



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών

Ψηφιακή Επεξεργασία Σήματος

1^η εργαστηριακή άσκηση

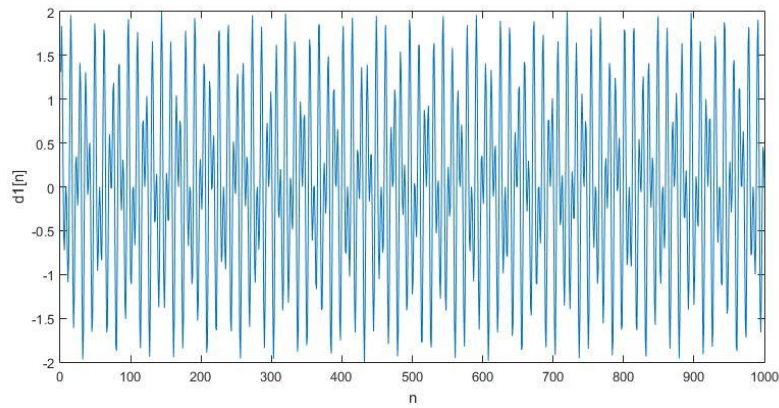
Βασιλοπούλου Φωτεινή ΑΜ: 03114854

Λαζαρίδης Κωνσταντίνος ΑΜ: 03114939

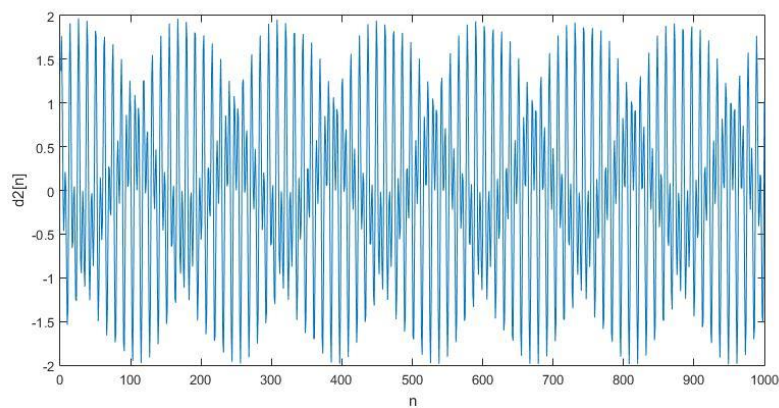
Ακαδημαϊκό έτος 2018-2019

Μέρος 1.

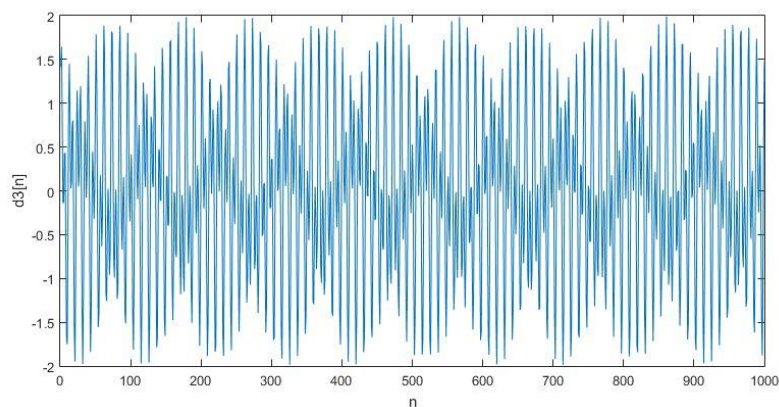
1.1. Αρχικά δημιουργούμε, σύμφωνα με τον πίνακα που δίνεται, τους 10 διαφορετικούς τηλεφωνικούς τόνους. Τα σήματα αυτά φαίνονται στις εικόνες που ακολουθούν. Η συχνότητα δειγματοληψίας των σημάτων είναι $f_s=8192Hz$ και αποτελούνται από 1000 δείγματα.



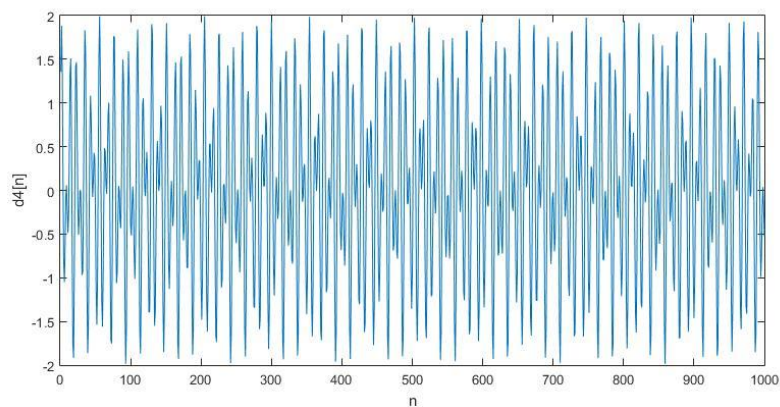
Εικόνα 1: Γραφική παράσταση του σήματος που παράγεται με το πάτημα του πλήκτρου 1.



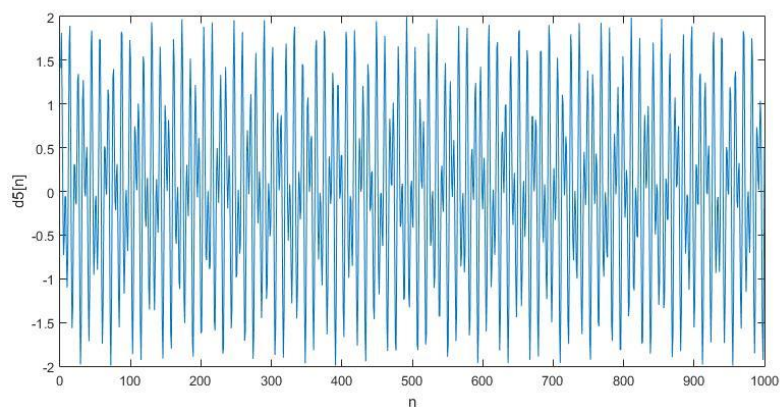
Εικόνα 2: Γραφική παράσταση του σήματος που παράγεται με το πάτημα του πλήκτρου 2.



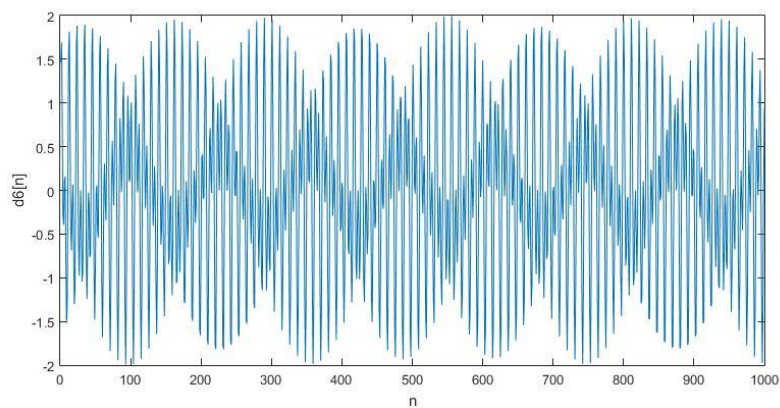
Εικόνα 3: Γραφική παράσταση του σήματος που παράγεται με το πάτημα του πλήκτρου 3.



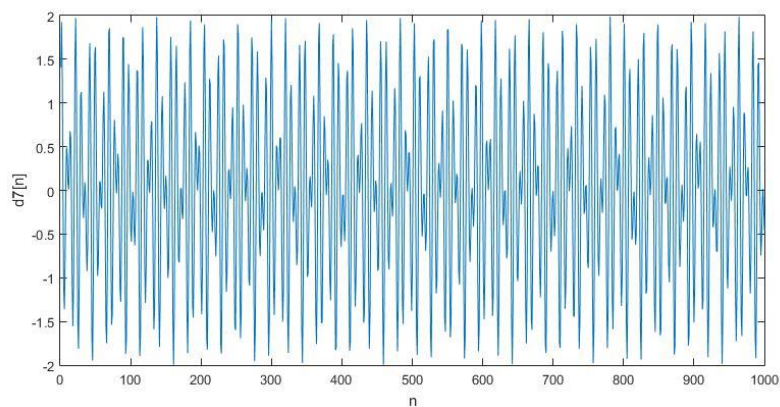
Εικόνα 4: Γραφική παράσταση του σήματος που παράγεται με το πάτημα του πλήκτρου 4.



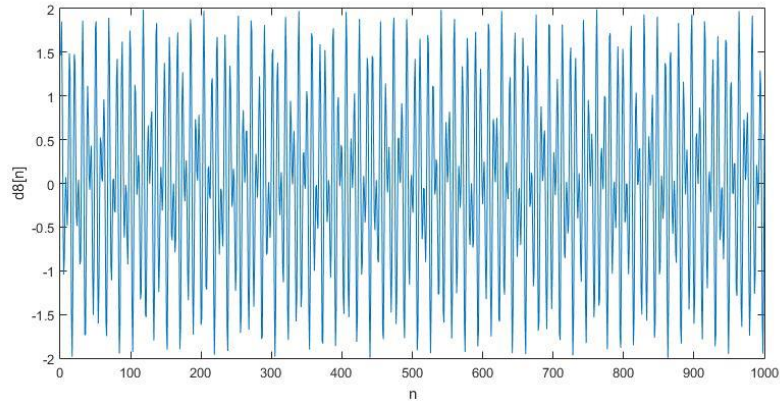
Εικόνα 5: Γραφική παράσταση του σήματος που παράγεται με το πάτημα του πλήκτρου 5.



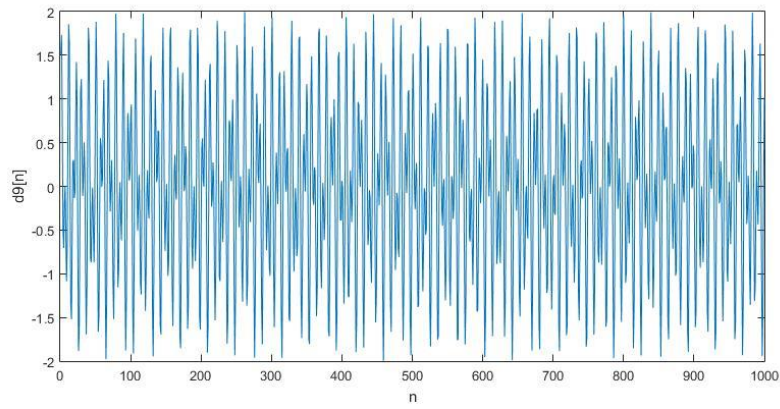
Εικόνα 6: Γραφική παράσταση του σήματος που παράγεται με το πάτημα του πλήκτρου 6.



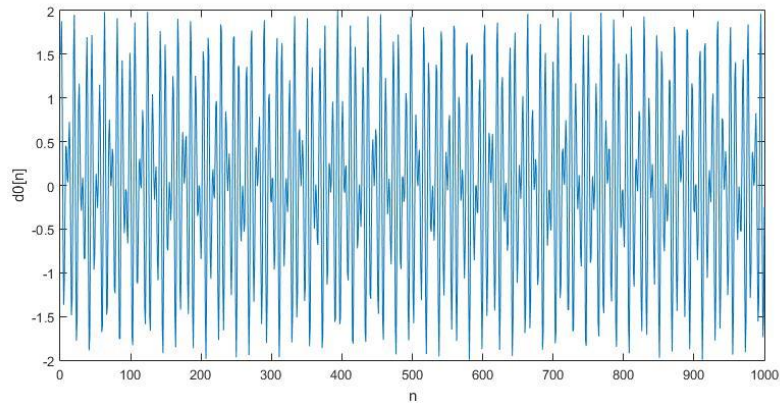
Εικόνα 7: Γραφική παράσταση του σήματος που παράγεται με το πάτημα του πλήκτρου 7.



Εικόνα 8: Γραφική παράσταση του σήματος που παράγεται με το πάτημα του πλήκτρου 8.

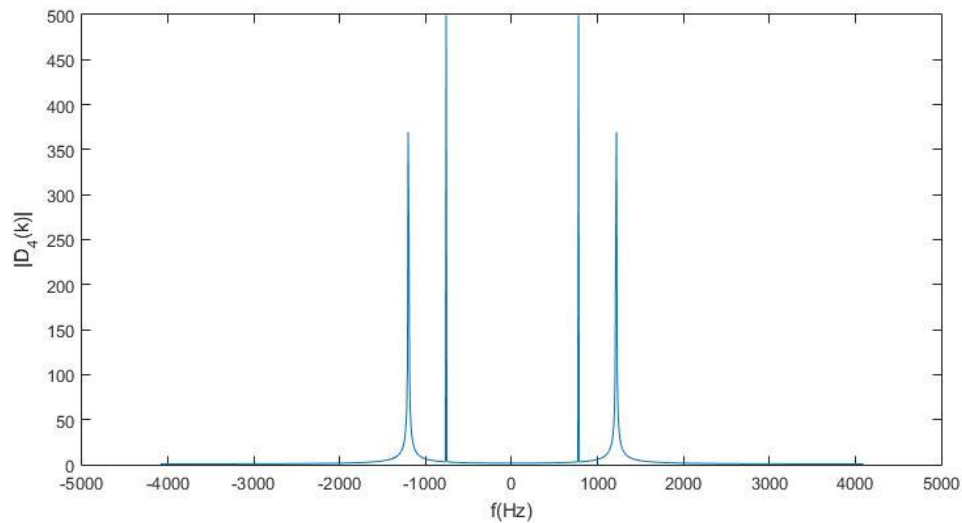


Εικόνα 9: Γραφική παράσταση του σήματος που παράγεται με το πάτημα του πλήκτρου 9.

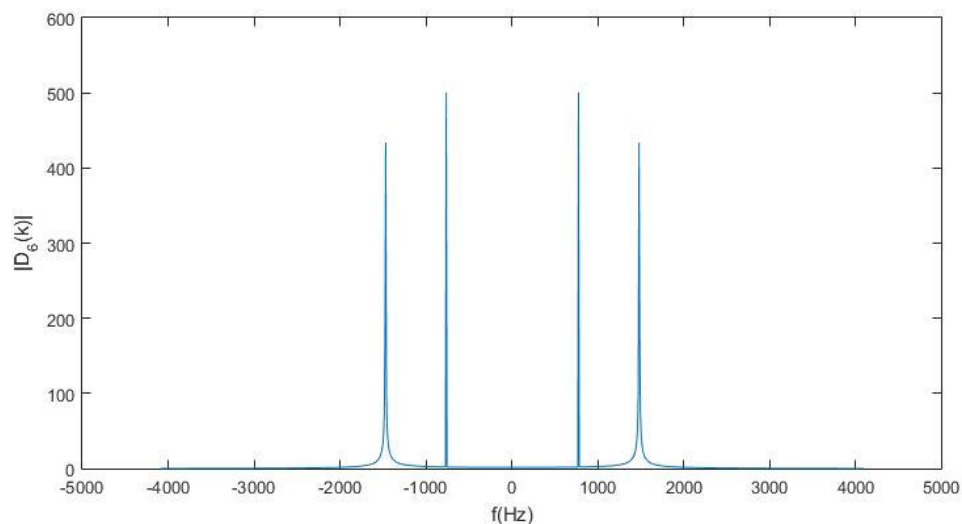


Εικόνα 10: Γραφική παράσταση του σήματος που παράγεται με το πάτημα του πλήκτρου 0.

2.2. Χρησιμοποιώντας την συνάρτηση $\text{fft}()$ του *Matlab* υπολογίζουμε τον DFT των σημάτων $d_4[n]$ και $d_6[n]$. Οι ζητούμενες γραφικές παραστάσεις φαίνονται στη συνέχεια.



Εικόνα 11: Γραφική παράσταση του Διακριτού Μετασχηματισμού Fourier του σήματος $d_4[n]$.

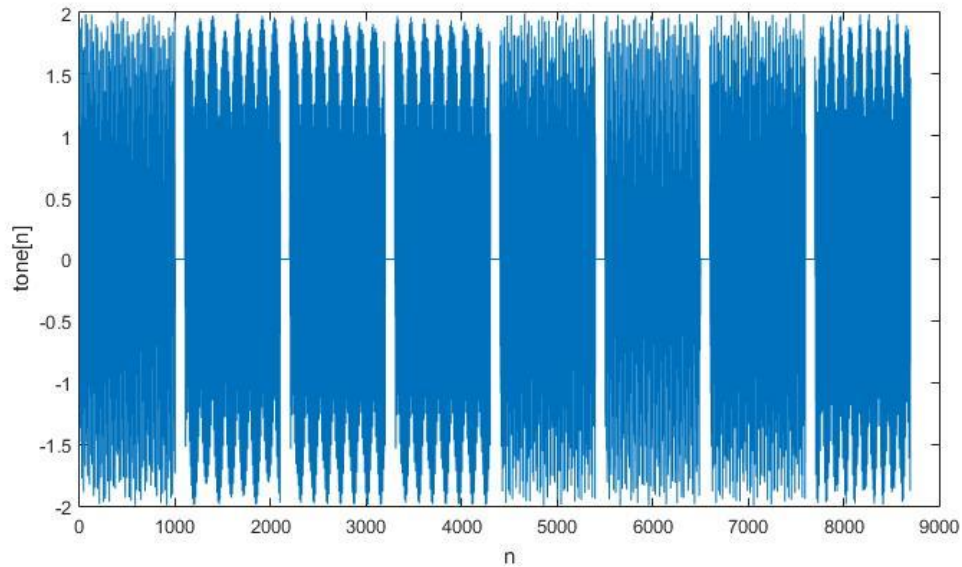


Εικόνα 12: Γραφική παράσταση του Διακριτού Μετασχηματισμού Fourier του σήματος $d_6[n]$.

1.3. Στο ερώτημα αυτό στόχος είναι η δημιουργία ενός σήματος που περιέχει διαδοχικούς τηλεφωνικούς τόνους που αντιστοιχούν στο άθροισμα των αριθμών μητρώου, δηλαδή:

$$03114854 + 03114939 = 06229793$$

Κάθε τόνος έχει διάρκεια 1000 δειγμάτα όπως ορίστηκε σε προηγούμενο ερώτημα και οι διαδοχικοί τόνοι διαχωρίζονται μεταξύ τους με διαστήματα 100 μηδενικών δειγμάτων. Το τελικό σήμα είναι αποθηκευμένο στο αρχείο “*tone_sequence.wav*”.



Εικόνα 13: Γραφική παράσταση του σήματος που προέκυψε από το άθροισμα των αριθμών μητρώου.

1.4-5. Δημιουργούμε τετραγωνικά και *Hamming* παράθυρα μήκους 1000 δειγμάτων και υπολογίζουμε τον μετασχηματισμό *Fourier* του σήματος που δημιουργήσαμε στο προηγούμενο ερώτημα. Στη συνέχεια, για κάθε ένα από τα διαστήματα που αντιστοιχούν σε τηλεφωνικό τόνο, υπολογίζουμε τις συχνότητες στις οποίες εντοπίζεται μεγάλο πλάτος (>120) του σήματος στο πεδίο της συχνότητας. Οι συχνότητες αυτές αντιστοιχούν στις συχνότητες των ημιτόνων από τα οποία αποτελείται κάθε τόνος. Τα αποτελέσματα φαίνονται στη συνέχεια.

List of k and touch-tones with hamming windows:

k	touch_tone	k	touch_tone
116.00000	0.72885	164.00000	1.03044
95.00000	0.59690	181.00000	1.13726
86.00000	0.54035	164.00000	1.03044
86.00000	0.54035	164.00000	1.03044
105.00000	0.65973	181.00000	1.13726
105.00000	0.65973	149.00000	0.93619
105.00000	0.65973	181.00000	1.13726
86.00000	0.54035	181.00000	1.13726

Εικόνα 14: Τα αποτελέσματα που προκύπτουν με χρήση του Matlab όπου φαίνονται τα ζητούμενα αποτελέσματα. Κάθε γραμμή αντιστοιχεί σε ένα τηλεφωνικό τόνο του σήματος (με τη σειρά που αυτοί εμφανίζονται στο σήμα) και k είναι ο δείκτης που υποδηλώνει τις συχνότητες στις εντοπίζονται τα ημίτονα, σύμφωνα με τον DFT κάθε τόνου. Φαίνονται επίσης οι touch-tone συχνότητες ($2\pi f/fs$), όπου $f = k \cdot fs/N$, στις οποίες αντιστοιχεί κάθε δείκτης k .

1.6. Στη συνέχεια ζητείται η δημιουργία μίας συνάρτησης *ttdecode()* η οποία να δέχεται σαν όρισμα ένα σήμα που αποτελείται από διαδοχικούς τηλεφωνικούς τόνους και να δίνει ως έξοδο ένα διάνυσμα που να περιλαμβάνει τους αριθμούς που αντιστοιχούν σε κάθε τόνο.

Αρχικά υλοποιούμε τη συνάρτηση *find_touch_tone_limits()* η οποία δέχεται το σήμα με την ακολουθία των τηλεφωνικών τόνων και επιστρέφει έναν πίνακα διαστάσεων $(n \times 2)$. Κάθε γραμμή αντιστοιχεί σε έναν τηλεφωνικό τόνο του σήματος και στις δύο στήλες περιλαμβάνονται τα σημεία έναρξης και λήξης, αντίστοιχα, του αντίστοιχου τόνου. Η συνάρτηση λειτουργεί για οποιοδήποτε πλήθος μηδενικών δειγμάτων που παρεμβάλλονται μεταξύ των τηλεφωνικών τόνων ή υπάρχουν στην αρχή ή στο τέλος του σήματος. Στην περίπτωση που μεταξύ δύο διαδοχικών τόνων δεν παρεμβάλλονται μηδενικά δείγματα, και εφόσον έχουμε θεωρήσει ότι κάθε τόνος αποτελείται από 1000 δείγματα, το διάστημα που έχει προσδιοριστεί ότι περιέχει περισσότερους από έναν τόνους χωρίζεται σε επιμέρους τμήματα μήκους 1000 δειγμάτων. Με αυτόν τον τρόπο η συνάρτηση επιστρέφει σε κάθε περίπτωση τα σημεία έναρξης και λήξης κάθε τηλεφωνικού τόνου.

Στη συνάρτηση *ttdecode()* υπολογίζεται ο αριθμός και τα όρια των τηλεφωνικών τόνων που περιέχονται στο σήμα εισόδου. Για κάθε διάστημα στο οποίο περιέχεται ένας τόνος χρησιμοποιούνται παράθυρα Hamming μήκους 1000 δειγμάτων και υπολογίζεται ο Διακριτός Μετασχηματισμός *Fourier*. Στη συνέχεια υπολογίζεται η ενέργεια του σήματος στο πεδίο της συχνότητας από τον τύπο:

$$E_k = |X[k]|^2$$

Οι συχνότητες στις οποίες η ενέργεια είναι μεγαλύτερη από μία ενδεικτική τιμή (εδώ επιλέχθηκε να είναι $E_k > 4500$) αποθηκεύονται σε έναν πίνακα (*frequency*). Από τον πίνακα αυτόν διατηρούμε μόνο τις συχνότητες στις οποίες εντοπίζονται οι τοπικά μέγιστες τιμές της ενέργειας, οι οποίες αντιστοιχούν στις ημιτονοειδείς συνιστώσες του σήματος.

Στη συνέχεια για την εύρεση του τηλεφωνικού τόνου στον οποίο αντιστοιχούν οι δύο συχνότητες που έχουν βρεθεί προηγουμένως χρησιμοποιείται η μορφή του Πίνακα 1 της εκφώνησης της εργασίας, ο οποίος φαίνεται στη συνέχεια. Συγκεκριμένα, δημιουργούμε τους πίνακες *rows* $_{(1 \times 4)}$ και *columns* $_{(1 \times 3)}$. Κάθε συχνότητα που έχει βρεθεί συγκρίνεται με τις τιμές των *touch-tone* συχνοτήτων του Πίνακα 1. Το στοιχείο είτε της γραμμής είτε της στήλης που βρίσκεται εγγύτερα στην συχνότητα που έχει υπολογιστεί γίνεται ίσο με τη μονάδα. Με αυτόν τον τρόπο για κάθε συχνότητα οι δύο πίνακες υποδηλώνουν τον σωστό τηλεφωνικό τόνο.

	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 των τηλεφωνικών τόνων, για συχνότητα δειγματοληψίας 8192 Hz.

Για παράδειγμα ένας τόνος για τον οποίο έχει υπολογιστεί ότι οι δύο συχνότητες στις οποίες υπάρχουν ημίτονα είναι 0.6597 και 1.1373 προκύπτουν οι πίνακες $rows=[0\ 0\ 1\ 0]$ και $columns=[0\ 0\ 1]$. Επομένως με βάση τον παραπάνω πίνακα ο τηλεφωνικός τόνος ανήκει στο πλήκτρο 9.

Η διαδικασία αυτή επαναλαμβάνεται για όλους τους τηλεφωνικούς τόνους του σήματος και η συνάρτηση *ttdecode()* επιστρέφει το διάνυσμα με τα ψηφία. Η ορθότητα της συνάρτησης επιβεβαιώνεται για το σήμα που δημιουργήθηκε στο ερώτημα 1.3.

```
Touch-tones from our register numbers:  
0      6      2      2      9      7      9      3
```

Εικόνα 15: Αποτελέσματα της συνάρτησης *ttdecode()* για το σήμα που δημιουργήθηκε στο ερώτημα 1.3 από το άθροισμα των αριθμών μητρώου.

1.7. Τέλος ακολουθούμε την ίδια διαδικασία για τα σήματα *easySig* και *hardSig* που δίνονται. Από την έξοδο της συνάρτησης *ttdecode()* προκύπτουν τα ψηφία στα οποία αντιστοιχούν οι δύο αυτές ακολουθίες τηλεφωνικών τόνων.

```
Touch-tones from easySig tone sequence:  
1      3      2      6      3      9      0      0  
  
Touch-tones from hardSig tone sequence:  
9      0      9      6      3      2      1      1      9      1
```

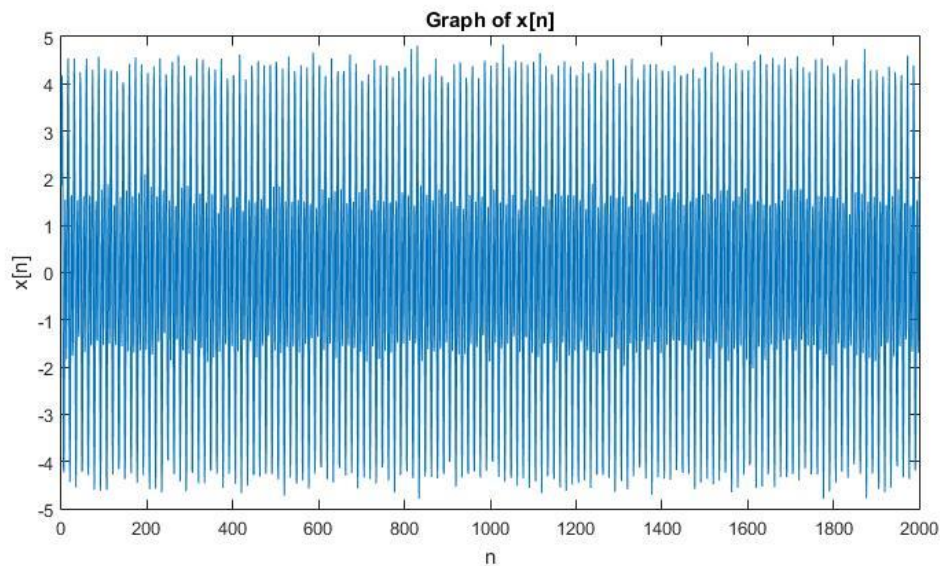
Εικόνα 16: Αποτελέσματα της συνάρτησης *ttdecode()* για τα σήματα *easySig* και *hardSig*.

Μέρος 2.

2.1. (α) Στις εικόνες που ακολουθούν φαίνονται η γραφική παράσταση του σήματος $x[n]$, όπως προκύπτει από τη δειγματοληψία του σήματος $x(t)$ με συχνότητα $fs = 1kHz$, στο διάστημα $[0,2]$ sec.

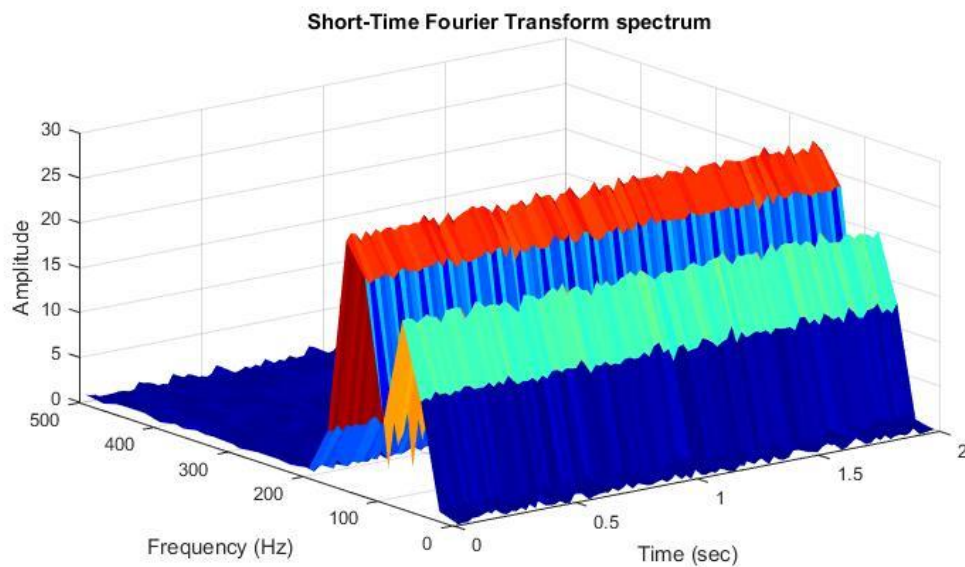
$$x(t) = 2 \cos(2\pi 70t) + 3 \sin(2\pi 140t) + 0.15v(t)$$

όπου $v(t)$ λευκός *Gaussian* θόρυβος μηδενικής μέσης τιμής.



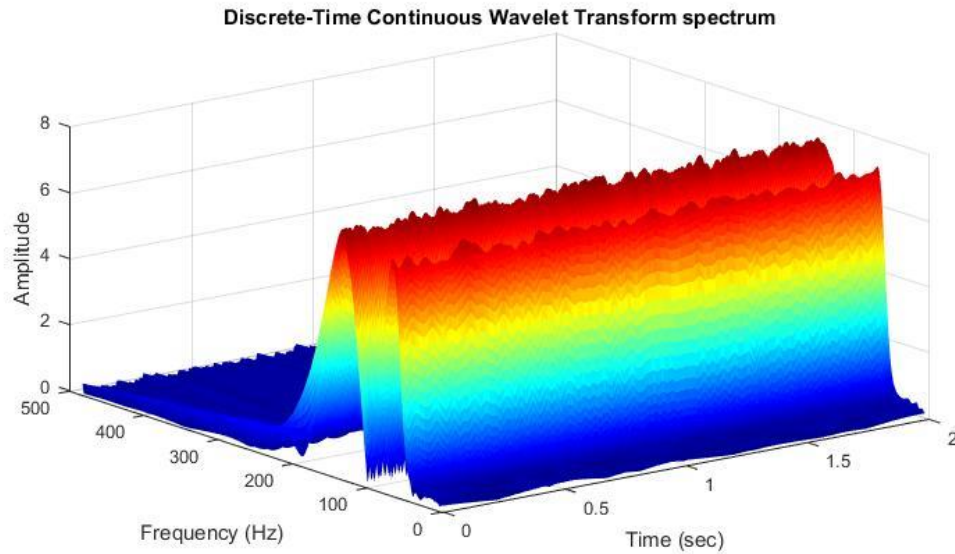
Εικόνα 17: Γραφική παράσταση του σήματος $x[n]$.

(β) Στη συνέχεια υπολογίζουμε τον *STFT* του σήματος με χρήση της ρουτίνας *spectrogram()*. Επιλέγουμε, σύμφωνα με τις οδηγίες, παράθυρα *Hamming* με μήκος ίσο με 40msec και επικάλυψη 20msec. Τα αποτελέσματα φαίνονται στη συνέχεια.



Εικόνα 18: Μετασχηματισμός *Fourier* βραχέος χρόνου (*STFT*) του σήματος $x[n]$.

(γ) Τέλος υπολογίζουμε τον $DT-CWT$ του σήματος $x[n]$, χρησιμοποιώντας “*Morlet*” *wavelet*.



Εικόνα 19: Μετασχηματισμός με χρήση Wavelets του σήματος $x[n]$.

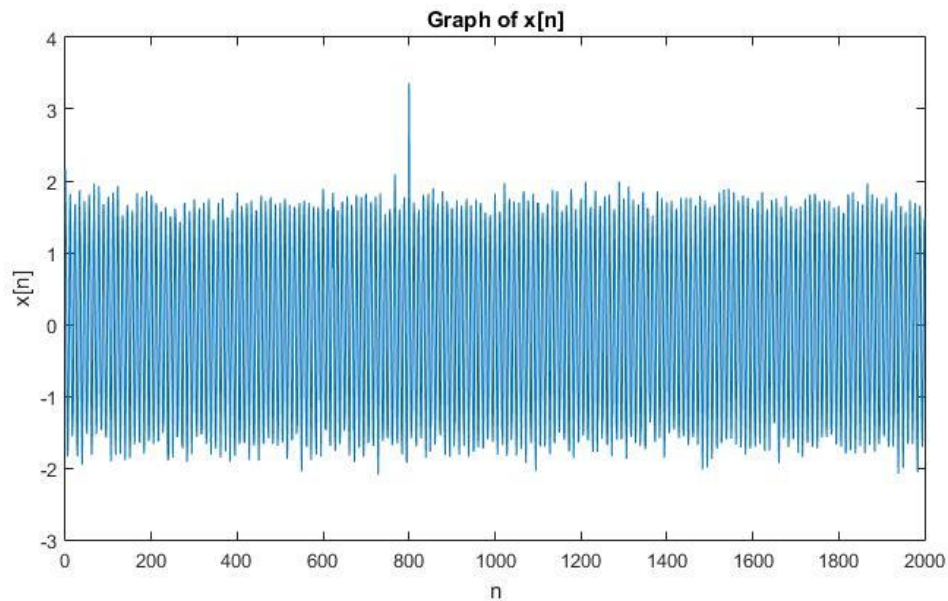
(δ) Παρατηρώντας τα δύο διαγράμματα μπορούμε να εντοπίσουμε με σχετική ακρίβεια τις συχνότητες των δύο ημιτονοειδών συνιστωσών του σήματος $x[n]$. Οι συχνότητες αυτές (70Hz και 140Hz) προσδιορίζονται με μεγαλύτερη ακρίβεια όταν χρησιμοποιείται $DT-CWT$ με *Morlet wavelets*. Το αποτέλεσμα αυτό είναι αναμενόμενο εφόσον στον μετασχηματισμό *Fourier* βραχέος χρόνου χρησιμοποιούνται σταθερού μήκους χρονικά παράθυρα, μειώνοντας τη διακριτική ικανότητα στο πεδίο της συχνότητας και διατηρώντας ικανοποιητικό επίπεδο ανάλυσης στο πεδίο του χρόνου. Αντίθετα χρησιμοποιώντας μετασχηματισμό με *Morlet wavelets*, η διακριτική ικανότητα στο πεδίο του χρόνου παραμένει ικανοποιητική αλλά το μήκος του παραθύρου αλλάζει σε συνάρτηση με το χρόνο. Με αυτόν τον τρόπο επιτυγχάνεται μεγαλύτερη ανάλυση (στο πεδίο των συχνοτήτων) με μείωση της διάρκειας του παραθύρου σε χαμηλές συχνότητες (όπου εντοπίζονται οι ημιτονοειδείς συνιστώσες) και μικρότερη ανάλυση στις υπόλοιπες, όπου υπάρχει κυρίως θόρυβος. Παρ' όλα αυτά η μορφή του σήματος δεν αλλάζει, εφόσον αυτό είναι στατικό.

2.2. (α) Στη συνέχεια ακολουθούμε την ίδια ακριβώς διαδικασία για διαφορετικό σήμα. Στην περίπτωση αυτή το σήμα είναι:

$$x(t) = 1.7 \cos(2\pi 90t) + 0.15v(t) + 1.7\delta(t - 0.625) + 1.7\delta(t - 0.800)$$

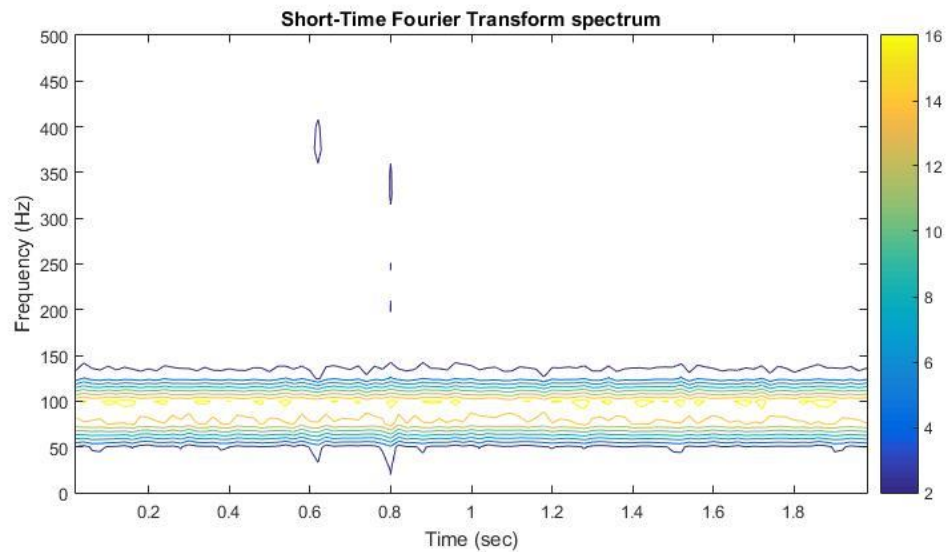
όπου $v(t)$ λευκός *Gaussian* θόρυβος μηδενικής μέσης τιμής.

Το συνεχές σήμα δειγματοληπτείται και πάλι με συχνότητα $fs = 1kHz$, στο διάστημα $[0,2]$ sec.



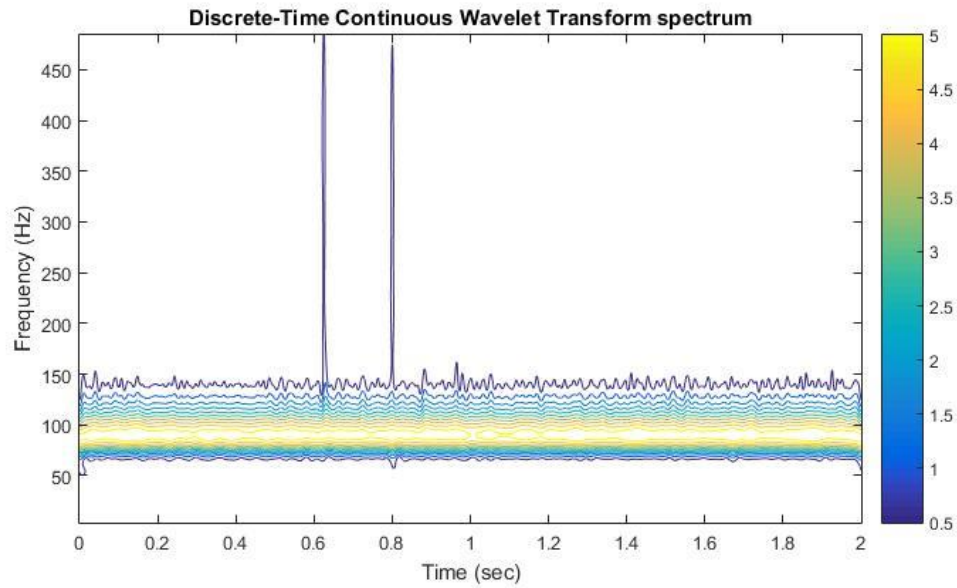
Εικόνα 20: Γραφική παράσταση

(β) Στη συνέχεια υπολογίζουμε τον *STFT* του σήματος με χρήση της ρουτίνας *contour()*. Επιλέγουμε, σύμφωνα με τις οδηγίες, παράθυρα *Hamming* με μήκος ίσο με 40msec και επικάλυψη 20msec. Τα αποτελέσματα φαίνονται στη συνέχεια.



Εικόνα 21: Μετασχηματισμός *Fourier* βραχέος χρόνου (*STFT*) του σήματος $x[n]$.

(γ) Τέλος υπολογίζουμε τον $DT-CWT$ του σήματος $x[n]$, χρησιμοποιώντας “*Morlet*” *wavelet*.



Εικόνα 22: Μετασχηματισμός με χρήση Wavelets του σήματος $x[n]$.

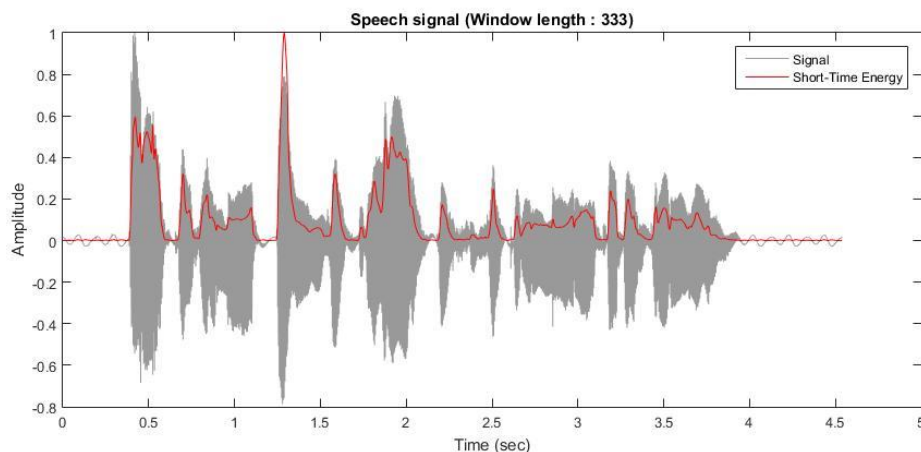
(δ) Όπως και στην προηγούμενη περίπτωση διαπιστώνουμε ότι και με τους δύο μετασχηματισμούς μπορούμε να ανιχνεύσουμε τις συχνότητες των δύο ημιτόνων. Καλύτερο αποτέλεσμα προκύπτει με τη χρήση μετασχηματισμού *wavelets* για τους λόγους που ήδη αναφέρθηκαν.

Μέρος 3.

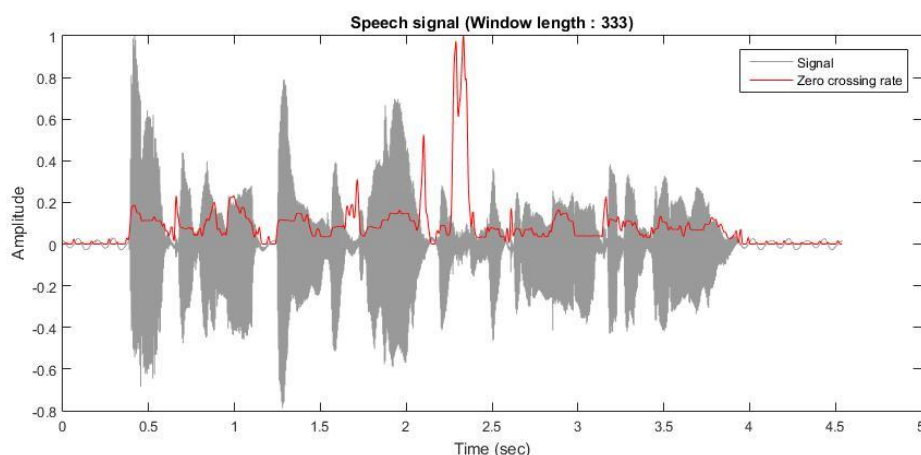
Στόχος των παρακάτω ερωτημάτων είναι ο υπολογισμός της ενέργειας βραχέως χρόνου και του ρυθμού εναλλαγής προσήμου ενός σήματος ομιλίας και ενός σήματος μουσικής. Στον κώδικα που υλοποιήθηκε δίνεται η δυνατότητα επιλογής από τον χρήστη μεταξύ των δύο παραπάνω σημάτων. Στη συνέχεια η επεξεργασία για την εξαγωγή των ζητούμενων χαρακτηριστικών είναι κοινή και στις δύο περιπτώσεις.

Τόσο τα σήματα μουσικής και ομιλίας όσο και τα ζητούμενα χαρακτηριστικά (*short-time energy*, *zero-crossing rate*) εμφανίζονται κανονικοποιημένα στα διαγράμματα που ακολουθούν εφόσον μας αφορά η γενικότερη συμπεριφορά των δύο αυτών μεγεθών και όχι οι ακριβείς τιμές του πλάτους.

3.1. Το σήμα ομιλίας είναι «Αλλά και για να επέμβει χρειάζεται συγκεκριμένη καταγγελία» Η συχνότητα δειγματοληψίας του σήματος είναι 16kHz. Για τον υπολογισμό της ενέργειας βραχέως χρόνου και του ρυθμού εναλλαγής προσήμου του σήματος ομιλίας επιλέγονται παράθυρα *Hamming* μήκους 20.8 msec (δηλαδή 333 δείγματα ανά παράθυρο). Τα αποτελέσματα φαίνονται στα διαγράμματα που ακολουθούν.

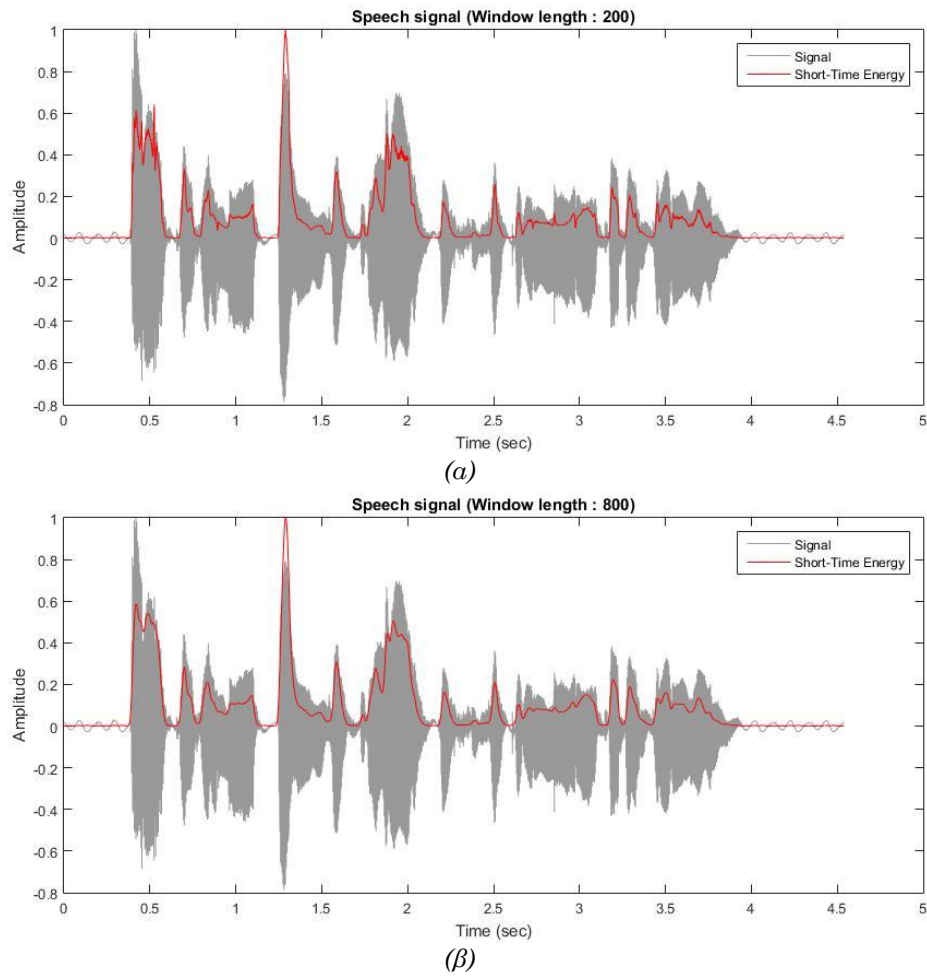


Εικόνα 23: Ενέργεια βραχέως χρόνου του σήματος ομιλίας, για μήκους παραθύρου 20.8msec και 333 δείγματα.

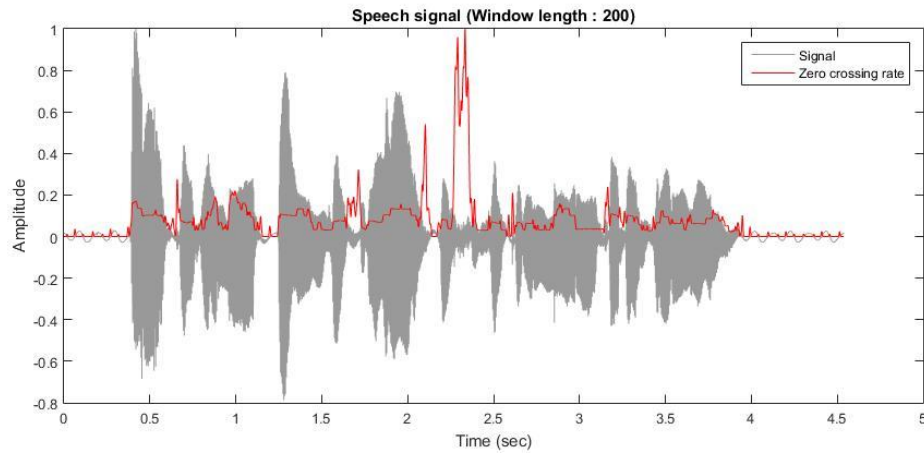


Εικόνα 24: Ρυθμός εναλλαγής προσήμου του σήματος ομιλίας, για μήκους παραθύρου 20.8msec και 333 δείγματα.

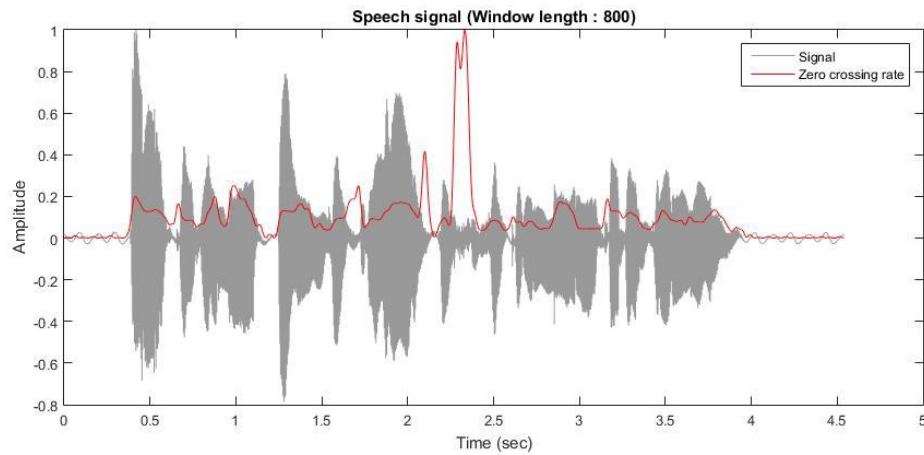
Στη συνέχεια μεταβάλλουμε τον αριθμό των δειγμάτων του παραθύρου *Hamming* ώστε η διάρκειά του να είναι εντός αλλά και εκτός του προτεινόμενου μήκους (δηλαδή 20-30msec). Τα αποτελέσματα φαίνονται στη συνέχεια.



Εικόνα 25: Ενέργεια βραχέος χρόνου του σήματος ομιλίας για παράθυρα *Hamming* μήκους (α) 200 και (β) 800 δειγμάτων.



(α)



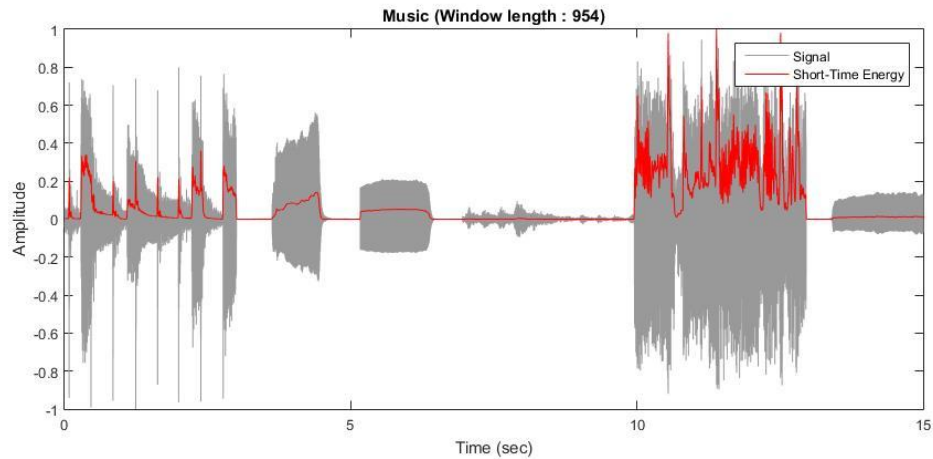
(β)

Εικόνα 26: Ρυθμός εναλλαγής προσήμου του σήματος ομιλίας για παράθυρα Hamming μήκους (α) 200 και (β) 800 δειγμάτων.

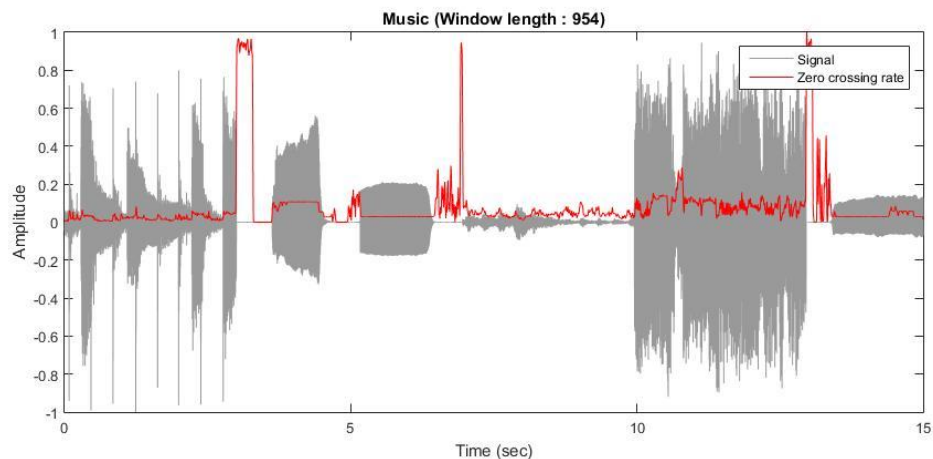
Από τις παραπάνω γραφικές παραστάσεις διαπιστώνουμε ότι με την αύξηση του μήκους του παραθύρου που χρησιμοποιείται, προκύπτει ομαλοποίηση τόσο του ρυθμού εναλλαγής προσήμου όσο και της ενέργειας βραχέως χρόνου.

Επίσης, παρατηρούμε ότι μπορούν εύκολα να διαχωριστούν τα φωνήεντα από τα σύμφωνα στο σήμα ομιλίας. Συγκεκριμένα, τα φωνήεντα λόγω της περιοδικότητας και της μικρής τους συχνότητας εμφανίζουν μεγάλη ενέργεια βραχέως χρόνου και μικρό ρυθμό εναλλαγής προσήμου. Αντίθετα, τα σύμφωνα είναι μη περιοδικά και υψηλότερης συχνότητας, με αποτέλεσμα να προκύπτει υψηλός ρυθμός εναλλαγής προσήμου και μικρή, αλλά όχι μηδενική, ενέργεια.

3.2. Στη συνέχεια επαναλαμβάνουμε την παραπάνω διαδικασία για το σήμα μουσικής που περιλαμβάνεται στο υλικό της εκφώνησης. Το σήμα αυτό έχει συχνότητα δειγματοληψίας 44.1kHz. Για τον υπολογισμό της ενέργειας βραχέως χρόνου και του ρυθμού εναλλαγής προσήμου του σήματος μουσικής επιλέγονται παράθυρα *Hamming* μήκους 21.6msec (δηλαδή 964 δείγματα ανά παράθυρο). Τα αποτελέσματα φαίνονται στα διαγράμματα που ακολουθούν.

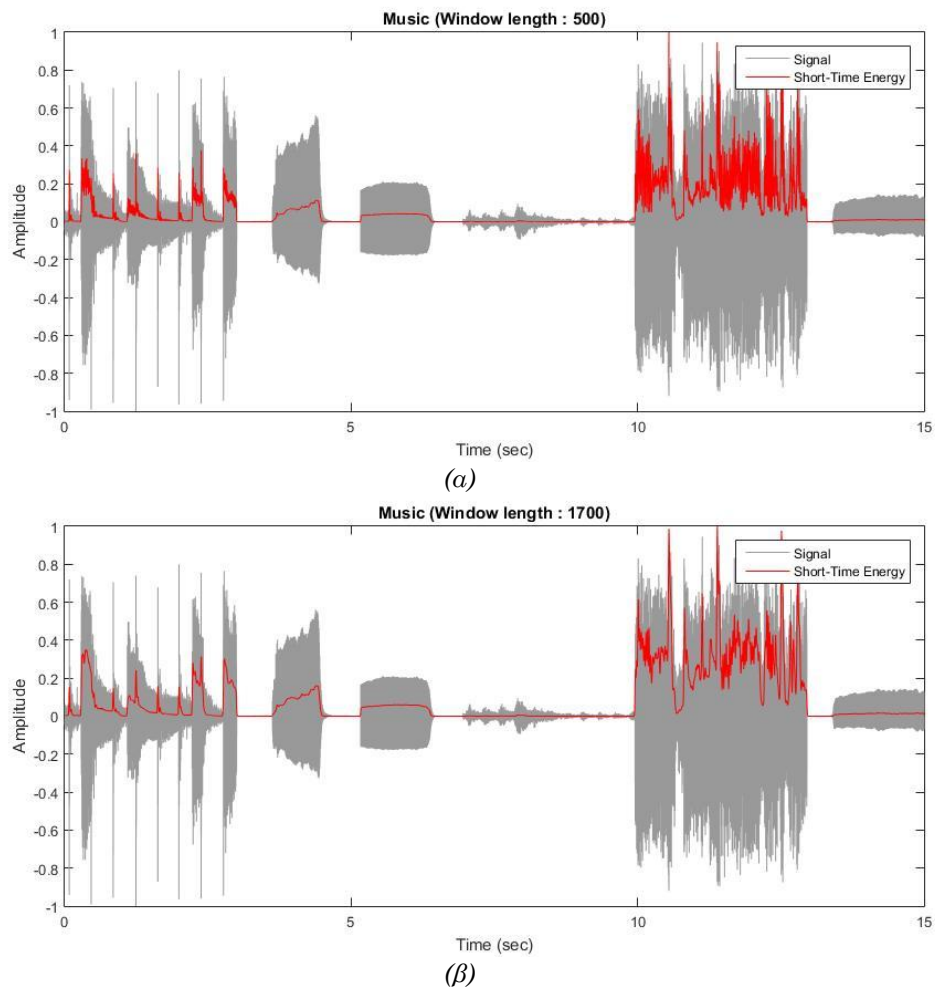


Εικόνα 27: Ενέργεια βραχέως χρόνου του σήματος μουσικής, για μήκους παραθύρου 21.6msec και 954 δείγματα.

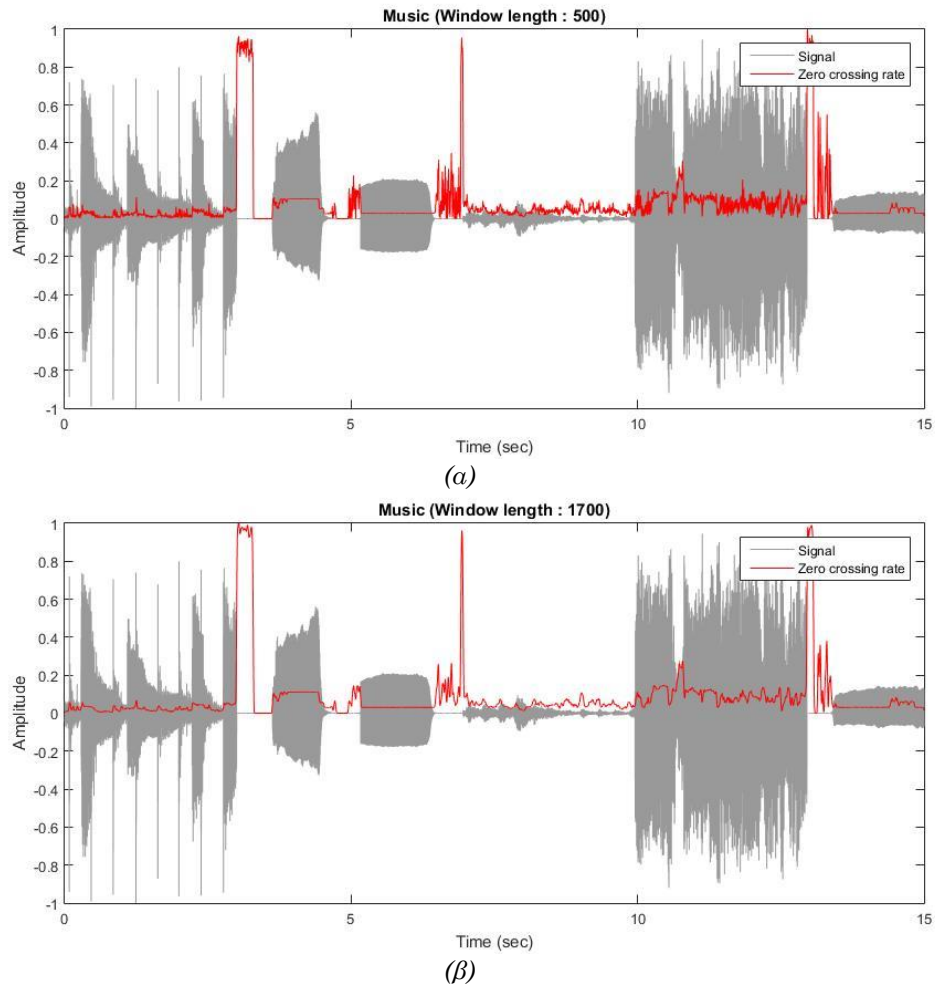


Εικόνα 28: Ρυθμός εναλλαγής προσήμου του σήματος μουσικής, για μήκους παραθύρου 21.6msec και 954 δείγματα.

Στη συνέχεια μεταβάλλουμε τον αριθμό των δειγμάτων του παραθύρου Hamming ώστε η διάρκειά του να είναι εντός αλλά και εκτός του προτεινόμενου μήκους (δηλαδή 20-30msec). Τα αποτελέσματα φαίνονται στη συνέχεια.



Εικόνα 29: Ενέργεια βραχέος χρόνου του σήματος μουσικής για παράθυρα Hamming μήκους (α) 500 και (β) 1700 δειγμάτων.



Εικόνα 30: Ρυθμός εναλλαγής προσήμου του σήματος μουσικής για παράθυρα Hamming μήκους (α) 500 και (β) 1700 δειγμάτων.

Από τις παραπάνω γραφικές παραστάσεις παρατηρούμε ότι, όπως και στην προηγούμενη περίπτωση, η αύξηση του μήκους του παραθύρου που χρησιμοποιείται προκαλεί ομαλοποίηση τόσο του ρυθμού εναλλαγής προσήμου όσο και της ενέργειας βραχέως χρόνου. Επίσης εντοπίζονται σημαντικές διαφορές μεταξύ των διαφορετικών ειδών μουσικής που περιέχονται στο σήμα, και στις δύο περιπτώσεις. Επομένως γνωρίζοντας την μορφή των δύο αυτών μεγεθών (*short-time energy*, *zero-crossing rate*) για διάφορα είδη μουσικής μπορεί να ανιχνευθεί μέσω λογισμικού το είδος της μουσικής σε τυχαία σήματα εισόδου.