



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ

1^Η ΕΡΓΑΣΤΗΡΙΑΚΗ ΆΣΚΗΣΗ

Αναστάσιος – Στέφανος Αναγνώστου 03119051

Σπυρίδων Παπαδόπουλος 03119058

Ακαδημαϊκό Έτος 2021-2022

Παραδοτέα: 20-04-2022

ΠΕΡΙΕΧΟΜΕΝΑ

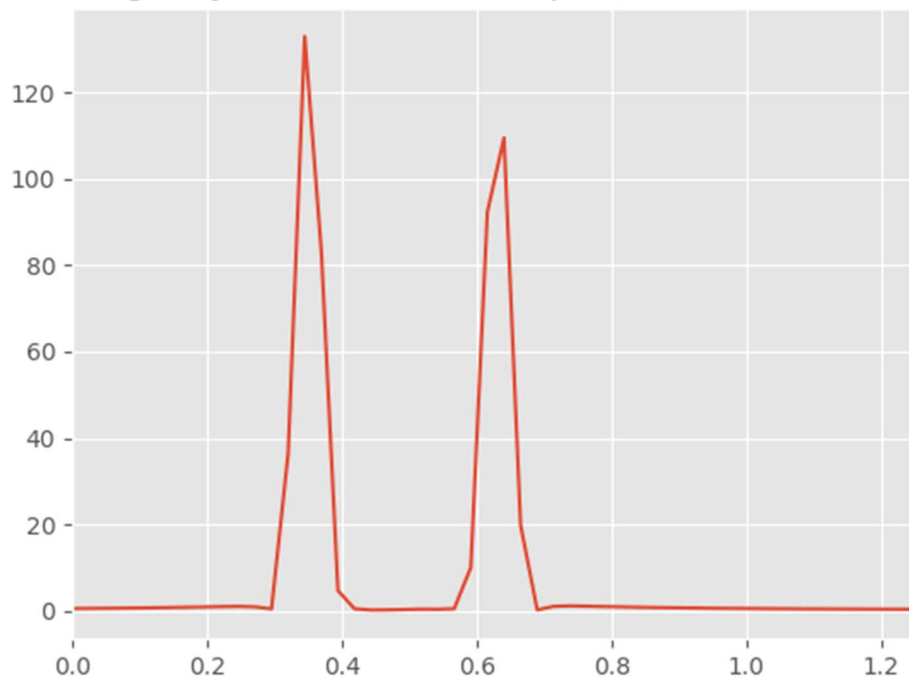
Μέρος 1 ^ο – Φασματική Ανάλυση και Ανίχνευση Ημιτονοειδών με τον Διακριτό Μετασχηματισμό Fourier (DFT).....	3
1.1.....	3
1.2.....	5
1.3.....	7
1.4.....	8
Μέρος 2 ^ο – Σύστημα Εντοπισμού Τηλεφωνικών Τόνων (Telephone Touch- Tones)	9
2.1.....	9
2.2.....	10
2.3.....	11
2.4.....	11
2.5.....	13
2.6.....	13
2.7.....	15
Μέρος 3 ^ο – Χαρακτηριστικά Βραχέος Χρόνου Σημάτων Φωνής και Μουσικής (Ενέργεια και Ρυθμός Εναλλαγής Προσήμου)	16
3.1.....	16
3.2.....	20

ΜΕΡΟΣ 1^ο – ΦΑΣΜΑΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΑΝΙΧΝΕΥΣΗ ΗΜΙΤΟΝΟΕΙΔΩΝ ΜΕ ΤΟΝ ΔΙΑΚΡΙΤΟ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟ FOURIER (DFT)

1.1

Η αρχική εικόνα του φάσματος του σήματος $y[n] = w_{\text{hamm}}[n] \cdot (x_1[n] + x_2[n])$, με $x_1[n] = A_1 \cdot e^{j \cdot (\omega_1 \cdot n + \phi_1)}$ και $x_2[n] = A_2 \cdot e^{j \cdot (\omega_2 \cdot n + \phi_2)}$, $w_{\text{hamm}}[n] = 0.54 - 0.46 \cdot \cos(2 \cdot \pi \cdot n \cdot N)$, $0 \leq n \leq N$, συχνότητες $\omega_1 = \frac{\pi}{9}$, $\omega_2 = \frac{\pi}{5}$ και πλάτη $A_1 = 1$, $A_2 = 0.9$ είναι:

DFT of signal y with $N = 256$ samples, $\omega_1 = 0.111\pi$, $\omega_2 = 0.2\pi$



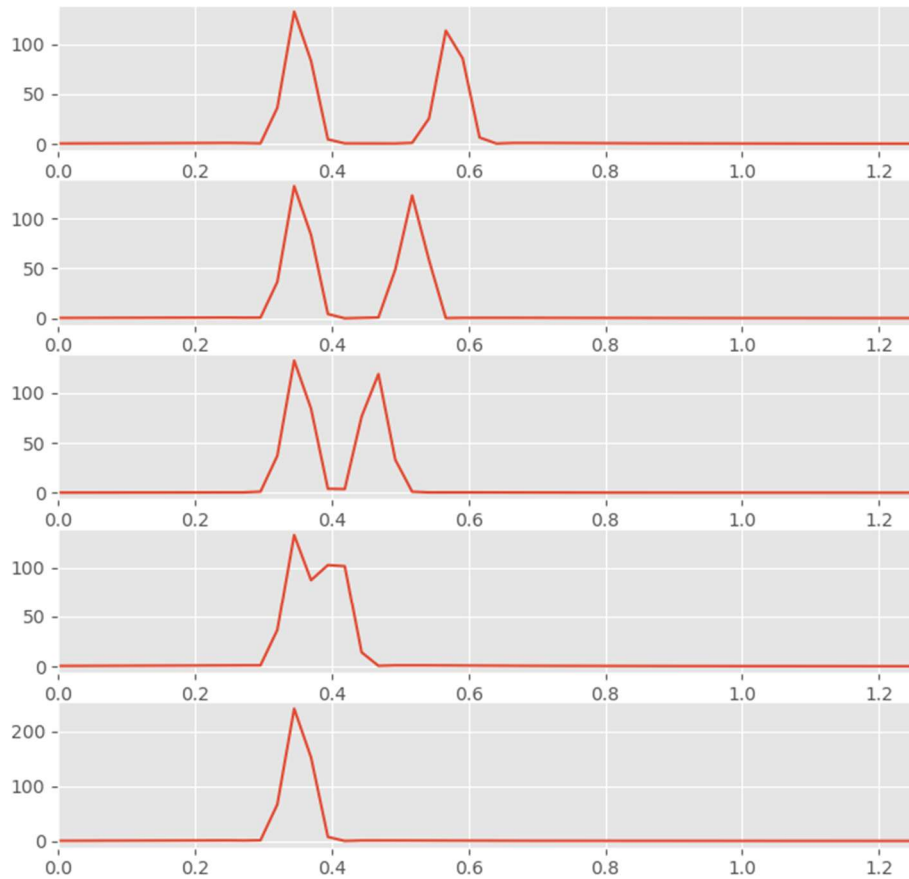
Εικόνα 1 Ο Διακριτός Μετ/σμός Fourier του δεδομένου σήματος $y[n]$

Σημειώνεται ότι στο συγκεκριμένο φάσμα αλλά και στα ακόλουθα, ο οριζόντιος άξονας είναι το εύρος γωνιακών συχνοτήτων $[0, 2\pi]$. Εκάστοτε υπόκειται σε μεγέθυνση προς καλύτερη επισκόπηση των αποτελεσμάτων.

Πράγματι, οι κορυφές του φάσματος σημειώνονται στις επιλεχθείσες συχνότητες του σήματος.

Για να εξεταστεί η διακριτική ικανότητα του συγκεκριμένου μετασχηματισμού $N = 256$ δειγμάτων, χωρίζεται το διάστημα μεταξύ των δύο δεδομένων συχνοτήτων σε 5 τμήματα. Άρα, μεταβάλλεται η συχνότητα ω_2 σύνολο 5 φορές, καθεμία φορά πλησιάζοντας περισσότερο την συχνότητα ω_1 . Κατά την μεταβολή της, εξετάζεται κατά πόσον παραμένουν ευδιάκριτες οι δύο κορυφές.

DFT of signal with a stable frequency $\omega_1=0.111\pi$
and a moving frequency starting from $\omega_2=0.182\pi$



Εικόνα 2 Διαδοχικοί Διακριτοί Μετασχηματισμοί Fourier για διάφορες τιμές της συχνότητας ω_2 κινούμενες προς την ω_1

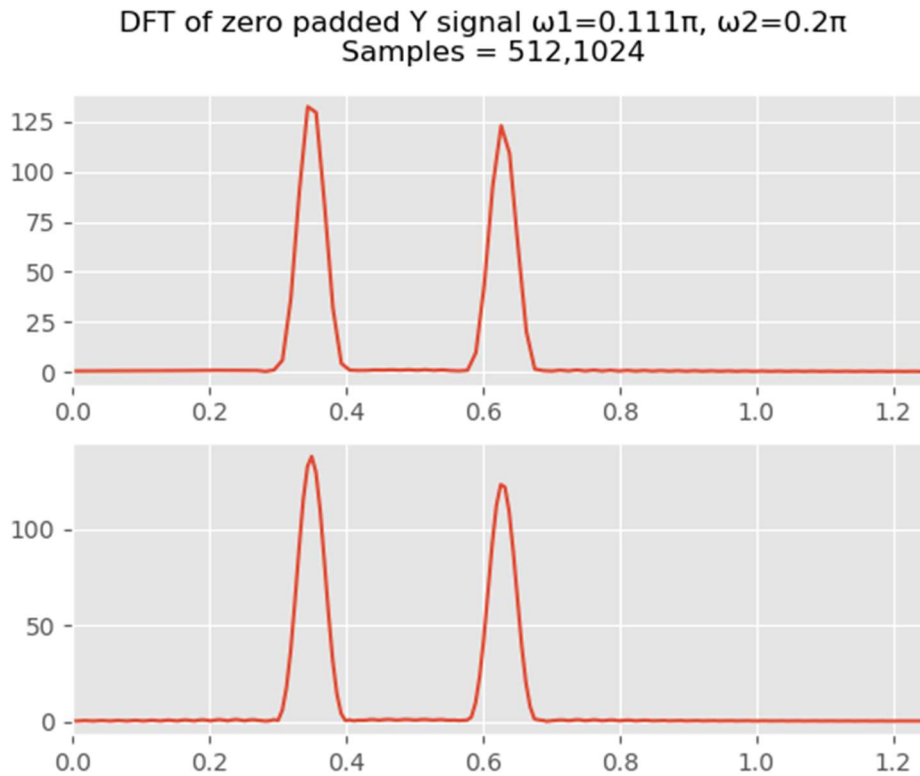
Παρατηρείται ότι, με την χρήση $L = 256$ δειγμάτων, η διακριτική ικανότητα του Μετ/σμού Fourier δεν είναι ικανοποιητική, καθώς φαίνεται ότι οι κορυφές του φάσματος συχνότητας συγχέονται χωρίς να πλησιάσουν πολύ οι τιμές των συχνοτήτων. Συγκεκριμένα, η αλληλεπικάλυψη συμβαίνει ήδη για τις τιμές $\omega_1 = 0,111\pi$ και $\omega_2 = 0,111\pi + 0,1117 = 0,1465\pi$ (τρίτο υποδιάγραμμα με αρίθμηση από πάνω προς κάτω) και φτάνει στο αμέσως επόμενο βήμα. Η ελάχιστη διαφορά συχνοτήτων, δηλαδή, μπορεί να είναι έως και $\Delta\omega = 0,1117$.

1.2

Επαναλαμβάνεται η προηγούμενη διαδικασία, αλλά αυτήν την φορά παρεμβάλλονται

- $N_1 = 512 - 256 = 256$ μηδενικά
- $N_2 = 1024 - 256 = 768$ μηδενικά

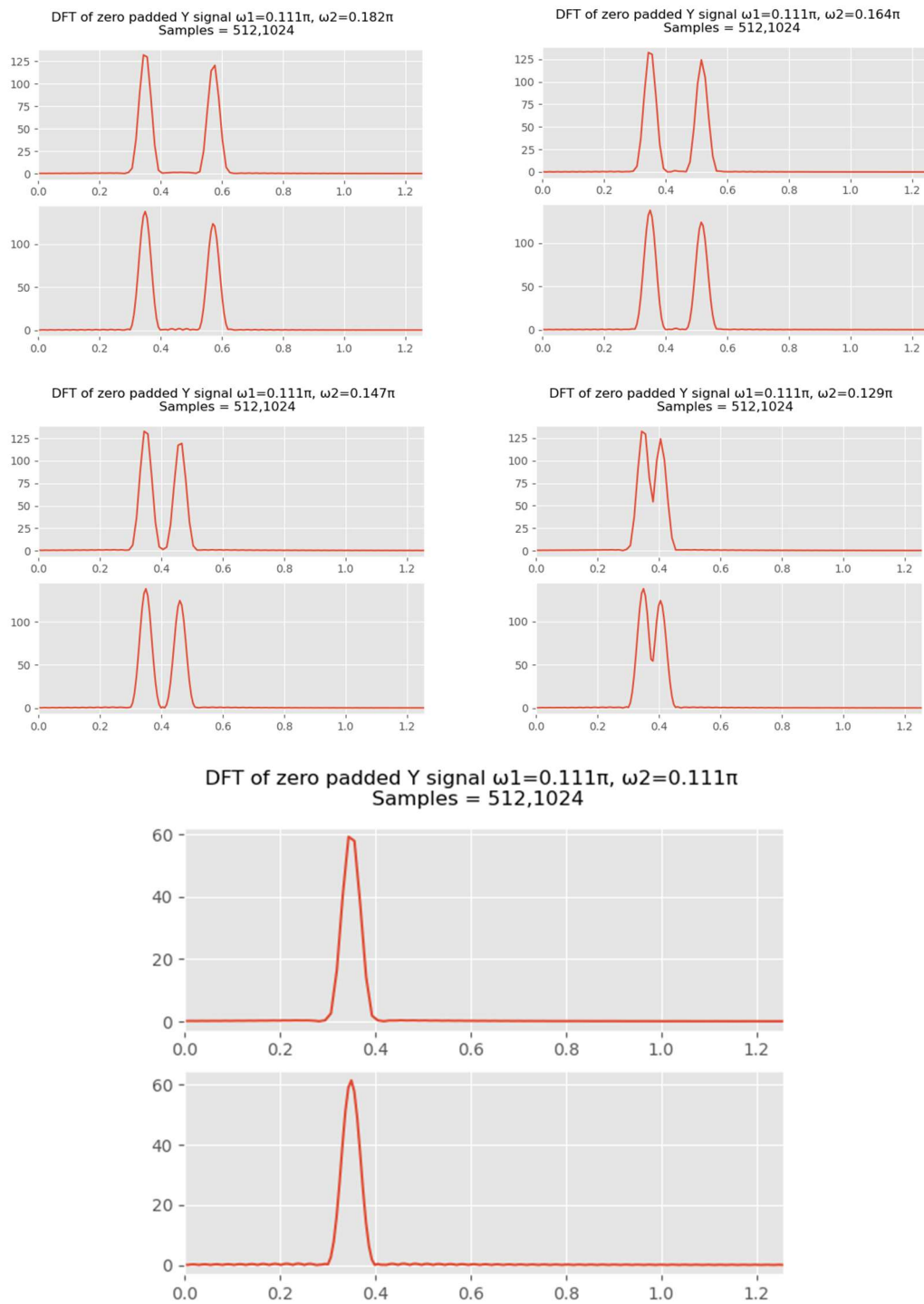
στο τέλος του σήματος, ώστε οι προκύπτοντες Διακριτοί Μετασχηματισμοί Fourier να έχουν αντιστοίχως μήκη των 512 και 1024 δειγμάτων. Τα αντίστοιχα φάσματα φαίνονται παρακάτω:



Εικόνα 3 Τα φάσματα συχνότητας του δεδομένου σήματος $y[n]$ χρήσει Διακριτών Μετ/σμών Fourier 512 και 1024 δειγμάτων αντιστοίχως

Παρατηρείται ότι ο μετασχηματισμός των 1024 δειγμάτων είναι ευκρινέστερος αυτού των 512 δειγμάτων, καθώς οι καμπύλες του φάσματος είναι ομαλότερες, χωρίς απότομες μεταβολές και γωνίες. Αυτό συμβαίνει γιατί ο Διακριτός Μετασχηματισμός Fourier δέχεται ως είσοδο μία ακολουθία συγκεκριμένου μήκους και παράγει μία έξοδο ίδιου μήκους. Άρα, με την προσθήκη μηδενικών σημείων στην είσοδο του μετασχηματισμού, εξαναγκάζεται η παραγωγή μίας εξόδου περισσότερων σημείων, δηλαδή μεγαλύτερης λεπτομέρειας, και ίδιου συχνοτικού περιεχομένου, αφού τα μηδενικά δεν επηρεάζουν τις συχνότητες του αρχικού σήματος. Ως αποτέλεσμα, παράγεται μία λεπτομερέστερη και ευκρινέστερη εικόνα του ίδιου φάσματος.

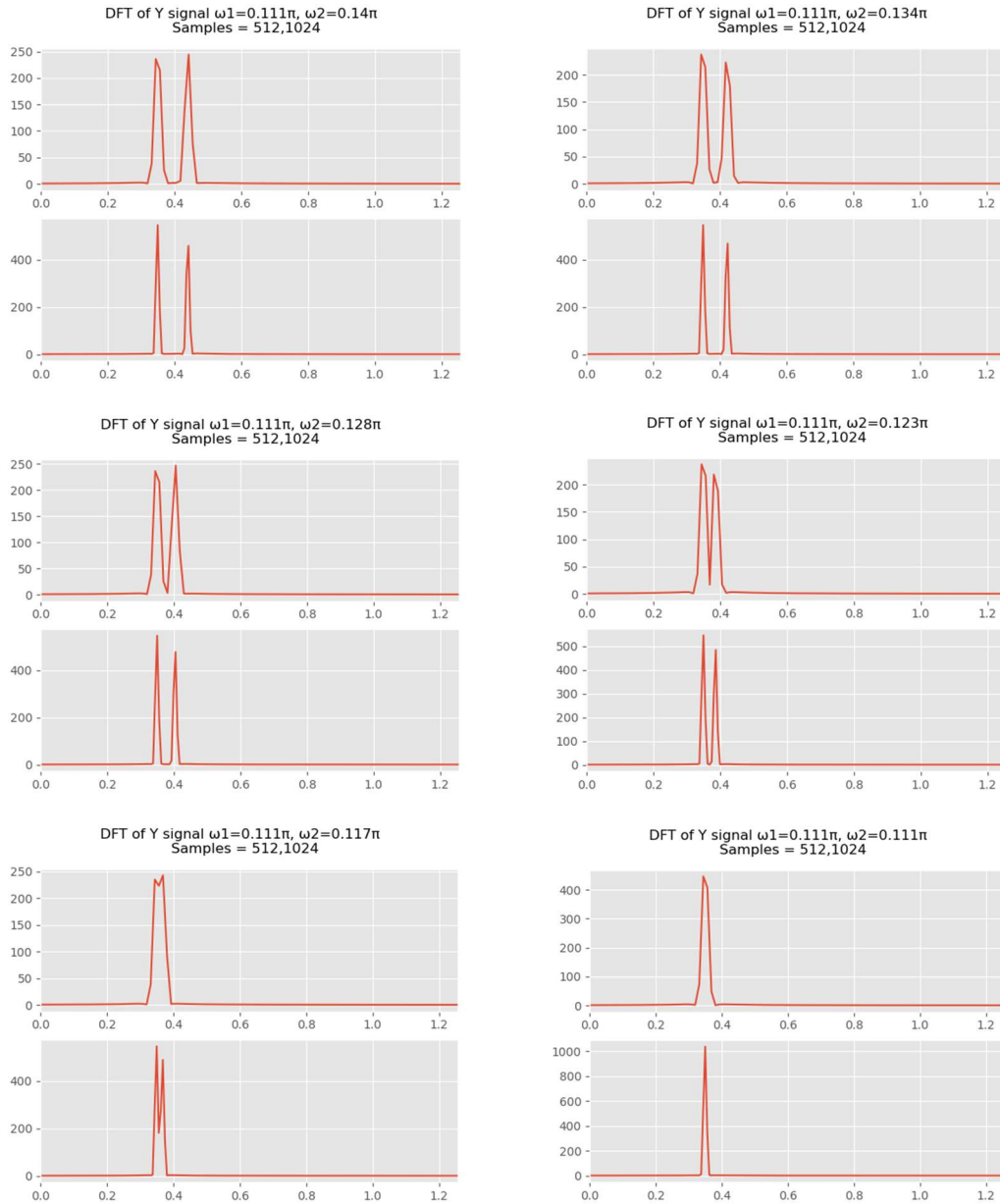
Εάν επιχειρηθεί η προσέγγιση των συχνοτήτων του σήματος όπως προηγουμένως, λαμβάνονται τα ακόλουθα γραφήματα:



Φαίνεται, δηλαδή, ότι το φάσμα των 1024 δειγμάτων, παρ' ότι ευκρινέστερο, δεν έχει υψηλότερη διακριτική ικανότητα από αυτό των 512 δειγμάτων, καθώς η αλληλεπικάλυψη των δύο κορυφών συμβαίνει στην ίδια συχνότητα σε αμφότερα τα φάσματα. Αξιοσημείωτο είναι ότι, παρά της ευκρίνειάς τους, δεν έχουν υψηλότερη διακριτική ικανότητα ούτε από το φάσμα του προηγούμενου πειράματος των 0 μηδενικών, καθώς οι κορυφές αλληλεπικαλύπτονται για την ίδια τιμή του ω_2 . Είναι σαφές, λοιπόν, ότι η διαδικασία προσθήκης μηδενικών στο αρχικό σήμα δεν αυξάνει την διακριτική ικανότητα του φάσματος παρά μόνο συνεισφέρει στην λεπτομερή απεικόνισή του.

1.3

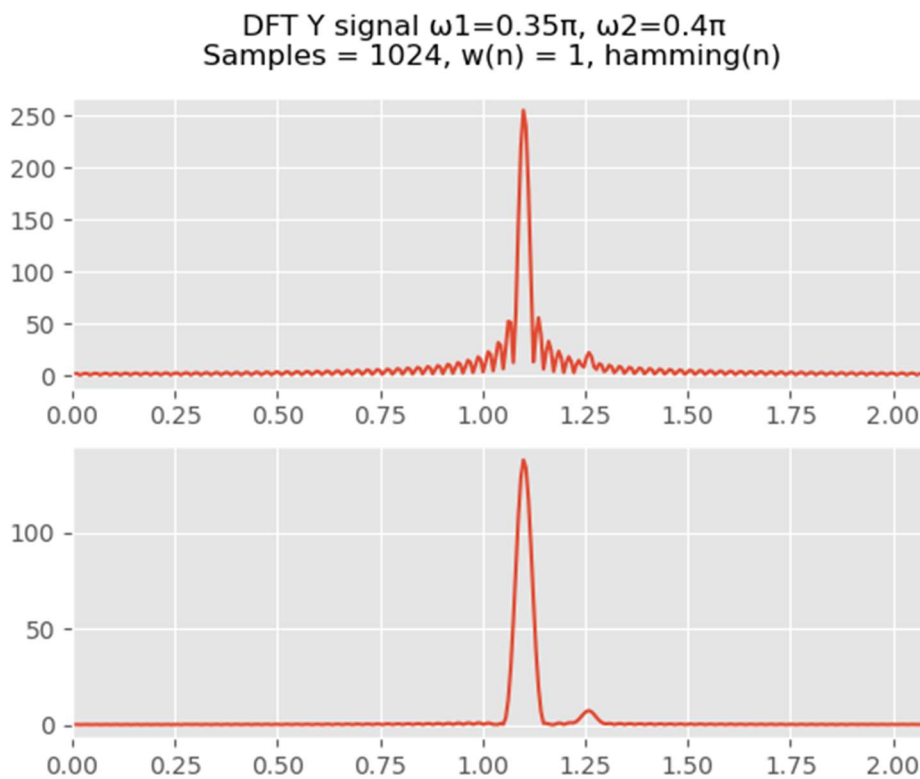
Στο παρόν πείραμα, χρησιμοποιούνται περισσότερα «ωφέλιμα» σημεία, αντί μηδενικών, για την απεικόνιση του φάσματος. Συγκεκριμένα, χρησιμοποιούνται αντιστοίχως $L_1 = 512$ και $L_2 = 1024$ δείγματα με σκοπό να μελετηθεί η επίδρασή τους σε αυτό. Επαναλαμβάνεται η διαδικασία προσεγγίσεως των συχνοτήτων, με αρχική τιμή της συχνότητας $\omega_2 = 0,1465\pi$, δηλαδή την συχνότητα οριακής αλληλεπικάλυψης κορυφών. Παρακάτω απεικονίζονται τα διαγράμματα:



Φαίνεται από το πρώτο κιόλας διάγραμμα ότι οι κορυφές των συχνοτήτων είναι πιο απότομες και ως εκ τούτου δεν υπάρχει αλληλεπικάλυψη. Κατά την ολίσθηση της συχνότητας ω_2 προς την ω_1 , παρατηρείται ότι η αλληλεπικάλυψη των κορυφών συμβαίνει στην τιμή $\omega_2 = 0,128\pi$ για το φάσμα των $L_1 = 512$ δειγμάτων, ενώ οι κορυφές του φάσματος των $L_2 = 1024$ δειγμάτων είναι ευδιάκριτες μέχρι και στην τιμή $\omega_2 = 0,123\pi$. Διαπιστώνεται σαφώς, λοιπόν, ότι ο καθοριστικός παράγοντας για την διακριτική ικανότητα του φάσματος του Διακριτού Μετασχηματισμού Fourier είναι το πλήθος των χρησιμοποιούμενων δειγμάτων από το αρχικό σήμα και όχι το πλήθος των προστιθέμενων μηδενικών στο τέλος του.

1.4

Στο πείραμα αυτό μελετάται η επίδραση των διαφόρων παραθύρων, ιδανικού τετραγωνικού και παραθύρου hamming, στο φάσμα ενός σήματος σαν το προηγούμενο αλλά με παραμέτρους συχνότητες $\omega_1 = 0,35\pi$ και $\omega_2 = 0,4\pi$, πλάτη $A_1 = 1$ και $A_2 = 0,05$ και τυχαίες φάσεις. Οι κύριες συχνότητές του, δηλαδή, είναι σχετικά κοντινές και εμφανίζονται με διαφορετικές εντάσεις · η μεν πρώτη με ένταση 1, η δε δεύτερη με ένταση 20 φορές μικρότερη. Στο φάσμα αναμένονται δύο κοντινές κορυφές με μεγάλη διαφορά στο ύψος τους. Για τον σχεδιασμό των φασμάτων χρησιμοποιείται σήμα μήκους $L = 256$ δειγμάτων με $Z = 1024 - 256 = 768$ μηδενικά στο τέλος του, ώστε να παραχθεί μετασχηματισμός μήκους $N = 1024$. Πριν τον μετασχηματισμό, εφαρμόζονται τα παράθυρα. Παρακάτω απεικονίζονται τα φάσματα:



Παρατηρείται ότι, στην περίπτωση όπου χρησιμοποιείται το τετραγωνικό παράθυρο, εμφανίζονται πλευρικοί λοβοί στις αρμονικές χαμηλοτέρων τάξεων γύρω από τις κεντρικές συχνότητες. Εν προκειμένω μάλιστα, επειδή η μία συχνότητα εμφανίζεται με 20 φορές μεγαλύτερη ένταση από την άλλη, οι πλευρικοί της λοβοί είναι τόσο μεγάλοι ώστε σχεδόν να απορροφούν την δεύτερη κυρίαρχη συχνότητα και να μην απεικονίζεται ευκρινώς.

Ωστόσο, όταν χρησιμοποιείται το παράθυρο hamming, παρατηρείται ότι οι πλευρικοί λοβοί απομονώνονται επιτυχώς και μειώνεται κατά πολύ το πλάτος τους, τόσο μάλιστα, ώστε η δεύτερη κυρίαρχη συχνότητα, παρά την χαμηλή της ένταση σε σχέση με την άλλη, να απεικονίζεται ευκρινέστατη.

ΜΕΡΟΣ 2^ο – ΣΥΣΤΗΜΑ ΕΝΤΟΠΙΣΜΟΥ ΤΗΛΕΦΩΝΙΚΩΝ ΤΟΝΩΝ (TELEPHONE TOUCH- TONES)

2.1

Προς προετοιμασία για το επόμενο πείραμα, δίδεται ο ακόλουθος πίνακας διακριτών συχνοτήτων και, βάσει αυτού, κατασκευάζονται δέκα ηχητικά σήματα, αριθμημένα από το 0 έως το 9, με τις αντίστοιχες συχνότητες. Χρησιμοποιούνται 1000 δείγματα για καθέναν τόνο.

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

	Ωσθλών		
Ωγραμμών	0,9273	1,0247	1,1328
0,5346	1	2	3
0,5906	4	5	6
0,6535	7	8	9
0,7217		0	

Λαμβάνονται, τελικά, οι εξής δέκα συναρτήσεις, μία ανά τόνο:

$$d_0[n] = \sin(0.7217 \cdot n) + \sin(1.0247 \cdot n)$$

$$d_1[n] = \sin(0.5346 \cdot n) + \sin(0.9273 \cdot n)$$

$$d_2[n] = \sin(0.5346 \cdot n) + \sin(1.0247 \cdot n)$$

$$d_3[n] = \sin(0.5346 \cdot n) + \sin(1.1328 \cdot n)$$

$$d_4[n] = \sin(0.5906 \cdot n) + \sin(0.9237 \cdot n)$$

$$d_5[n] = \sin(0.5906 \cdot n) + \sin(1.0247 \cdot n)$$

$$d_6[n] = \sin(0.5906 \cdot n) + \sin(1.1328 \cdot n)$$

$$d_7[n] = \sin(0.6535 \cdot n) + \sin(0.9237 \cdot n)$$

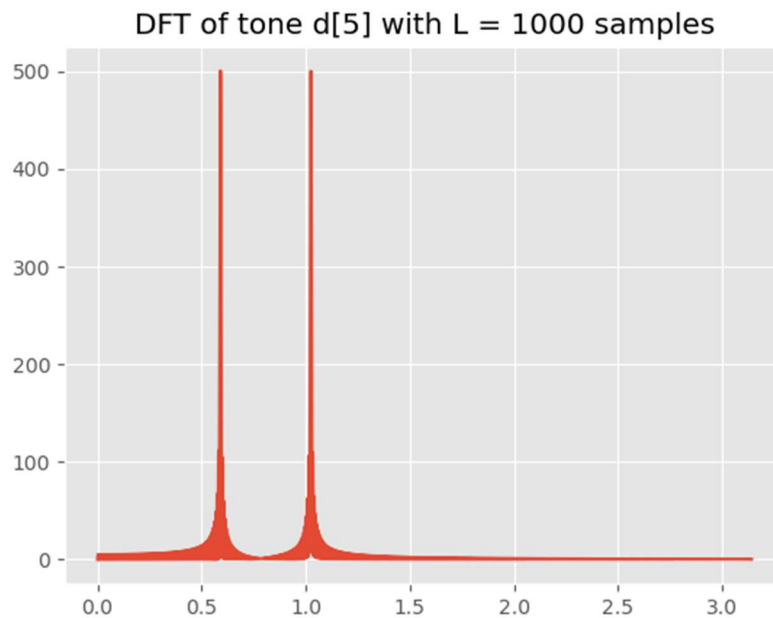
$$d_8[n] = \sin(0.6535 \cdot n) + \sin(1.0247 \cdot n)$$

$$d_9[n] = \sin(0.6535 \cdot n) + \sin(1.1328 \cdot n)$$

Για $n \in \mathbb{Z} \cap n \in [0,1000)$.

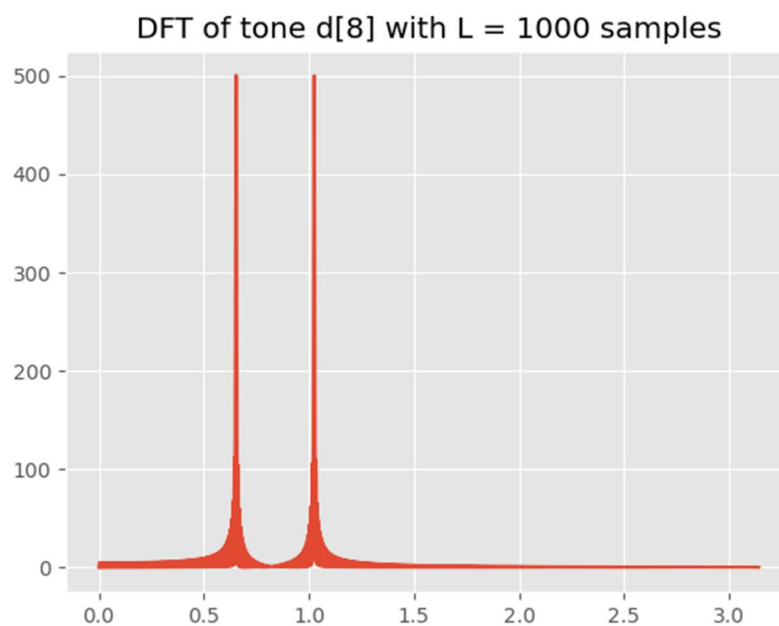
2.2

Υπολογίζοντας τον Διακριτό Μετασχηματισμό Fourier των τόνων υπ' αριθμόν 5 και 8, αναμένονται για κάθε φάσμα δύο κορυφές, καθεμία στην αντίστοιχη διακριτή συχνότητα των ημιτόνων του σήματος.



Εικόνα 4 Διακριτός Μετασχηματισμός Fourier του τόνου 5

Πράγματι, οι κορυφές των φασμάτων βρίσκονται στις αναμενόμενες συχνότητες. Μάλιστα, μία από τις κορυφές είναι κοινή και στα δύο φάσματα, αφού οι τόνοι υπ' αριθμόν 5 και 8 βρίσκονται στην ίδιαν στήλη και ως εκ τούτου έχουν μία από τις δύο συχνότητες ίδια.



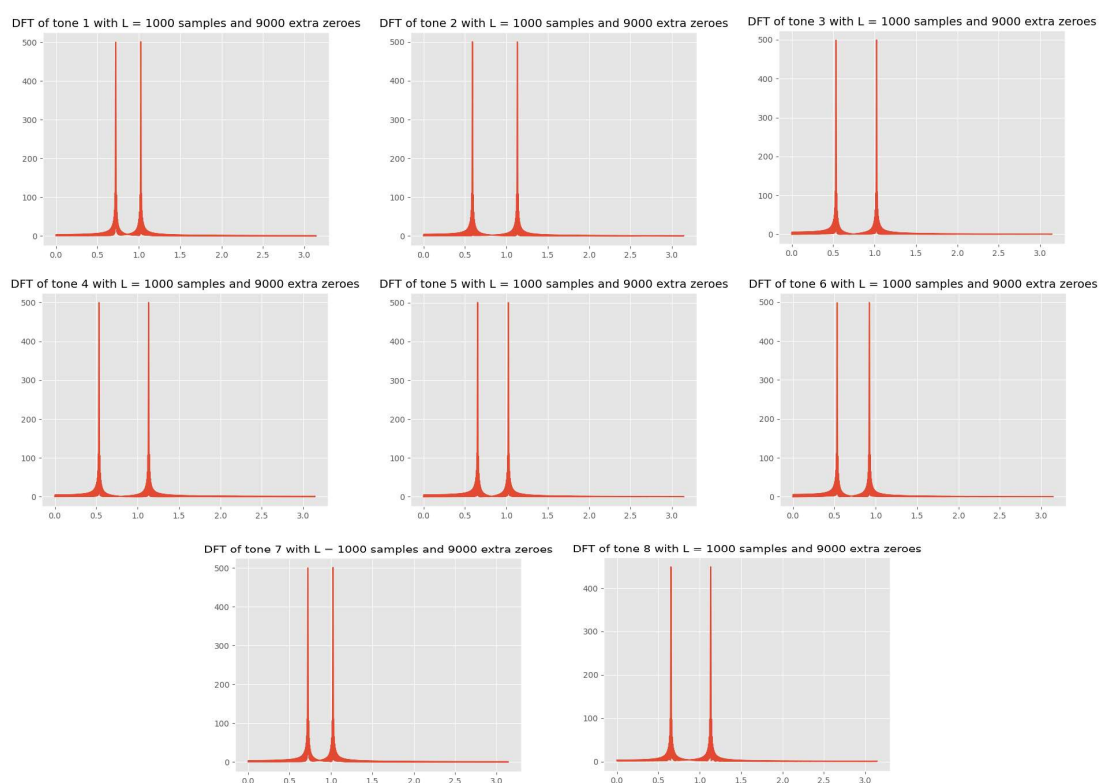
Εικόνα 5 Διακριτός Μετασχηματισμός Fourier του τόνου 8

2.3

Στο σημείο αυτό δημιουργείται μία ακολουθία των παραπάνω τόνων απεχόντων μεταξύ τους κατά 100 μηδενικά δείγματα. Η ακολουθία προκύπτει από το άθροισμα των αριθμών μητρώου των μελών της παρούσας ομάδας, δηλαδή $03119051 + 03119058 = 06238109$. Άρα, ο πρώτος τόνος της ακολουθίας είναι ο υπ' αριθμόν 0 με 1000 δείγματα και 100 μηδενικά δείγματα. Ο δεύτερος είναι ο υπ' αριθμόν 6 με 1000 δείγματα και 100 μηδενικά δείγματα και ούτω καθεξής. Στην συγκεκριμένη υλοποίηση της ακολουθίας, δεν παρεμβλήθηκαν 100 μηδενικά δείγματα στο τέλος του τελευταίου τόνου.

2.4

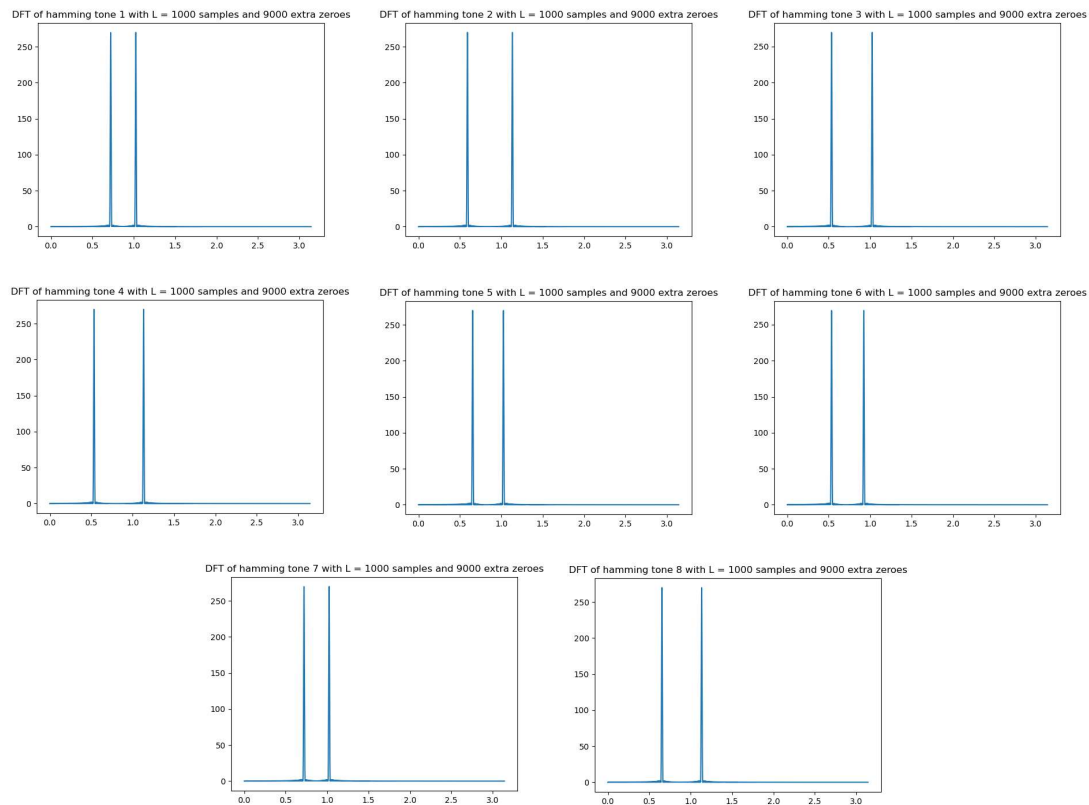
Τώρα, η δημιουργηθείσα ακολουθία θα διασπαστεί στα επί μέρους σήματά της χρήσει καταλλήλων χρονικών παραθύρων και θα υπολογιστεί ο Διακριτός Μετασχηματισμός Fourier για καθένα παραθυρωμένο σήμα. Η διαδικασία θα επαναληφθεί δύο φορές, μία για ιδανικό τετραγωνικό παράθυρο και μία για παράθυρο hamming. Αφού απομονώθηκαν, χρήσει παραθύρου, οι τόνοι, παρεμβλήθηκαν 9000 μηδενικά στο τέλος τους προς υψηλότερη ευκρίνεια των γραφημάτων. Παρακάτω βρίσκονται τα αποτελέσματα¹:



Εικόνα 6 Τα φάσματα των απομονωμένων τόνων, χρήσει ιδανικού τετραγωνικού παραθύρου, της δημιουργηθείσας ακολουθίας.

¹ Σημειώνεται ότι στον τίτλο καθενός γραφήματος αριθμείται η σειρά του τόνου στην ακολουθία και όχι ο αύξων αριθμός του τόνου. Παραδείγματος χάριν, ο τόνος 1 είναι εν προκειμένω ο d_0 , όχι ο d_1 , και ο τόνος 2 είναι ο d_6 , όχι ο d_2

Ομοίως, χρήσει του παραθύρου hamming:



Εικόνα 7 Τα φάσματα των απομονωμένων τόνων, χρήσει hamming παραθύρου, της δημιουργηθείσας ακολουθίας.

Όπως και στο [πείραμα 1.4](#), αξιοσημείωτη είναι η σαφώς υψηλότερη διακριτική ικανότητα των φασμάτων όταν αυτά δημιουργούνται χρήσει του παραθύρου hamming, σε σχέση με τα αντίστοιχα του τετραγωνικού παραθύρου. Αξιοσημείωτη είναι επίσης η διαφορά στο ύψος των κορυφών μεταξύ των φασμάτων των hamming παραθυρωμένων τόνων και των τετραγωνικώς παραθυρωμένων τόνων. Συγκεκριμένα, τα πρώτα έχουν περίπου το μισό ύψος των δευτέρων. Αυτό συμβαίνει διότι το τετραγωνικό παράθυρο έχει μέση τιμή 1, άρα δεν αλλοιώνει το πλάτος των δειγμάτων επί των οποίων εφαρμόζεται, ενώ το παράθυρο hamming

$$w_{\text{hamm}}[n] = 0.54 - 0.46 \cdot \cos(2 \cdot \pi \cdot n \cdot N) \approx \frac{1}{2} \cdot (1 - \cos(2 \cdot \pi \cdot n \cdot N))$$

έχει μέση τιμή περίπου $\frac{1}{2}$, με αποτέλεσμα να υποδιπλασιάζει το πλάτος των δειγμάτων επί των οποίων εφαρμόζεται.

2.5

Για τον υπολογισμό των δεικτών k και των αντιστοιχών συχνοτήτων εγγυτέρων στις touch-tone συχνοτήτες, χρησιμοποιείται ο εξής τύπος:

$$\omega_{\text{διακριτό}} = \frac{2 \cdot \pi \cdot k}{N}$$

όπου N το πλήθος των δειγμάτων του Διακριτού Μετασχηματισμού Fourier.

Άρα, οι δείκτες k υπολογίζονται ως:

$$k = \frac{\omega_{\text{διακριτό}} \cdot N}{2 \cdot \pi}$$

Εάν θέσουμε ως τιμή του N την συχνότητα δειγματοληψίας 8192Hz, τότε λαμβάνονται οι πραγματικές συχνοτήτες σε Hz:

$$k_{\text{γραμμών}} = [697, 770, 852, 940]$$

$$k_{\text{στηλών}} = [1204, 1336, 1476]$$

Οι δείκτες αυτοί έχουν αποκτηθεί διά προσεγγίσεως στον κοντινότερο ακέραιο.

2.6

Για την υλοποίηση της συνάρτησης `ttdecode()` χρησιμοποιούνται δύο βοηθητικές συναρτήσεις: η συνάρτηση `find_digit(frequencies, row_freqs, col_freqs)` η οποία δέχεται ένα διάνυσμα συχνοτήτων `frequencies` όπου σημειώνονται οι κορυφές υπό εξέταση φάσμα και δύο διανύσματα δεικτών συχνοτήτων, γραμμών και στηλών, βάσει των οποίων θα αντιστοιχήσει τις συχνοτήτες του διανύσματος σε έναν από τους δεδομένους δείκτες. Η συνάρτηση επιστρέφει έναν ακέραιο αριθμό, τον αύξοντα αριθμό του ανιχνευμένου τόνου.

```
def find_digit(frequencies, row_freqs, col_freqs):
    row_ind = -1
    col_ind = -1
    error_margin = 0.02 #acceptable error
    for freq in frequencies:
        if freq < (row_freqs[0]+col_freqs[2])/2:
            for index, row in enumerate(row_freqs, start=0):
                if abs(row-freq)/row < error_margin:
                    row_ind = index
        else:
            for index, col in enumerate(col_freqs, start=0):
                if abs(col-freq)/col < error_margin:
                    col_ind = index
    if row_ind == -1 or col_ind == -1:
        return -1
    if row_ind == 3 and col_ind == 1:
        return 0
    return row_ind*3+col_ind+1
```

Χάριν ευκολίας, χρησιμοποιείται επίσης η συνάρτηση `mean_energy(signal)` για τον υπολογισμό της μέσης ενέργειας (ενέργεια ανά δείγμα) του σήματος εισόδου. Το μέγεθος αυτό είναι αναγκαίο για να καθοριστεί αν η εξεταζόμενη περιοχή του σήματος εισόδου της `ttdecode` είναι πράγματι τόνος ή θόρυβος / σιγή.

```
def mean_energy(signal):
    return sum([x*x for x in signal])/np.size(signal)
```

Η ίδια η συνάρτηση αποκωδικοποίησης χρησιμοποιεί ένα κυλιόμενο παράθυρο hamming επάνω στο σήμα και, εφόσον αποφασιστεί ότι η υπό εξέταση περιοχή είναι τόνος, ανιχνεύει τον τόνο και συνεχίζει την κύλιση. Έχει γίνει η παραδοχή ότι κάθε τόνος ακολουθείται από σιγή οποιουδήποτε μεγέθους. Συγκεκριμένα:

- Ορίζεται το μέγεθος του παραθύρου, το βήμα του και το πλήθος των προσθέτων μηδενικών δειγμάτων προς ευκρίνεια του φάσματος. Το άθροισμα αυτών των μεγεθών είναι το μέγεθος του Διακριτού Μετασχηματισμού Fourier που θα χρησιμοποιηθεί. Βάσει αυτού του μεγέθους αποδίδονται δείκτες στις συχνότητες εισόδου, οι οποίοι θα χρησιμοποιηθούν για την αντιστοίχιση με τις διακριτές συχνότητες.
- Εισέρχεται σε έναν επαναληπτικό βρόχο, μέσα στον οποίον υπολογίζει την ενέργεια ανά δείγμα του εξεταζόμενης περιοχής του σήματος και ελέγχει αν είναι πράγματι τόνος ή όχι.
 - Αν είναι, τότε εκτελεί τον Διακριτό Μετασχηματισμό Fourier, εντοπίζει σε αυτόν τα σημεία όπου σημειώνονται κορυφές, αντιστοιχεί αυτά τα σημεία σε έναν αριθμό τόνου και παραθέτει τον αριθμό στο αποτέλεσμα. Όσο δεν εντοπίζεται σιγή, κυλάει το παράθυρο, υπολογίζει την ενέργεια του σήματος εντός αυτού και συνεχίζει τους ελέγχους.
 - Αν δεν είναι τόνος, τότε, ομοίως, κυλάει το παράθυρο, υπολογίζει την νέα ενέργεια και συνεχίζει τους ελέγχους.

```
def ttdecode(signIn, row_freqs = wrow , col_freqs = wcol):
    vector = '' #initialisation of the result
    num = 100 #number of windows that 'fit' into the signal
    length = np.size(signIn)
    width = int(length/num) #width of the window, 1/100 of signal
    zero_pads = 20*width #extra zeroes to ensure good quality of DFT
    window = np.hamming(width)
    k_index_row = [int(x) for x in row_freqs*(width+zero_pads)/(2*np.pi)]
    k_index_col = [int(x) for x in col_freqs*(width+zero_pads)/(2*np.pi)]
    #indexes for the given frequencies
    step = max(int(width/10),1) #size of window step
    threshold = mean_energy(signIn) #threshold to qualify as a tone
    index = 0
    tone = []
    tone_fft = []
    peak_freqs = []
    window_energy=0
    while index + width <= length:
        window_energy = mean_energy(signIn[index:index+width])
        if window_energy < threshold: #not recognised as a tone
            while window_energy < threshold and index+width+step <= length:
                #still not on a tone
                index += step
                window_energy = mean_energy(signIn[index:index+width])
                #keep looping until a tone is found

            if window_energy >= threshold: #recognised a tone
                tone = signIn[index:index+width] #isolate the tone
                tone_fft = np.fft.fft(np.append(window*tone,np.zeros(zero_pads)))
                #perform the DFT on the zero-padded windowed signal
                tone_fft = abs(tone_fft[0:int((width+zero_pads)/2)])
                #keep the absolute value of the first half of the signal
                peak_freqs, _ = find_peaks(tone_fft, distance = 10, height =
0.5*max(abs(tone_fft))) #find the samples where the peaks occur
                digit = find_digit(peak_freqs, k_index_row, k_index_col)
                vector += str(digit) + ' '
                while window_energy >= threshold and index+width+step <= length:
                    #still on the same tone
                    index += step
                    window_energy = mean_energy(signIn[index:index+width])
                    #keep looping until noise or gap is found
                index+=step
    return vector
```

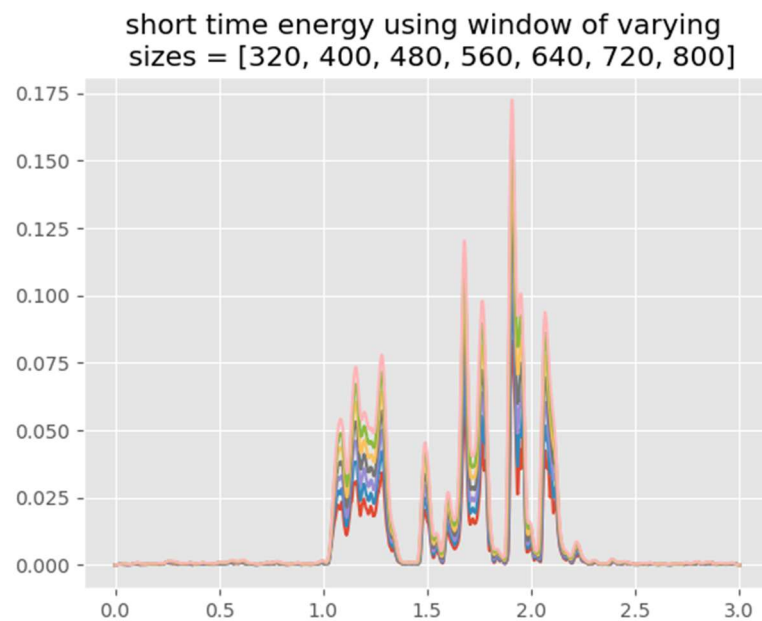
2.7

Οι τόνοι των 2 σημάτων όπως ανιχνεύθηκαν από την συνάρτηση `ttdecode` είναι οι:

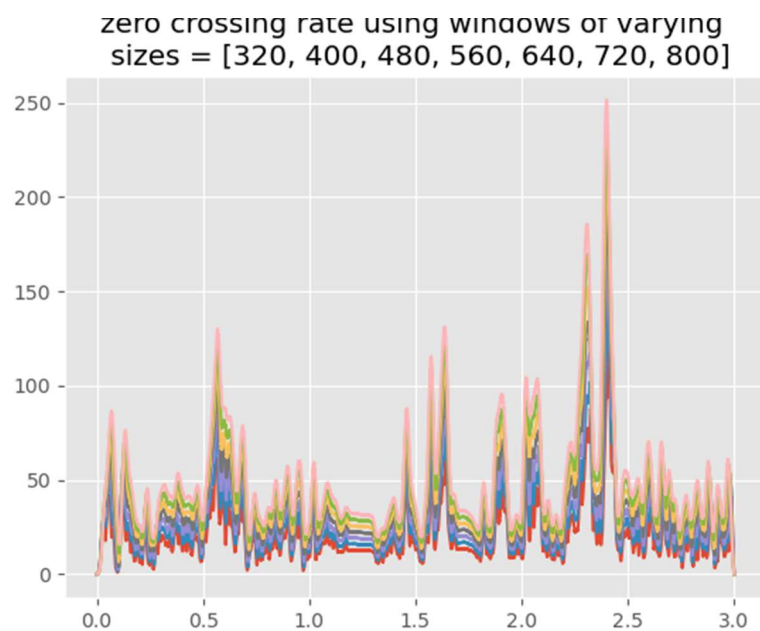
- `easySig` = 7 3 5 8 2 8 0 2.
- `hardSig` = 4 3 5 6 6 2 0 9 9 5.

3.1

Στο πείραμα αυτό θα μετρηθεί η ενέργεια βραχέος χρόνου και ο ρυθμός εναλλαγής προσήμου ενός σήματος φωνής. Οι μετρήσεις γίνονται χρήσει διαφόρων μηκών παραθύρων, μεταξύ 20-50 ms. Τα παράθυρα αυτά ολισθαίνουν επάνω στο σήμα με βήμα ενός δείγματος και κάνουν τις μετρήσεις εντός του εύρους τους. Παρατίθενται πρώτα τα διαγράμματα ενέργειας βραχέος χρόνου και ρυθμού εναλλαγής προσήμου για διάφορα μεγέθη παραθύρων. Ο οριζόντιος άξονας είναι αριθμημένος σε δευτερόλεπτα.



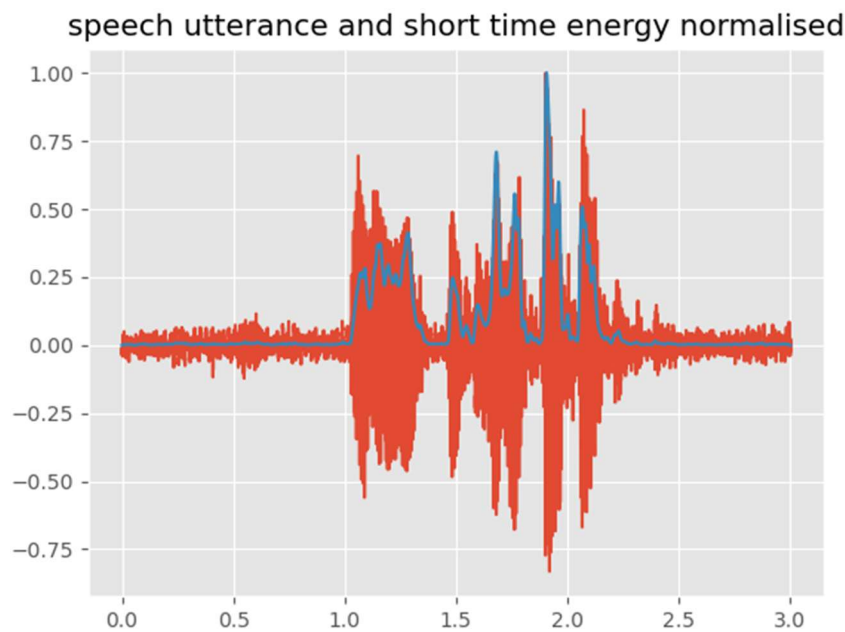
Εικόνα 8 Διαγράμματα Ενέργειας Βραχέος Χρόνου σήματος φωνής χρήσει παραθύρων διαφόρων μηκών.



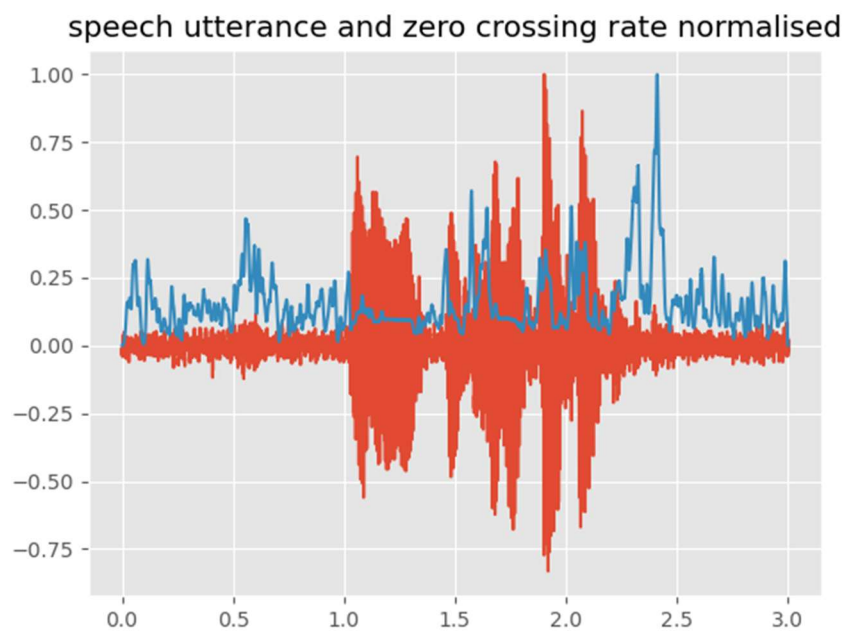
Εικόνα 9 Διαγράμματα Ρυθμού Εναλλαγής Προσήμου σήματος φωνής για παράθυρα διαφόρων μηκών.

Παρατηρείται ότι, ποιοτικά, η μορφή των διαγραμμάτων ενέργειας βραχέος χρόνου είναι πανομοιότυπη ανεξαρτήτως μεγέθους παραθύρου. Το ίδιο ισχύει και για τα διαγράμματα ρυθμού εναλλαγής προσήμου. Η μόνη αλλαγή είναι στο πλάτος. Συγκεκριμένα, διαπιστώνεται ότι, όσο μεγαλύτερο το μήκος του χρησιμοποιούμενου παραθύρου τόσο μεγαλύτερο το πλάτος της ενέργειας βραχέος χρόνου και του ρυθμού εναλλαγής προσήμου.

Παρατίθενται τα ίδια μεγέθη σχεδιασμένα επάνω στο αρχικό σήμα φωνής. Στα ακόλουθα διαγράμματα όλα τα σήματα είναι κανονικοποιημένα. Σημειώνεται ότι, ακριβώς επειδή είναι κανονικοποιημένα, δεν έχει σημασία, εν προκειμένω, το μήκος του παραθύρου, αφού επηρεάζει μονάχα το πλάτος και όχι την μορφή του σχεδίου.



Εικόνα 10 Κανονικοποιημένες απεικονίσεις σήματος φωνής και ενέργειας βραχέος χρόνου



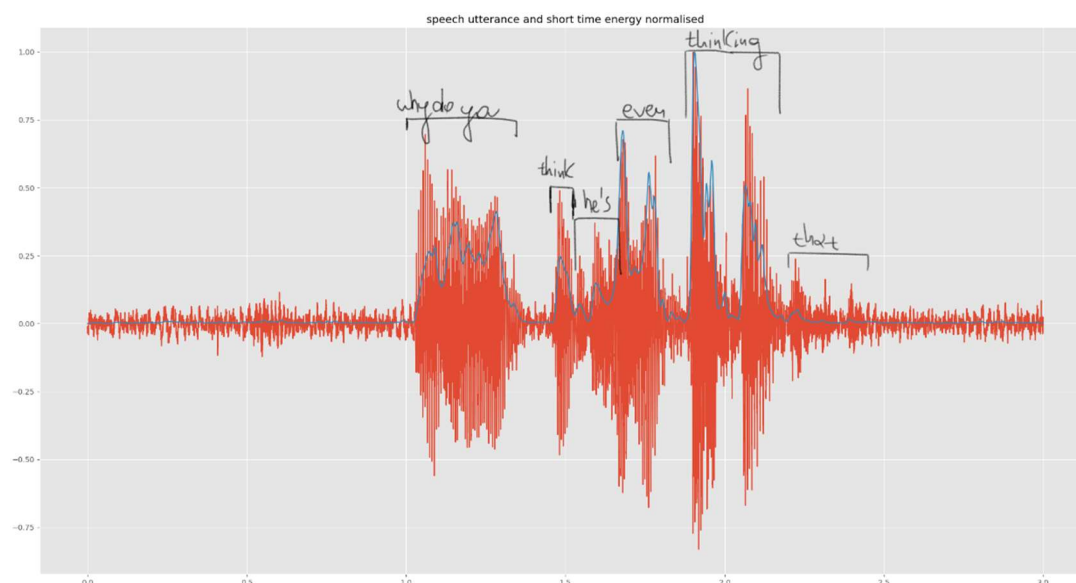
Εικόνα 11 Κανονικοποιημένες απεικονίσεις σήματος φωνής και ρυθμού εναλλαγής προσήμου

Με μία πρώτη εποπτεία διαπιστώνεται ότι η ενέργεια ακολουθεί το σήμα φωνής. Δηλαδή, όπου στο σήμα φωνής αυξάνεται το πλάτος, αυξάνεται και η ενέργεια. Το πλάτος στο σήμα φωνής αυξάνεται, γενικώς, όταν υπάρχει ομιλία και ιδιαιτέρως όταν ακούγονται έμφωνοι ήχοι (όπως το 'α' και το 'ε' στην ελληνική γλώσσα). Μάλιστα, σημασία έχει και η έμφαση με την οποία εκφέρονται, η οποία επηρεάζει αισθητά την αποτυπωμένη ενέργεια.

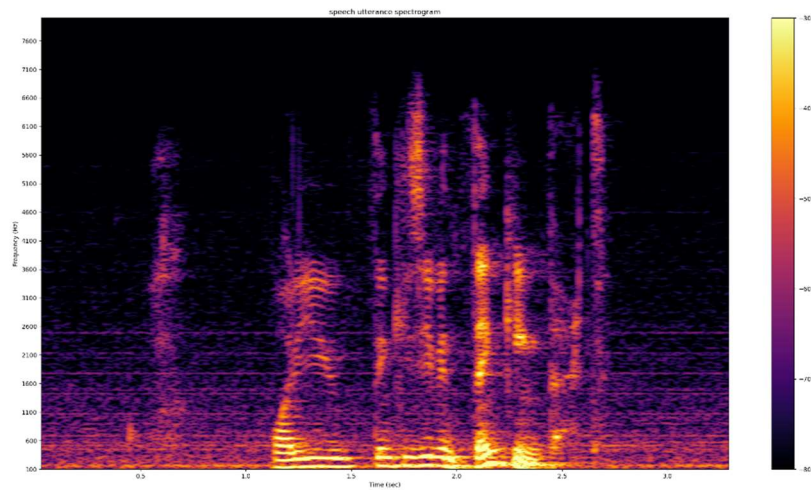
Παραδείγματος χάριν, οι δύο τελευταίες υψηλές κορυφές της κυματομορφής είναι οι συλλαβές της λέξης "thinking", συγκεκριμένα τα φωνήεντα /i/, ενώ οι ακόλουθες τρεις μικρές κορυφές είναι η λέξη "that". Ωστόσο, η λέξη "thinking" έχει σημαντικό ενεργειακό αποτύπωμα ενώ η "that" μετά βίας φαίνεται. Ο λόγος για τον οποίον συμβαίνει αυτό είναι ότι, ενώ η μεν πρώτη εκφέρεται με έμφαση, η δε δεύτερη εκφέρεται με μηδαμινή ένταση, σχεδόν σαν απλή εκπνοή. Οι τρεις τελευταίες κορυφές είναι αντιστοίχως η συλλαβή 'tha-', η διάρκεια της με μία αχνή ανάσα '-a' και στο τέλος το δασύ σύμφωνο /t^h/² '-t'.

Σχετικά με τον ρυθμό εναλλαγής προσήμου, οι διαπιστώσεις φαίνονται να αντιστρέφονται. Συγκεκριμένα, παρατηρείται ότι, όπου το πλάτος του σήματος φωνής αυξάνεται, ο ρυθμός εναλλαγής προσήμου μειώνεται. Άρα, γενικώς κατά την διάρκεια σιγής, ο ρυθμός εναλλαγής προσήμου φαίνεται να λαμβάνει υψηλές τιμές οι οποίες μειώνονται όταν αρχίζει η ομιλία. Έτσι εξακριβώνεται μάλιστα η ύπαρξη των συμφώνων στην ομιλία (αύξηση του ρυθμού).

Η πρόταση επάνω στην απεικόνιση του σήματος φωνής μοιάζει κάπως έτσι:



² Συμβολισμός Διεθνούς Φωνητικού Αλφαβήτου.

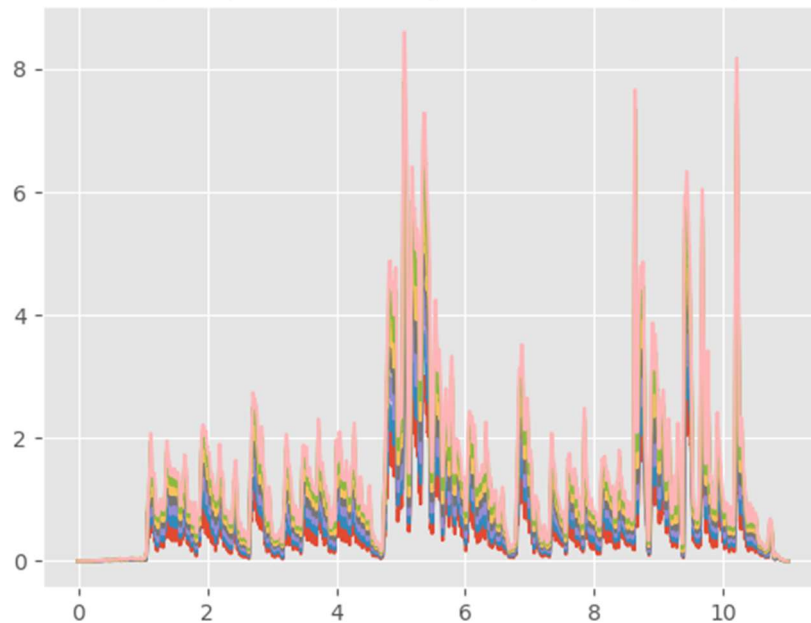


Παρατίθεται και το φασματογράφημα του σήματος `speech_utterance` προς επαλήθευση των ανωτέρω διαπιστώσεων. Όπως φαίνεται, οι υψηλότερες κορυφές του διαγράμματος ενέργειας αντιστοιχούν σε εντονότερο χρωματισμό του φασματογραφήματος και αυτό συμβαίνει όταν στην ομιλία εκφέρονται έμφωνοι ήχοι. Σημειώνεται, ότι υψηλό συχνοτικό περιεχόμενο δεν παρουσιάζουν μόνο οι ήχοι αυτοί, αλλά και οι άφωνοι ήχοι, μολονότι με μικρότερη ένταση, όπως φαίνεται, παραδείγματος χάριν, από την αποτύπωση του δασέος /t^h/ στο τέλος του φασματογραφήματος.

3.2

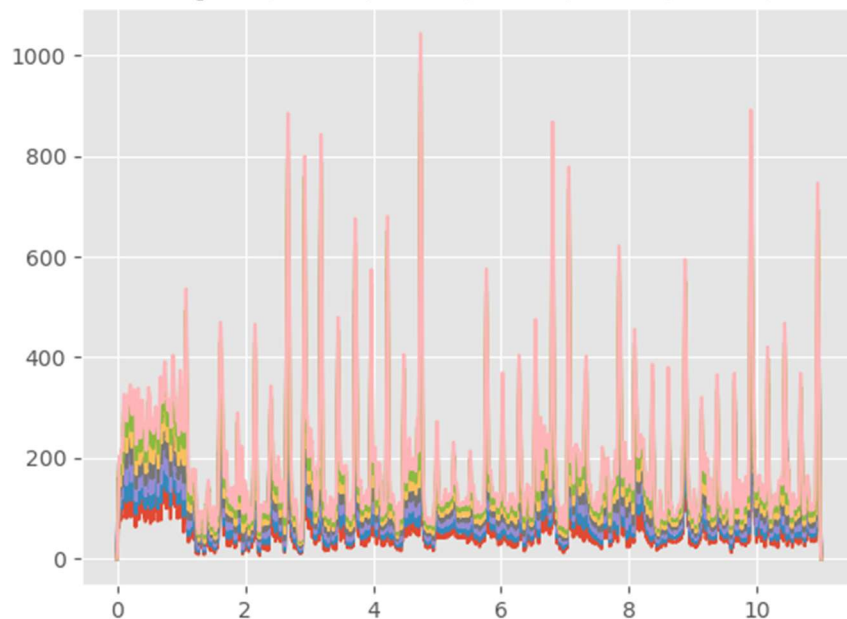
Επαναλαμβάνεται η ίδια διαδικασία, αυτήν την φορά σε ένα μουσικό σήμα στέρεο. Ωστόσο, απαιτείται πρώτα η μετατροπή του σε μονό. Η μετατροπή αυτή γίνεται αθροίζοντας τα δύο επί μέρους σήματα στέρεο σε ένα ενιαίο και λαμβάνοντας τον μέσον όρο. Παρατίθενται οι ίδιες ομάδες διαγραμμάτων όπως πάνω:

short time energy using window of varying
sizes = [960, 1200, 1440, 1680, 1920, 2160, 2400]



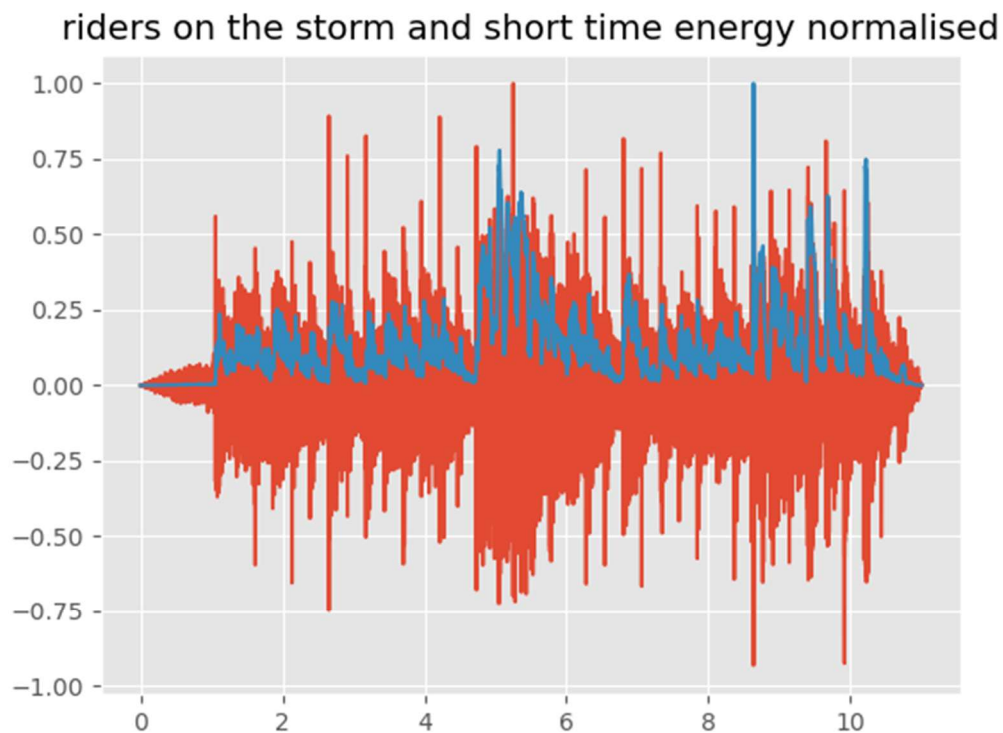
Εικόνα 12 Διαγράμματα Ενέργειας Βραχέος Χρόνου μουσικού σήματος χρήσει παραθύρων διαφόρων μηκών

zero crossing rate using windows of varying
sizes = [960, 1200, 1440, 1680, 1920, 2160, 2400]

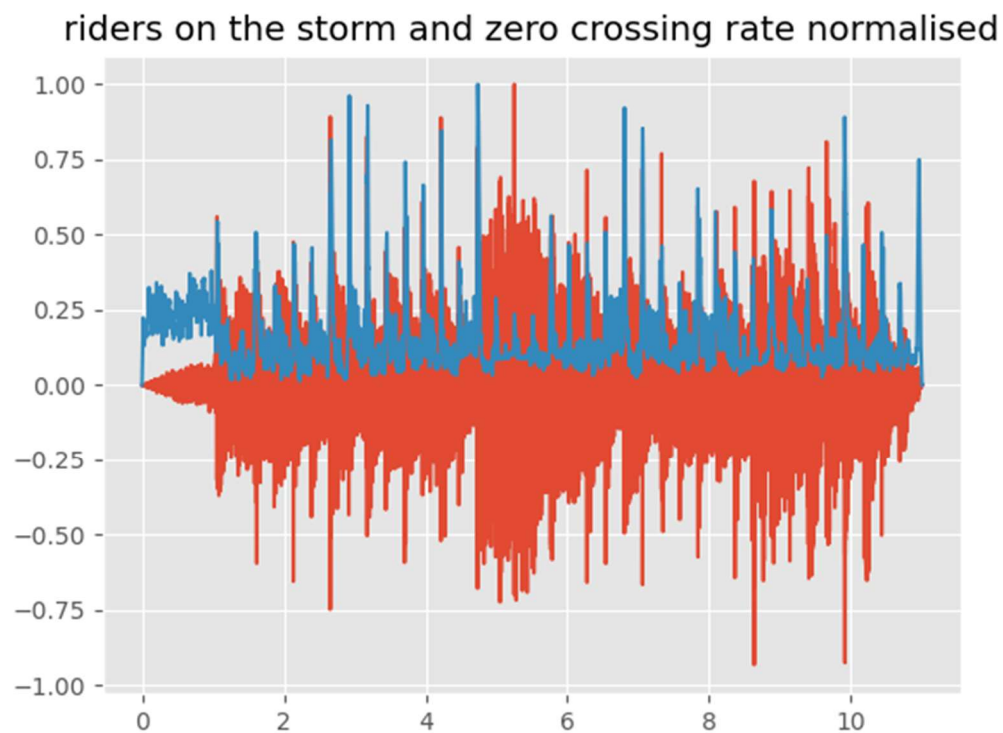


Εικόνα 13 Διαγράμματα Ρυθμού Εναλλαγής Προσήμου μουσικού σήματος χρήσει παραθύρων διαφόρων μηκών

Πάλι διαπιστώνεται ότι το μήκος των παραθύρων επηρεάζει μονάχα το πλάτος των υπολογιζόμενων μεγεθών και όχι την ποιοτική μορφή των διαγραμμάτων. Ακολουθούν τα αντίστοιχα διαγράμματα με τις απεικονίσεις των μεγεθών επάνω στο αρχικό σήμα φωνής.



Εικόνα 14 Κανονικοποιημένες απεικονίσεις μουσικού σήματος και ενέργειας βραχέος χρόνου.



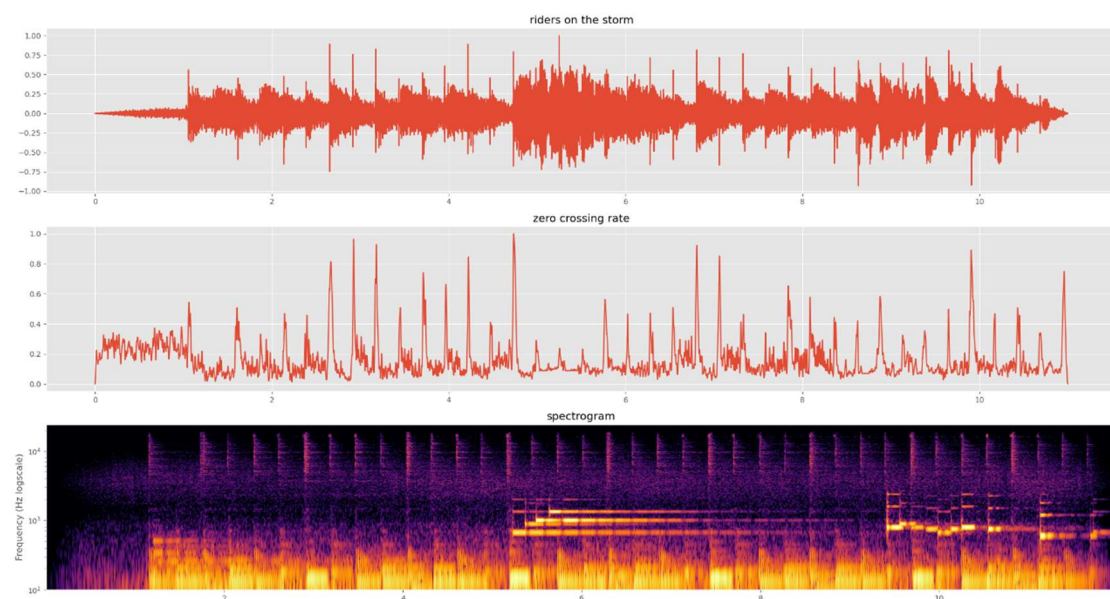
Εικόνα 15 Κανονικοποιημένες απεικονίσεις μουσικού σήματος και ρυθμού εναλλαγής προσήμου

Στο συγκεκριμένο παράδειγμα μουσικής, εφόσον δεν υπάρχει ομιλία, οι μετρήσεις αξιοποιούνται με διαφορετικόν τρόπο.

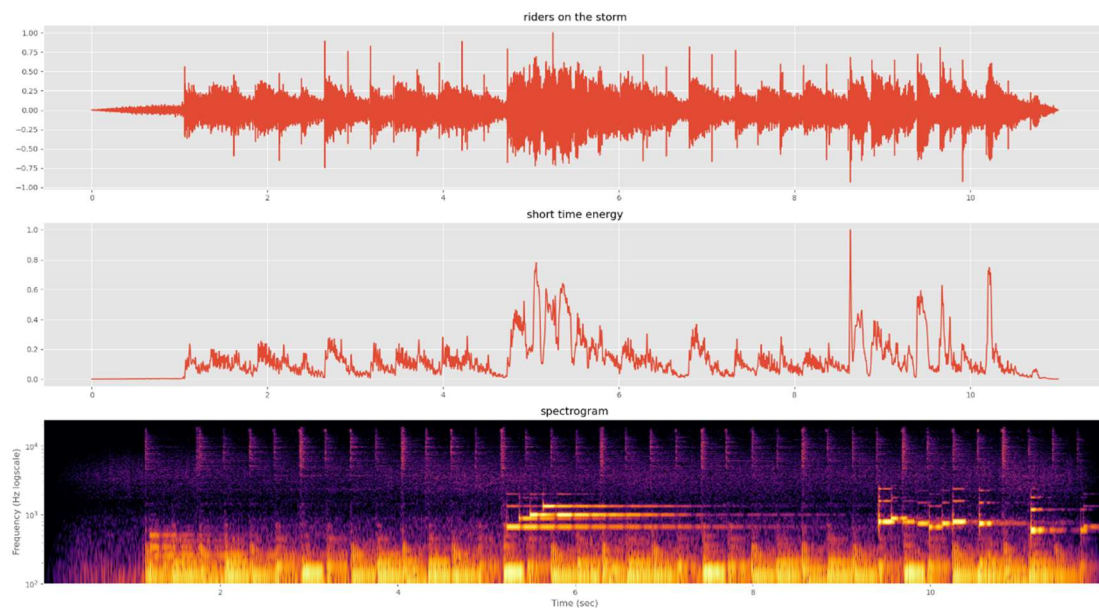
Κατ' αρχάς, κατά την θορυβώδη, «βροχερή» εισαγωγή, επαληθεύονται οι προηγούμενες διαπιστώσεις ότι η ενέργεια βραχέος χρόνου είναι ελάχιστη ενώ ο ρυθμός εναλλαγής προσήμου είναι υψηλός. Κατά την έναρξη της μουσικής όμως, παρατηρείται ότι οι λοφίσκοι στο διάγραμμα ενέργειας ακολουθούν κατά κάποιον τρόπο τις νότες της μελωδίας. Συγκεκριμένα, αποτυπώνονται οι νότες του μπάσου (1^ο έως και μεταξύ 4^{ου} και 5^{ου} δευτερολέπτου), ενώ στην συνέχεια αποτυπώνονται με ευκρίνεια οι νότες των πλήκτρων (περίπου 5^ο έως 6^ο δευτερόλεπτο), ξανά το μπάσο (6^ο έως 8^ο δευτερόλεπτο) και τέλος πάλι οι νότες των πλήκτρων. Να σημειωθεί ότι, με τον όρο νότες δεν εννοείται το ακριβές περιεχόμενο τους αλλά ουσιαστικά η ύπαρξή τους.

Το δε διάγραμμα ρυθμού εναλλαγής προσήμου πάλι λειτουργεί με αντίστροφο τρόπο. Ειδικότερα, φαίνεται να αποτυπώνει πολύ πιστά τα χτυπήματα των πιατινιών με απότομες υψηλές κορυφές, ενώ οι τιμές του πέφτουν όταν αρχίζουν να κυριαρχούν νότες οι οποίες επικαλύπτουν τα πιατίνια.

Οι παρατηρήσεις αυτές επαληθεύονται περαιτέρω αν εξετάσει κανείς το φασματογράφημα³ του σήματος μουσικής:



³ Μέσω της συνάρτησης `specgram` της βιβλιοθήκης `matplotlib` της `Python`.



Πράγματι, φαίνεται ότι όταν αρχίζουν να απεικονίζονται οι υψηλές συχνότητες (λόγω των πλήκτρων) στο φασματογράφημα, η ενέργεια βραχέος χρόνου αυξάνεται ενώ ο ρυθμός εναλλαγής προσήμου μειώνεται. Επιπλέον, φαίνεται η αντιστοίχιση των υψηλών και στιγμιαίων συχνοτήτων των πιατινιών με τα αντίστοιχα υψηλά και απότομα καρφιά του ρυθμού εναλλαγής προσήμου.