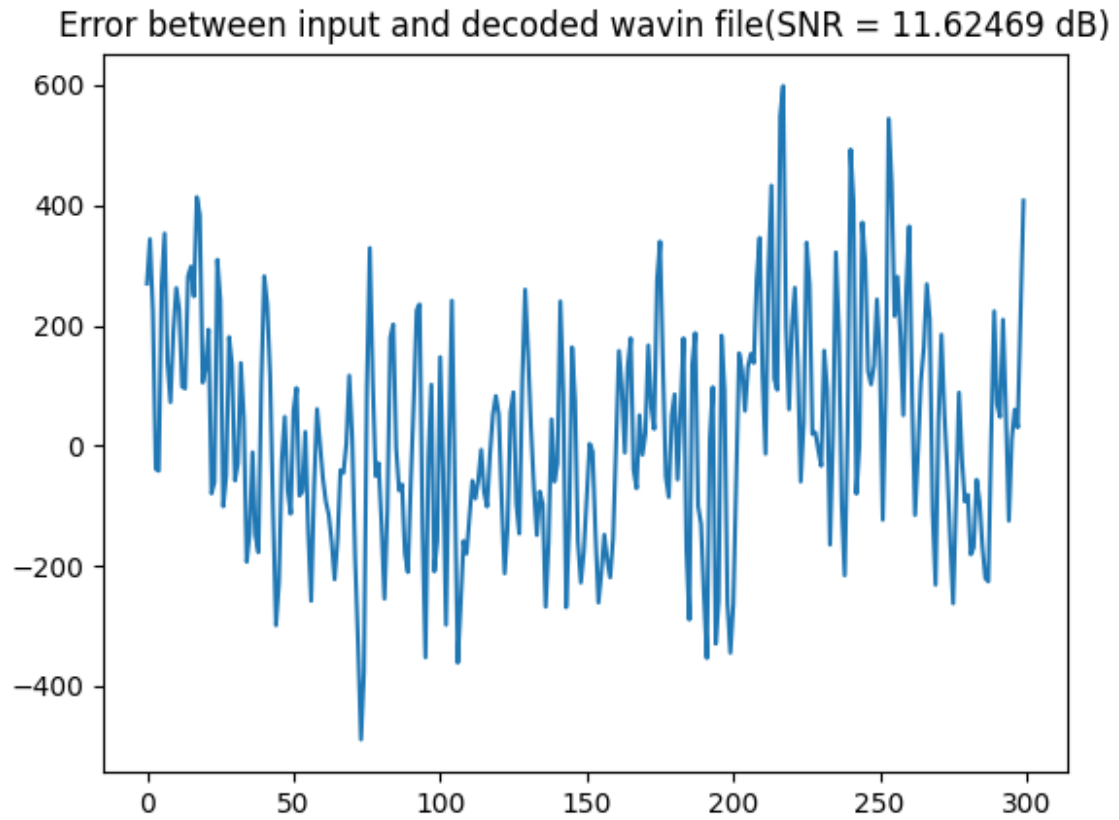


Εικόνα 5: Απεικόνιση του σφάλματος των δύο προαναφερθέντων σημάτων(Εικόνα 4) και του υπολογισμένου SNR





Εικόνα 6: Απεικόνιση του αρχικού σήματος και του σήματος μετά το συνολικό pipeline κωδικοποίησης-αποκωδικοποίησης MP3(και κατάλληλης ολίσθησης)



Εικόνα 7: Απεικόνιση του σφάλματος των δύο προαναφερθέντων σημάτων(Εικόνα 6) και του υπολογισμένου SNR