Εξηγήστε περιεκτικά και επαρκώς την εργασία σας. Επιτρέπεται προαιρετικά η συνεργασία εντός ομάδων των 2 ατόμων. Κάθε ομάδα 2 ατόμων υποβάλλει μια κοινή αναφορά που αντιπροσωπεύει μόνο την προσωπική εργασία των μελών της. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός του βιβλίου και του εκπαιδευτικού υλικού του μαθήματος, πρέπει να το αναφέρετε. Η παράδοση της αναφοράς και του κώδικα της εργασίας θα γίνει ηλεκτρονικά στο moodle του μαθήματος: https://courses.pclab.ece.ntua.gr/course/view.php?id=16. Στη σελίδα αυτή, στην ενότητα 'Απορίες Εργαστηρίων', μπορείτε επίσης να υποβάλετε απορίες και ερωτήσεις δημιουργώντας issues.

Επισημαίνεται ότι απαγορεύεται η ανάρτηση των λύσεων των εργαστηριακών ασκήσεων στο github, ή άλλες ιστοσελίδες. Η σχεδίαση και το περιεχόμενο των εργαστηριακών projects αποτελούν αντικείμενο πνευματικής ιδιοκτησίας της διδακτικής ομάδας του μαθήματος.

Θέμα: Κωδικοποίηση σημάτων Μουσικής βάσει του ψυχοακουστικού μοντέλου (Perceptual Audio Coding)[1]

Πολλά σύγχρονα συστήματα κωδικοποίησης μουσικής (π.χ. MP3) βασίζονται στις ιδιότητες του συστήματος ακοής του ανθρώπου με στόχο να συμπιέσουν μία ηχογράφηση 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 γίνεται 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. Στόχος της άσκησης είναι να συμπιέσουμε το ηχητικό σήμα μουσικής διάρκειας περίπου 14 sec που είναι αποθηκευμένο στο αρχείο *music.wav*. Το σήμα έχει ηχογραφηθεί με συχνότητα δειγματοληψίας $F_s = 44100$ Hz και έχει κωδικοποιηθεί με PCM χρησιμοποιώντας 16 bits ανά δείγμα. Παρακάτω περιγράφονται οι βασικές έννοιες του ψυχοακουστικού μοντέλου:

Absolute Threshold of hearing

Το *κατώφλι ακοής* (Absolute Threshold of Hearing) 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 μη γραμμική συνάρτηση εκτίμησής του φαίνεται στο Σχήμα 1 και ισούται με:

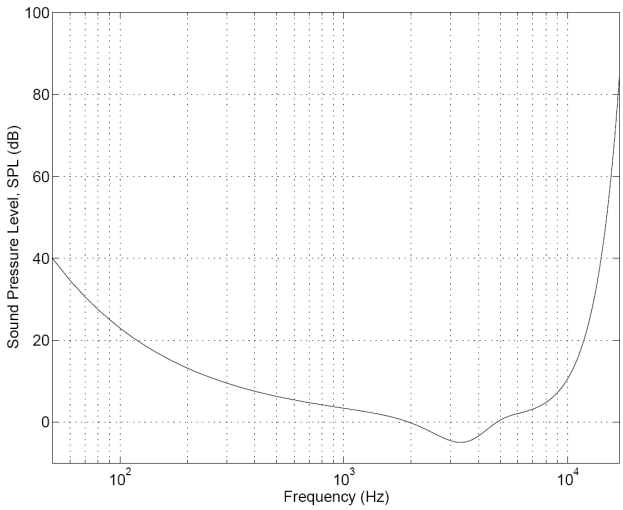
$$T_q(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4 \text{ (dB SPL)}. \quad (1)$$

Όταν το κατώφλι ακοής χρησιμοποιείται σε εφαρμογές συμπίεσης, τότε το $T_q(f)$ θα μπορούσε να ερμηνευθεί ως το μέγιστο επιτρεπόμενο ποσό ενέργειας για την κωδικοποίηση των παραμορφώσεων που εισάγονται στο πεδίο της συχνότητας.

Critical Bands

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

[1] Αυτή η άσκηση υλοποιεί περίπου το Layer I του MPEG-1 (Η τεχνολογία audio coding με MP3 είναι το Layer III)



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

συχνοτήτων της κλίμακας Hz στην κλίμακα Bark βάση του μη γραμμικού μοντέλου αντίληψης του ήχου δίνεται από την Εξ. (2):

$$b(f) = 13 \arctan(.00076 f) + 3.5 \arctan[(f/7500)^2] \text{ (Bark)}, \tag{2}$$

όπου f το διάνυσμα συχνοτήτων σε Hz.

Παραθυροποίηση
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.

Μέρος 1. Ψυχοακουστικό Μοντέλο 1

Σκοπός του Μέρους 1 είναι η δημιουργία μιας συνάρτησης που υλοποιεί το Ψυχοακουστικό Μοντέλο 1, η οποία παίρνει σαν είσοδο το παραθυροποιημένο πλαίσιο ανάλυσης και επιστρέφει το συνολικό κατώφλι κάλυψης T_g.

Βήμα 1.0: Κανονικοποίηση του σήματος
Κανονικοποιήστε το σήμα μουσικής αφού το διαβάσετε στη python, 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 χρησιμοποιώντας $N = 512$ δείγματα, και ανάλυση σε κάθε πλαίσιο.

Βήμα 1.1: Φασματική Ανάλυση
Ο στόχος σε αυτό το βήμα είναι να αποκτήσουμε μία υψηλής ανάλυσης εκτίμηση του φάσματος του σήματος εκφρασμένο σε μονάδες SPL (Sound Pressure Level) όπως εκφράζεται η πίεση του αέρα στο τύμπανο του αυτιού. Αρχικά ορίζουμε την κλίμακα Bark σύμφωνα με την Εξ. (2), 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.

$$P(k) = PN + 10\log_{10}\left|\sum_{n=0}^{N-1} w(n)x(n)e^{-j\frac{2\pi kn}{N}}\right|^2, 0 \leq k \leq \frac{N}{2}. \quad (3)$$

Η σταθερά κανονικοποίησης PN ισούται με 90.302 dB ενώ ως παράθυρο $w(n)$ χρησιμοποιούμε το Hanning το οποίο ορίζεται ως:

$$w(n) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N}\right)\right]. \quad (4)$$

Χρήσιμες python υπορουτίνες: **fft(), hanning()** της numpy.

Βήμα 1.2: Εντοπισμός μασκών τόνων και θορύβου (Maskers)

Αφού υπολογιστεί το φάσμα ισχύος $P(k)$, στη συνέχεια θέλουμε να εντοπίσουμε 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, όπως φαίνεται στην Εξ.(6). Στις υψηλές συχνότητες οι μάσκες καλύπτουν ευρύτερες γειτονιές.

Άρα η παρακάτω συνάρτηση $S_T(k)$, Εξ. (5), επιστρέφει boolean τιμές $\{0, 1\}$ που προσδιορίζουν αν στη θέση k υπάρχει τονική μάσκα. (Κάθε θέση k στην οποία η S_T έχει λογική τιμή 1 αντιστοιχεί σε μία τονική μάσκα που καλύπτει τις γειτονικές συχνότητες). Η συνάρτηση εντοπίζει τις τονικές μάσκες ελέγχοντας για τοπικά μέγιστα στις διαφορετικές συχνοτικές περιοχές, όπως ορίζονται παρακάτω:

$$S_T(k) = \begin{cases} 0, & \text{αν } k \notin [3, 250] \\ P(k) > P(k \pm 1) \bigwedge P(k) > P(k \pm \Delta_k) + 7\text{dB}, & \text{αν } k \in [3, 250] \end{cases} \quad (5)$$

όπου το $\bigwedge$ ισοδυναμεί με boolean and και το Δ_k:

$$\Delta_k \in \begin{cases} 2, & 2 < k < 63 \quad (0.17 - 5.5\text{kHz}) \\ [2, 3] & 63 \leq k < 127 \quad (5.5 - 11\text{kHz}) \\ [2, 6] & 127 \leq k \leq 250 \quad (11 - 20\text{kHz}) \end{cases} \quad (6)$$

Αφού βρείτε τις θέσεις των τονικών μασκών, υπολογίστε την ισχύ τους. Η ισχύς[2] $P_{TM}(k)$ της μάσκας στη θέση k υπολογίζεται με βάση τις τιμές του φάσματος ισχύος στις διακριτές συχνότητες $(k-1), k, (k+1)$ ως:

$$P_{TM}(k) = \begin{cases} 10\log_{10}(10^{0.1(P(k-1))} + 10^{0.1(P(k))} + 10^{0.1(P(k+1))})(\text{dB}), & \text{αν } S_T(k) = 1 \\ 0, & \text{αν } S_T(k) = 0 \end{cases} \quad (7)$$

Για την εύρεση των *μασκών του θορύβου* (noise maskers) σας δίνεται ο προυπολογισμένος πίνακας P_NM για το κάθε παράθυρο ανάλυσης, τον οποίο θα χρησιμοποιήσετε για να τις αναπαραστήσετε(**)[3].

Βήμα 1.3: Μείωση και αναδιοργάνωση των μασκών(**) [3]

Σε αυτό το βήμα μειώνουμε τον αριθμό των μασκών, χρησιμοποιώντας δυο διαφορετικά κριτήρια. Και σε αυτό το σημείο σας δίνονται 2 πίνακες, συγκεκριμένα οι P_TMc, P_NMc οι οποίοι περιέχουν τις τονικές μάσκες P_TMc και τις μάσκες θορύβου P_NMc για το κάθε παράθυρο ανάλυσης, μετά τη μείωση και την αναδιοργάνωση.

[2] P_{TM}, TM = Tone Masker, P_{NM}, NM = Noise Masker,

[3](**) Για όσους επιθυμούν, προαιρετικά, να υλοποιήσουν και να υπολογίσουν μόνοι τους τις μάσκες θορύβου καθώς και να υλοποιήσουν τη συνάρτηση για τη μείωση και την αναδιοργάνωση των μασκών, θα υπάρξει βαυμολογικό bonus μέχρι και 15% επί του βαθμού της άσκησης. Περαιτέρω πληροφορίες υπάρχουν στο Παράρτημα Α και Β.

Βήμα 1.4: Υπολογισμός των δυο διαφορετικών κατωφλίων κάλυψης (Individual Masking Thresholds)
Μετά την μείωση του αριθμού των μασκών (Βήμα 1.3), υπολογίζουμε τα δύο διαφορετικά κατώφλια κάλυψης. Το κάθε κατώφλι αντιπροσωπεύει το ποσοστό κάλυψης στο σημείο i το οποίο προέρχεται από την μάσκα τόνου ή θορύβου στο σημείο j. Το δύο κατώφλια υπολογίζονται ως:

$$T_{TM}(i,j) = P_{TM}(j) - 0.275b(j) + SF(i,j) - 6.025(\text{dB SPL}) \tag{8}$$

$$T_{NM}(i,j) = P_{NM}(j) - 0.175b(j) + SF(i,j) - 2.025(\text{dB SPL}) \tag{9}$$

όπου το $P_{TM}(j)$ και $P_{NM}(j)$ η ισχύς των μασκών (τόνων και θορύβου αντίστοιχα) στο σημείο j, και $b(j)$ οι συχνότητες στη κλίμακα Bark, Εξ.(2). Η συνάρτηση $SF(i,j)$ υπολογίζει την έκταση της κάλυψης από το σημείο j στο οποίο βρίσκεται η μάσκα έως το σημείο i το οποίο υφίσταται κάλυψη και μοντελοποιείται ως εξής:

$$SF(i,j) = \begin{cases} 17\Delta_b - 0.4P_{TM}(j) + 11, & -3 \le \Delta_b < -1 \\ (0.4P_{TM}(j) + 6)\Delta_b, & -1 \le \Delta_b < 0 \\ -17\Delta_b, & 0 \le \Delta_b < 1 \\ (0.15P_{TM}(j) - 17)\Delta_b - 0.15P_{TM}(j), & 1 \le \Delta_b < 8 \end{cases} \tag{10}$$

Η συνάρτηση $SF(i,j)$ προσεγγίζει το ελάχιστο επίπεδο ισχύος το οποίο πρέπει να έχουν οι γειτονικές συχνότητες έτσι ώστε να γίνουν αντιληπτές από τον άνθρωπο. Η $SF(i,j)$ υπολογίζεται για κάθε μάσκα τόνου και θορύβου, οι θέσεις j των οποίων μπορούν να εντοπιστούν ως $j : P_{TM}(j) > 0$ και $j : P_{NM}(j) > 0$. Στο συγκεκριμένο μοντέλο θεωρούμε πως η κάλυψη περιορίζεται σε μία γειτονιά των 12-Bark δηλαδή στις θέσεις $i : b(i) \in [b(j) - 3, b(j) + 8]$. Το $\Delta_b = b(i) - b(j)$ είναι η διαφορά των συχνοτήτων σε κλίμακα Bark μεταξύ της θέσης j της μάσκας και του κάθε σημείου i της γειτονιάς. Για τον υπολογισμό των κατωφλίων T_{NM}, στην Εξ.(10) αλλάζουμε το $P_{TM}(j)$ σε $P_{NM}(j)$.

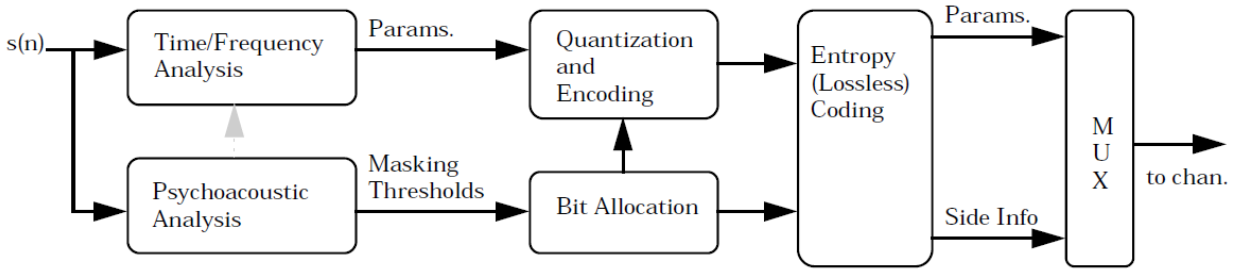
Βήμα 1.5: Υπολογισμός του συνολικού κατωφλίου κάλυψης (Global Masking Threshold)
Τα ξεχωριστά κατώφλια κάλυψης τα οποία υπολογίστηκαν στο Βήμα 1.4 συνδυάζονται για την δημιουργία του συνολικού κατωφλίου σε κάθε διακριτή συχνότητα ξεχωριστά. Το συνολικό κατώφλι $T_g(i)$ υπολογίζεται αθροιστικά με την παρακάτω Εξίσωση:

$$T_g(i) = 10\log_{10}\left(10^{0.1T_q(i)} + \sum_{l=1}^{L} 10^{0.1T_{TM}(i,\ell)} + \sum_{m=1}^{M} 10^{0.1T_{NM}(i,m)}\right) \text{ dB SPL}, \tag{11}$$

όπου το $T_q(i)$ είναι το Absolute Threshold of Hearing (ATH) σε κάθε διακριτή συχνότητα i, $T_{TM}(i,l)$ και $T_{NM}(i,m)$ τα ξεχωριστά κατώφλια κάλυψης των τόνων και του θορύβου αντίστοιχα, όπως υπολογίστηκαν στο Βήμα 1.4, και τα L και M ο αριθμός των μασκών (Βήμα 1.3). Ουσιαστικά, για τον υπολογισμό των επιμέρους κατωφλίων κάλυψης $T_{TM}(i,\ell)$ και $T_{NM}(i,m)$ για κάθε critical band, που αντιστοιχεί στις διαφορετικές μάσκες, αθροίζονται τα επιμέρους κατώφλια σε κάθε πλαίσιο ανάλυσης ξεχωριστά.

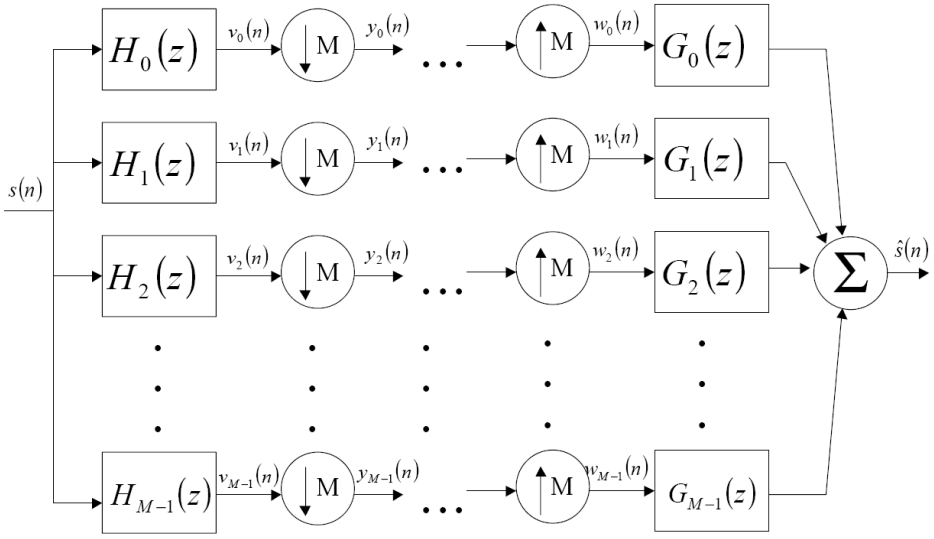
Μέρος 2. Χρονο-Συχνοτική Ανάλυση με Συστοιχία Ζωνοπερατών Φίλτρων

Όλοι οι κωδικοποιητές ήχου, δείτε για παράδειγμα το block diagram του Σχ. 2, βασίζονται σε κάποιο είδος χρονο-συχνοτικής ανάλυσης 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 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.



Σχήμα 2: Generic Perceptual Audio Encoder.

Στο Μέρος 2 φιλτράρουμε τα παράθυρα $x(n)$ του σήματος 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)
Όπως ήδη αναφέραμε για την ανάλυση του σήματος στις κρίσιμες συνιστώσες του, ανά χρονικό πλαίσιο, χρησιμοποιούνται συστοιχίες ζωνοπερατών φίλτρων. Σε αρκετά συστήματα συμπίεσης, 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. Οι κρουστικές αποκρίσεις τους είναι:

$$h_k(n) = \sin\left[\left(n + \frac{1}{2}\right)\frac{\pi}{2M}\right]\sqrt{\frac{2}{M}}\cos\left[\frac{(2n + M + 1)(2k + 1)\pi}{4M}\right] \tag{12}$$

και

$$g_k(n) = h_k(2M - 1 - n) \tag{13}$$

αντίστοιχα, με μήκος $L = 2M$, όπου $\sin[(n + \frac{1}{2})\frac{\pi}{2M}]$ (για $0 \leq n \leq L - 1$ και $0 \leq k \leq M - 1$) ένα βαθυπερατό φίλτρο γνωστό ως "ημιτονικό παράθυρο".

Βήμα 2.1: Ανάλυση με Συστοιχία Φίλτρων
Στο Βήμα 2.1 κάνουμε (i) συνέλιξη του σήματος $x(n)$ με τα φίλτρα σύνθεσης $h_k(n)$, όπου $k = 1, \ldots, M$ και (ii) υποδειγματοληψία του φιλτραρισμένου σήματος. Η ακολουθία εξόδου:

$$v_k(n) = h_k(n) * x(n) = \sum_{m=0}^{L-1} x(n - m)h_k(m),\ k = 0, 1, \ldots, M - 1, \tag{14}$$

από τη συνέλιξη υποδειγματοληπτείται κατά παράγοντα M για να διαιρεθεί το αρχικό σήμα στις χρονικές του συνιστώσες $y_k(n) = v_k(Mn)$.
Ο *αποδεκατισμός* (decimation) που επιδέχονται γίνεται στο μέγιστο βαθμό χωρίς να υπάρχουν φαινόμενα *επικάλυψης* (aliasing). Ωστόσο, τα μη ιδανικά ζωνοπερατά φίλτρα εισάγουν επικαλύψεις μεταξύ των συνιστωσών τις οποίες και θεωρούμε αμελητέες.

Βήμα 2.2: Κβαντοποίηση
Ένα απαραίτητο στοιχείο κάθε ψηφιακού κωδικοποιητή είναι ο κβαντιστής ο οποίος αντιστοιχεί τις τιμές των δειγμάτων μίας διακριτής ακολουθίας σε αριθμημένα επίπεδα κβάντισης. Για την κωδικοποίηση μουσικής συνήθως χρησιμοποιούνται μη-γραμμικοί κβαντιστές αλλά για τις ανάγκες του παρόντος εργαστηρίου θα υλοποποιήσετε 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 θα προσαρμόζεται σε κάθε πλαίσιο ανάλυσης. 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:

$$B_k = \log_2\left(\frac{R}{min(T_g(i))}\right) - 1, \tag{15}$$

όπου R το πλήθος των βαθμίδων έντασης του αρχικού σήματος $s(n)$. Το $T_g(i)$ ορίζεται στο διάστημα ορισμού του κάθε φίλτρου ανάλυσης, δηλαδή $i : f(i) \in \left[f_k - \frac{F_s\pi}{2M}, f_k + \frac{F_s\pi}{2M}\right]$, $k = 1, \ldots, M$ όπου f_k η κεντρική συχνότητα του κάθε φίλτρου. Στον προσαρμόζομενο κβαντιστή, για μικρότερο λάθος κβαντισμού, η θέση του πρώτου επιπέδου προσαρμόζεται με βάση την

ελάχιστη τιμή του σήματος σε κάθε πλαισίο ανάλυσης. Το βήμα κβαντισμού Δ ρυθμίζεται βάσει του B_k και του εκάστοτε πεδίου τιμών $[x_{min}, x_{max}]$ στο τρέχον πλαίσιο ανάλυσης.
Επίσης, πειραματιστείτε εξετάζοντας το αποτέλεσμα της κβαντοποίησης και της σύνθεσης αλλά χρησιμοποιώντας αυτή τη φορά έναν μη-προσαρμοζόμενο κβαντιστή με σταθερό αριθμό bit του κβαντιστή, όπου $B_k = 8$ και σταθερό βήμα κβαντισμού Δ, το οποίο καθορίζεται από ένα υποτιθέμενο σταθερό πεδίο τιμών του σήματος $[-1, 1]$.
Για περισσότερες διευκρινίσεις ως προς την υλοποίηση των κβαντιστών δείτε το **Παράρτημα Γ**.

Βήμα 2.3: Σύνθεση
Στη συνέχεια, οι κβαντισμένες ακολουθίες $\hat{y}_k(n)$ στέλνονται στον αποκωδικοποιητή όπου παρεμβάλλονται με M μηδενικά και υπερδειγματοληπτούνται, σύμφωνα με τη σχέση

$$w_k(n) = \begin{cases} \hat{y}_k(n/M), & n = 0, M, 2M, 3M, \dots \\ 0, & \text{αλλιώς.} \end{cases} \tag{16}$$

Εν συνεχεία, τα $w_k(n)$ συνελίσονται με τα φίλτρα σύνθεσης, Εξ.(13), που λόγω της αντιστροφής τους στο χρόνο ικανοποιούν τον περιορισμό ως προς τη γραμμική φάση.
Η έξοδος $\hat{x}(n)$ της συστοιχίας φίλτρων ισούται με τη μετατοπισμένη είσοδο $x(n - n_0)$ αν το σφάλμα της κβάντισης είναι μηδενικό, κάτι που δε συμβαίνει στις περισσότερες περιπτώσεις.
Το συνολικό πλαίσιο ανάλυσης-σύνθεσης φαίνεται στο Σχήμα 3.
Σημειώνεται ότι στον αποκωδικοποιητή εκτός από τους δείκτες των σταθμών στις οποίες κβαντίζονται τα δείγματα της ακολουθίας $y_k(n)$, στέλνονται επίσης η θέση της πρώτης στάθμης και το Δ και τα δύο κωδικοποιημένα σε 16 bits.

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.

ΠΑΡΑΔΟΤΕΑ: Αρχεία python και αναφορά που να περιέχει:

- Image plots (σχήματα εικόνων) από όλα τα ενδιάμεσα στάδια στο Μέρος 1 (Ψυχοακουστικό μοντέλο), συγκεκριμένα plots του κάθε βήματος για κάποιο πλαίσιο ανάλυσης. Το T_g υπολογίζεται για όλα τα πλαίσια.
- Δυο διαφορετικά αρχεία .wav μετά την τελική ανακατασκευή του σήματος της μουσικής για τις δυο διαφορετικές μεθόδους κβαντοποίησης: 1) ψυχοακουστικό μοντέλο με προσαρμοζόμενο κβαντιστή 2) σταθερό αριθμό bit κβαντισμού, $B_k = 8$ και μη-προσαρμοζόμενο κβαντιστή.
- Αποτελέσματα συγκρίσεων μεταξύ των δυο διαφορετικών μεθόδων (ποσοστά συμπίεσης ως προς το αρχικό σήμα) και το μέσο τετραγωνικό λάθος.
- Σχήματα εικόνων από το αρχικό και ανακατασκευασμένο σήμα (και με τις δύο μεθόδους), όπως και απεικόνιση του λάθους.
- Συνοπτική επεξήγηση τον αλγορίθμων και σχολιασμός των αποτελεσμάτων.

ΠΑΡΑΡΤΗΜΑ Α: Πληροφορίες σχετικά με το Βήμα 2 (Μέρος 1)

Στο Βήμα 2 σας δίνεται ο προυπολογισμένος πίνακας όπου και έχουν ήδη βρεθεί οι μάσκες θορύβου. Οι μάσκες θορύβου για κάθε critical band, $P_{NM}(\bar{k})$ υπολογίζονται στα σημεία τα οποία δεν ανήκουν μέσα στις γειτονιές $\pm\Delta$ των τονικών μασκών:

$$NoiseMember_i(k) = \begin{cases} 1, & \text{αν } k \notin j \pm \Delta_j \text{ όπου } j \in [\ell, u] \text{ με } S_T(j) = 1 \\ 0, & \text{αν } k \in j \pm \Delta_j \text{ όπου } j \in [\ell, u] \text{ με } S_T(j) = 1 \end{cases} \tag{17}$$

για όλα τα k που ανήκουν σε κάθε critical band (i), όπου $i = 1, ..., 25$. Και άρα

$$P_{NM}(\bar{k}) = 10\log_{10}\sum_j 10^{0.1P(j)} \text{ dB}, \text{ όπου } \bar{k} = (\prod_{j=\ell}^{u} j)^{1/(\ell-u+1)} \tag{18}$$

Για όσους επιθυμούν προαιρετικά να υλοποιήσουν την συνάρτηση για τον υπολογισμό των μασκών θορύβου θα χρειαστεί να υλοποιήσουν την συνάρτηση $[P_{NM}]$ όπου:

$$[P_{NM}] = findNoiseMaskers(P, P_{TM}, b).$$

Είσοδος της συνάρτησης είναι το φάσμα ισχύος του σήματος από το Βήμα 1.1, η ισχύς των τονικών μασκών P_{TM} τις οποίες υπολογίσατε και η κλίμακα συχνοτήτων Bark όπως ορίστηκε στην Εξ.(2). Έξοδος της συνάρτησης θα είναι η ισχύς των μασκών θορύβου P_{NM}.

ΠΑΡΑΡΤΗΜΑ Β: Πληροφορίες σχετικά με το Βήμα 3 (Μέρος 1)

Όσον αφορά την μείωση του αριθμό των μασκών, σας δόθηκαν και πάλι οι προυπολογισμένοι πίνακες. Η βασική ιδέα είναι για την μείωση και την αναδιοργάνωση των μασκών είναι η εξής: Χρησιμοποιούμε 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

   $$P_{TM,NM}(k) \geq T_q(k) \tag{19}$$

   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.

Για όσους επιθυμούν προαιρετικά να υλοποιήσουν την συνάρτηση για την μείωση και αναδιοργάνωση των μασκών θα χρειαστεί να υλοποιήσουν την συνάρτηση:

$$[P_{TM}, P_{NM}] = checkMaskers(P_{TM}, P_{NM}, T_q, b).$$

Είσοδος της συνάρτησης είναι το P_{TM} το οποίο υπολογίστηκε στο Βήμα 1.2, το P_{NM} (πληροφορίες για το οποίο βρίσκετε στο Παράρτημα Α, το Absolute Threshold of Hearing, Εξ. (1) και ο πίνακας των συχνοτήτων σε Bark $b(f)$, Εξ. (2). Έξοδος της συνάρτησης είναι τα καινούρια διανύσματα P_{TM} και P_{NM}, τα οποία θα χρησιμοποιήσετε στο Βήμα 1.4.

Σημείωση: Η επίλυση των δύο ερωτημάτων (Παράρτημα Α και Β) θα συνεισφέρει βαθμολογικό bonus μέχρι και 15% επί του βαθμού της άσκησης.

ΠΑΡΑΡΤΗΜΑ Γ: ΠΑΡΑΤΗΡΗΣΕΙΣ / ΔΙΕΥΚΡΙΝΙΣΕΙΣ

1. **Κανονικοποίηση**: Κανονικοποιήστε το σήμα μουσικής, διαιρώντας τα δείγματα του με την απόλυτη μέγιστη τιμή του σήματος, έτσι ώστε καθ΄ όλη τη διάρκεια του να είναι στο διάστημα $[-1, 1]$.

2. **Παραθυροποίηση**: Η επεξεργασία με το Ψυχοακουστικού Μοντέλου (Μέρος 1) και τη Συστοιχία Φίλτρων (Μέρος 2) θα εκτελεστεί για πλαίσια ανάλυσης $N = 512$ δειγμάτων **χωρίς επικάλυψη**.

3. **Κβαντοποίηση** (Βήμα 2.2): Ο κβαντιστής του ψυχοακουστικού μοντέλου που θα υλοποιήσετε θα προσαρμόζεται σε κάθε πλαίσιο ανάλυσης. Εκτός του αριθμού των bits που ορίζονται από το ψυχοακουστικό μοντέλο, ο κβαντιστής θα προσαρμόζεται και βάσει της ελάχιστης και μέγιστης τιμής του σήματος σε κάθε πλαίσιο ανάλυσης. Άρα, το βήμα κβαντισμού Δ ρυθμίζεται βάσει του B_k και του εκάστοτε πεδίου τιμών $[x_{min}, x_{max}]$ στο τρέχον πλαίσιο ανάλυσης.

   Στο βήμα αυτό σας ζητείται επίσης να πειραματιστείτε και να συμπιέσετε το σήμα χρησιμοποιώντας και σταθερό αριθμό bits του κβαντιστή, όπου $B_k = 8$ bits σε κάθε πλαίσιο ανάλυσης, και σταθερό βήμα κβαντισμού Δ, το οποίο καθορίζεται από ένα υποτιθέμενο σταθερό πεδίο τιμών του σήματος $[-1, 1]$.

4. **Τελική Ανασύνθεση**: Η τελική ανασύνθεση του σήματος μουσικής, θα γίνει με την τεχνική OverLap-Add, χωρίς επικάλυψη μεταξύ των διαδοχικών πλαισίων ανάλυσης. Διευκρίνιση: λόγω της συνέλιξης, το φιλτραρισμένο σήμα έχει περισσότερα δείγματα από το αρχικό πλαίσιο ανάλυσης, όπου $N = 512$. Τα επιπλέον αυτά δείγματα θα πρέπει να ληφθούν υπόψη στην τελική ανακατασκευή και να προστεθούν ανάλογα.

5. **Αποτελέσματα συγκρίσεων** μεταξύ των δυο διαφορετικών μεθόδων: Εκτός από τα ποσοστά συμπίεσης των δύο συμπιεσμένων σημάτων ως προς το αρχικό, σας ζητείται να υπολογίσετε και το *μέσο τετραγωνικό λάθος*, Mean Square Error (MSE), το οποίο ορίζεται ως η μέση τιμή του τετραγώνου της διαφοράς του αρχικού από το συμπιεσμένο σήμα. ΠΡΟΣΟΧΗ: Τα φίλτρα της σύνθεσης εισάγουν μία καθυστέρηση στο συμπιεσμένο σήμα, η οποία θα πρέπει να βρεθεί και να αφαιρεθεί πριν τον υπολογισμό του MSE. Απεικονίστε το αρχικό, το ανακατασκευασμένο σήμα, αλλά και το λάθος που προκύπτει (και με τις δύο μεθόδους).

6. **Προσοχή**: Κατεβάστε το μουτικό σήμα καθώς και τους προυπολογισμένους πίνακες που σας δίνονται έτοιμοι και βρίσκονται στο συμπληρωματικό υλικό στο moodle.

Λεπτομέρειες για το ψυχοακουστικό μοντέλο, τη συστοιχία φίλτρων, καθώς επίσης και για το μοντέλο κωδικοποίσης στο οποίο βασίζεται η συγκεκριμένη άσκηση θα βρείτε στα άρθρα (τα οποία μπορείτε να κατεβάσετε από το http://cvsp.cs.ntua.gr/courses/dsp/material.shtm:

[1] T. Painter, A. Spanias, "Perceptual Coding of Digital Audio", IEEE Proceedings 2000.

[2] T. Painter, A. Spanias, "A Review of Algorithms for Perceptual Coding of Digital Audio Signals", in Proc. of 13th Int'l Conf. on Digital Signal Processing, 1997.