

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 στη σελίδα του μαθήματος: <https://helios.ntua.gr/course/view.php?id=872>. 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: Εισαγωγή στην Ψηφιακή Επεξεργασία Σημάτων με Python και Εφαρμογές σε Ακουστικά Σήματα

Μέρος 1ο - Σύστημα Εντοπισμού Τηλεφωνικών Τόνων (Telephone Touch – Tones)

Στόχος αυτής της άσκησης είναι να παρουσιάσει πως λειτουργεί το τηλεφωνικό τονικό σύστημα χρησιμοποιώντας σήματα διαφορετικών συχνοτήτων για να εντοπίζει ποιο πλήκτρο έχει πατηθεί. Ο εντοπισμός αυτών των συχνοτήτων μπορεί να γίνει με την χρήση του Διακριτού Μετ/σμού Fourier (DFT) $X[k]$ του τηλεφωνικού σήματος $x[n]$. Με το πάτημα ενός πλήκτρου στο τηλέφωνο ακούγεται ένας ήχος που είναι το άθροισμα 2 ημιτόνων, το υψίσυχο ημίτονο δείχνει την στήλη που ανήκει το πλήκτρο στο touch-pad της τηλεφωνικής συσκευής και το χαμηλόσυχο ημίτονο δείχνει την αντίστοιχη γραμμή του touch-pad σύμφωνα με τον Πίνακα 1 που ακολουθεί. Παραδείγματος χάρη, το πλήκτρο 5 αντιστοιχεί στο σήμα $d_5[n] = \sin(0.5906n) + \sin(1.0247n)$.

	Ω_{column}		
Ω_{row}	0.9273	1.0247	1.1328
0.5346	1	2	3
0.5906	4	5	6
0.6535	7	8	9
0.7217		0	

Πίνακας 1: Διακριτές Συχνότητες για Τηλεφωνικούς Τόνους για Συχνότητα Δειγματοληψίας 8192 Hz.

- 1.1. Δημιουργήστε τους 10 διαφορετικούς τόνους σύμφωνα με την παραπάνω εκφώνηση. Κάθε τόνος πρέπει να έχει μήκος 1000 δείγματα. (Υπόδειξη: Μπορείτε να χρησιμοποιήσετε την ρουτίνα **Audio()** από το πακέτο `IPython.display` για να ακούσετε τους ήχους.)
- 1.2. Υπολογίστε τον DFT, $N = 1024$ σημείων, των σημάτων $d_3[n]$, $d_8[n]$ και $d_9[n]$ και δημιουργήστε τις γραφικές παραστάσεις των $|D_3[k]|$, $|D_8[k]|$ και $|D_9[k]|$. Τι παρατηρείτε ως προς τη μορφή τους;
- 1.3. Δημιουργήστε και αποθηκεύστε σε αρχείο “tone_sequence.wav” ένα σήμα με διαδοχικούς τηλεφωνικούς τόνους ‘μεταφράζοντας’ το άθροισμα των αριθμών μητρώου των μελών της

κάθε ομάδας σε τονικά σήματα. Κάθε ένα ψηφίο του αθροίσματος πρέπει να διαχωρίζεται από το προηγούμενο με 100 μηδενικά δείγματα. Για παράδειγμα αν οι αριθμοί μητρώου των δύο μελών είναι $AM_1 = 03092432$ και $AM_2 = 03093543$ τότε τα ζητούμενα ψηφία είναι: 0 6 1 8 5 9 7 5 ($= 03092432 + 03093543$). Αν η εργασία υλοποιείται από ένα άτομο τότε τα ζητούμενα ψηφία είναι ο αριθμός μητρώου.

Χρήσιμες Συναρτήσεις: **write()** από τη βιβλιοθήκη soundfile, ή από το πακέτο scipy.io.wavfile.

- 1.4. Χρησιμοποιώντας την συνάρτηση **fft()** της numpy και κατάλληλα χρονικά παράθυρα, (i.) τετραγωνικά, και (ii.) Hamming, μήκους $L = 1000$, υπολογίστε τον Μετ/σμό Fourier των παραθυροποιημένων σημάτων, $N = 1024$ σημείων, όπως προκύπτουν από το προηγούμενο σήμα. (Υπόδειξη: Δημιουργήστε τόσα παραθυροποιημένα σήματα όσοι και οι τόνοι που περιέχονται στο αρχικό σήμα).
- 1.5. Υπολογίστε μια λίστα από δείκτες k και τις αντίστοιχες συχνότητες που θεωρείτε ότι βρίσκονται εγγύτερα στις touch-tone συχνότητες.
- 1.6. Δημιουργήστε μια συνάρτηση με το όνομα **ttdecode()**, που θα δέχεται σαν όρισμα εισόδου ένα τονικό σήμα (όπως περιγράφηκε στο **Ερωτ. 1.3**) και επιστρέφει ένα διάνυσμα με τα αντίστοιχα ψηφία. Για παράδειγμα αν το σήμα εισόδου **signIn** περιέχει τους τόνους για το νούμερο 210 – 3434120, η έξοδος της συνάρτησης **Vector** θα είναι $Vector = 2\ 1\ 0\ 3\ 4\ 3\ 4\ 1\ 2\ 0$. Επιβεβαιώστε την ορθή λειτουργία της ρουτίνας θέτοντας σαν είσοδο, το σήμα του **Ερωτ. 1.3**.

Υπόδειξη: Η συνάρτηση θα πρέπει πρώτα να υπολογίζει την ενέργεια κάθε ενός από τους τόνους του σήματος εισόδου με την χρήση της ρουτίνας **fft()**. Έπειτα να εντοπίζει ποιες είναι εκείνες οι συχνότητες που έχουν την μεγαλύτερη ενέργεια και να τις αντιστοιχίζει στις αρχικές συχνότητες του Πίνακα 1. Με βάση αυτές τις αντιστοιχίσεις, εντοπίστε ποιο είναι το αντίστοιχο ψηφίο. Υπενθυμίζεται ότι η ενέργεια E_k του σήματος γύρω από τη συχνότητα με δείκτη k ισούται με $E_k = |X[k]|^2$.

- 1.7. Με χρήση της εντολής **load()** της numpy φορτώστε τα αρχεία *easy_sig.npy*, *medium_sig.npy*, *hard_sig.npy* από το συμπληρωματικό υλικό της άσκησης “dsp25_lab1_Data.zip” που βρίσκεται στο Helios, τα οποία και αντιστοιχούν σε τρία διαφορετικά σήματα, τα *easySig*, *mediumSig* και *hardSig*. Προσδιορίστε τα ψηφία στα οποία αντιστοιχούν οι τόνοι των 3 σημάτων με την χρήση της ρουτίνας **ttdecode()**.

Σημειώνεται ότι σε κάθε περίπτωση οι απαντήσεις σας πρέπει να συνοδεύονται με τις σχετικές γραφικές παραστάσεις και σχόλια ώστε να είναι όσο το δυνατό τεκμηριωμένες.

Μέρος 2ο - Χαρακτηριστικά Βραχέος Χρόνου Σημάτων Φωνής και Μουσικής

Οι μετρήσεις βραχέος χρόνου είναι μετρήσεις που γίνονται σε ένα μετακινούμενο παράθυρο του σήματος. Μετρήσεις αυτής της κατηγορίας μπορούν να ληφθούν απευθείας από το πεδίο του χρόνου, όπως η ενέργεια βραχέος χρόνου που ορίζεται ως:

$$E_n = \sum_{m=-\infty}^{\infty} |x[m]|^2 w[n-m], \quad (1)$$

όπου w ένα παράθυρο της επιλογής μας, το οποίο συνήθως είναι το Hamming παράθυρο (συνάρτηση **hamming()** στη numpy). Αντίστοιχα, ο ρυθμός εναλλαγής προσήμου (Zero Crossing Rate), για την περίπτωση που χρησιμοποιήσουμε τετραγωνικό παράθυρο N δειγμάτων, με

πλάτος $1/2N$, ορίζεται ως:

$$Z_n = \sum_{m=-\infty}^{\infty} |\text{sgn}(x[m]) - \text{sgn}(x[m-1])| w[n-m]. \quad (2)$$

Αντίστοιχες μετρήσεις βραχέος χρόνου μπορούν να ληφθούν και στο πεδίο της συχνότητας, όπως για παράδειγμα το φασματικό κέντρο (spectral centroid - SC), το οποίο και ορίζεται ως η μέση συχνότητα ενός σήματος σε ένα χρονικό παράθυρο, και η φασματική ροή (spectral flux - SF), η οποία ορίζεται ως η μεταβολή του φασματικού περιεχομένου ανάμεσα σε δύο διαδοχικά χρονικά παράθυρα. Συγκεκριμένα, αν $|X[k, m]|$ είναι το μέτρο του διακριτού μετασχηματισμού Fourier βραχέος χρόνου (Short Time Fourier Transform - STFT), N σημείων, ενός σήματος $x[n]$, οι παραπάνω μετρήσεις για το m -οστό παράθυρο του σήματος δίνονται από τις εξισώσεις:

$$SC_m = \frac{\sum_{k=0}^{N/2} k |X[k, m]|^2}{\sum_{k=0}^{N/2} |X[k, m]|^2} \cdot \frac{f_s}{N} \quad (3)$$

$$SF_m = \left\| \frac{|X[k, m+1]|^2}{\sum_{k=0}^{N/2} |X[k, m+1]|^2} - \frac{|X[k, m]|^2}{\sum_{k=0}^{N/2} |X[k, m]|^2} \right\|_2, k = 0, 1, \dots, N/2. \quad (4)$$

- 2.1. Θεωρήστε το σήμα φωνής της πρότασης “Oh mother, I saw them” που περιέχεται στο αρχείο “speech_utterance.wav” (συχνότητα δειγματοληψίας: 16 kHz) του συμπληρωματικού υλικού “dsp25_lab1_Data.zip” της άσκησης στο Helios. Απεικονίστε το σήμα στο πεδίο του χρόνου.
- 2.2. Υπολογίστε την ενέργεια βραχέος χρόνου και το ρυθμό εναλλαγής προσήμου του σήματος, βάσει των τύπων (1) και (2), και παραθέστε τις μετρήσεις αυτές σε κοινό διάγραμμα με το αρχικό σας σήμα. Χρησιμοποιήστε παράθυρα μήκους 20-50 ms. Τι παρατηρείτε μεγάλωνοντας το μήκος του παραθύρου; Μπορείτε να χρησιμοποιήσετε αυτές τις μετρήσεις για να διαχωρίσετε φωνή από σιωπή ή έμφωνους (π.χ. /aa/, /ih/) από άφωνους ήχους (π.χ. /f/, /p/);
- 2.3. Υπολογίστε το μετασχηματισμό Fourier βραχέος χρόνου (χρήσιμες συναρτήσεις: librosa.stft(), scipy.signal.stft()) του σήματος, και σχεδιάστε το πλάτος του σε λογαριθμική κλίμακα. Έπειτα, χρησιμοποιήστε τον υπολογισμένο STFT ώστε να εξάγετε τις τιμές των μετρήσεων του φασματικού κέντρου και της φασματικής ροής. Για τον STFT, πειραματιστείτε με το εύρος του παραθύρου (ενδεικτικές τιμές: 20-50 ms) και χρησιμοποιήστε $N = 2048$ σημεία. Τι παρατηρείτε;
- 2.4. Επαναλάβετε την ίδια διαδικασία για το σήμα μουσικής “music.wav” (συχνότητα δειγματοληψίας: 48 kHz) που επίσης βρίσκεται στο συμπληρωματικό υλικό, και σχολιάστε τη μεταβολή των τιμών των χαρακτηριστικών ανά το χρόνο.

Μέρος 3ο - Ψηφιακή Επεξεργασία και Φασματική Ανάλυση Βιοσημάτων Κητωδών

Στόχος της άσκησης που ακολουθεί είναι η ψηφιακή επεξεργασία και φασματική ανάλυση δύο βιοσημάτων που προέρχονται από θαλάσσια θηλαστικά, έναν φουσητήρα (Sperm whale) και μια megάπτερη φάλαινα (Humpback whale). Οι φουσητήρες εκπέμπουν παλμικούς ασυνεχείς ήχους (clicks - codas) για ηχοεντοπισμό ενώ οι megάπτερες φάλαινες εκπέμπουν συνεχή σήματα (songs) που συχνά συνοδεύονται από αρμονικές. Τα σήματα ηχοεντοπισμού που εκπέμπουν τα κητώδη έχουν συλλεχθεί από υδρόφωνα. Μπορείτε να τα βρείτε σε μορφή αρχείων wav στην ανοιχτή βάση δεδομένων: ‘Watkins Marine Mammal Sound Database, Woods Hole Oceanographic

Institution and the New Bedford Whaling Museum.¹ αλλά για ευκολία τα παραθέτουμε και στο συμπληρωματικό υλικό “dsp25_lab1_Data.zip” της άσκησης στο Helios.



Σχήμα 1: Φυσητήρας (αριστερά) και megάπτερη φάλαινα (δεξιά)².

- 3.1. Το πρώτο βήμα για την επεξεργασία των raw βιοσημάτων κητωδών (‘85005006.wav’ (Sperm whale) και ‘9220100Q.wav’ (Humpback whale)) είναι η μετατροπή τους από σήματα Τάσης σε σήματα Πίεσης. Για την μετατροπή του σήματος Τάσης σε σήμα Πίεσης χρησιμοποιούμε τα χαρακτηριστικά λειτουργίας του υδρόφωνου, δηλαδή τον συντελεστή ενίσχυσης (Gain Factor – GF) και την ευαισθησία του οργάνου (Sensitivity) ενώ ο υπολογισμός γίνεται από την ακόλουθη σχέση:

$$pressure = \frac{1.5 \cdot y \cdot GF}{10^{(sensitivity/20)}}, \quad (5)$$

όπου η μεταβλητή y αντιπροσωπεύει τα raw βιοσήματα των δύο wav αρχείων, τα οποία φορτώνετε με τη βοήθεια της συνάρτησης `librosa.load()`, και η μεταβλητή GF αντιπροσωπεύει το gain factor του υδρόφωνου που δίνεται ίσο με 0.16. Η ευαισθησία του οργάνου είναι ίση με -155 dB relative to $1V/\mu Pa$. Μετατρέψτε τα raw σήματα σε σήματα Πίεσης και σχεδιάστε το διάγραμμα Πίεσης-Χρόνου για τα δύο θηλαστικά.

- 3.2. Υπολογίστε τη μέση τετραγωνική ρίζα (rms) της πίεσης που δίνεται από τη σχέση:

$$p_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N pressure_i^2}, \quad (6)$$

όπου N είναι το μήκος του σήματος.

Υπολογίστε επίσης το επίπεδο ηχητικής πίεσης (Sound Pressure Level – SPL) που ορίζεται ως:

$$SPL_{RMS} = 20 \log_{10} \frac{p_{rms}}{p_{ref}}, \quad (7)$$

όπου: $p_{ref} = 1 \mu Pa$ και η μονάδα μέτρησης του SPL_{RMS} είναι σε dB relative to $1 \mu Pa$.

¹<https://whoicf2.who.edu/science/B/whalesounds/index.cfm>

²Φυσητήρας: <https://www.discoverwildlife.com/animal-facts/marine-animals/sperm-whale-facts>
Μεγάπτερη φάλαινα https://en.wikipedia.org/wiki/Humpback_whale

- 3.3. Επιβεβαιώστε για ένα από τα δύο βιοσήματα με τη βοήθεια της συνάρτησης `numpy.fft.fft()` το θεώρημα του Parseval [1]:

$$\sum_{n=0}^{N-1} |p[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |P[k]|^2, \quad (8)$$

για το μέγεθος της πίεσης p που υπολογίσατε στο **Ερώτ. 3.1**.

- 3.4. Για να μελετήσουμε το βιοσήμα ενός θαλάσσιου θηλαστικού συχνά απαλείφουμε από το αρχικό σήμα τις χαμηλές συχνότητες στις οποίες εμφανίζεται ο ανθρωπογενής θόρυβος (κυρίως θόρυβος που προέρχεται από πλοία). Σχεδιάστε και εφαρμόστε στα δύο βιοσήματα ένα υπερβατό Butterworth φίλτρο, τάξης $N = 3$ (συναρτήσεις `scipy: scipy.signal.butter()` και `scipy.signal.lfilter()` ή `scipy.signal.filtfilt()`), το οποίο να απομονώνει συχνότητες μεγαλύτερες από 200 Hz. Σχεδιάστε για κάθε βιοσήμα σε κοινό διάγραμμα το γράφημα Πίεσης-Χρόνου και το γράφημα του σήματος Πίεσης-Χρόνου μετά την εφαρμογή του Butterworth φίλτρου. Εξηγήστε τα αποτελέσματά σας.

- 3.5. Ας υποθέσουμε ότι οι μεγάπτερες φάλαινες ακούνε σε συχνότητες από $f_{lower} = 300\text{Hz}$ έως $f_{higher} = 3\text{kHz}$ ενώ οι φυσητήρες σε όλο το φάσμα. Κάνοντας την παραδοχή ότι επίπεδα ηχητικής πίεσης SPL_{rms} που υπερβαίνουν συστηματικά (όχι στιγμιαία) τα 100 dB relative to $1\mu\text{Pa}$ επηρεάζουν τη συμπεριφορά των κητωδών εξετάστε αν ο ανθρωπογενής υποθαλάσσιος ήχος που δίνεται στο επεξεργασμένο αρχείο που προέρχεται από την N.O.A.A [2] με τίτλο ‘Pile driving.wav’ και έχει συλλεχθεί από υδρόφωνο με χαρακτηριστικά λειτουργίας $GF = 0.16$ και $Sensitivity = -125$, μπορεί να επηρεάσει τη συμπεριφορά τους.

- 3.6. Υπολογίστε την ενέργεια βραχέος χρόνου (Short-Time Energy) του αρχικού σήματος Πίεσης-Χρόνου. Σχεδιάστε σε κοινό διάγραμμα το αρχικό σήμα Πίεσης-Χρόνου και αυτό της ενέργειας και για τα δύο βιοσήματα. Σχολιάστε τα αποτελέσματα.

- 3.7. Υλοποιήστε τον μη-γραμμικό ενεργειακό τελεστή Teager-Kaiser Energy Operator (TEO), ο οποίος θα εφαρμοστεί στα σήματα Πίεσης-Χρόνου $p[n]$ των δύο βιοσημάτων. Ο ενεργειακός τελεστής Teager-Kaiser προτάθηκε για πρώτη φορά από τον Teager και παρουσιάστηκε συστηματικά από τον Kaiser και έχει χρησιμοποιηθεί εκτενώς ως εργαλείο για αποδιαμόρφωση διαφορών κατηγοριών σημάτων έχοντας ως βασική ιδιότητα να ανιχνεύει την ενέργεια ενός γραμμικού ταλαντωτή³.

Για σήματα συνεχούς χρόνου, ο ενεργειακός τελεστής Teager στην γενική του περίπτωση ορίζεται ως:

$$\Psi[x(t)] = (x'(t))^2 - x(t)x''(t). \quad (9)$$

Για την περίπτωση του διακριτού-χρόνου σήματος Πίεσης-Χρόνου, $p[n]$, αντικαθιστώντας τις παραγώγους με τους ισοδύναμους τελεστές διακριτού χρόνου, η ενέργεια Teager παίρνει την ακόλουθη μορφή:

$$\Psi[p[n]] = p^2[n] - p[n-1]p[n+1]. \quad (10)$$

³Για λεπτομέρειες σχετικά με τον ενεργειακό τελεστή Teager δείτε το [3], ενώ για διάφορες εφαρμογές με χρήση του AM-FM μοντέλου και του TEO δείτε το [4].

Υπόδειξη: Στα όρια του σήματος θα πρέπει να επαναλάβετε τις τιμές.

Αφού την υλοποιήσετε σχεδιάστε όπως και στο προηγούμενο ερώτημα την γραφική της παράσταση για κάθε ένα από τα δύο βιοσήματα μαζί με τα σήματα της Πίεσης-Χρόνου. Τι παρατηρείτε σε σχέση με την βραχυπρόθεσμη ενέργεια που υπολογίσατε στο **Ερωτ. 3.6**;

- 3.8. Ένας εκτιμητής της φασματικής πυκνότητας μπορεί να προσδιοριστεί από το μέσο τροποποιημένο περιοδόγραμμα του Welch [5] μέσω των σχέσεων:

$$PSD_r[k] = \frac{1}{NB} |P_r[k]|^2, \overline{PSD}[k] = \frac{1}{K} \sum_{r=0}^{K-1} PSD_r[k] \quad (11)$$

όπου $P_r[k]$ ο DFT N -σημείων του παραθυρωμένου τμήματος r του σήματος Πίεσης-Χρόνου $p_r[n] = p[rR + n]w[n]$, N το μήκος του παραθύρου, $0 \leq R \leq N - 1$ και B μια σταθερά κανονικοποίησης που δίνεται από τη σχέση: $B = \frac{1}{N} |\sum_{n=0}^{N-1} w^2(n)|$. Τέλος \overline{PSD} είναι μέσος όρος των Περιοδογραμμάτων των K τμημάτων του σήματος και η μονάδα μέτρησής του είναι σε dB relative to $1\mu Pa^2/Hz$. Σχεδιάστε το περιοδόγραμμα του Welch σε λογαριθμική κλίμακα με τη βοήθεια της συνάρτησης `scipy.signal.welch()`. Τι παρατηρείτε σχετικά με την κατανομή της ενέργειας στο φάσμα στα δύο σήματα;

- 3.9. Σχεδιάστε το φασματογράφημα (spectrogram) των δύο βιοσημάτων. Για να το κάνετε αυτό υπολογίστε τον Fourier μετασχηματισμό των σημάτων με τη βοήθεια της συνάρτησης `librosa.stft()`. Απεικονίστε στη συνέχεια το τετράγωνο του πλάτους του μετασχηματισμού Fourier σε κλίμακα dB με τη βοήθεια των συναρτήσεων `librosa.amplitude_to_db()` και `librosa.display.specshow`. Παρατηρήστε την κατανομή της ενέργειας των βιοσημάτων στο φάσμα και σχολιάστε αν συμφωνεί με το αποτέλεσμα του προηγούμενου ερωτήματος.
- 3.10. Μια τεχνική που επιτρέπει στην ενίσχυση της αντίθεσης ανάμεσα στον background θόρυβο και τα foreground μεταβατικά γεγονότα (όπως είναι τα σήματα που μελετούμε) είναι η Per Channel Energy Normalization (PCEN) [6]. Σχεδιάστε τα mel-φασματογραφήματα (mel-spectrograms) των σημάτων μέσω και των συναρτήσεων `librosa.feature.melspectrogram()`, `librosa.amplitude_to_db()` και στη συνέχεια με τη βοήθεια της συνάρτησης `librosa.pcen()` απεικονίστε το αποτέλεσμα για κάθε ένα από τα βιοσήματα. Σχολιάστε τα αποτελέσματα, συγκρίνοντας τα mel-φασματογραφήματα πριν και μετά την εφαρμογή της μεθόδου PCEN.

ΠΑΡΑΔΟΤΕΑ Ηλεκτρονική παράδοση του κώδικα Python, της ακολουθίας touch-tone υπο μορφή .wav που φτιάξατε στο **Ερώτημα 1.3** του Μέρους 1, καθώς και συνοπτικής αναφοράς που θα απαντάει στα δοθέντα ερωτήματα και θα περιλαμβάνει τις ζητούμενες γραφικές αναπαραστάσεις. Η αναφορά προτείνεται να είναι σε μορφή .pdf.

References

- [1] James H. McClellan, Ronald W. Schafer, and Mark A. Yoder, *DSP First (2nd Edition)*, Prentice-Hall, Inc., USA, 2007.

- [2] “National Oceanic and Atmospheric Administration, Passive Acoustics group. stf Multi-sound NOAA PAGroup 01,” 2021.
- [3] Petros Maragos, James F Kaiser, and Thomas F Quatieri, “Energy separation in signal modulations with application to speech analysis,” *IEEE Transactions on Signal Processing*, vol. 41, pp. 3024–3051, 1993.
- [4] Petros Maragos, “Summer School on Speech Signal Processing, S4P Lecture Slides: Lecture I: Nonlinear aspects of speech production: Modulations and energy operators, Interspeech-2018. <http://cvsp.cs.ntua.gr/interspeech2018.shtm>,” 2018.
- [5] Oppenheim A. Schafer R., *Discrete Signal Processing*, Prentice Hall Signal Processing Series, 1999.
- [6] Vincent Lostanlen, Justin Salamon, Mark Cartwright, Brian McFee, Andrew Farnsworth, Steve Kelling, and Juan Pablo Bello, “Per-channel energy normalization: Why and how,” *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 39–43, 2019.