Ψηφιακή Επεξεργασία Σημάτων
2η Εργαστηριακή Άσκηση
Θέμα: Κωδικοποίηση σημάτων Μουσικής βάσει ψυχοακουστικού μοντέλου (Perceptual Audio Coding)

ΠΑΝΑΓΙΩΤΑΡΑΣ ΗΛΙΑΣ ΑΜ: 03115746

Ακαδ. Έτος 2017-18 | Ημ. Παράδοσης 18/05/18 | 6ο Εξάμηνο

Μέρος 1) Ψυχοακουστικό Μοντέλο 1

Το παρόν σύστημα κωδικοποίησης μουσικής βασίζεται στις ιδιότητες του συστήματος ακοής του ανθρώπου όπως πολλά την σήμερον εποχή (όπως π.χ. MP3). Στόχος μας να συμπιέσουμε ένα σήμα μουσικής διάρκειας περίπου 14 sec που είναι αποθηκευμένο στο αρχείο **music_0.wav** δίνοντας έμφαση κυρίως στις αντιλήψιμες συχνότητες των κρίσιμων συχνοτικών περιοχών όπως αυτές ορίζονται από το ψυχοακουστικό μοντέλο. Τα διαθέσιμα bits κβαντισμού, ανάλογα με τον επιθυμητό βαθμό συμπίεσης, κατανέμονται ανά χρονικό τμήμα και κρίσιμη συχνοτική ζώνη με στόχο : α) το λάθος κβαντισμού να γίνει όσο το δυνατό λιγότερο αντιληπτό και β) οι χρονο-συχνοτικές συνιστώσες του σήματος που ακούγονται περισσότερο να λαμβάνουν περισσότερο χώρο στην κωδικοποίηση από αυτές που επικαλύπτονται και χάνονται στη διαδικασία της ακοής. Το σήμα έχει ηχογραφηθεί με συχνότητα δειγματοληψίας $F_s = 44100\ Hz$ και έχει κωδικοποιηθεί με PCM χρησιμοποιώντας 16 bits ανά δείγμα.

Η παραθυροποίηση γίνεται σύμφωνα με τα πρότυπα του **MPEG-1.** Η ανάλυση που ακολουθεί γίνεται σε πλαίσια ανάλυσης $x(n)$ του αρχικού σήματος $s(n)$. Το μήκος των παραθύρων ισούται με $N = 512$ δείγματα. Η επεξεργασία με το Ψυχοακουστικό Μοντέλο (Μέρος 1) και τη Συστοιχία Φίλτρων (Μέρος 2) εκτελείται σε κάθε πλαίσιο ανάλυσης χωρίς επικάλυψη. Τα πλαίσια ανάλυσης για το στάδιο του Ψυχοακουστικού Μοντέλου παραθυρώνονται με παράθυρο **Hanning**.

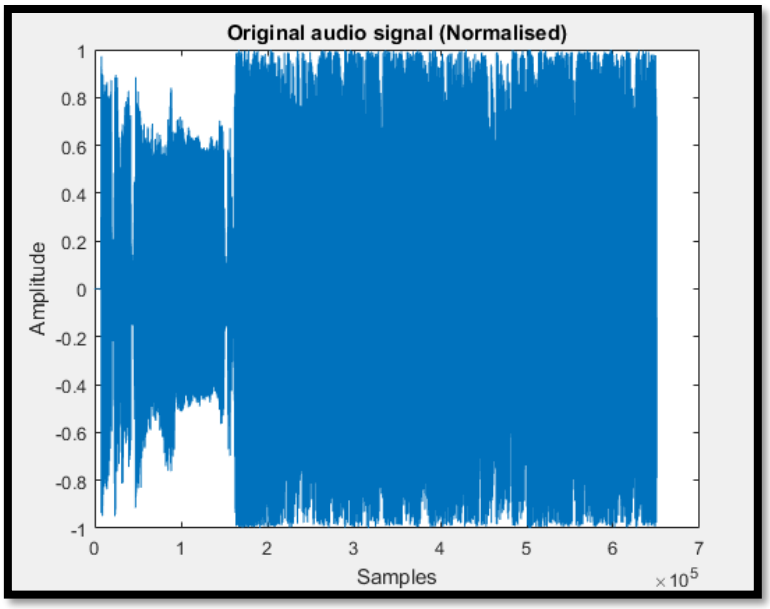
Στο Μέρος 1 (Ψυχοακουστικό μοντέλο) θα παρουσιαστούν συγκεκριμένα plots του κάθε βήματος για το 60^o πλαίσιο ανάλυσης. Παράλληλα θα περιγράφονται και οι αλγόριθμοι που χρησιμοποιούνται. Σκοπός του Μέρους 1 είναι η υλοποίηση του Ψυχοακουστικού Μοντέλου 1, υπολογίζοντας το συνολικό κατώφλι κάλυψης T_g για κάθε πλαίσιο.

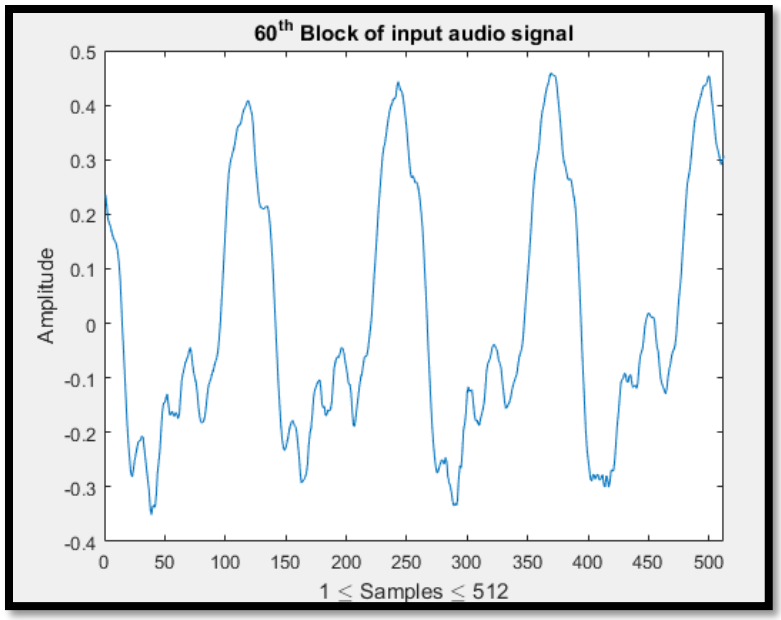
Παραδοτέα είναι:

- Αναφορά (pdf μορφή)
- Αρχεία κώδικα σε MATLAB (**ΕΠΕΞΗΓΗΣΗ ΤΟΥ ΚΩΔΙΚΑ ΥΠΑΡΧΕΙ ΣΤΑ ΑΡΧΕΙΑ .m**)
- Δυο διαφορετικά αρχεία .wav μετά την τελική ανακατασκευή του σήματος της μουσικής για τις δυο διαφορετικές μεθόδους κβαντοποίησης

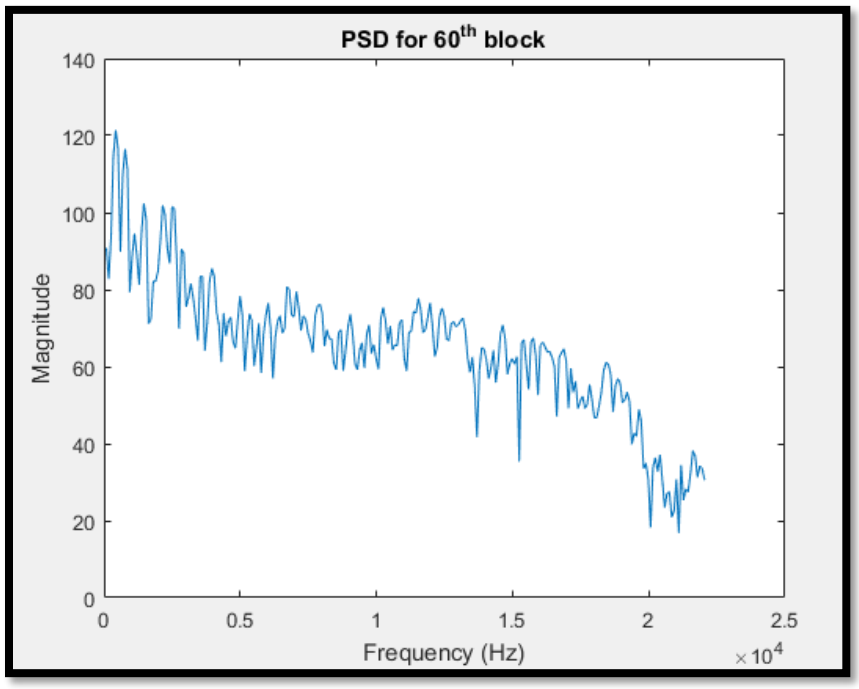
Βήμα 1.0: Κανονικοποίηση του σήματος

Κανονικοποιούμε το σήμα μουσικής αφού το διαβάσουμε στο matlab, διαιρώντας τα δείγματα του με την απόλυτη μέγιστη τιμή του σήματος, έτσι ώστε καθ' όλη τη διάρκεια του να έχει τιμές μεταξύ [−1, 1]. Για τα επόμενα βήματα, χρειάζεται παραθυροποιήση του σήματος (N = 512 δείγματα) και ανάλυση σε κάθε πλαίσιο. Το μουσικό σήμα που μας δόθηκε προφανώς δεν έχει μήκος πολλαπλάσιο του 512, άρα επιλέγουμε το τελευταίο παράθυρο να απορριφθεί. Παρατίθεται το κανονικοποιημένο αρχικό σήμα μουσικής καθώς και το 60^o πλαίσιο:





Βήμα 1.1: Φασματική Ανάλυση

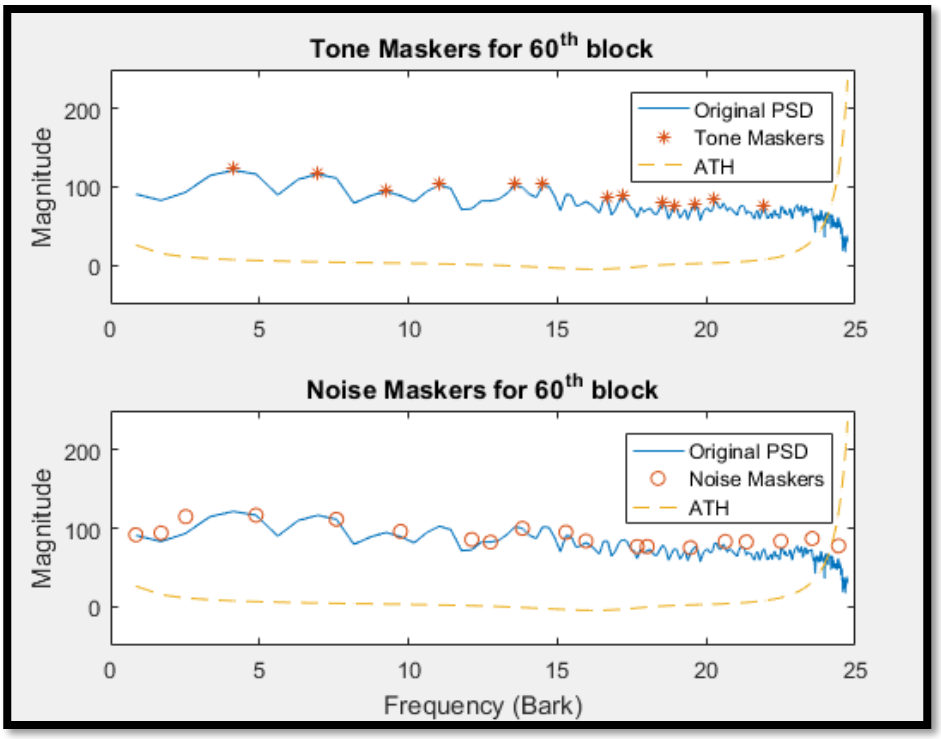


Αρχικά ορίζουμε την κλίμακα Bark σύμφωνα με την εξίσωση που δίνεται, και στη συνέχεια υπολογίζουμε το N-σημείων φάσμα ισχύος P(k) του σήματος όπου N = 512 δείγματα όπως έχει καθιερωθεί στο πρότυπο MPEG Layer-1. Ο fft για τον υπολογισμό του Power Spectral Density (PSD) γίνεται σε παράθυρα σήματος 512 και άρα το αποτέλεσμα αρχικά είναι ένα διάνυσμα 512, αλλά επειδή το αποτέλεσμα του fft είναι συμμετρικό κρατάμε από το 1:256. Οι κρίσιμες συχνοτικές περιοχές (critical bands) του ψυχοακουστικού μοντέλου έχουν σχεδιαστεί καθ΄ ομοίωση των περιοχών στις οποίες συντονίζονται οι νευροδέκτες του ακουστικού φλοιού. Οι 25 πρώτες κρίσιμες συχνοτικές περιοχές μοντελοποιούνται με την ψυχοακουστική κλίμακα συχνοτήτων Bark η οποία έχει πεδίο τιμών στο διάστημα [1, 25]. Η μετατροπή των συχνοτήτων της κλίμακας Hz στην κλίμακα Bark βάση του μη γραμμικού μοντέλου αντίληψης του ήχου δίνεται από την συνάρτηση **hz2bark**. Παρατίθεται παραπάνω η PSD για το 60^{ο} μπλοκ.

Βήμα 1.2: Εντοπισμός μασκών τόνων και θορύβου (Maskers)

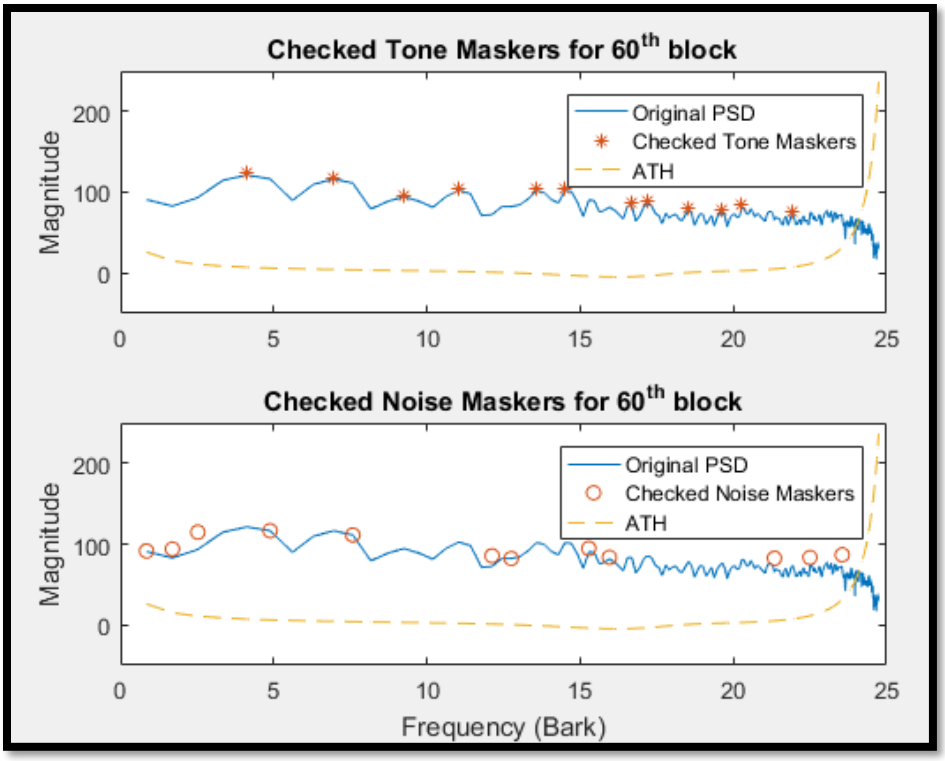
Σε αυτό το βήμα εντοπίζουμε ανά critical band τοπικά μέγιστα (μάσκες) τα οποία είναι μεγαλύτερα από τις γειτονικές τους συχνότητες τουλάχιστον κατα 7 dB. Το εύρος της γειτονιάς υπολογισμού των μασκών διαφέρει ανά διακριτή συχνότητα k (1:256) και υπολογίζεται από την συνάρτηση **delta_k** με βάση την σχέση που δίνεται.Θα παρατηρήσουμε πως στις υψηλές συχνότητες οι μάσκες καλύπτουν ευρύτερες γειτονιές. Σκοπός μας είναι να βρούμε τις θέσεις των τονικών μασκών και να υπολογίσουμε την ισχύ τους. Η ισχύς PTM(k) της μάσκας στη θέση k υπολογίζεται με βάση τις τιμές του φάσματος ισχύος στις διακριτές συχνότητες (k − 1),(k + 1) με βάση την σχέση που δίνεται. Η συνάρτηση **S_T** επιστρέφει boolean τιμές [0, 1] που προσδιορίζουν για κάθε θέση k αν υπάρχει τονική μάσκα. Η συνάρτηση εντοπίζει τις τονικές μάσκες ελέγχοντας για τοπικά μέγιστα στις διαφορετικές συχνοτικές περιοχές, όπως ορίζονται στις σχέσεις που δίνονται. Κάθε θέση k στην οποία ο πίνακας S_T έχει λογική τιμή 1 αντιστοιχεί σε μία τονική μάσκα που καλύπτει τις γειτονικές συχνότητες. Για την εύρεση των μασκών του θορύβου (noise maskers) μας δίνεται η έτοιμη συνάρτηση **findNoiseMaskers.** Παρατίθενται οι μάσκες θορύβου και οι τονικές μάσκες για το 60^{o} πλαίσιο, ενώ φαίνεται και η θέση τους πάνω στην αντίστοιχη PSD.

- Σημειώνεται πως φαίνεται και η απόλυτη τιμή κατωφλίου ακοής (Absolute Threshold of Hearing). Το κατώφλι ακοής χαρακτηρίζει το ποσό της ενέργειας σε dB- Sound Pressure Level (dB SPL) που πρέπει να έχει ένας τόνος (π.χ. ημίτονο) συχνότητας f ώστε να γίνει αντιληπτός σε περιβάλλον πλήρους ησυχίας. Η συνάρτηση που την υλοποιεί είναι η **hz2dBSPL.**



Βήμα 1.3: Μείωση και αναδιοργάνωση των μασκών

Σε αυτό το βήμα μειώνουμε τον αριθμό των μασκών, χρησιμοποιώντας δυο διαφορετικά κριτήρια: 1) Κάθε μάσκα τόνου και θορύβου η οποία βρίσκεται κάτω από το κατώφλι απόλυτης ακοής (ATH) απορρίπτεται, άρα μόνο οι μάσκες οι οποίες πληρούν την σχέση $P_{TM,NM}(k) \geq T_q(k)$ θα παραμείνουν, όπου $T_q(k)$ είναι το SPL του κατωφλίου σε περιβάλλον ησυχίας.
2) Σε κινούμενα πάραθυρα του 0.5 Bark βρίσκουμε τις μάσκες και τις αντικαθιστούμε με την πιο δυνατή σε ένταση. Η συνάρτηση για το βήμα αυτό μας δίνεται έτοιμη και είναι η **checkMaskers.** Παρατίθενται οι αναδιοργανωμένες μάσκες θορύβου και οι τονικές μάσκες για το 60^{o} πλαίσιο, ενώ φαίνεται και η θέση τους πάνω στην αντίστοιχη PSD:



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

Σε αυτό το βήμα υπολογίζουμε τα δύο διαφορετικά κατώφλια κάλυψης. Το κάθε κατώφλι αντιπροσωπεύει το ποσοστό κάλυψης στο σημείο I (1:256) το οποίο προέρχεται από την μάσκα τόνου ή θορύβου στο σημείο j (1:256).

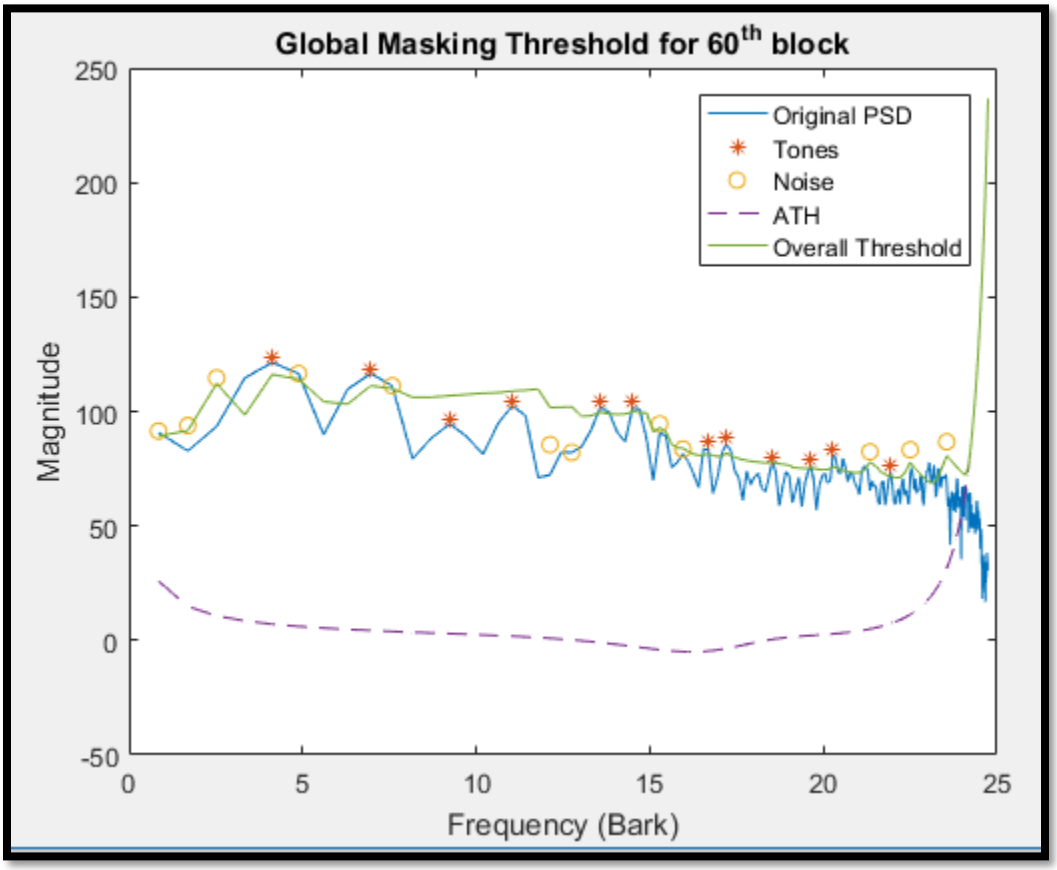
Η συνάρτηση **S_F** υπολογίζει την έκταση της κάλυψης από το σημείο j στο οποίο βρίσκεται η μάσκα έως το σημείο i το οποίο υφίσταται κάλυψη. Προσεγγίζει το ελάχιστο επίπεδο ισχύος το οποίο πρέπει να έχουν οι γειτονικές συχνότητες έτσι ώστε να γίνουν αντιληπτές από τον

άνθρωπο. Υπολογίζεται για κάθε μάσκα τόνου και θορύβου, οι θέσεις j των οποίων μπορούν να εντοπιστούν ως j: $P_{TM} > 0$ και $P_{NM} > 0$. Στο συγκεκριμένο μοντέλο θεωρούμε πως η κάλυψη περιορίζεται σε μία γειτονιά των 12-Bark. Έτσι, σε κάθε μάσκα υπολογίζουμε το SF χρησιμοποιώντας την γειτονιά της μάσκας. Αν μια συχνότητα i δεν ανήκει στο περιθώριο των 12 bark σε σχέση με τη συχνότητα j όπου και βρίσκεται μία μάσκα, μηδενίζουμε κατευθείαν τους πίνακες κατωφλίων κάλυψης **TTM** και **TNM**.

Σημειώνεται πως οι δείκτες j, i, k αναφέρονται στις διακριτές συχνότητες. Στο Βήμα 1.2 ψάχνουμε για όλα τα k που υπάρχουν μάσκες, στο βήμα 1.3 βρίσκουμε τα j δηλαδή τις τελικές θέσεις των μασκών και στο βήμα 1.4 για κάθε j βρίσκουμε το T_{TM}(i,j) όπου το i ορίζεται στη γειτονιά των 12 Barks με κέντρο το j. Επίσης, τονίζεται πως το **delta_b** στον κώδικα είναι η απόσταση σε barks του i από το j.

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

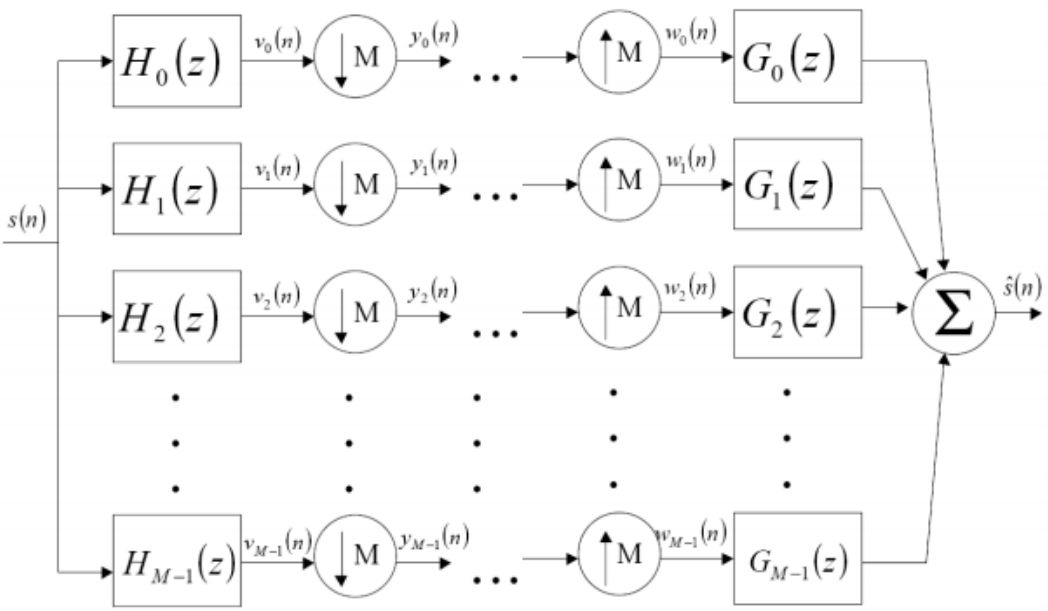
Τα ξεχωριστά κατώφλια κάλυψης τα οποία υπολογίστηκαν στο Βήμα 1.4 συνδυάζονται για την δημιουργία του συνολικού κατωφλίου σε κάθε διακριτή συχνότητα i ξεχωριστά. Το συνολικό κατώφλι T_g(i) υπολογίζεται αθροιστικά με τον τύπο που δίνεται. Για να υπολογιστεί το T_g απλά αθροίζονται τα επιμέρους κατώφλια σε κάθε πλαίσιο ανάλυσης ξεχωριστά. Παρατίθεται το συνολικό κατώφλι κάλυψης για το 60^o πλαίσιο ανάλυσης:



Μέρος 2. Χρονο-Συχνοτική Ανάλυση με Συστοιχία Ζωνοπερατών Φίλτρων

Η χρονο-συχνοτική ανάλυση χρησιμοποιείται για την εξαγωγή ενός συνόλου παραμέτρων, οι οποίες χρησιμοποιούνται για την κβαντοποίηση και την κωδικοποίηση του ηχητικού σήματος. Για την ανάλυση αυτή συνήθως χρησιμοποιούνται συστοιχίες ζωνοπερατών φίλτρων, οι οποίες και καλύπτουν όλο το φάσμα συχνοτήτων. Η συστοιχία των ζωνοπερατών φίλτρων διαιρεί το φάσμα σε υποζώνες συχνοτήτων και με αυτό τον τρόπο παρέχονται πληροφορίες σχετικά με την συχνοτική κατανομή του σήματος, οι οποίες βοηθούν στην ταυτοποίηση των αντιληπτικά περιττών σημείων του σήματος. Με άλλα λόγια, η συστοιχία ζωνοπερατών φίλτρων διευκολύνει την ανάλυση με το ψυχοακουστικό μοντέλο καθώς επίσης η αποσύνθεση αυτή του σήματος στις διαφορετικές συχνοτικές περιοχές βοηθά στη μείωση των στατιστικών redundancies.

Στο Μέρος 2 φιλτράρουμε τα παράθυρα x(n) του σήματος με τη συστοιχία φίλτρων ανάλυσης $h_k(n)$ και σύνθεσης $g_k(n), 0 \leq k \leq M-1, M = 32$ (ο αριθμός των φίλτρων) . Στόχος είναι η δημιουργία μιας συνάρτησης που υλοποιεί τη διαδικασία του σχήματος ακολούθως, 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.



Τα φίλτρα σχεδιάζονται βάσει μίας τροποποιημένης εκδοχής του γνωστού διακριτού μετασχηματισμού συνημιτόνων. Ο εν λόγω Modified Discrete Cosine Transform (MDCT) είναι πλήρως αντιστρέψιμος και δεν εισάγει λάθη στην κωδικοποίηση του σήματος. Στο στάδιο της κωδικοποίησης και αποκωδικοποίησης του συστήματος συμπίεσης που υλοποιείται στην άσκηση χρησιμοποιούνται M = 32 φίλτρα ανάλυσης και σύνθεσης αντίστοιχα με μήκος $L = 2M = 64$.

Έτσι, για κάθε πλαίσιο του αρχικού μας σήματος και για κάθε ένα από τα 32 φίλτρα ανάλυσης και σύνθεσης ακολουθούμε την ακόλουθη διαδικασία. Αρχικά, συνελίσουμε το σήμα μας με το k-οστό φίλτρο ανάλυσης και ύστερα κάνουμε downsampling κατά έναν παράγοντα M για να διαιρεθεί το αρχικό σήμα στις χρονικές του συνιστώσες.

Για τις ανάγκες του παρόντος εργαστηρίου υλοποποιούμε έναν προσαρμοζόμενο ομοιόμορφο κβαντιστή 2^{B_k} επιπέδων, όπου B_k ο αριθμός των bits κωδικοποίησης ανά δείγμα της ακολουθίας $y_k(n)$ στο τρέχον πλαίσιο ανάλυσης x(n) του σήματος. Το βήμα, Δ, αυτού του κβαντιστή προσαρμόζεται σε κάθε πλαίσιο ανάλυσης, καθώς τα επίπεδα του κβαντιστή πρέπει να είναι αρκετά έτσι ώστε το σφάλμα κβαντισμού να μη γίνεται αντιληπτό από τον άνθρωπο μετά τη συμπίεση του σήματος μουσικής. Το μέγιστο ανεκτό σφάλμα συνδέεται με το συνολικό κατώφλι κάλυψης T_g(i) του ψυχοακουστικού μοντέλου όπως προέκυψε από το Μέρος 1 και η σχέση υπολογισμού είναι :

$$B_k = \log_2\left(\frac{R}{\min(T_g(i)} - 1\right),$$

$$όπου = 2^{16}\ το\ πλήθος\ των\ βαθμίδων\ έντασης\ του\ αρχικού\ σήματος$$

Ουσιαστικά χωρίζουμε το T_g σε τμήματα των 8 δειγμάτων ώστε κάθε φίλτρο να έχει την δική του περιοχή (32*8=256). Το βήμα κβαντισμού Δ ρυθμίζεται βάσει του Bk και του εκάστοτε πεδίου τιμών $[x_{min}, x_{max}]$στο τρέχον πλαίσιο ανάλυσης.

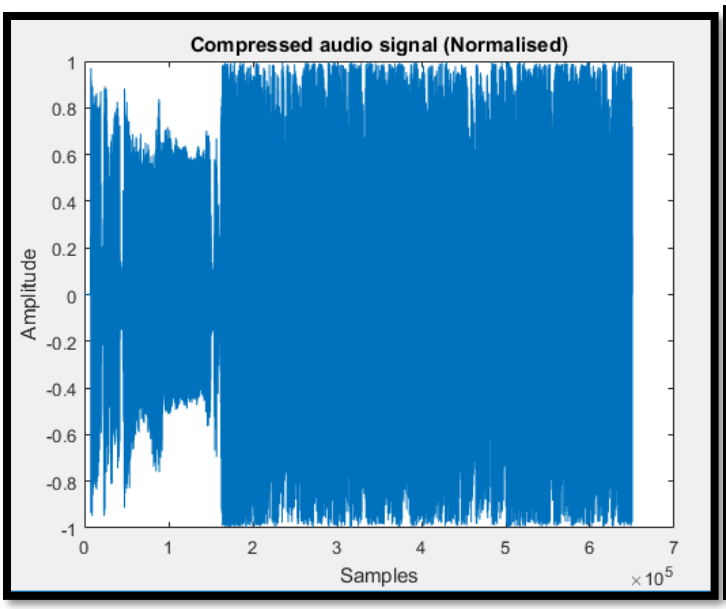
Στη συνέχεια, οι κβαντισμένες ακολουθίες $\hat{y}_k(n)$ στέλνονται στον αποκωδικοποιητή όπου παρεμβάλλονται με M μηδενικά και υπερδειγματοληπτούνται. Στη συνέχεια τα $w_k(n)$ συνελίσονται με το κάθε φίλτρο σύνθεσης. Τέλος, προσθέτουμε τα σήματα που έχουν προκύψει από το κάθε φίλτρο για να πάρουμε τελικό.

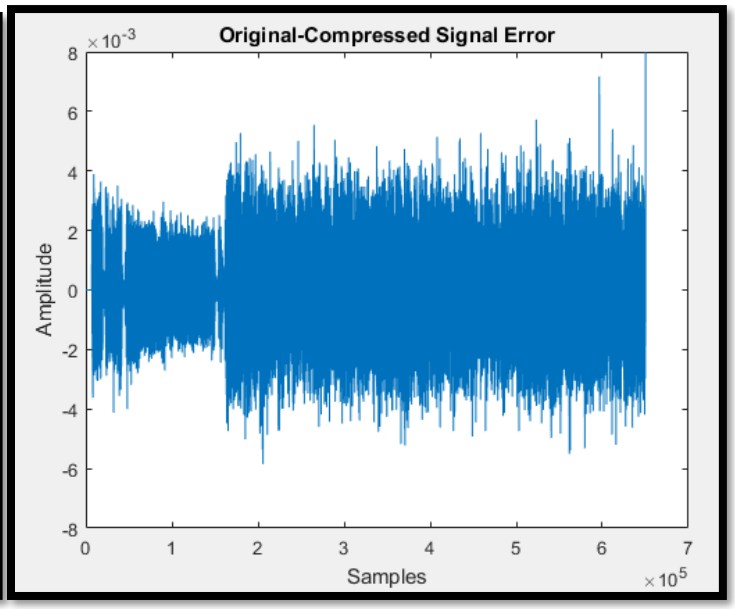
Τονίζεται πως για συμπιεστεί το σήμα χρησιμοποιώντας μη-προσαρμοζόμενο κβαντιστή πρέπει η τιμή της μεταβλητής *adap_or_not* να γίνει 0 από 1. Το τελικό αρχείο στην περίπτωση του adaptive quantizer θα λέγεται music_1.wav και στην περίπτωση του not-adaptive music_2.wav .

Η τελική ανακατασκευή του σήματος μουσικής $\hat{s}(n)$ γίνεται με εφαρμογή της τεχνικής OverLap-Add, όπως εξηγήθηκε στη θεωρία του Block Convolution, χωρίς επικάλυψη μεταξύ των διαδοχικών πλαισίων ανάλυσης, με τη χρήση της συνάρτησης **overlap_add**. Παρατηρούμε πως έχουμε το τελικό συμπιεσμένο σήμα μας σε μορφή μπλοκ αλλά με μέγεθος 639 αντί 512. Αυτό συμβαίνει λόγω της συνέλιξης του σήματος με τα φίλτρα. Έτσι, αυτό που κάνουμε είναι, καθώς φτιάχνουμε το τελικό μας σήμα **s** σε μορφή κατάλληλη για αποθήκευση σε αρχείο, να ξεκινάμε το επόμενο μπλοκ στο δείγμα 513 και τα ενδιάμεσα δείγματα να προσθέτονται μεταξύ τους. Τελικά, καταλήγουμε σε ένα σήμα με ελαφρώς λιγότερα δείγματα από το αρχικό. (650336 αντι 650415).

Αποτελέσματα και παρατηρήσεις

Παρατίθεται το τελικό σήμα για προσαρμοζόμενο κβαντιστή και το αντίστοιχο λάθος:

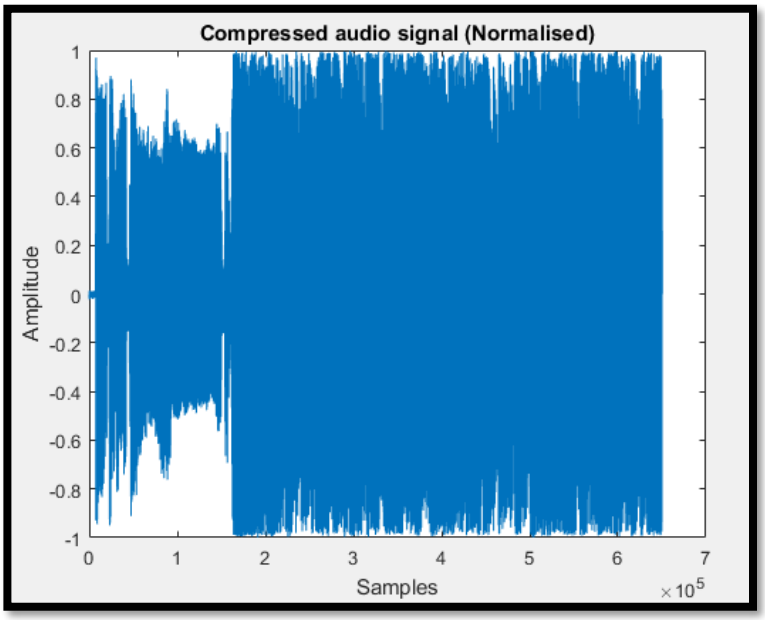


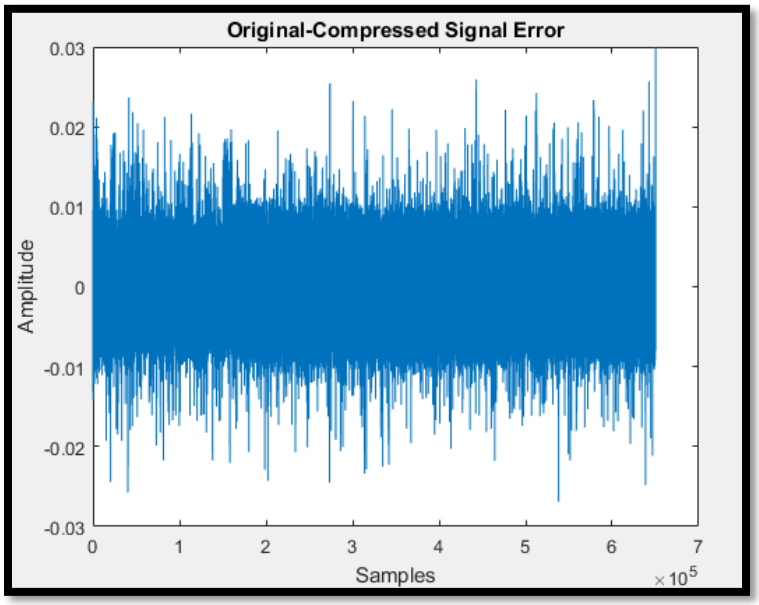


Το ποσοστό συμπίεσης είναι: 0,556 = 55,6%

Το μέσο τετραγωνικό λάθος είναι: 9.106384357737000e-06

Παρατίθεται το τελικό σήμα για μη-προσαρμοζόμενο κβαντιστή και το αντίστοιχο λάθος:





Το ποσοστό συμπίεσης είναι: 0,5 = 50%

Το μέσο τετραγωνικό λάθος είναι: 1.753761699514532e-05

Το ποσοστό συμπίεσης βρίσκεται με το διαιρέσουμε το άθροισμα όλων των bits που χρησιμοποιήθηκαν με τον αριθμό των φίλτρων, τον αριθμό των πλαισίων και τον αριθμό 16 (αριθμός bits που χρησιμοποιήθηκαν στο αρχικό σήμα).

Παρατηρούμε ακούγοντας τα και τα δύο ανακατασκευασμένα σήματα πως η διαφορά στην ποιότητα δεν είναι ιδιαίτερα αισθητή. Ωστόσο, μπορούμε με λίγο παραπάνω παρατήρηση να ακούσουμε μια ελαφρώς μεγαλύτερη διαφορά ανάμεσα στο αρχικό και το τελικό σήμα στην περίπτωση του μη-προσαρμοσμένου κβαντιστή. Αυτό αιτιολογείται από το μέσο τετραγωνικό σφάλμα το οποίο στην περίπτωση του μη-προσπαρμοζόμενου κβαντιστή είναι μία τάξη μεγαλύτερο από το αντίστοιχο στην περίπτωση του προσαρμοζόμενου κβαντιστή. Ο προσαρμοζόμενος κβαντιστής χρησιμοποιεί το μοντέλο που υλοποιήσαμε στο εργαστήριο με τους διάφορους αλγορίθμους που περιγράφηκαν και με σκοπό την βελτίωση της ποιότητας (ή έστω την βελτίωση της ανθρώπινης εμπειρίας στο άκουσμα του κομματιού).

Επίσης, είναι άξιο να σημειωθεί πως το ποσοστό συμπίεσης έιναι καλύτερο στην περίπτωση του μη-προσαρμοζόμενου κβαντιστή, καθώς εκεί χρησιμοποιούνται τα μισά bits απότι στο αρχικό σήμα μουσικής, εξού και το 50%. Έτσι, καταλήγουμε πως ενώ ο μη-προσαρμοζόμενος κβαντιστής κάνει καλύτερη συμπίεση οδηγεί εν γένει σε χειρότερ ποιότητα, ενώ το αντίστροφο ισχύει για τον προσαρμοζόμενο κβαντιστή, ο οποίος προσδίδει και μία παραπάνω πολυπλοκότητα αφού χρειάζεται την υλοποίηση του Ψυχοακουστικού μοντέλου.