



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Ψηφιακή Επεξεργασία Σήματος

1^η Εργαστηριακή Άσκηση

Εισαγωγή στην Ψηφιακή Επεξεργασία Σημάτων με Python και Εφαρμογές σε Ακουστικά Σήματα

Ονοματεπώνυμο
ΑΜ

Γεώργιος Αλέξανδρος Γεωργαντζάς
03120017

Βλάσιος Σκουλής
03120065

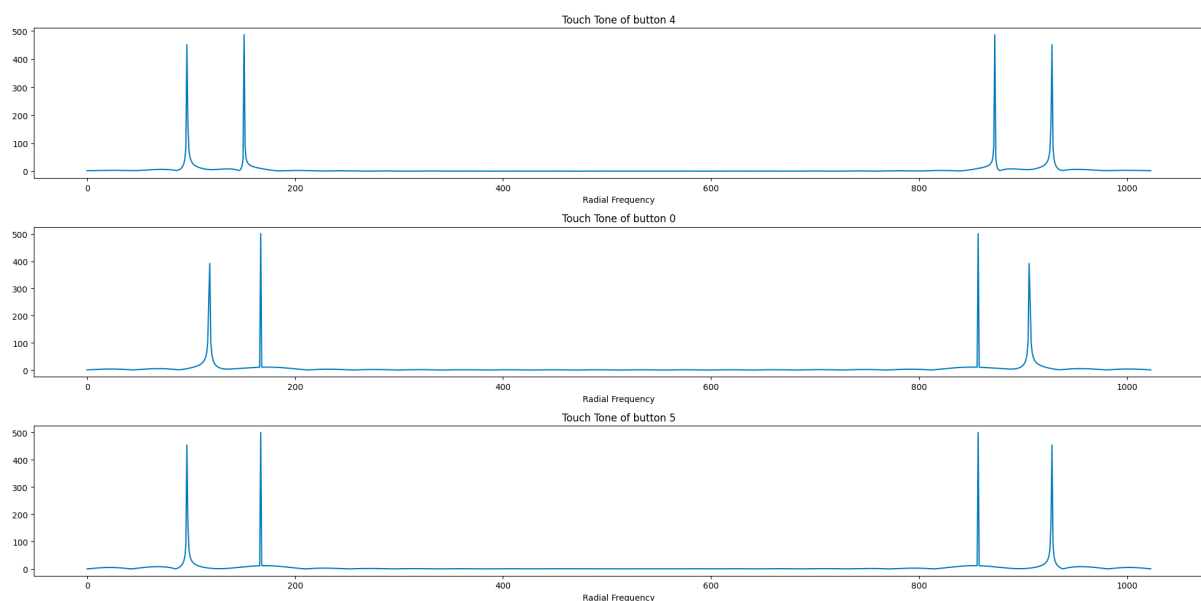
Κυριακή 14 Απριλίου 2024

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

1.1

Δημιουργούμε τα 10 σήματα του touch-pad.

1.2

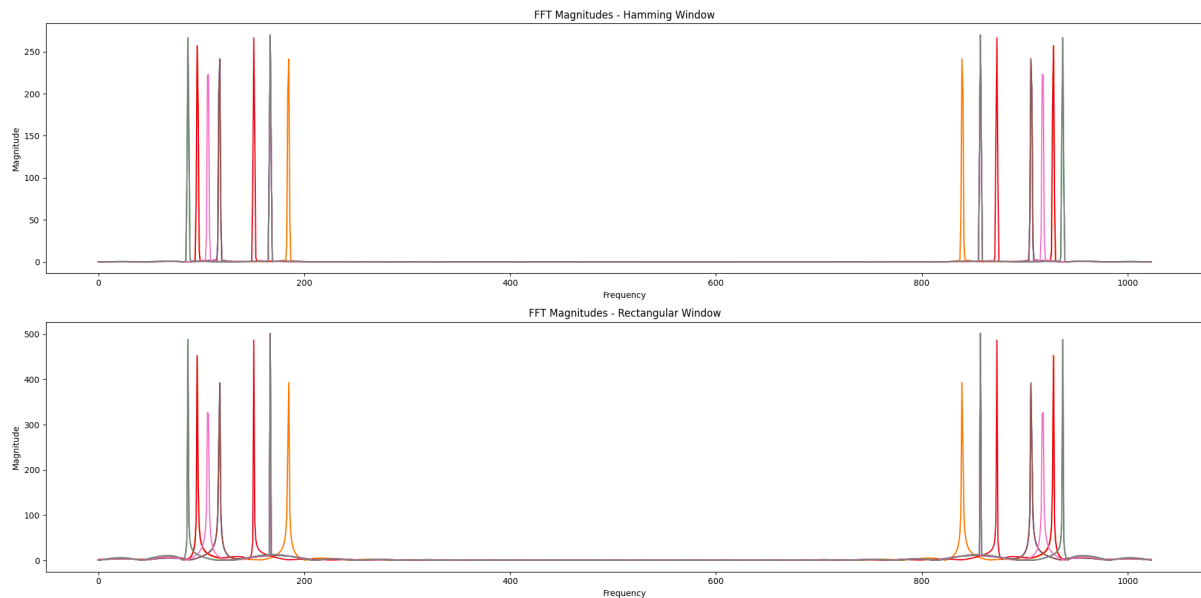


Η συμμετρία που βλέπουμε σε κάθε ένα από τα διαγράμματα οφείλεται στις ιδιότητες του Μετασχηματισμού Fourier που εφαρμόζεται σε πραγματικά σήματα. Όταν εφαρμόζουμε έναν Μετασχηματισμό Fourier σε ένα πραγματικό σήμα (όπως είναι τα σήματά μας), το αποτέλεσμα είναι ένα φάσμα που αποτελείται από φανταστικούς αριθμούς. Αυτό το φάσμα είναι Hermitian Symmetric, πράγμα που σημαίνει ότι οι αρνητικοί όροι συχνότητας είναι ο συζυγής των θετικών όρων συχνότητας. Σε όρους μέτρου, αυτό οδηγεί σε ένα συμμετρικό φάσμα. Γι' αυτό βλέπουμε συμμετρία στα διαγράμμάτα.

1.3

Το άθροισμα των αριθμών μητρώου μας οδηγεί στο 06240082. Δημιουργούμε λοιπόν μία λίστα, την **totalsignal**, στην οποία κάνουμε append ένα προς τα ένα τα σήματα των παραπάνω ψηφίων με βάση το ερώτημα 1.1. Χωρίζουμε κάθε ψηφίο από το επόμενο με 100 μηδενικά χρησιμοποιώντας την **np.zeros()**

1.4



Πολλαπλασιάζουμε με παράθυρα μήκους $L = 1000$ το **totalsignal** προκειμένου να διαχωρίσουμε τα διαδοχικά ψηφία του σήματος και έπειτα παίρνουμε το Μετασχηματισμό Fourier κάθε παραθυροποιημένου σήματος με $N = 1024$.

Σε κάθε ένα από τα παραπάνω διαγράμματα, φαίνεται με διαφορετικό χρώμα ο Fourier Μετασχηματισμός κάθε ψηφίου. Η διαφορά μεταξύ των δύο διαγραμμάτων οφείλεται στις διαφορετικές χαρακτηριστικές του παραθύρου Hamming και του Τετραγωνικού παραθύρου. Παράθυρο Hamming: Το παράθυρο Hamming είναι ένα ομαλό παράθυρο που μειώνει τα side lobes και το spectral leakage, που είναι η διάδοση της ενέργειας του σήματος σε συχνότητες διαφορετικές από αυτήν στην οποία είναι κεντραρισμένο. Αυτό έχει ως αποτέλεσμα ένα πιο ομαλό διάγραμμα μετατροπής Fourier (FFT) στην ένταση, όπως φαίνεται στο πρώτο διάγραμμα. Το αντάλλαγμα είναι ένας ελαφρώς ευρύτερος κύριος λοβός, που σημαίνει μια ελαφρά μειωμένη ανάλυση συχνότητας.

Τετραγωνικό Παράθυρο: Το Τετραγωνικό Παράθυρο, από την άλλη, δεν κάνει πολλά για τη μείωση της διαρροής φάσματος, με αποτέλεσμα πιο έντονους πλευρικούς λοβούς στο διάγραμμα Μετασχηματισμού Fourier στην ένταση. Ωστόσο, έχει έναν πιο στενό κύριο λοβό, παρέχοντας καλύτερη ανάλυση συχνότητας.

1.7

Η είσοδος των τριών ζητούμενων σημάτων μας στην συνάρτηση `ttdecode()` μας δίνει τα εξής output:

[9, 3, 9, 2, 6, 8, 4, 8]

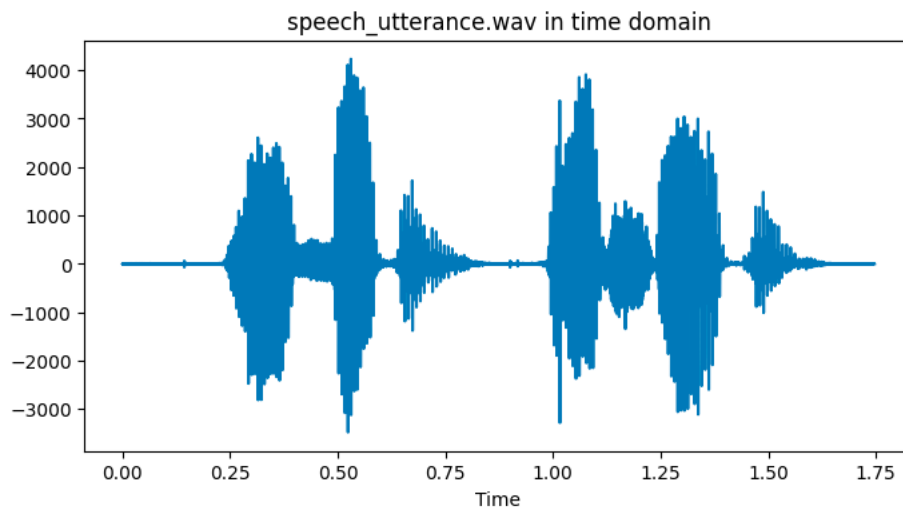
[2, 7, 3, 3, 2, 5, 4, 2, 5, 4]

[6, 9, 2, 6, 3, 1, 9, 9, 1, 0]

Μέρος 2: Χαρακτηριστικά Βραχέος Χρόνου Σημάτων Φωνής και Μουσικής

2.1

Διαβάζουμε το σήμα “speech_utterance.wav” και το πλοττάρουμε σε συνάρτηση με τον χρόνο:



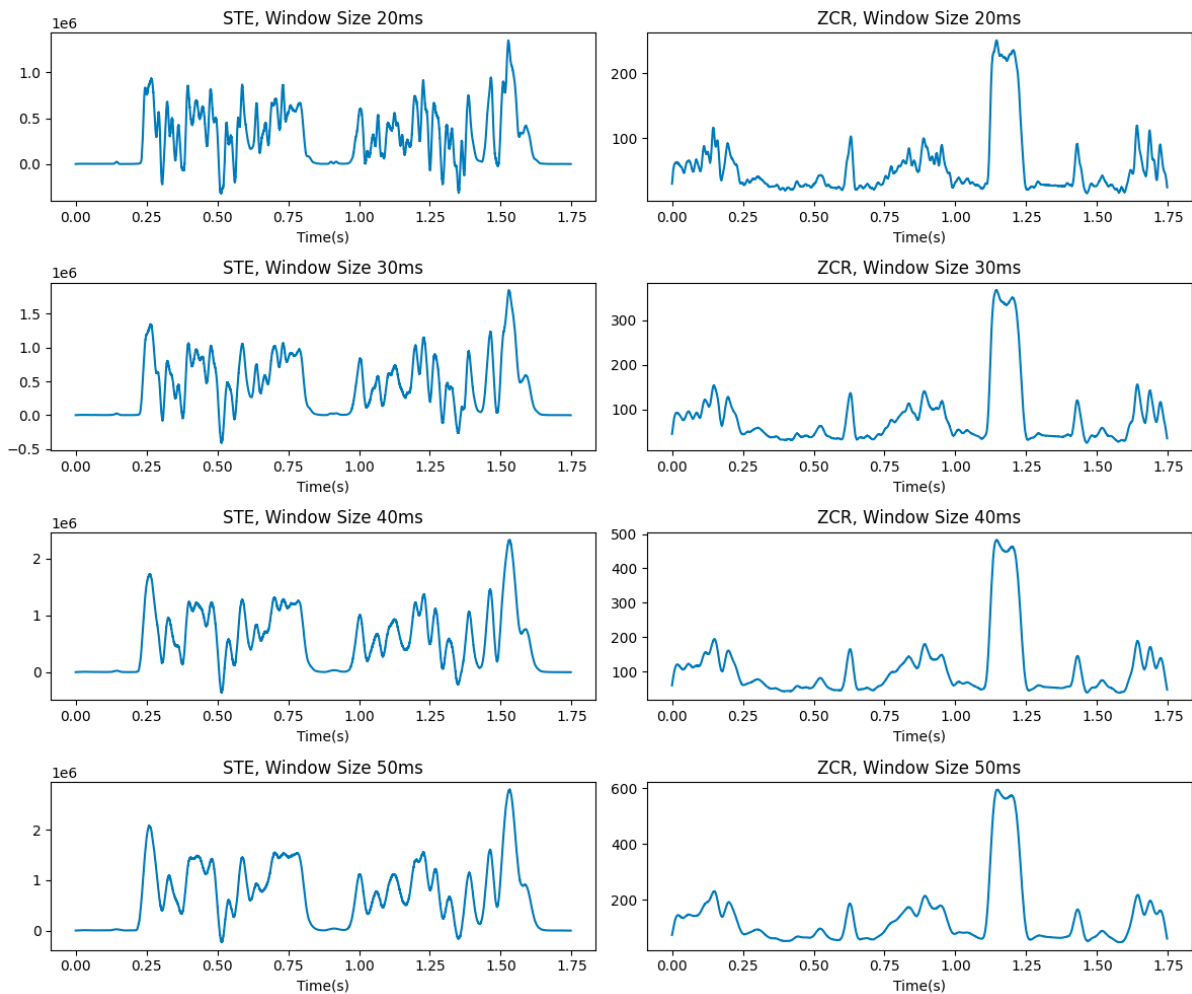
2.2

Εφαρμόζουμε τους παρακάτω τύπους για τον υπολογισμό των Short Time Energy και Zero Crossing Rate αντίστοιχα:

$$E_n = \sum_{m=-\infty}^{\infty} |x[m]|^2 w[n-m]$$

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

Και παίρνουμε τα εξής διαγράμματα για διαφορετικά μήκη παραθύρων:



Παρατηρούμε πως για μεγαλύτερα παράθυρα, οι υπολογισμοί των STE και ZCR υποκύπτουν λιγότερο σε ραγδαίες μεταβολές που μπορεί να οφείλονται σε θόρυβο, οδηγώντας σε πιο ομαλά διαγράμματα. Αυτό όμως έχει ως αντάλλαγμα, να χάνεται περισσότερη λεπτομέρεια και να μειώνεται η ευκρίνεια.

Η STE και ZCR μπορούν γενικά να χρησιμοποιηθούν για να διακρίνουν μεταξύ σιωπής, έμφωνων και άφωνων ήχων σε σήματα που περιέχουν ομιλία.

Σιωπή: Οι περίοδοι σιωπής ή συνήθως θα έχουν πολύ χαμηλή STE επειδή υπάρχει ελάχιστη ή καθόλου ενέργεια σε αυτά τα τμήματα. Ο ZCR θα μπορούσε να ποικίλει ανάλογα με το θόρυβο περιβάλλοντος.

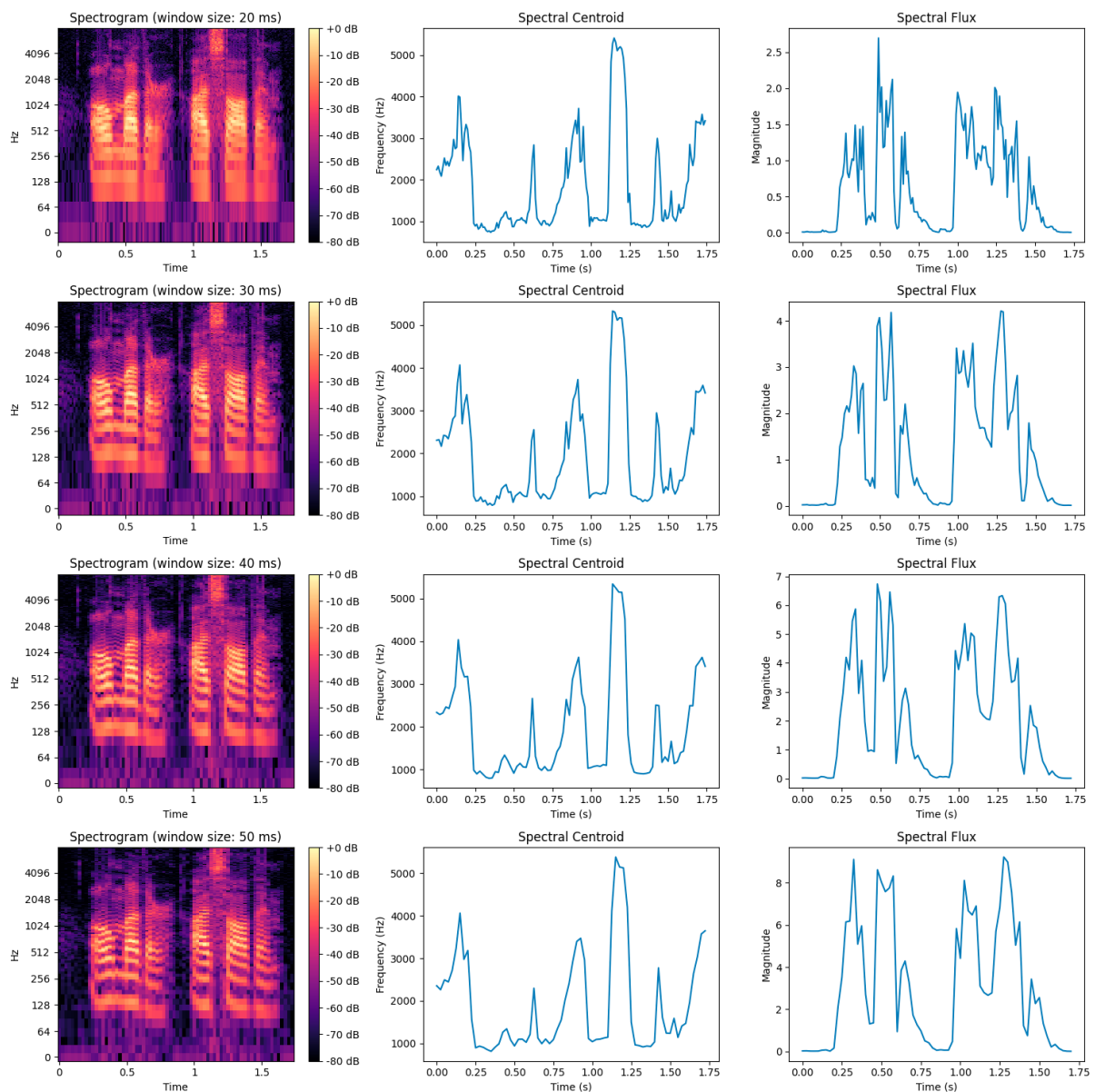
Έμφωνοι Ήχοι: Οι έμφωνοι ήχοι, όπως τα φωνήεντα ή τα φωνητικά σύμφωνα, παράγονται από την δόνηση των φωνητικών χορδών, ο οποίος οδηγεί σε σχετικά υψηλή ενέργεια. Επομένως, συνήθως θα έχουν υψηλότερη STE. Επίσης, έχουν χαμηλότερο ZCR λόγω της περιοδικής τους φύσης.

Άφωνοι Ήχοι: Οι άφωνοι ήχοι, όπως ορισμένα σύμφωνα, παράγονται χωρίς τη δόνηση των φωνητικών χορδών. Έχουν μικρότερη ενέργεια (χαμηλότερη από τους φωνητικούς ήχους αλλά υψηλότερη από τη σιωπή) και υψηλότερο ZCR λόγω της πιο τυχαίας, θορυβώδους φύσης αυτών των ήχων, που οδηγεί σε πιο συχνά zero crossings.

Παρ'όλα αυτά, ενώ η STE και ο ZCR μπορούν να παρέχουν κάποιο επίπεδο διάκρισης, δεν είναι επαρκείς για μια αξιόπιστη κατηγοριοποίηση των ήχων της ομιλίας. Μπορούμε στο περίπου και διαισθητικά, κοιτώντας τα STE διαγράμματα να κάνουμε εκτίμηση για το που αντιστοιχεί κάθε σημείο της ομιλίας του “speech_utterance.wav”, αλλά δεν είναι κάτι τόσο αξιόπιστο.

2.3

Χρησιμοποιώντας την βιβλιοθήκη **librosa** παίρνουμε τα παρακάτω διαγράμματα



Αυτά τα διαγράμματα παρέχουν πληθώρα πληροφοριών σχετικά με το αρχικό σήμα.

Spectrogram: Τα spectrograms (πρώτη στήλη) παρέχουν μια οπτική αναπαράσταση του φάσματος συχνοτήτων του σήματος καθώς αυτές ποικίλουν με τον χρόνο. Μας βοηθούν να κατανοήσουμε το περιεχόμενο συχνοτήτων του σήματος σε κάθε στιγμή. Για ένα σήμα ομιλίας σαν το δικό μας, βοηθάει στον εντοπισμό διαφορετικών φωνημάτων (διακριτές μονάδες ήχου) καθώς έχουν διαφορετικά πρότυπα συχνοτήτων.

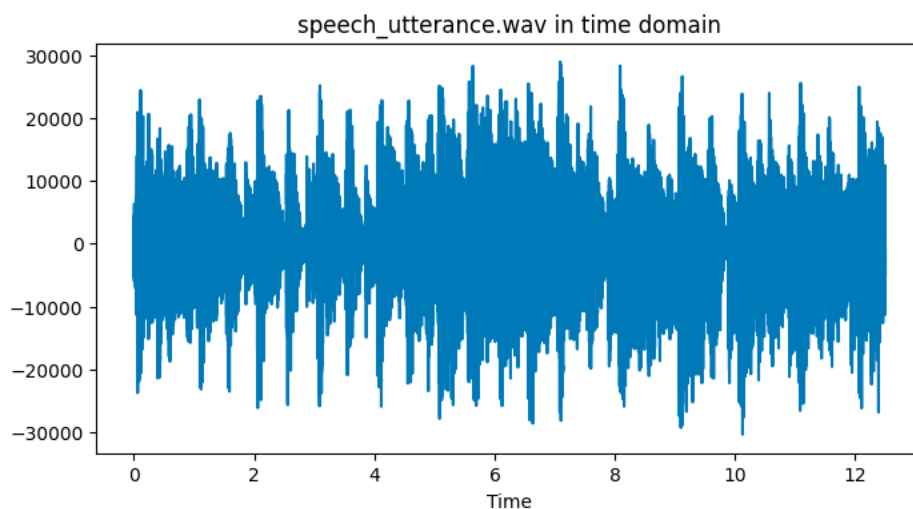
Φασματικό Κέντρο: Το φασματικό κέντρο (δεύτερη στήλη) είναι ένα μέτρο που υποδεικνύει πού βρίσκεται το "κέντρο βάρους" του φάσματος. Συνήθως συσχετίζεται με την αντιληπτή "φωτεινότητα" ενός ήχου. Στο πλαίσιο του σήματος μας μπορεί να διακρίνει μεταξύ διαφορετικών τύπων ομιλίας. Τα φωνήεντα φαίνεται να έχουν υψηλότερο φασματικό κέντρο από τα σύμφωνα.

Φασματική Ροή: Η φασματική ροή (τρίτη στήλη) μετρά πόσο γρήγορα αλλάζει το φάσμα ισχύος ενός σήματος, υπολογίζοντας το ρυθμό μεταβολής του πλάτους του φάσματος. Στο σήμα μας, υψηλές τιμές φασματικής ροής μπορεί να υποδεικνύουν την παρουσία μεταβάσεων ή την ταχεία αλλαγή στο σήμα, οι οποίες συνήθως αντιστοιχούν στην έναρξη ενός νέου φωνήματος.

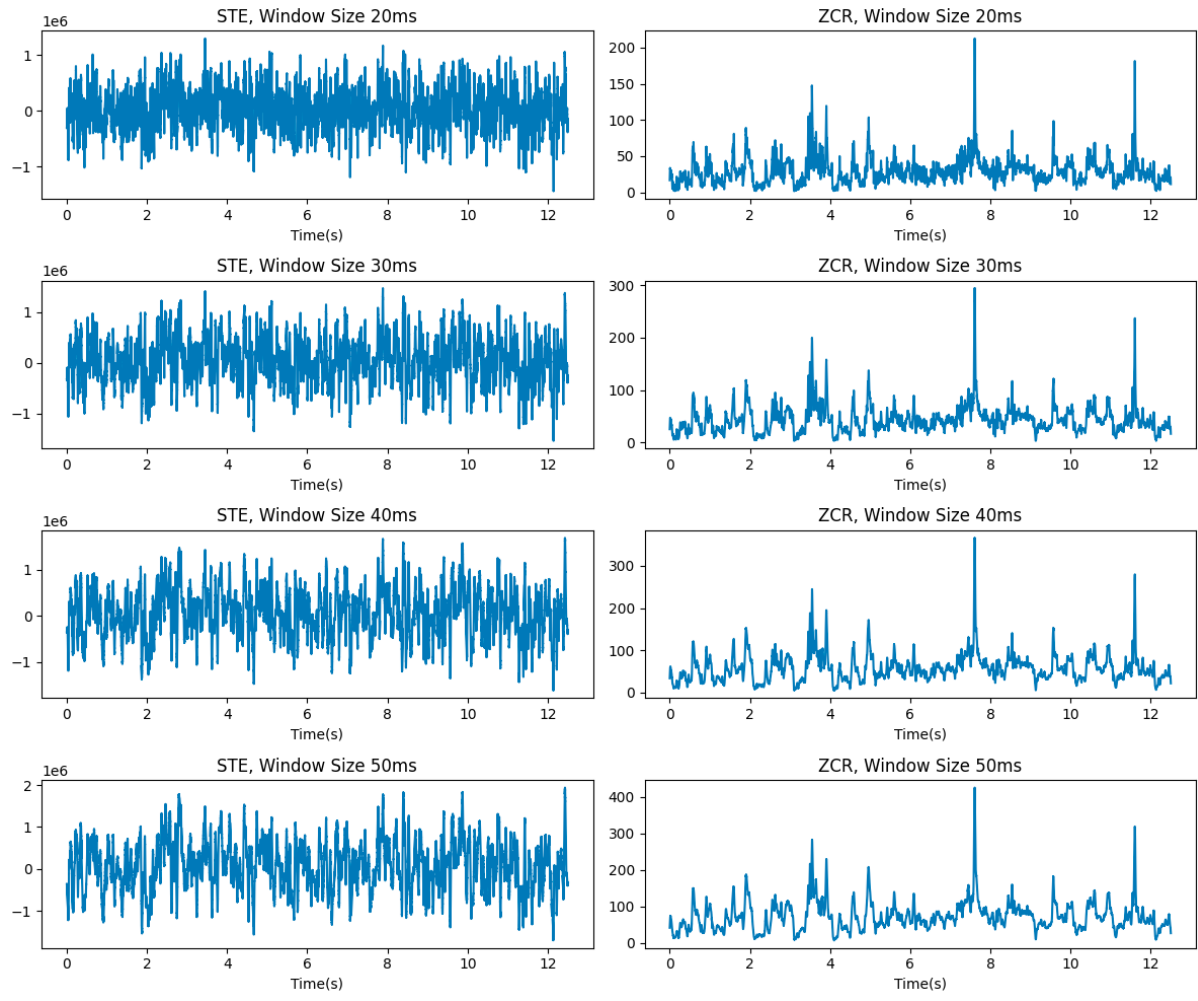
Αναλύοντας αυτά τα διαγράμματα, μπορούμε να αποκτήσουμε μια πιο βαθιά κατανόηση των χαρακτηριστικών του ομιλητικού σήματος, όπως η κατανομή της ενέργειας σε διαφορετικές συχνότητες, η "φωτεινότητα" ή "βαρύτητα" της ομιλίας, και πώς αυτές οι ιδιότητες αλλάζουν με τον χρόνο. Αυτές οι πληροφορίες μπορούν να είναι πολύ χρήσιμες σε διάφορες εφαρμογές, όπως η αναγνώριση ομιλίας, η αναγνώριση ομιλητή και η επεξεργασία ήχου.

2.4.1

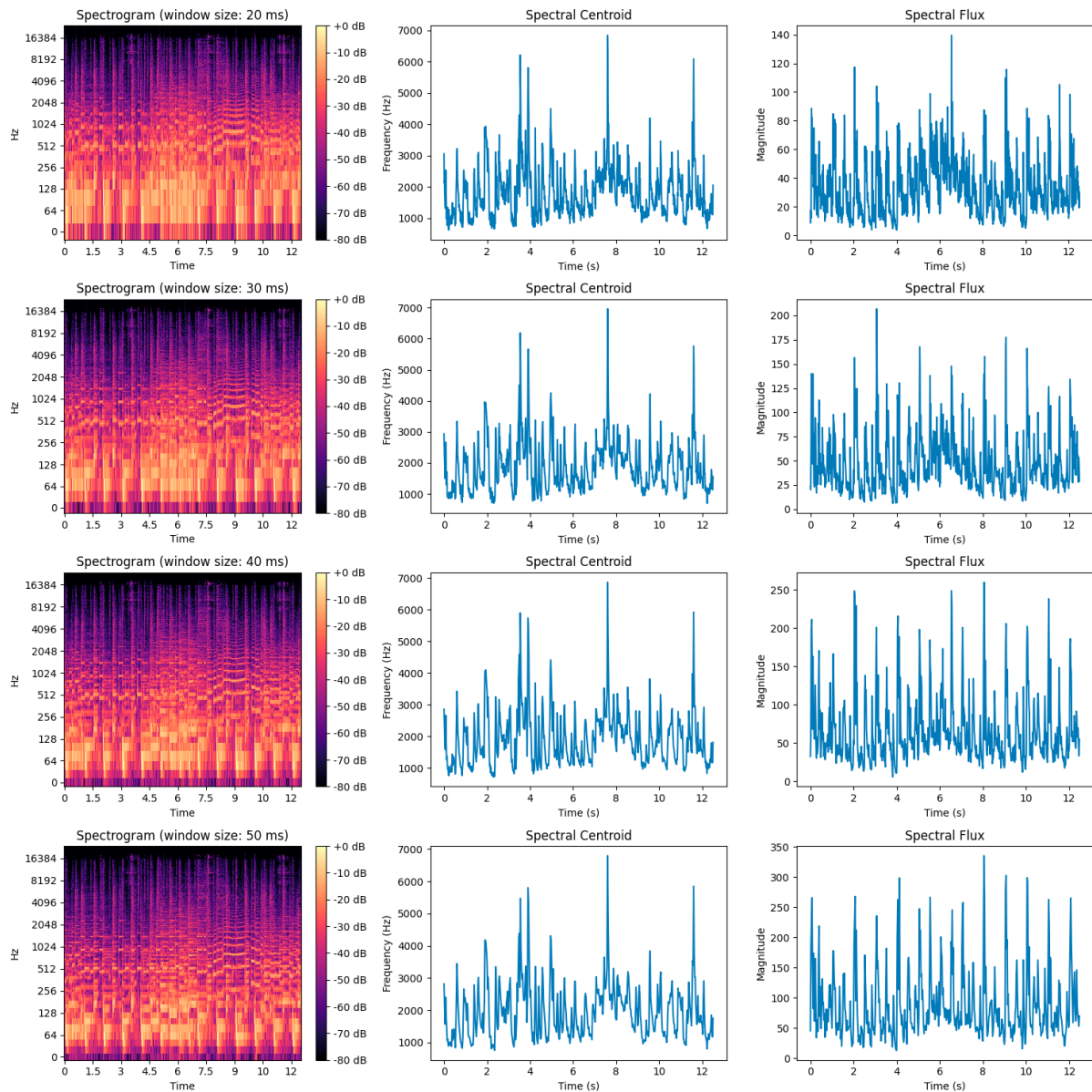
Επαναλαμβάνουμε τη διαδικασία για το αρχείο "music.wav":



2.4.2



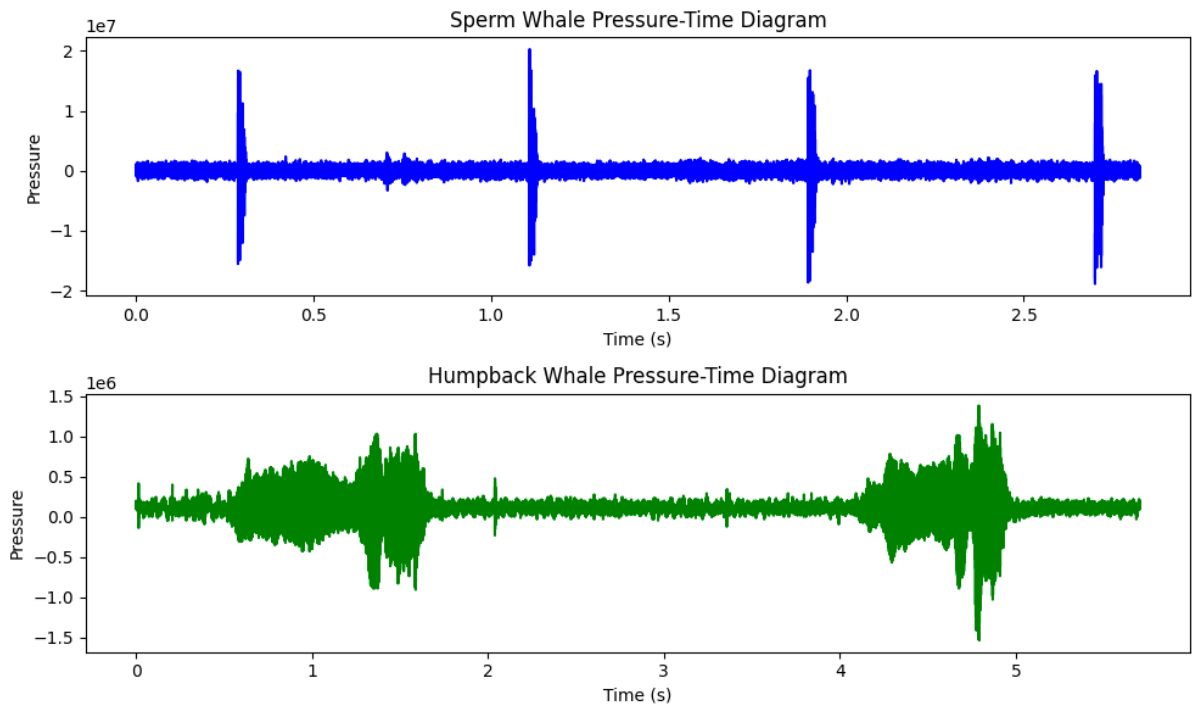
2.4.3



Μέρος 3: Ψηφιακή Επεξεργασία και Φασματική Ανάλυση Βιοσημάτων Κητωδών

3.1

Τα διαγράμματα πίεσης χρόνου για τα θηλαστικά:



3.2

Υπολογίζουμε για τα σήματα τα εξής:

RMS Pressure for Sperm Whale: 784971.1

RMS Pressure for Humpback Whale: 195801.56

SPL for Sperm Whale: 237.8970736321329 dB

SPL for Humpback Whale: 225.83632306317492 dB

3.3

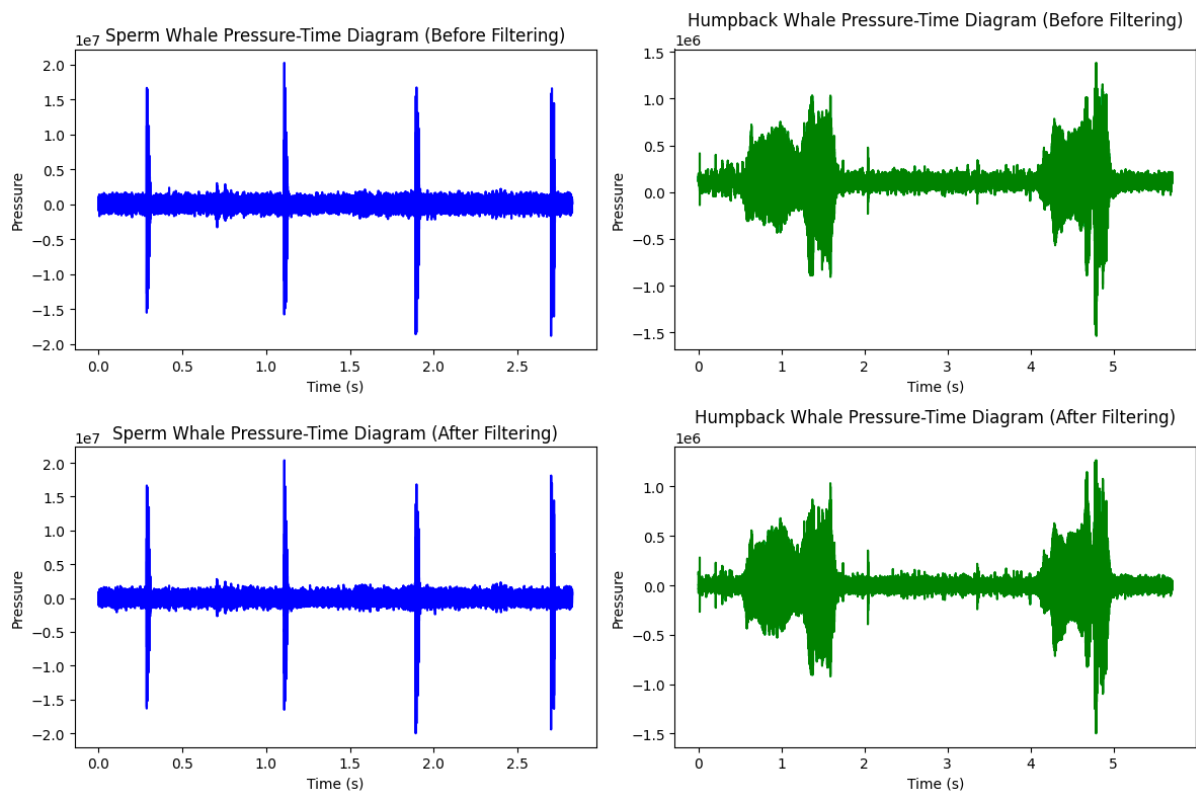
Επιβεβαιώνουμε το θεώρημα Parseval:

Sum of squared magnitudes in time domain: 1.4267764×10^{17}

Sum of squared magnitudes in frequency domain: $1.42677626774277 \times 10^{17}$

3.4

Εφαρμόζουμε το φίλτρο butterworth.



Παρατηρούμε πως οι κυματομορφές δεν έχουν αλλάξει ιδιαίτερα ως προς τη μορφή τους. Τα σημαντικά peaks παραμένουν εκεί που ήταν. Παρόλα αυτά υπάρχει ένα αισθητό smoothing που μπορεί να οφείλεται στην απομόνωση του ανθρωπογενή θορύβου.

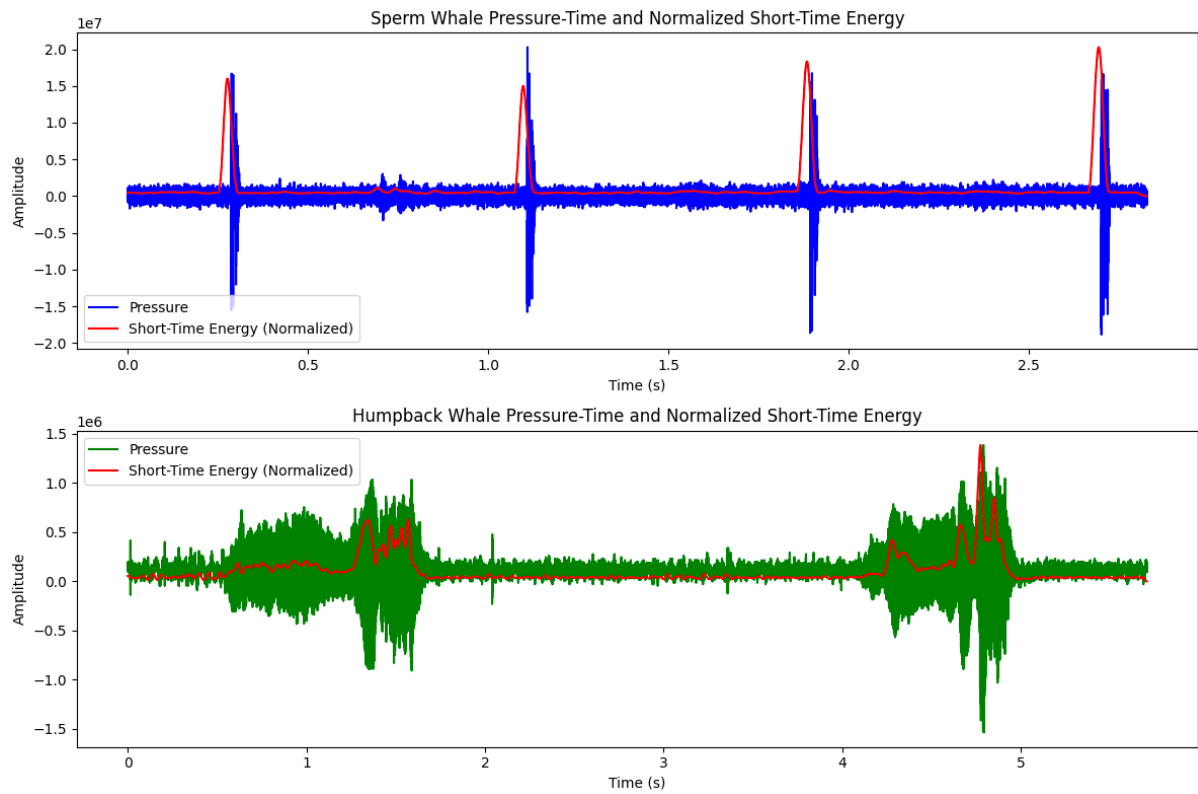
3.5

Output:

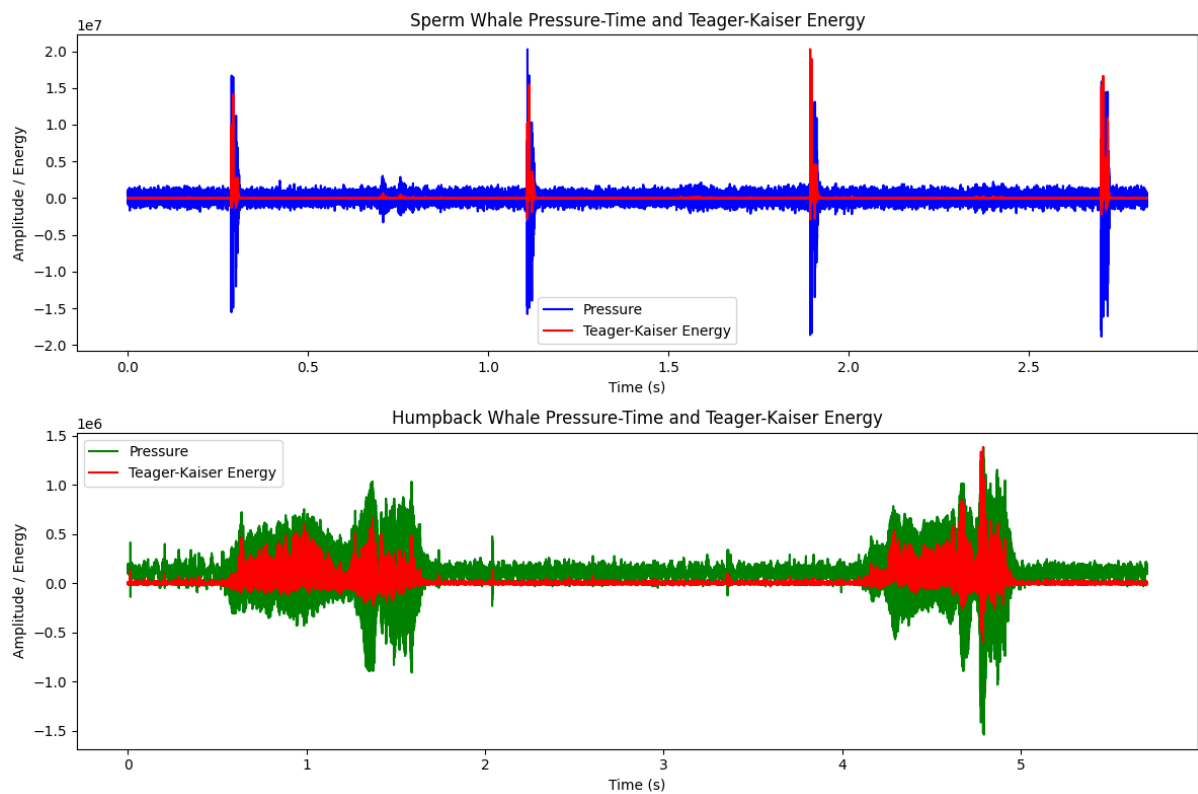
The SPLrms of the processed audio file is equal to 285.15361603044954 dB and exceeds the threshold (100 dB).

This could potentially affect the behavior of cetaceans

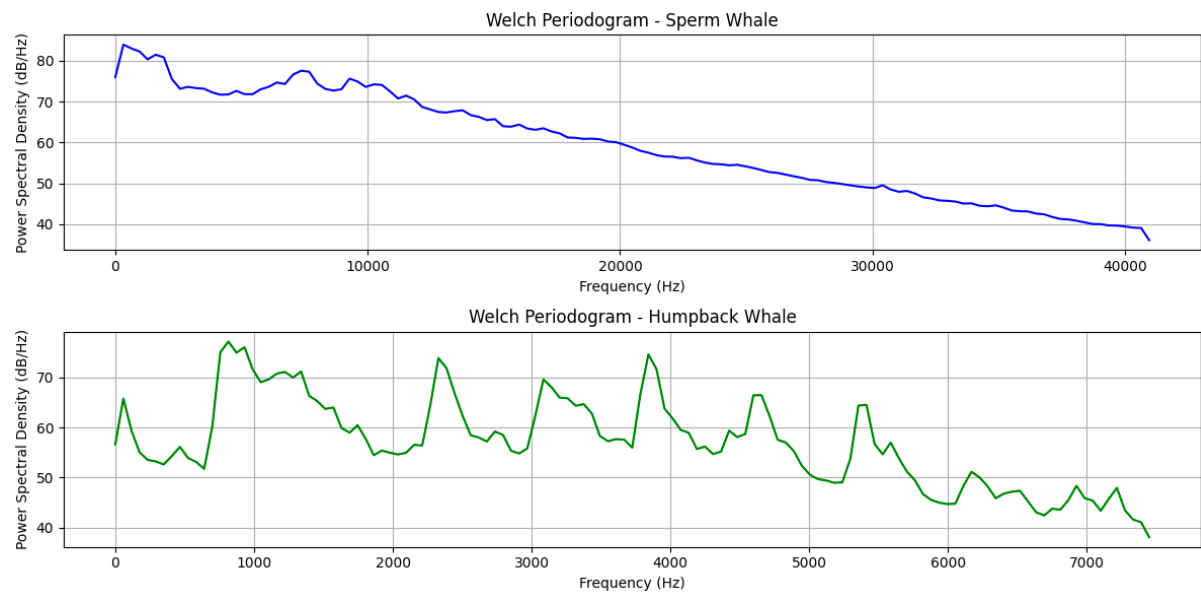
3.6



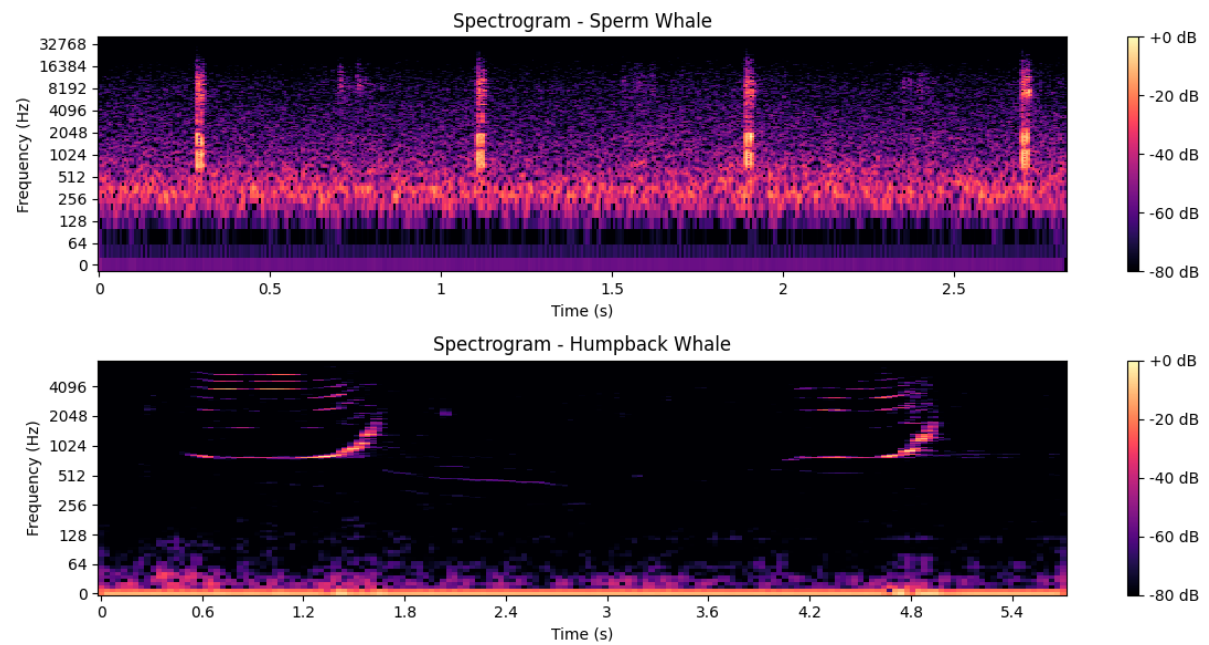
3.7



3.8



3.9



3.10

