



Ψηφιακή Επεξεργασία Σημάτων

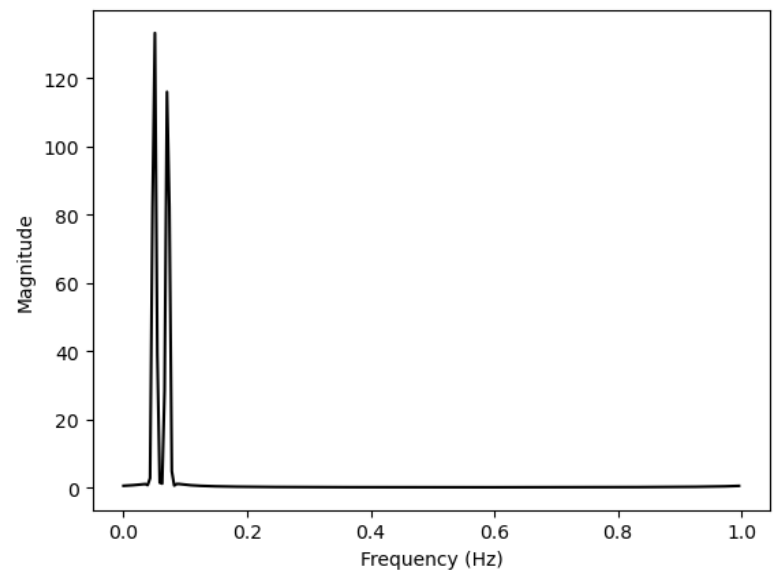
1η Εργαστηριακή Άσκηση

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

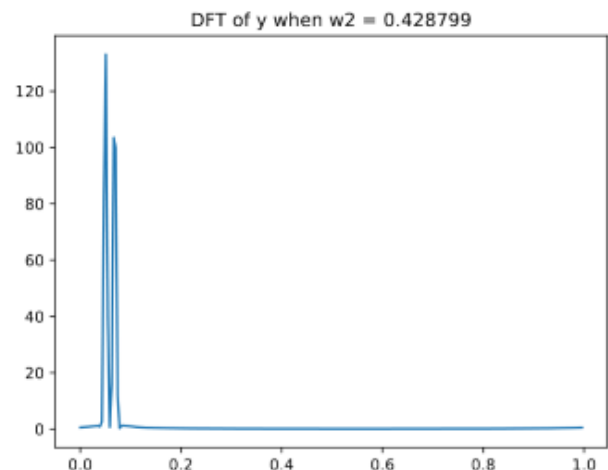
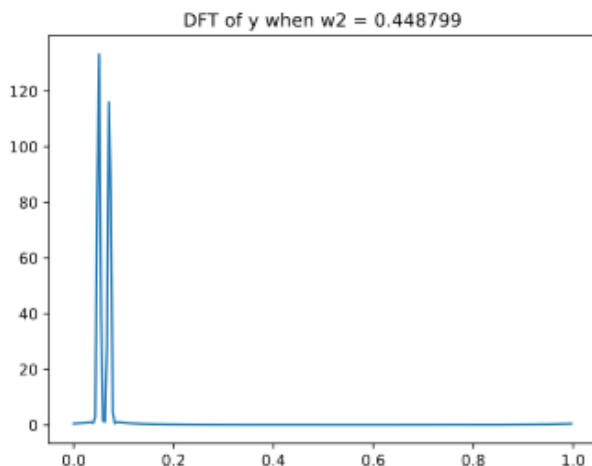
Μέρος 1ο: Φασματική Ανάλυση και Ανίχνευση Ημιτονοειδών με τον Διακριτό Μετ/σμό Fourier (DFT)

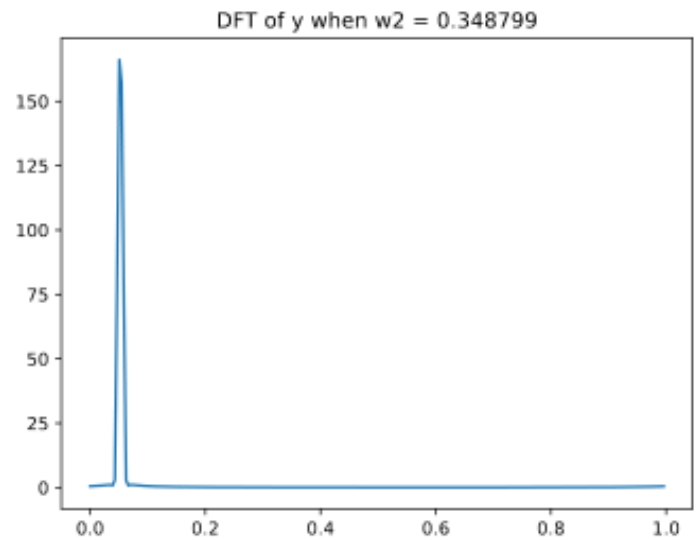
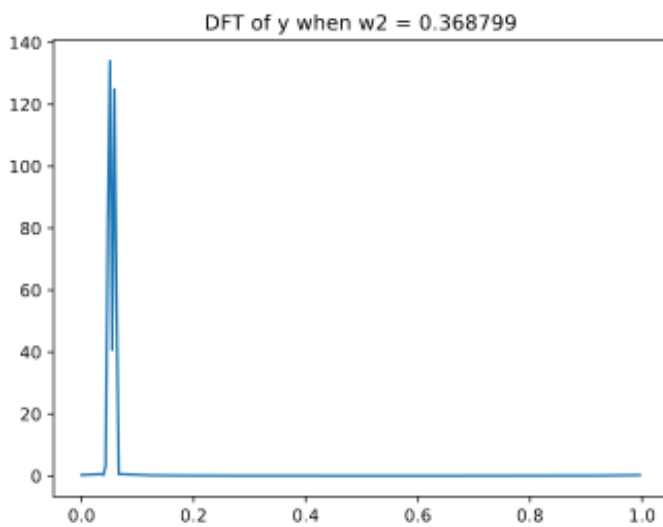
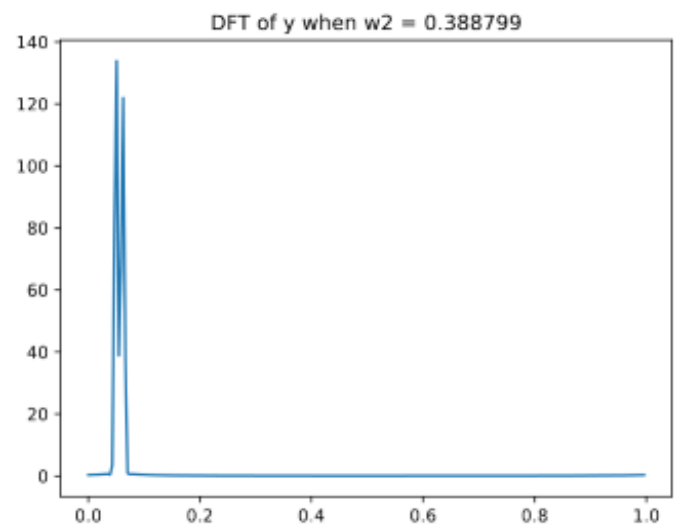
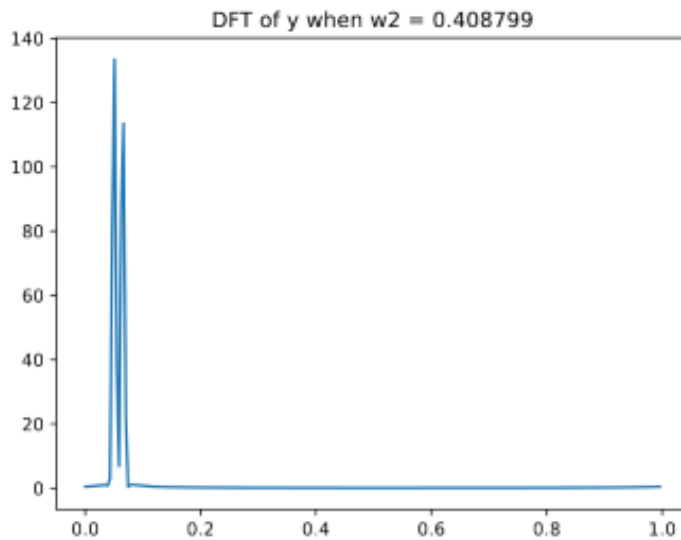
1.1.

Αρχικά υπολογίζουμε τον DFT του σήματος $y[n]$ για τις αρχικές τιμές των ω_1 και ω_2 :



Μειώνουμε σταδιακά την τιμή του ω_2 , πλησιάζοντας έτσι την τιμή του ω_1 και σχεδιάζουμε σε κάθε βήμα τον DFT



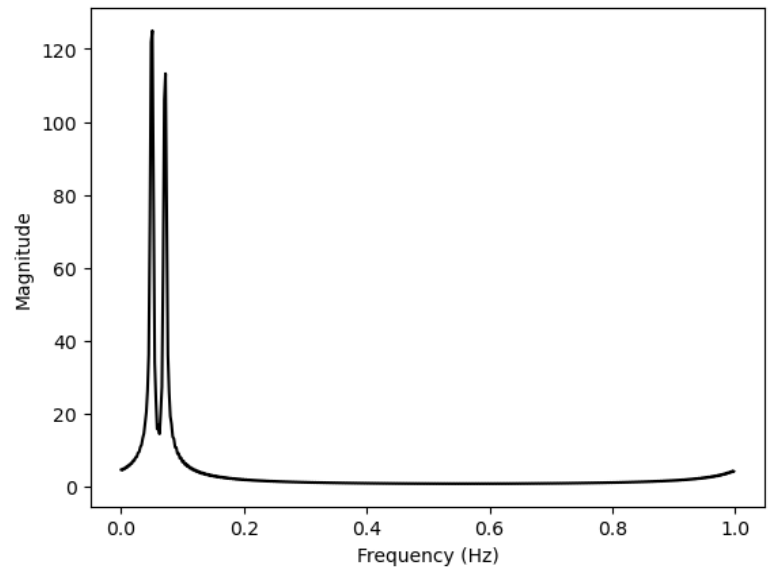


Όπως ήταν αναμενόμενο, πλησιάζοντας την τιμή του ω_1 οι δύο κορυφές έρχονται όλο και πιο κοντά. Παρατηρούμε ότι η διαφορά των δύο γωνιακών συχνοτήτων μπορεί να γίνει μέχρι και $\Delta\omega=0.0746$ (όταν $\omega_2 = 0.3888$) χωρίς να χάνεται η δυνατότητα διάκρισης μεταξύ των δύο κορυφών.

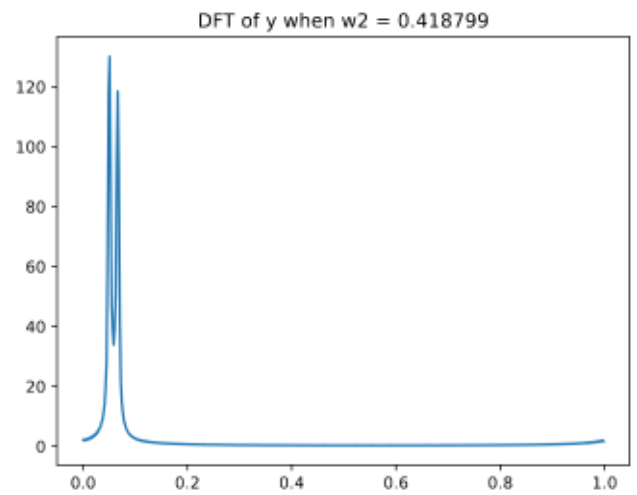
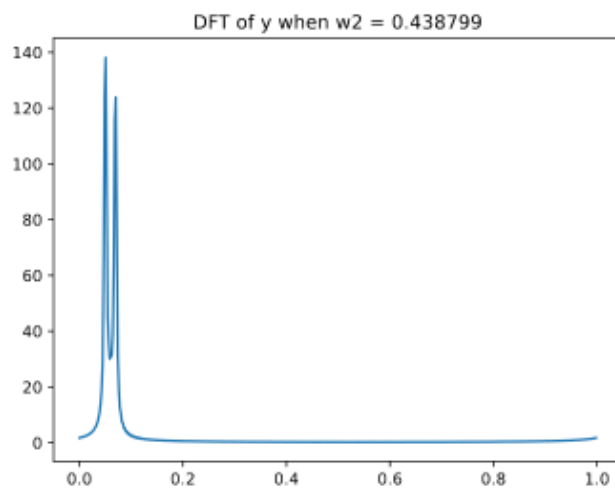


1.2.

Παρακάτω είναι η αναπαράσταση του DFT με $N = 512$ και $L = 256$:

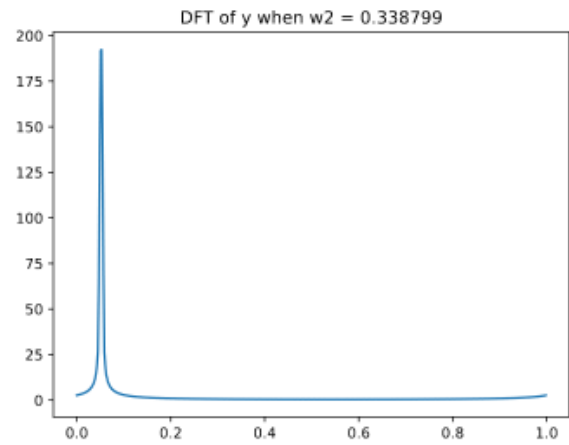
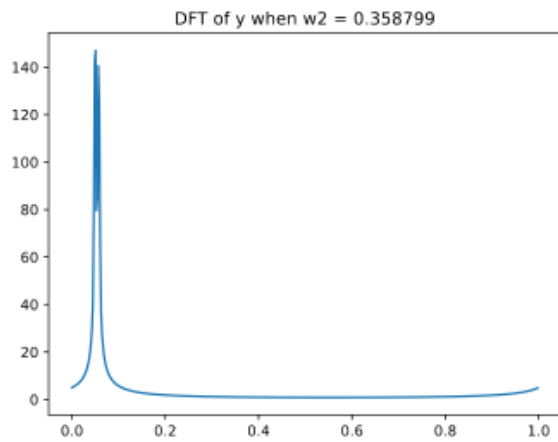
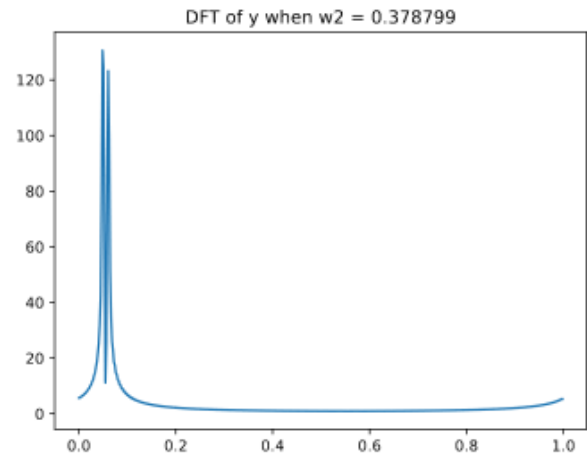
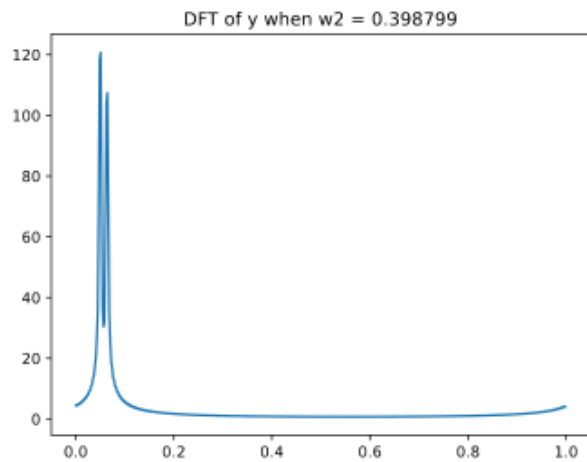


Ακολουθούμε την ίδια διαδικασία με το ερώτημα 1.1, δηλαδή μειώνουμε σταδιακά το ω_2





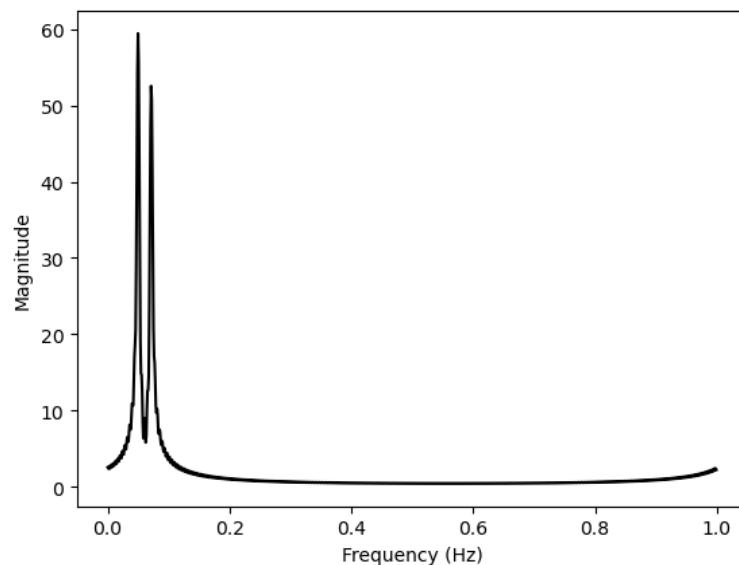
Χριστόδουλος Στυλιανίδης 03120614
Γιάννης Γεωργιάδης 03120615



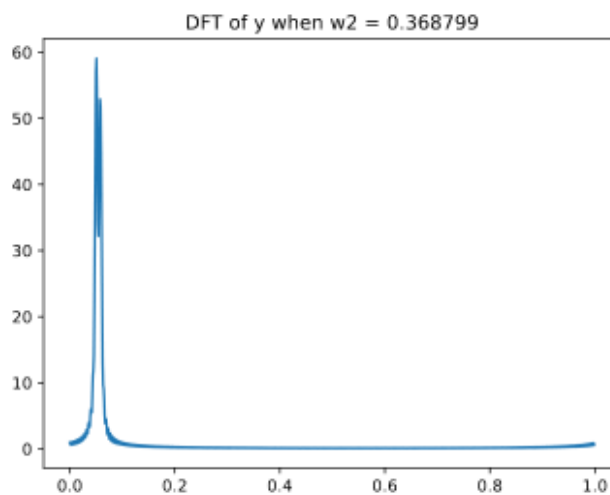
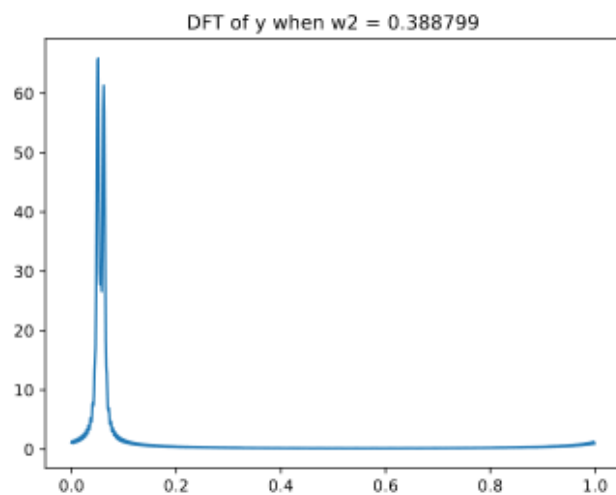
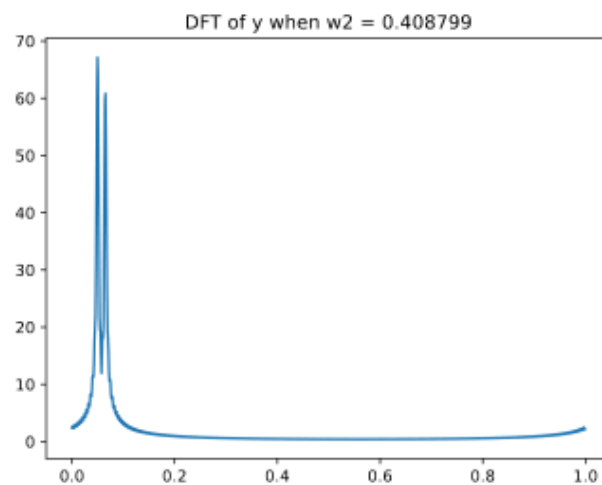
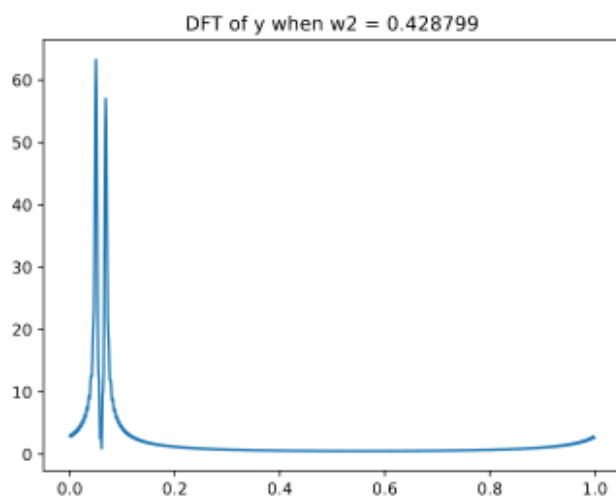
Σε αυτή την περίπτωση παρατηρούμε ότι η διαφορά των δύο γωνιακών συχνοτήτων μπορεί να γίνει μέχρι και $\Delta\omega=0.0646$ (όταν $\omega_2 = 0.3788$) χωρίς να χάνεται η δυνατότητα διάκρισης μεταξύ των δύο κορυφών.

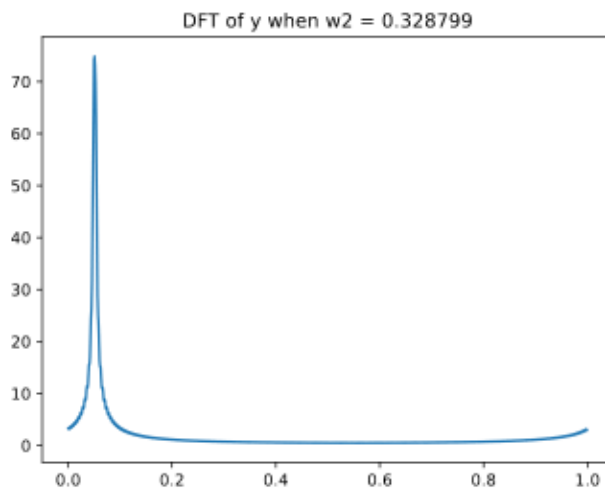
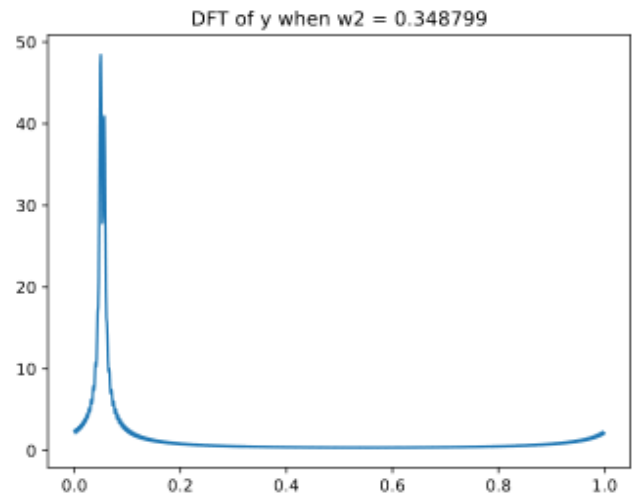
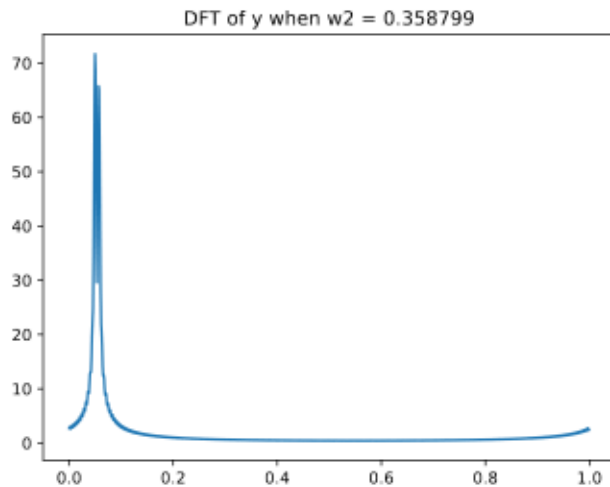


Παρακάτω ακολουθούμε την ίδια
διαδικασία για τον DFT με
 $N = 1024$ και $L = 256$:



Μειώνουμε σταδιακά το ω_2 :





Σε αυτή την περίπτωση παρατηρούμε ότι η διαφορά των δύο γωνιακών συχνοτήτων μπορεί να γίνει μέχρι και $\Delta\omega=0.0446$ (όταν $\omega_2 = 0.3588$) χωρίς να χάνεται η δυνατότητα διάκρισης μεταξύ των δύο κορυφών.

Η δυνατότητα φασματικής διάκρισης αυξάνεται όσο μεγαλώνει το N (δηλαδή το μήκος του DFT). Επομένως, στον DFT 1024 σημείων οι 2 κορυφές μπορούν να διακριθούν για μικρότερα $\Delta\omega$ απ' ότι σε DFTs λιγότερων σημείων. Προφανώς το ίδιο ισχύει αν συγκρίνουμε τον DFT-512 σημείων και τον DFT-256 σημείων: ο πρώτος παρουσιάζει μεγαλύτερη δυνατότητα φασματικής διάκρισης από τον δεύτερο.

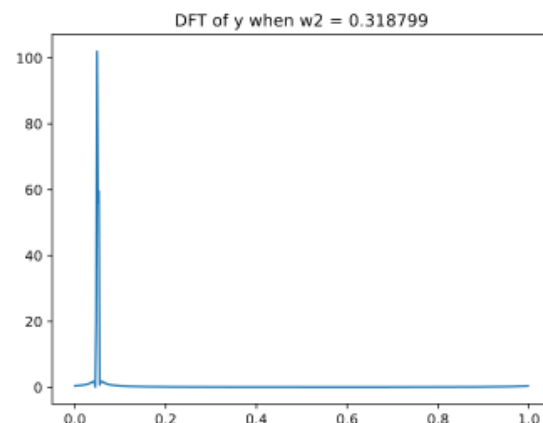
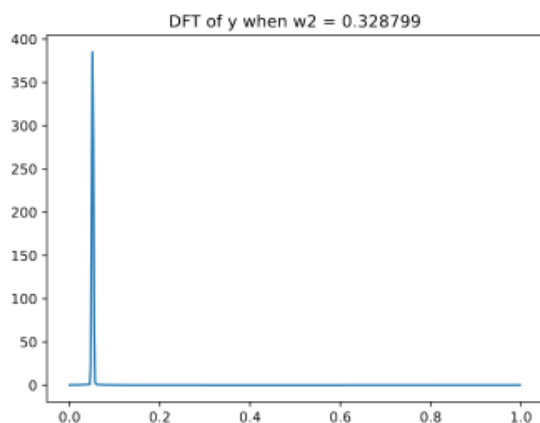
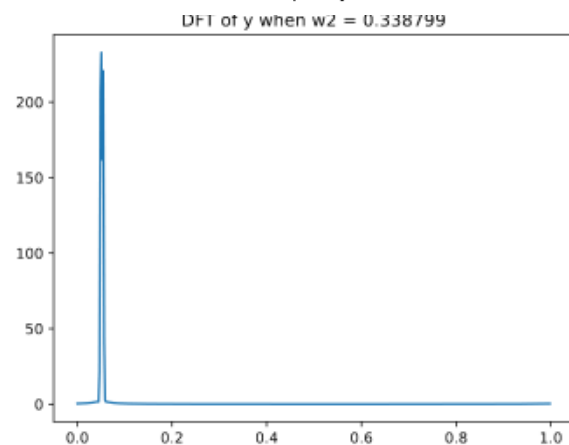
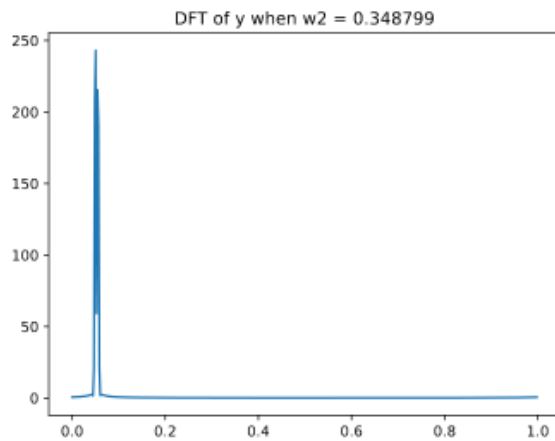
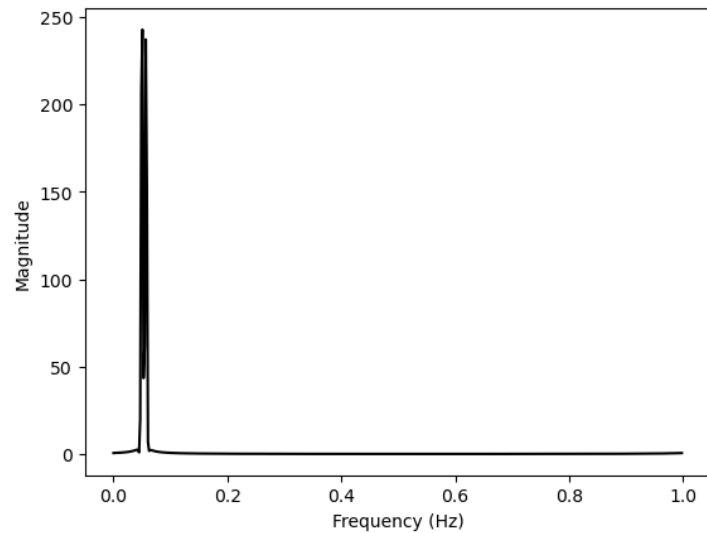
Βλέπουμε λοιπόν, αυτό που είναι γνωστό από την θεωρία, ότι δηλαδή, αν αυξηθεί το μέγεθος του DFT με πρόσθεση μηδενικών (zero-padding) αλλάζει το διάστημα των δειγμάτων συχνότητας επειδή το σήμα μας βασίζεται σε σήμα μεγαλύτερου μήκους. Με άλλα λόγια επιτυγχάνεται μια πιο λεπτομερής ανάλυση της συχνότητας του σήματος και καλύτερη ανάλυση μεταξύ κοντινών συχνοτήτων.



1.3.

Παρακάτω ακολουθούμε την ίδια διαδικασία για τον DFT με $N = 512$ και $L = 512$:

Μειώνοντας σταδιακά το ω_2 :

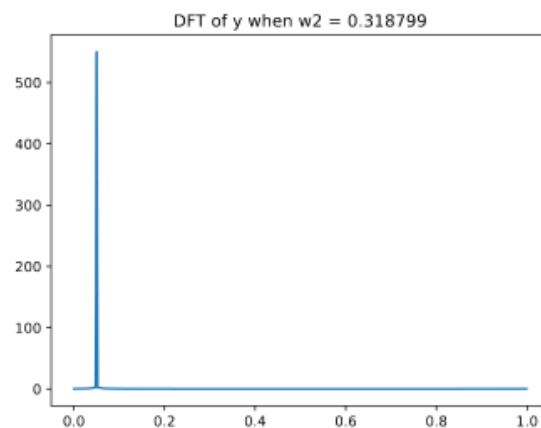
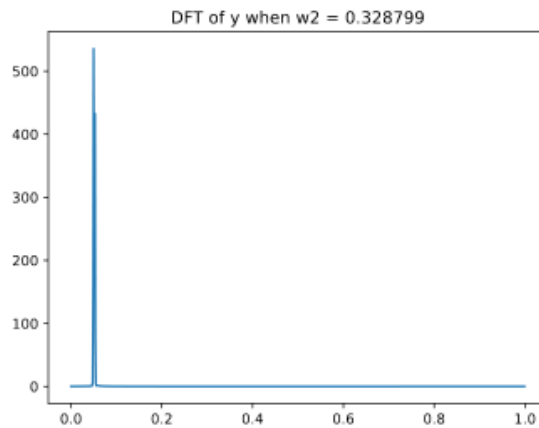
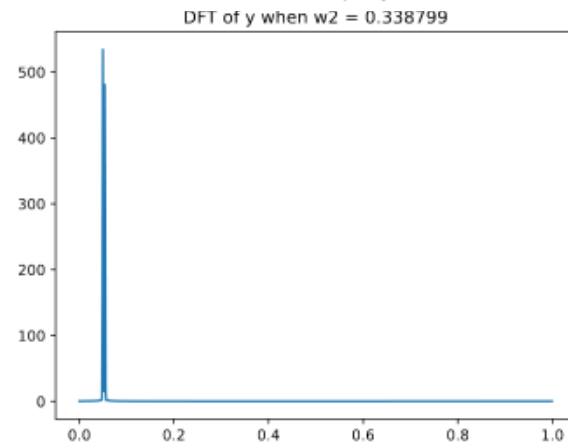
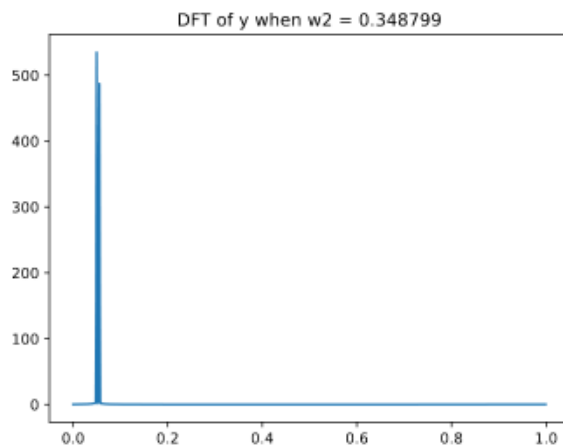
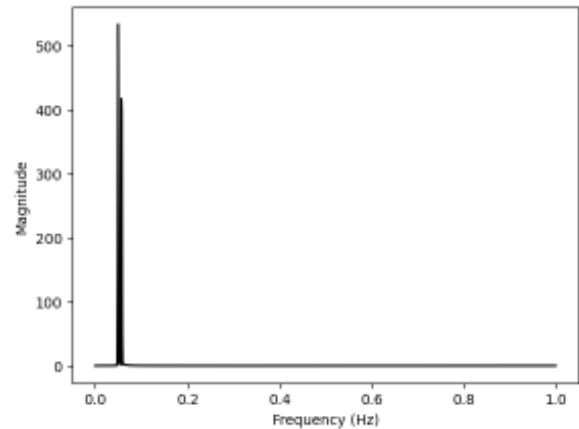


Σε αυτή την περίπτωση παρατηρούμε ότι η διαφορά των δύο γωνιακών συχνοτήτων μπορεί να γίνει μέχρι και $\Delta\omega = 0.0446$ (όταν $\omega_2 = 0.3588$) χωρίς να χάνεται η δυνατότητα διάκρισης μεταξύ των δύο κορυφών.



Παρακάτω ακολουθούμε την ίδια διαδικασία για τον DFT με $N = 1024$ και $L = 1024$:

Μειώνοντας σταδιακά το ω_2 :



Σε αυτή την περίπτωση παρατηρούμε ότι η διαφορά των δύο γωνιακών συχνοτήτων μπορεί να γίνει μέχρι και $\Delta\omega = 0.0246$ (όταν $\omega_2 = 0.3388$) χωρίς να χάνεται η δυνατότητα διάκρισης μεταξύ των δύο κορυφών.

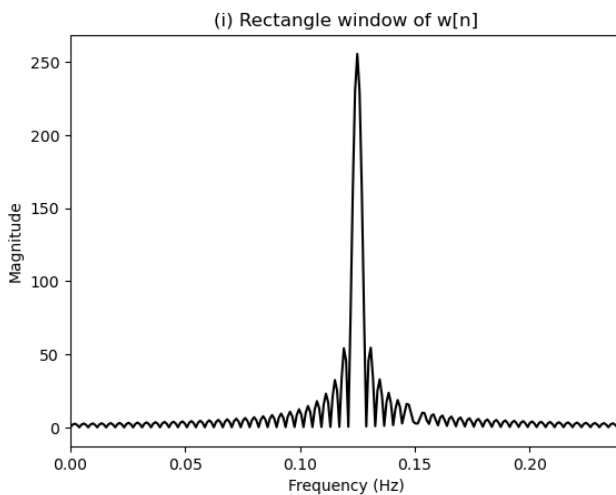
Είναι εμφανές ότι ο DFT-1024 σημείων παρέχει μια λεπτομερέστερη φασματική ανάλυση του σήματος σε σύγκριση με τον DFT-512 σημείων. Πρακτικά αυτό σημαίνει ότι οι δύο κορυφές μπορούν να πλησιάσουν περισσότερο χωρίς να συγχωνευθούν.



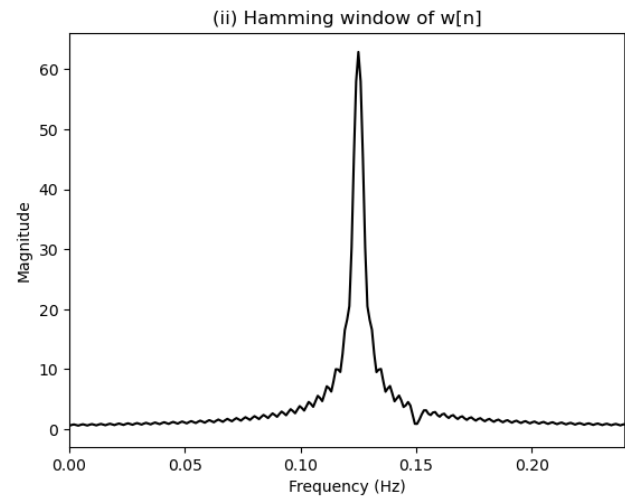
1.4.

Πάρακάτω, DFT μήκους $N = 1024$ και $L = 256$:

(i) για τετραγωνικό παράθυρο



(ii) παράθυρο Hamming



Διακρίνεται πολύ πιο έντονα η κορυφή που αντιστοιχεί στην συχνότητα ω_1 διότι έχει σημαντικά μεγαλύτερο πλάτος από την άλλη ημιτονοειδή συνιστώσα.

Όσον αφορά τα παράθυρα μπορούμε να κάνουμε τις εξής διαπιστώσεις:

Το τετραγωνικό παράθυρο έχει αισθητά στενότερο κύριο λοβό συγκριτικά με το παράθυρο hamming, το οποίο επηρεάζει θετικά την ανάλυση του φάσματος. Όμως το παράθυρο hamming προτιμάται διότι όπως είναι φανερό από τους παραπάνω DFTs έχει σημαντικά μικρότερο πλάτος πλευρικών λοβών (συγκριτικά με τον κύριο) γεγονός που μειώνει την διαρροή (leakage)



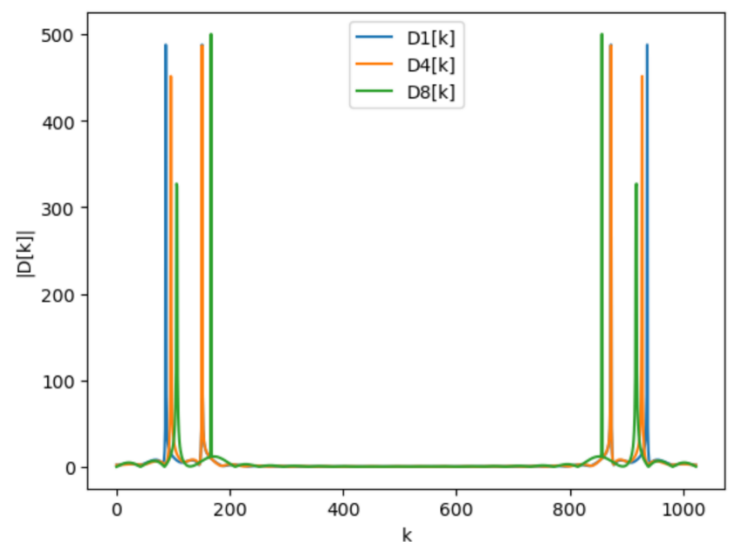
Μέρος 2ο - Σύστημα Εντοπισμού Τηλεφωνικών Τόνων (Telephone Touch Tones)

2.2.

Δεξιά φαίνονται οι DFTs των τόνων για τα πλήκτρα 1,4 και 8.

Αν περιοριστούμε στο διάστημα 0-512 (δηλαδή στο εύρος συχνοτήτων $[0, \pi)$) παρατηρούμε ότι κάθε τόνος εμφανίζει δύο κορυφές, που αντιπροσωπεύουν τις δύο ημιτονοειδείς συνιστώσες του. Η αντιστοίχιση μεταξύ του δείκτη k στον οποίο εμφανίζεται μια κορυφή και της συχνότητας που αντιπροσωπεύει η εν λόγω κορυφή δίνεται μέσω της σχέσης:

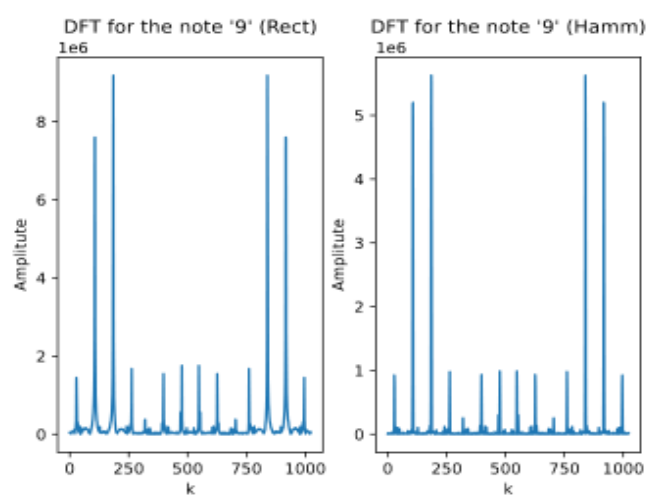
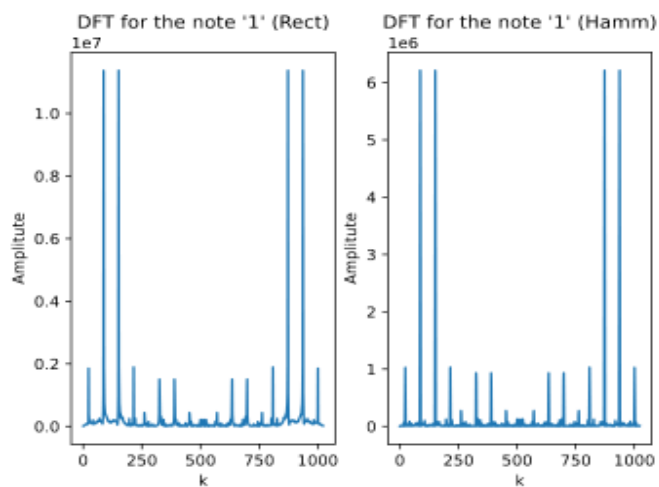
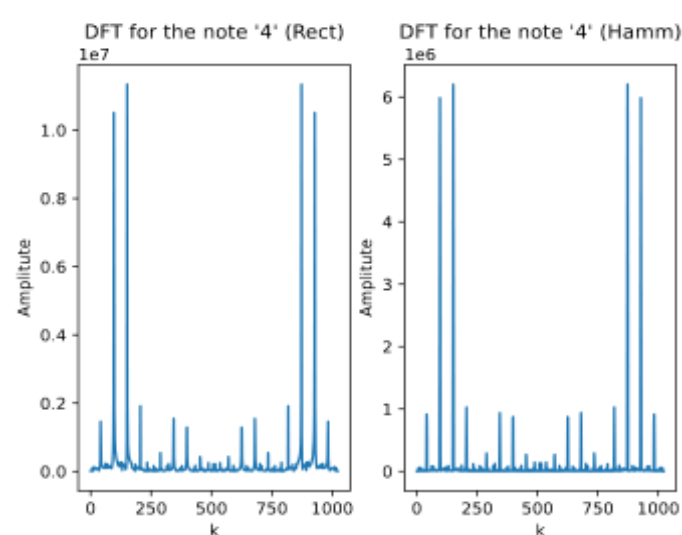
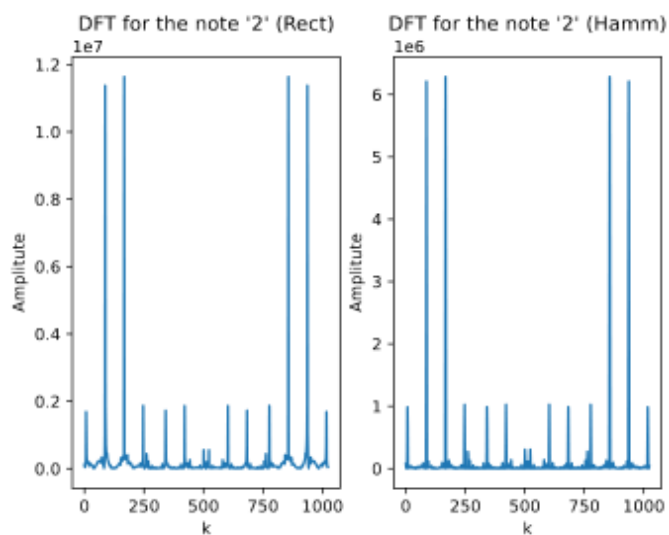
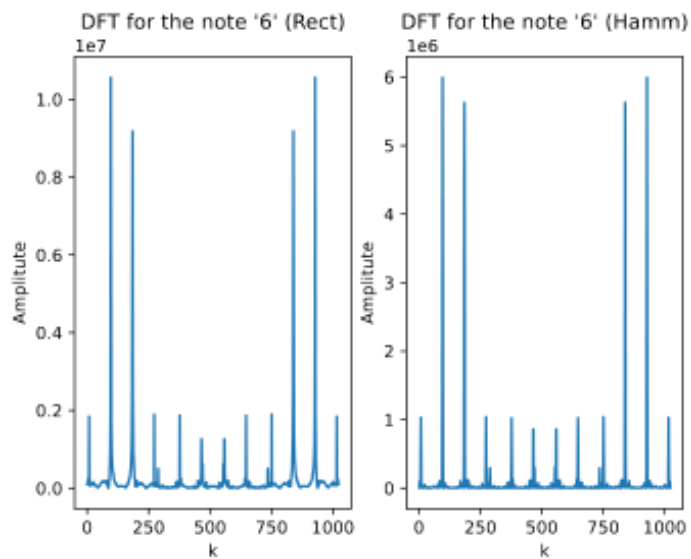
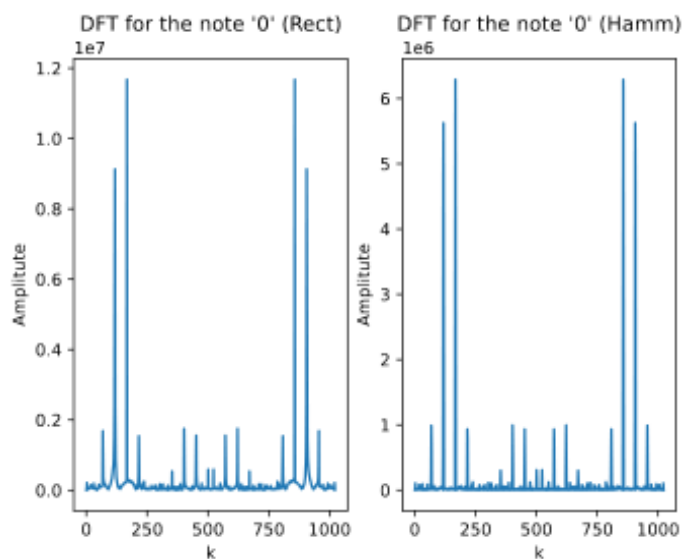
$$k = \text{int}\left(N \frac{\omega}{2\pi}\right)$$



2.4 Παρακάτω παρουσιάζονται οι DFTs των παραθυροποιημένων σημάτων με χρονικά παράθυρα (i) τετραγωνικά και (ii) Hamming για το σήμα tone_sequence.wav το οποίο αντιστοιχεί στο άθροισμα των αριθμών μητρώων μας (06241229).



Χριστόδουλος Στυλιανίδης 03120614
Γιάννης Γεωργιάδης 03120615





Χριστόδουλος Στυλιανίδης 03120614
Γιάννης Γεωργιάδης 03120615

2.5.

Περιορίζομαστε στο αριστερό μισό του φάσματος (εύρος από 0 μέχρι π) και υπολογίζουμε τους δείκτες k που αντιστοιχούν σε κάθε συχνότητα από τη σχέση

$$k = \text{int}\left(N \frac{\omega}{2\pi}\right)$$

Τα αποτελέσματα φαίνονται στον διπλανό πίνακα.

ω	k
0.5346	87
0.5906	96
0.6535	107
0.7217	118
0.9273	151
1.0247	167
1.1328	185

2.7.

Καλούμε την `ttdecode()` και λαμβάνουμε τα ψηφία που στα οποία αντιστοιχούν οι τόνοι κάθε .npy αρχείου

easy: 31404815

medium: 3248821096

hard: 2044963764

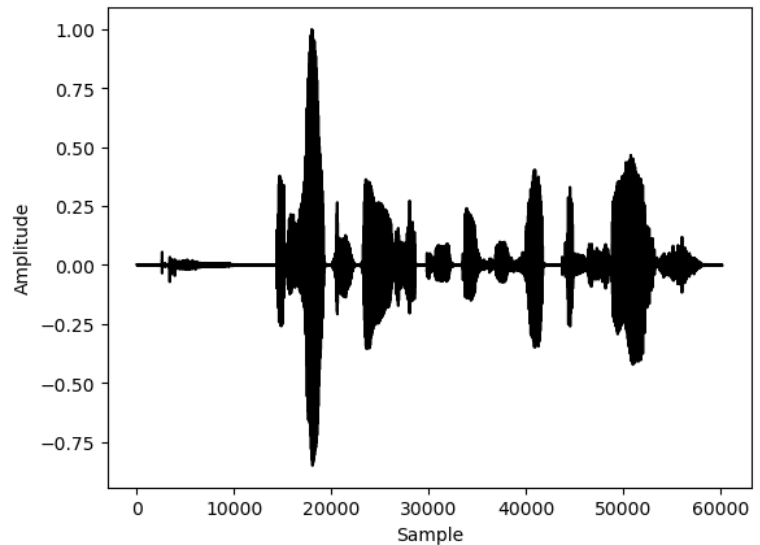


Μέρος 3ο - Χαρακτηριστικά Βραχέος Χρόνου Σημάτων Φωνής και Μουσικής

3.1.

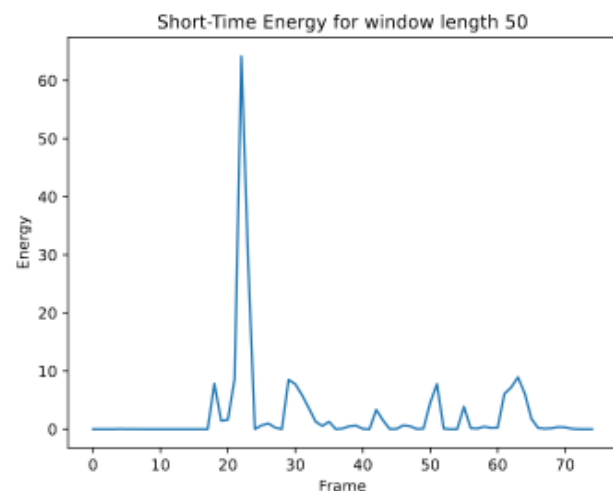
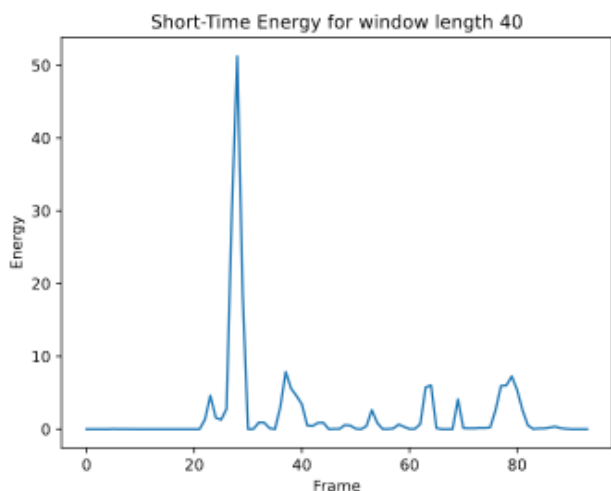
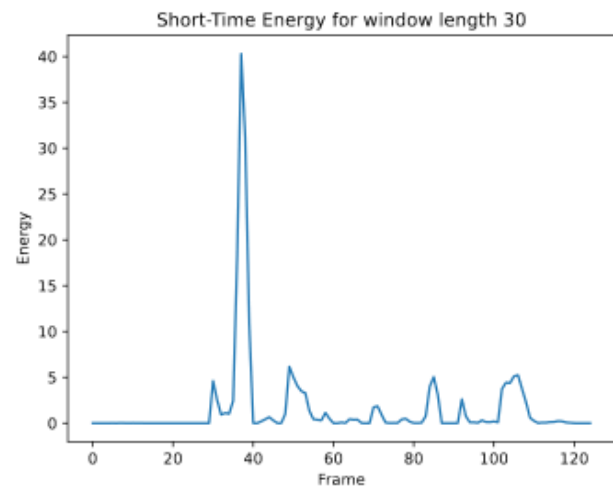
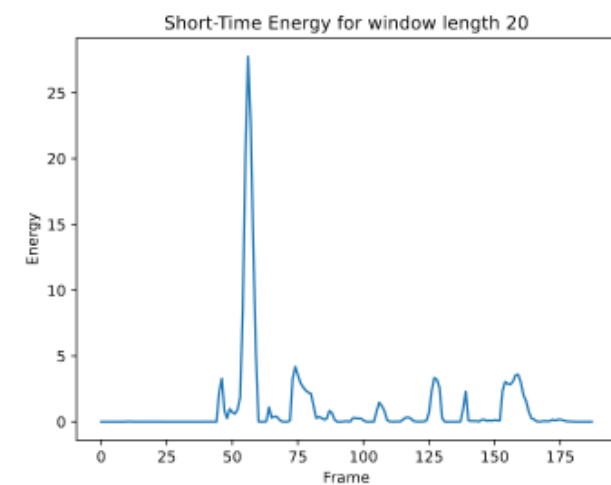
Απεικόνιση του σήματος στο πεδίο του χρόνου:

“It helps those people who help themselves”



3.2.

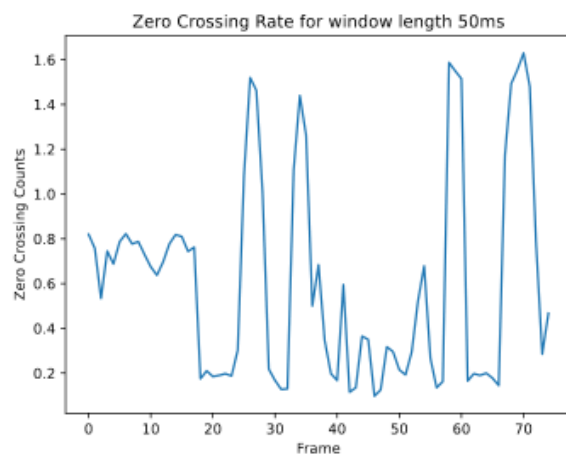
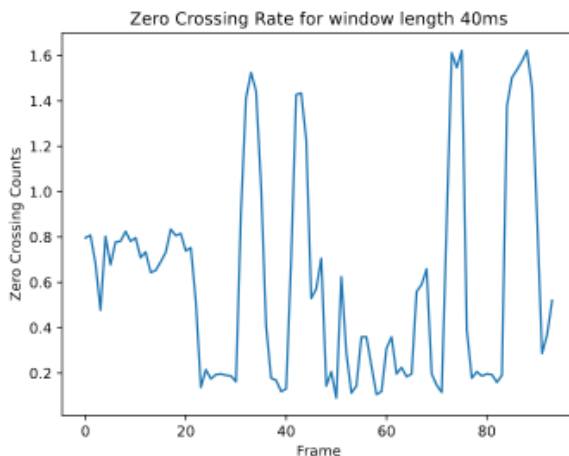
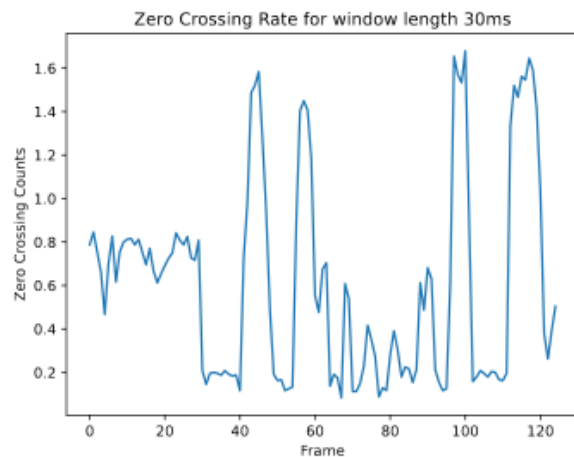
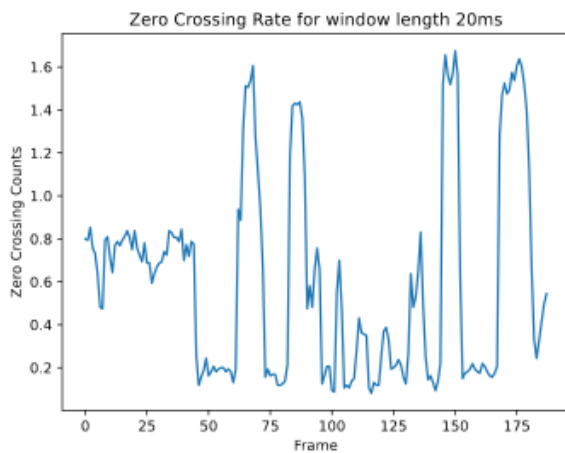
Ενέργεια βραχέος χρόνου για παράθυρα μήκους 20-50ms:





Χριστόδουλος Στυλιανίδης 03120614
Γιάννης Γεωργιάδης 03120615

Zero crossings του ίδιου σήματος για παράθυρα 20-50ms:



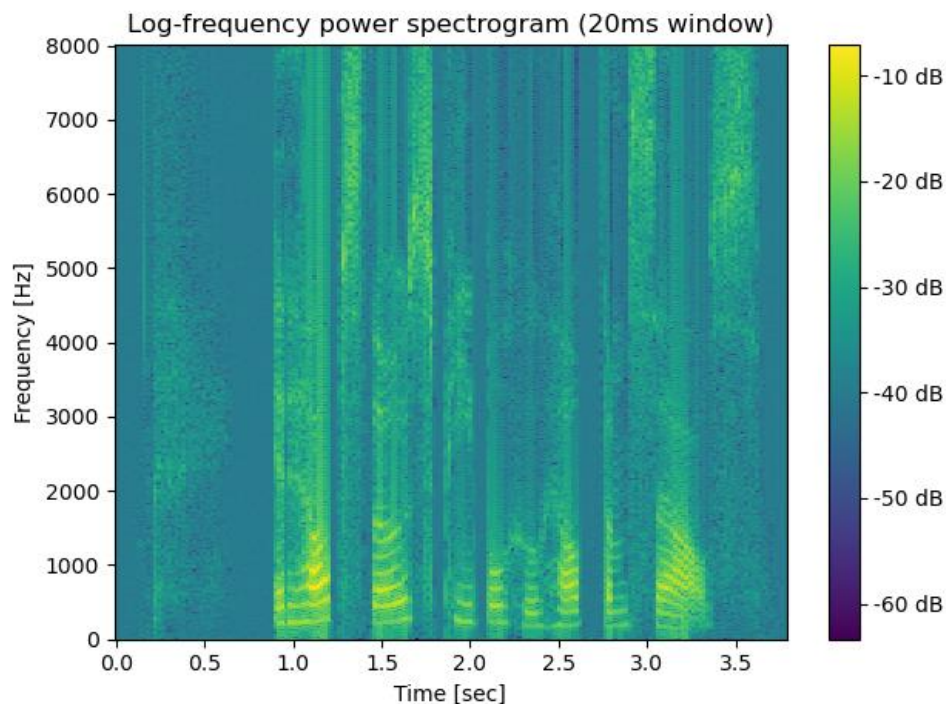
Χρησιμοποιώντας αυτές τις μετρήσεις μπορούμε να διαχωρίσουμε έμφωνους από άφωνους ήχους καθώς η υψηλή ενέργεια και το χαμηλό zero crossing rate αντιστοιχούν στα φωνήεντα, ενώ αντίθετα η χαμηλή ενέργεια και το ψηλό zero crossing rate στα σύμφωνα. Για παράδειγμα στο παράθυρο 20ms, κοντά στο 50ο frame, τα zero-crossings είναι λιγότερα, καθώς η ενέργεια είναι μέγιστη (σε όλο το σήμα), αυτό αντιστοιχεί στον έμφωνο ήχο “ee” (h**e**lps).



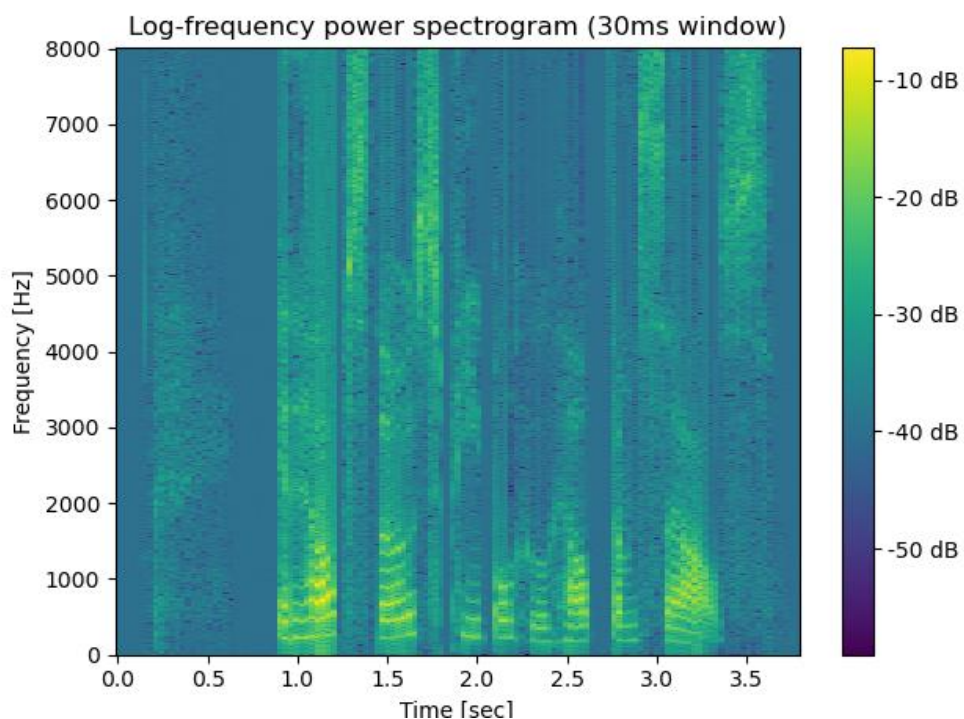
Χριστόδουλος Στυλιανίδης 03120614
Γιάννης Γεωργιάδης 03120615

3.3. SFT σε λογαριθμική κλίμακα:

For the window of 20ms:
Average Spectral Centroid: 296.20
Average Spectral Flux: 0.04



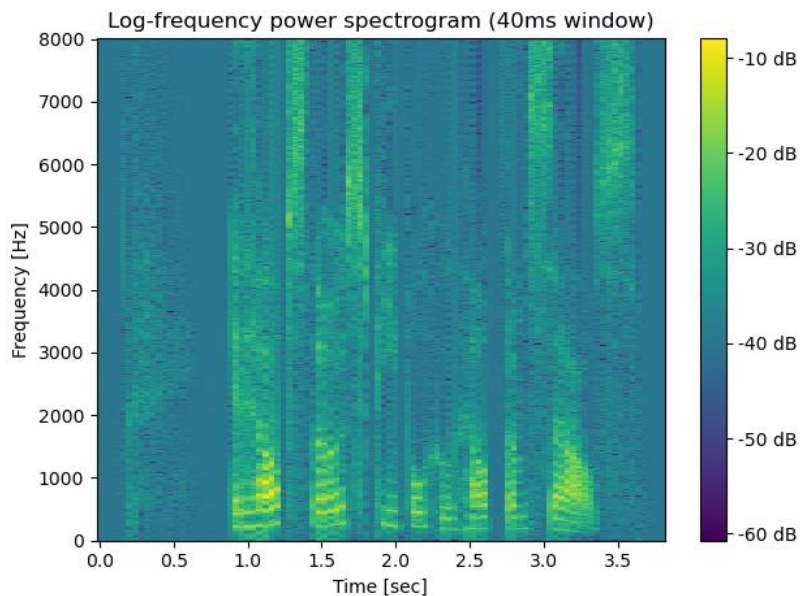
For the window of 30ms:
Average Spectral Centroid: 289.41
Average Spectral Flux: 0.03



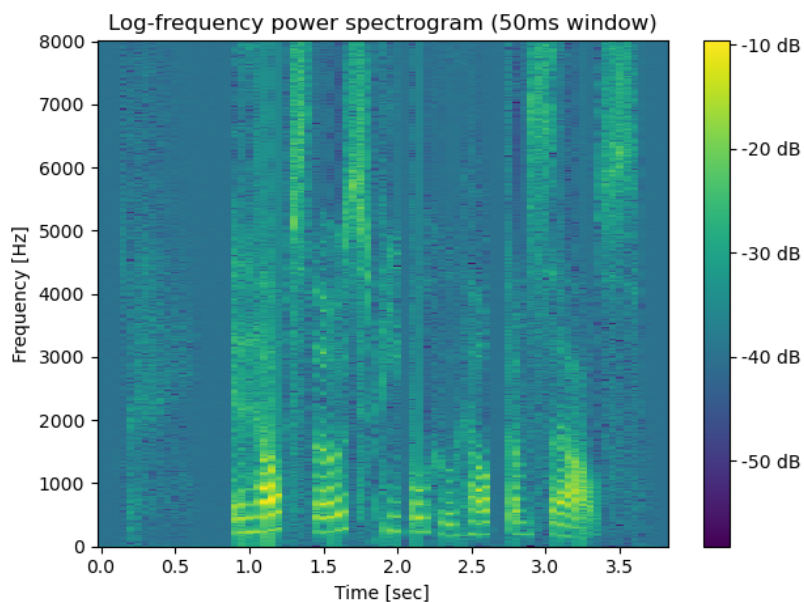


Χριστόδουλος Στυλιανίδης 03120614
Γιάννης Γεωργιάδης 03120615

For the window of 40ms:
Average Spectral Centroid: 286.36
Average Spectral Flux: 0.02



For the window of 50ms:
Average Spectral Centroid: 276.28
Average Spectral Flux: 0.01



Όσο αυξάνεται το μέγεθος του παραθύρου, η μεν ανάλυση στον άξονα της συχνότητας γίνεται καλύτερη, όμως έχουμε μεγαλύτερη ανακρίβεια στον χρόνο. Εντελώς ανάλογα, όταν μειώνεται το μέγεθος του παραθύρου, ισχύουν τα αντίθετα. Στον φάκελο υπάρχει το αρχείο spectrogram_speech.gif στο οποίο φαίνονται οι διαφορές στα spectrograms.



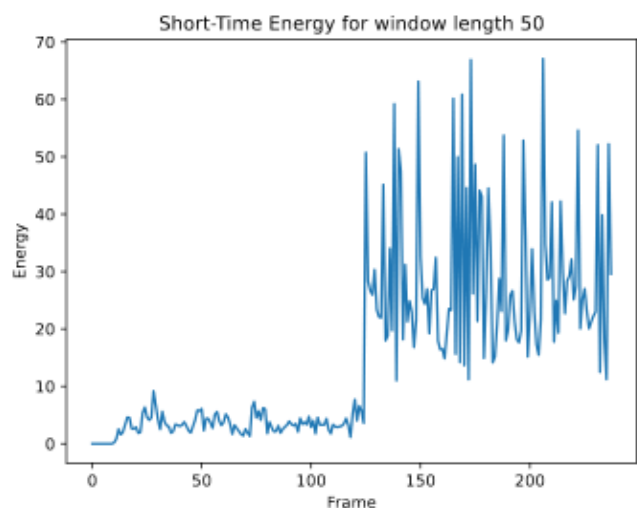
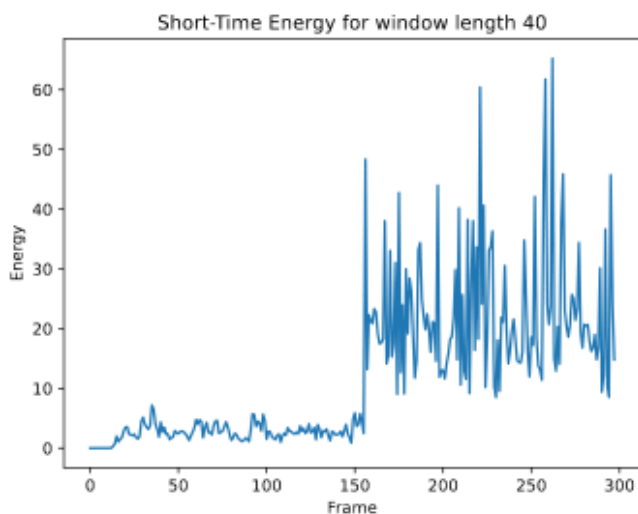
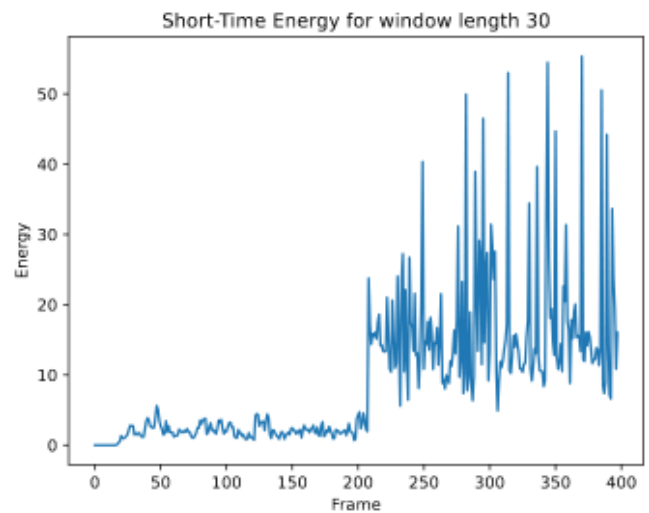
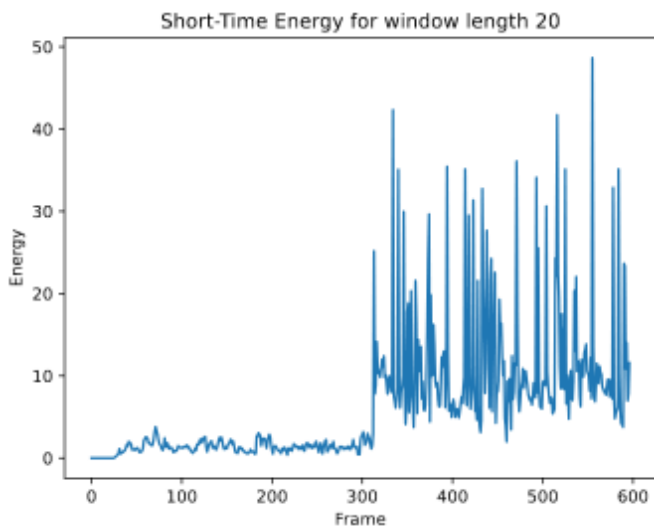
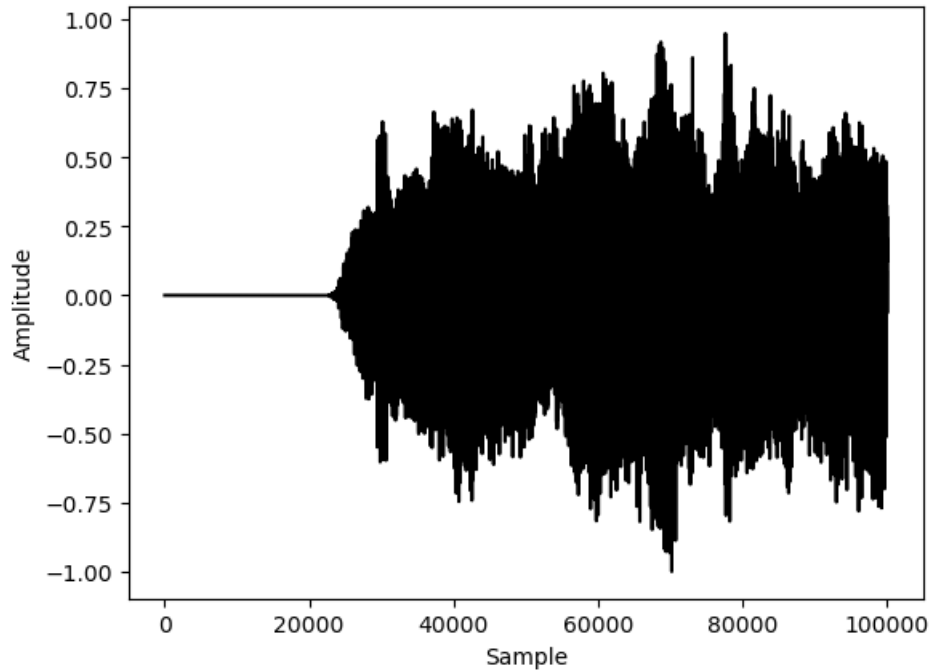
Χριστόδουλος Στυλιανίδης 03120614
Γιάννης Γεωργιάδης 03120615

3.4.

Απεικόνιση του σήματος στο
πεδίο του χρόνου:

12 δευτερόλεπτα από το
Island (Astro Remix)

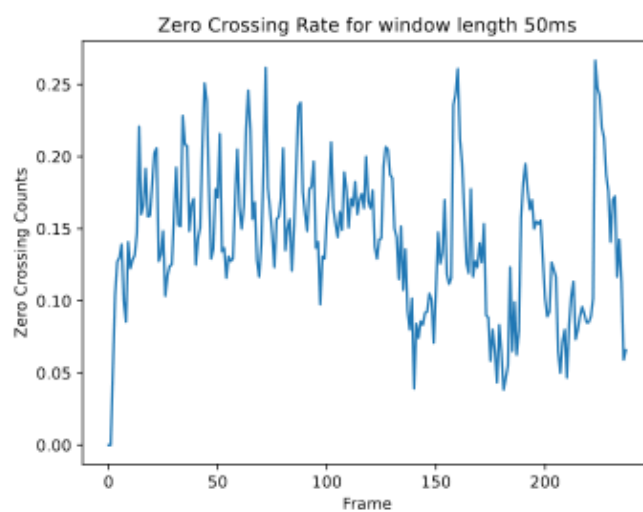
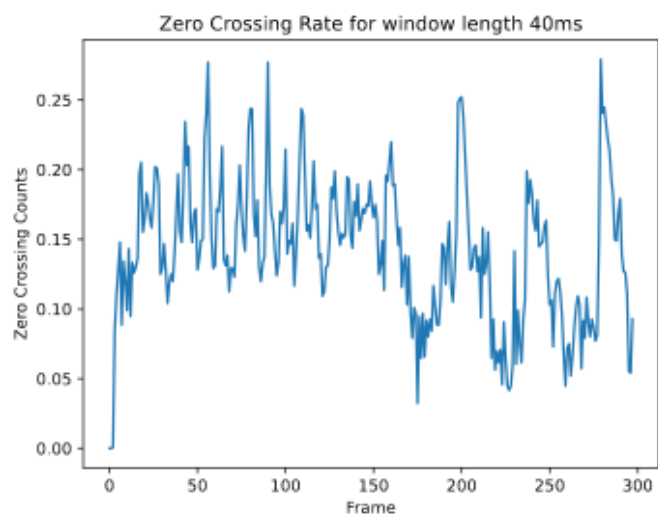
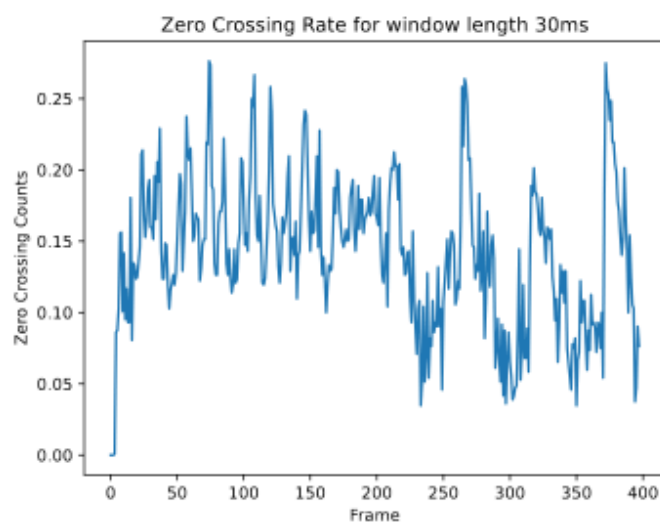
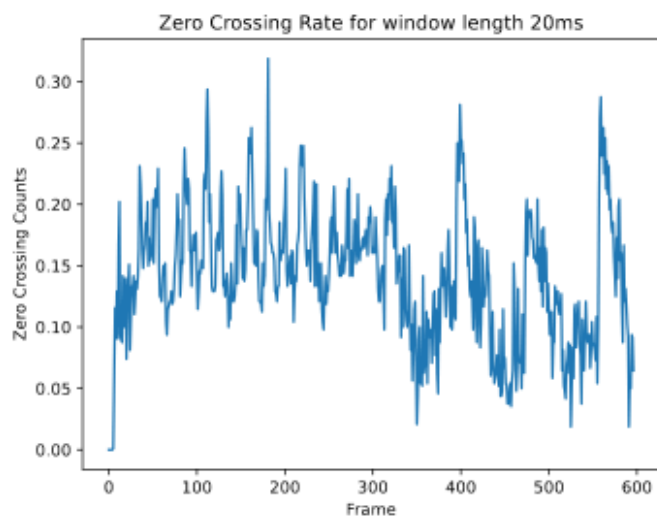
Ενέργεια βραχέος χρόνου για
παράθυρα μήκους 20-50ms:





Χριστόδουλος Στυλιανίδης 03120614
Γιάννης Γεωργιάδης 03120615

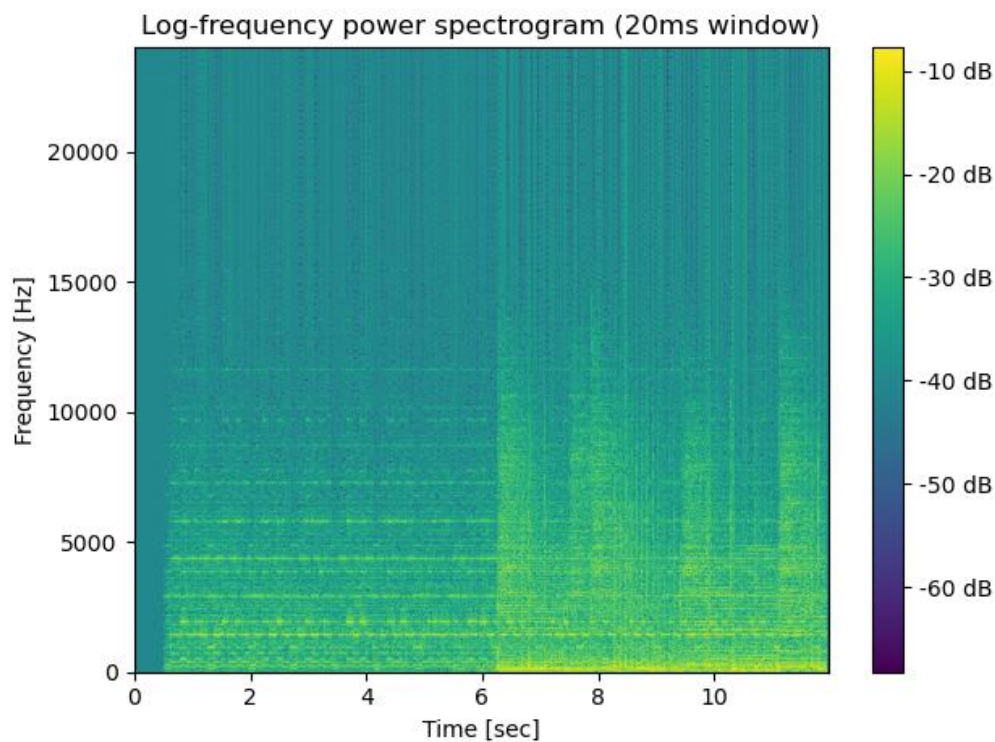
Zero crossings του ίδιου σήματος για παράθυρα 20-50ms:



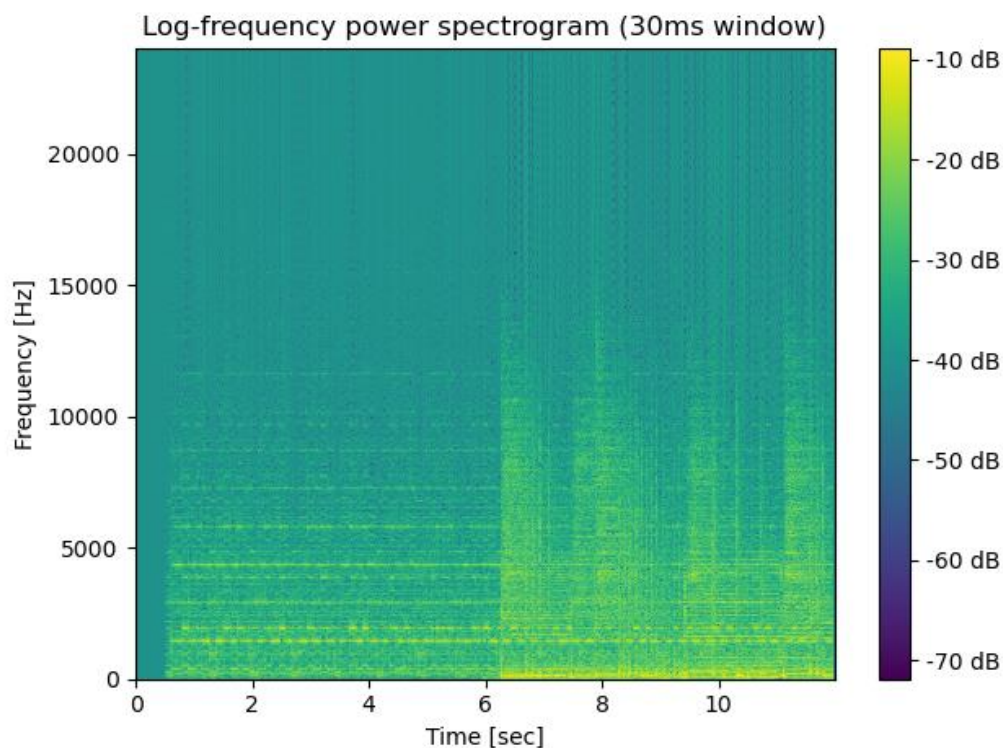


Χριστόδουλος Στυλιανίδης 03120614
Γιάννης Γεωργιάδης 03120615

For the window of 960 samples:
Average Spectral Centroid: 109.66
Average Spectral Flux: 0.05



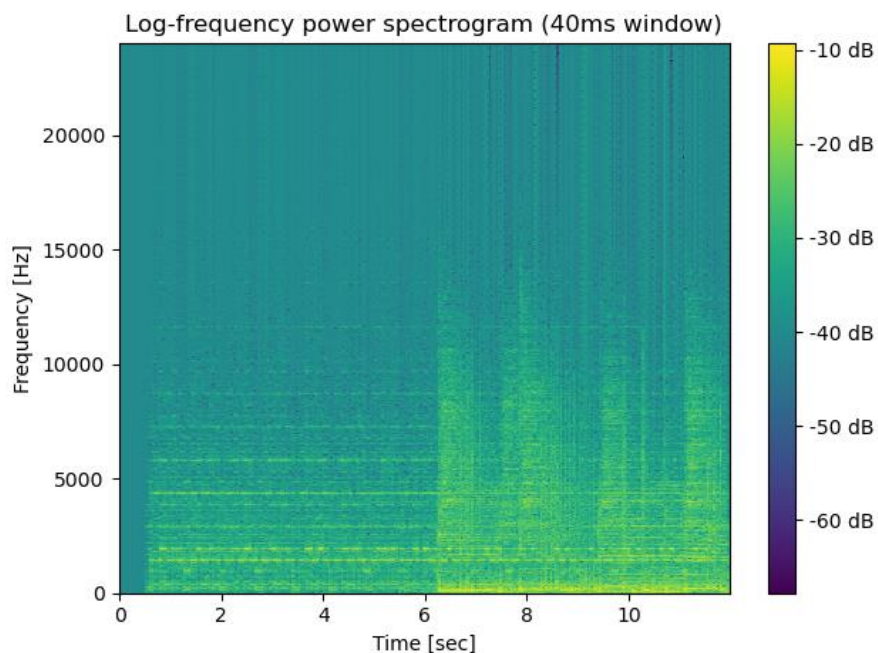
For the window of 1440 samples:
Average Spectral Centroid: 108.05
Average Spectral Flux: 0.04



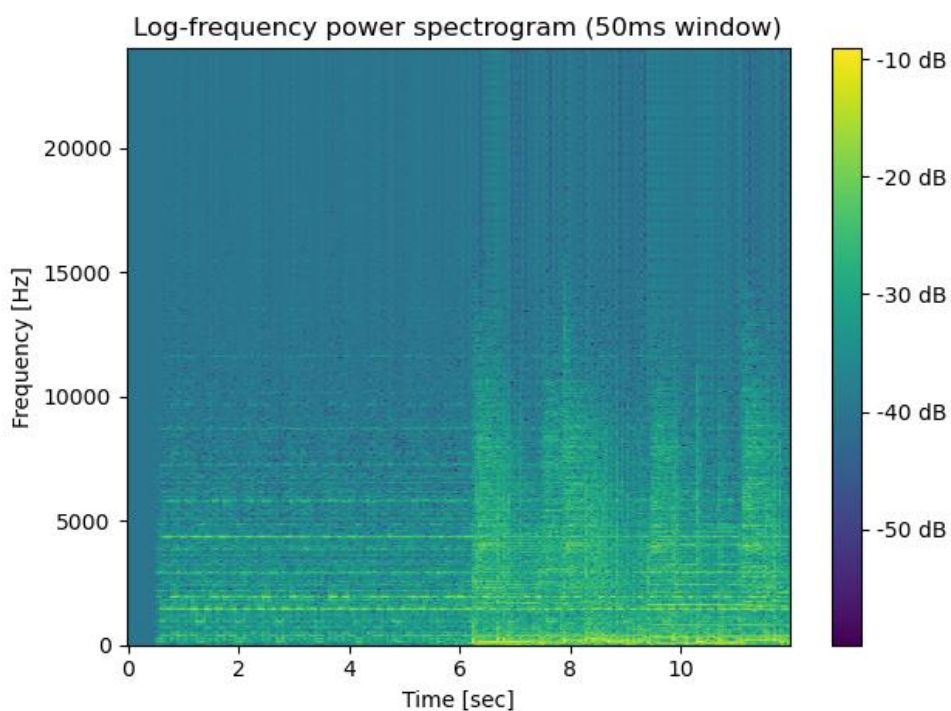


Χριστόδουλος Στυλιανίδης 03120614
Γιάννης Γεωργιάδης 03120615

For the window of 1920 samples:
Average Spectral Centroid: 108.01
Average Spectral Flux: 0.02



For the window of 2400 samples:
Average Spectral Centroid: 106.99
Average Spectral Flux: 0.02



Ομοίως με το speech_utterance, έχουμε τις ίδιες παρατηρήσεις για την ανάλυση. Το αρχείο spectrogram_music.gif περιέχει τις διαφορές στα spectrograms.