

ΕΜΠ

ΣΗΜΜΥ

ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΩΝ

1^η ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ

ΑΚ.ΕΤΟΣ : 2014-2015

ΠΑΡΑΔΟΤΕΑ: 7-3-15

ΟΝΟΜΑΤΑ:

ΚΑΛΑΒΡΥΤΙΝΟΣ ΧΑΡΑΛΑΜΠΟΣ ΑΜ: 3112024

ΚΑΝΕΛΛΟΠΟΥΛΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ ΑΜ: 3112193

ΘΕΜΑ:

«Εισαγωγή στη Ψηφιακή Επεξεργασία Σημάτων
με MatLab και Εφαρμογές σε Ακουστικά
Σήματα»

Part 1. Σύστημα εντοπισμού τηλεφωνικών τόνων (telephone touch-tones).

Αρχικά, χρησιμοποιώντας το διπλανό πίνακα και τον τύπο $t = \sin(w1 \cdot n) + \sin(w2 \cdot n)$;

Δημιουργούμε τους 10 διαφορετικούς τόνους.

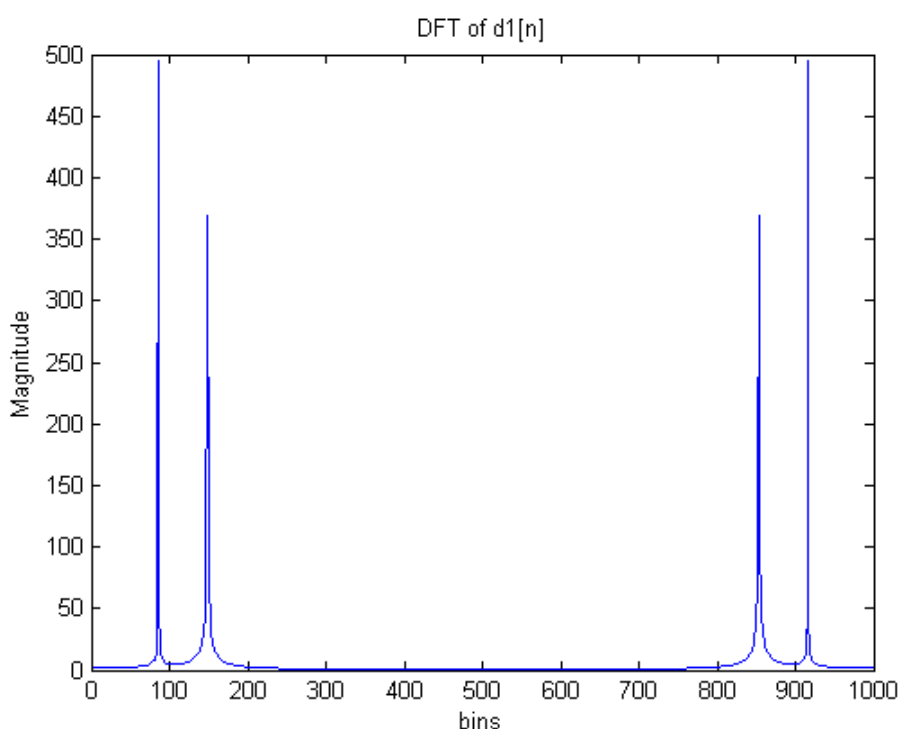
Κάθε τόνος έχει μήκος **N=1000** δειγμάτων.

Σημειώση: Τα ψηφία *, # δεν ελήφθησαν υπόψιν στη συγκεκριμένη άσκηση. $F_s=8192\text{Hz}$ η συχνότητα δειγματοληψίας

	1209 Hz	1336 Hz	1477 Hz
697 Hz	1	2	3
770 Hz	4	5	6
852 Hz	7	8	9
941 Hz	*	0	#

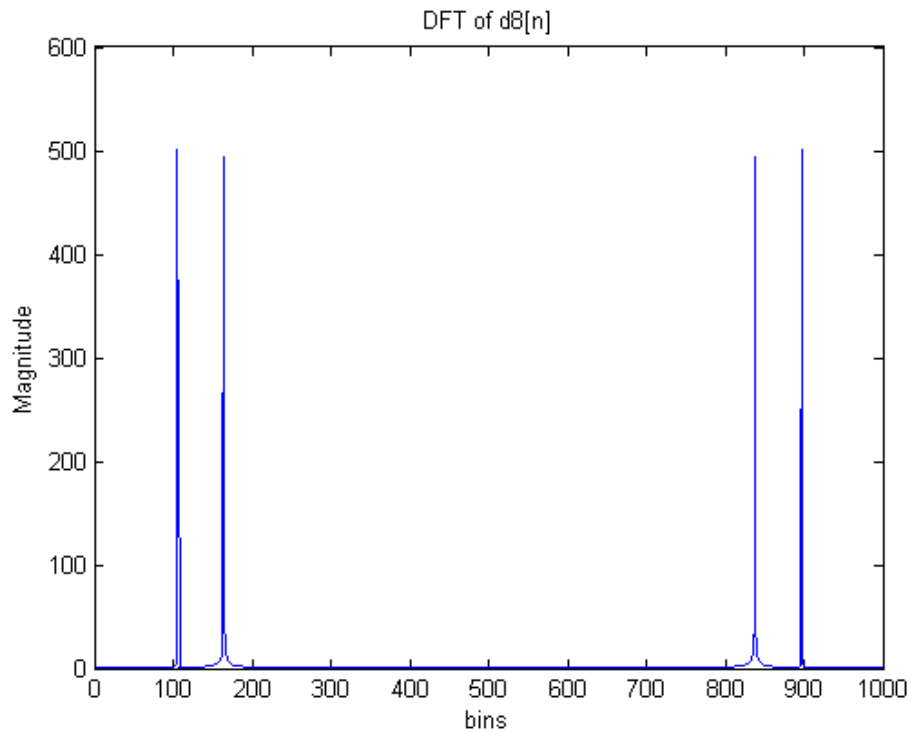
Σχ.1.1 Διακριτές συχνότητες για τηλεφωνικούς τόνους

Με χρήση της ενσωματωμένης σηνάρτησης του Matlab `fft()` βρίσκουμε τον DFT μετασχηματισμό των σημάτων **d1[n]**, **d8[n]**. Με χρήση της γνωστής συνάρτησης `plot()` δημιουργούμε τις γραφικές παραστάσεις που φαίνονται ακολούθως.



Σχ.1.2 Γραφική παράσταση του DFT d1[n]

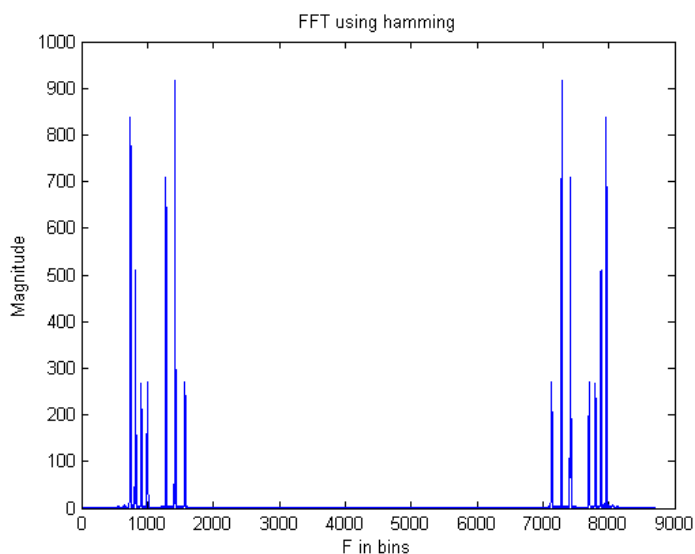
Μεγεθυνόντας στις παραστάσεις μπορούμε να δούμε ότι οι δύο κορυφές βρίσκονται στις θέσεις **86 και 149 bins** (μας ενδιαφέρει μόνο το αριστερό τμήμα της παράστασης) που ισοδυναμούν σε συχνότητες $86 \cdot F_s / N = 86 \cdot 8,192 \text{Hz} = 704 \text{Hz}$ και $149 \cdot F_s / N = 1220 \text{Hz}$. Από το Σχ.1. 1 βλέπουμε ότι οι συχνότητες αυτές, αν εξαιρέσουμε μια μικρή απόκλιση, αντιστοιχούν στον αριθμό 1, όπως ήταν αναμενόμενο. Ομοίως, παρατηρούμε ότι και από το Σχ.1.3 προκύπτει ο αριθμός 8, χρησιμοποιώντας ίδια λογική.



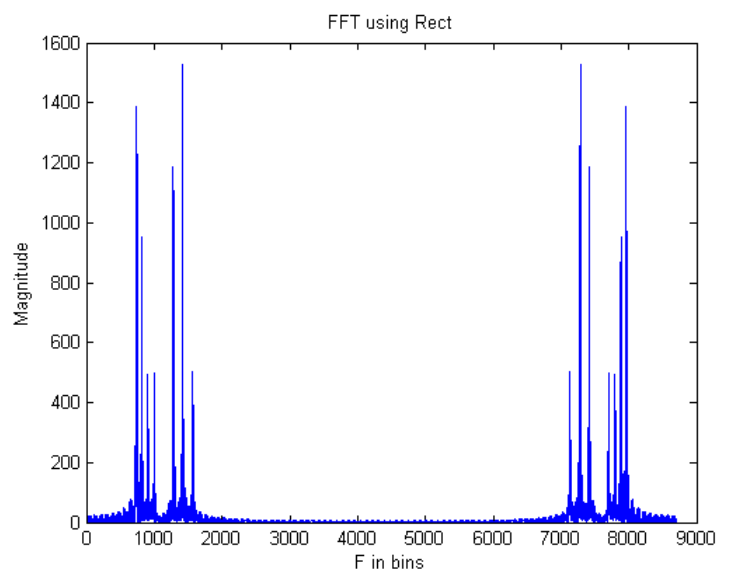
Σχ.1.3 Γραφική παράσταση του DFT $d8[n]$

Το σήμα με τους τηλεφωνικούς τόνους του αθροίσματος των 2 AM βρίσκεται στο φάκελο με το όνομα `'tone_sequence.wav'`.

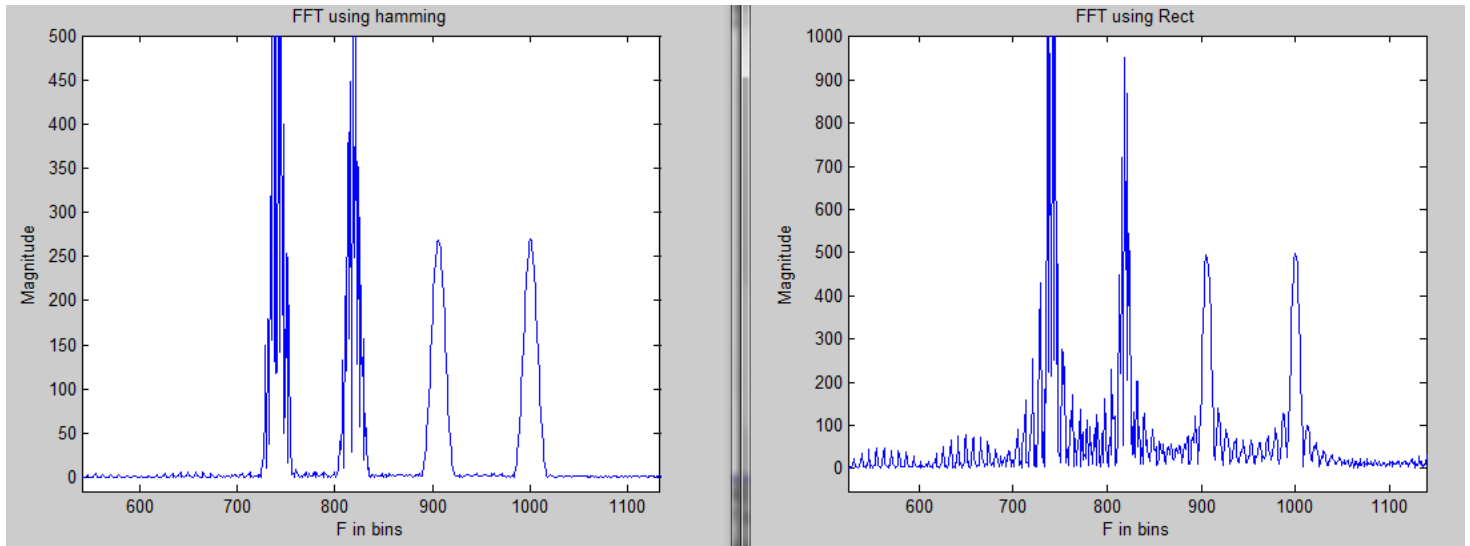
Χρησιμοποιώντας την συνάρτηση `fft()` και παράθυρα `hamming` και τετραγωνικό στο σήμα που δημιουργήθηκε στο 1.3, βρίσκουμε το εξής αποτέλεσμα. Παρατηρούμε ότι ο DFT του παραθυροποιημένου σήματος με `hamming` περιορίζει το πλάτος των `side lobes` που προκύπτουν λόγω των μηδενικών (`zero-padding`), σε αντίθεση με το τετραγωνικό παράθυρο, που είναι πολύ εμφανείς. Ως αποτέλεσμα η παρασταση του DFT με `hamming` είναι πιο λεία (`smooth`)



Σχ.1.4.α Γραφική παράσταση του σήματος χρησιμοποιώντας παράθυρο `hamming`



Σχ.1.4.β Γραφική παράσταση του σήματος χρησιμοποιώντας τετραγωνικό παράθυρο



Σχ.1.5 Λεπτομέρεια των δύο προηγούμενων σχημάτων. Είναι εμφανές ότι οι κύριοι λοβοί διακρίνονται πιο καθαρά στην περίπτωση παραθυροποίησης με hamming. Διακρίνεται εύκολα ότι το πλάτος των side lobes είναι πολύ μικρότερο στην πρώτη (αριστερή) εικόνα.

Σχετικά με τα 3 τελευταία ερωτήματα του πρώτου μέρους, έπειτα από δοκιμές βλέπουμε ότι η συνάρτηση `ttdecode()` λειτουργεί ορθά. Παρατηρούμε ότι το διάνυσμα `vector` για τα σήματα `easySig` και `hardSig` μας επιστρέφει αντίστοιχα

```
>>vector = 2      1      1      8      5      0      8      0
6      9

>>vector =

Columns 1 through 16

6      6      2      2      6      6      2      2      6      6
0      0      0      0      9      9

Columns 17 through 28

3      3      9      9      3      3      9      9      0      0
2      2
```

Η συνάρτηση είναι φτιαγμένη έτσι ώστε να λειτουργεί για τόνους διάρκειας $N=1000$ δειγμάτων. Επομένως, παρατηρούμε ότι για το `hardSig` βρίσκει **28 τόνους** και όχι 14. Αυτό προκύπτει διότι η διάρκεια του κάθε τόνου είναι $N=2000$, και γι αυτό εμφανίζει διπλούς τους αριθμούς. Η πραγματική σειρά είναι 62626009393902

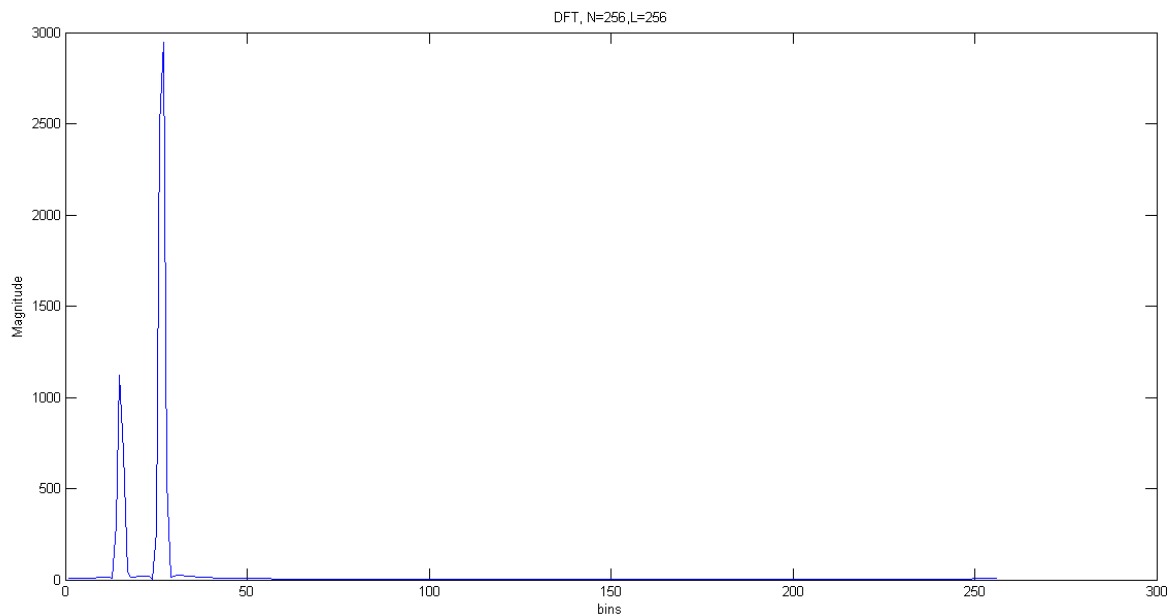
Υπήρξε μια προσπάθεια να επιλυθεί το πρόβλημα, ώστε η συνάρτηση να λειτουργεί σωστά ανεξάρτητα από τη διάρκεια των παλμών. Η βασική ιδέα είναι να βρεθεί το σύνολο των τόνων που υπάρχουν, και (προϋποθέτοντας ότι όλοι οι τόνοι έχουν ίση διάρκεια) να προσδιοριστεί το μήκος του παραθύρου που κόβεται το σήμα ως `length(x)/tone_number`, όπου `x` το σήμα που μας ενδιαφέρει.

Part2. Φασματική ανάλυση και Ανίχνευση Ημιτονοειδών χρησιμοποιώντας DFT

2.1. Παρατηρούμε ότι η συνάρτηση του DFT που υλοποιήσαμε βρίσκεται πολύ κοντά στην built-in function `fft()` του Matlab. Για τυχαίο σήμα **βρίσκουμε τη διαφορά σχεδόν αμελητέα, με το σφάλμα να βρίσκεται σε πολύ μικρά επίπεδα της τάξης του e^{-12}**

2.2. $N=256$, $L=256$ και σήμα ανάλυσης το $y[n]=w[n](x_1[n]+x_2[n])$ με τις εξής προδιαγραφές

```
A1=1; A2=0.8;  
%ph1=rand(1)*(2*pi); ph2=rand(1)*(2*pi); Για τυχαία παραγωγή  
αριθμών φ1,φ2 από το [0,2π]  
  
ph1=2.1314; ph2=3.4122;  
w1=pi/9; w2=pi/5;  
  
x1=A1*exp(1i*n*w1+ph1);  
x2=A2*exp(1i*n*w2+ph2);
```



Σχ.2.1 DFT για $N=256$, $L=256$ $w_1=\pi/5$, $w_2=\pi/9$

Χρησιμοποιούμε μια επανάληψη `for` ώστε να καταφέρουμε να διακρίνουμε για ποια τιμή της διαφοράς $dw=abs(w_1-w_2)$ διακρίνονται ικανοποιητικά για πρώτη φορά οι δύο κορυφές. Κρατάμε σταθερή την τιμή w_1 , και αλλάζουμε με μικρό βήμα $\pi/500$ τη w_2 . Προσεγγιστικά διακρίνουμε πως οι δύο κυματομορφές ξεχωρίζουν για διαφορά $dw=0.1068$. **Το παράθυρο hamming που χρησιμοποιούμε έχει μήκος N .** Παρατηρούμε ότι χρησιμοποιώντας αυτό το μήκος παραθύρου, οι δύο κορυφές έρχονται αρκετά κοντά η μία με την άλλη και ο διαχωρισμός γίνεται σε λίγα bins.

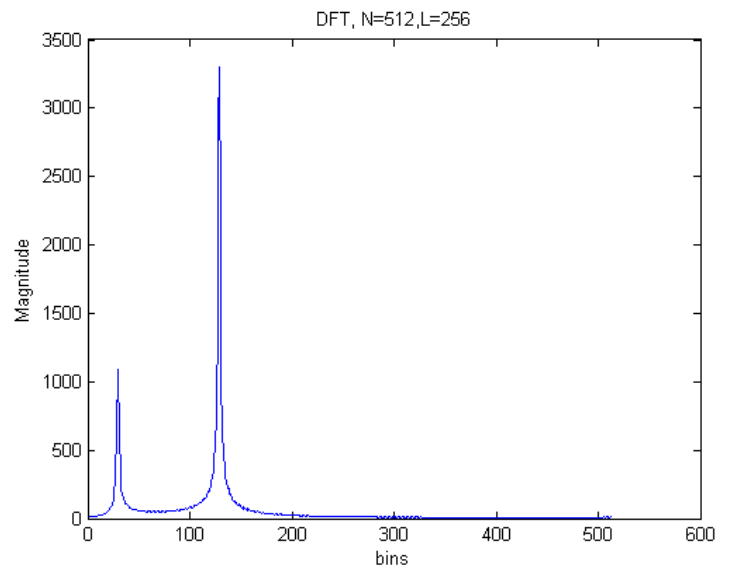
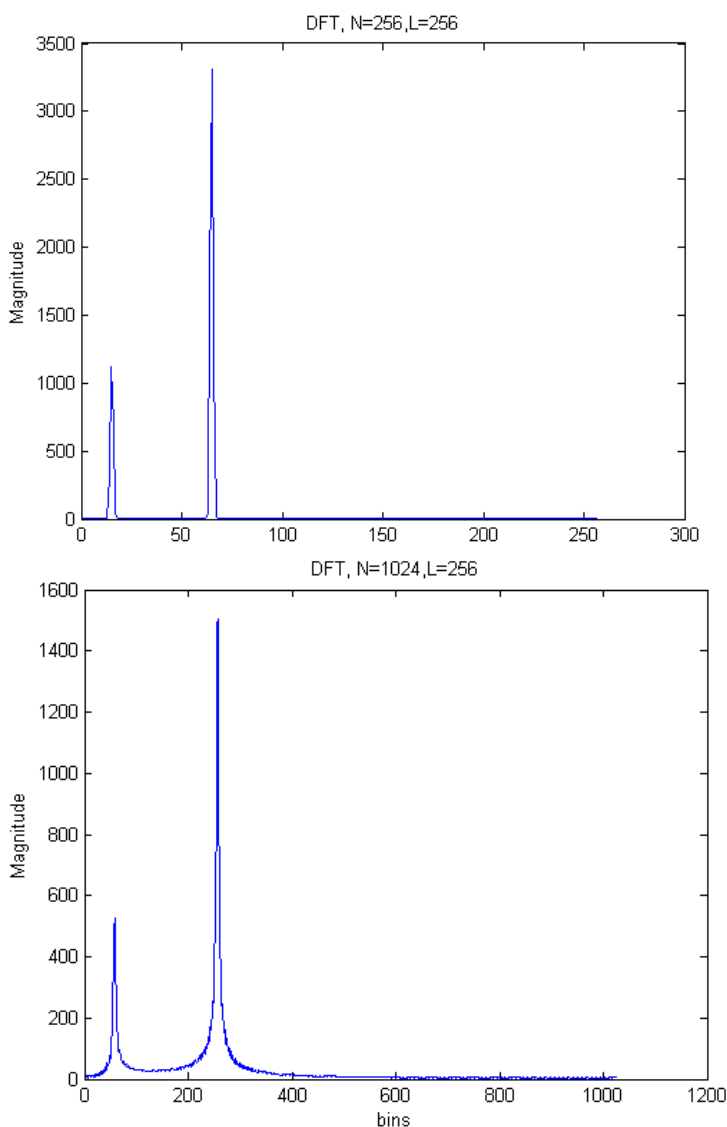
2.3. Επαναλαμβάνουμε το ίδιο πείραμα αλλάζοντας το N σε $N=512$, και $N=1024$ και κρατώντας το μήκος στο $L=256$. Συνεπώς και τα παράθυρα hamming θα έχουν αντίστοιχα μήκος 512 και 1024 κάθε φορά. Πριν υπολογίσουμε το DFT κάνουμε **zero padding** ώστε το σήμα μας να έχει **νέο μήκος L** ίσο με το N .

Για τη φασματική διακριτότητα (αναφέρεται ως *frequency resolution*):
*“A common misconception is that more zero padding results in better spectral resolution. However, zero padding and increasing the size of the DFT does not lead to better resolution, Notice that increasing the DFT size by zero-padding results in finer frequency spacing but does not change **the ability to resolve the two frequency components**. In order to achieve better frequency resolution, we need to change the length and shape of the window. **As the window length increases, we see improvements in our ability to distinguish the two frequency components**, and the relative amplitude of the two components becomes closer to the correct value”*

from Oppenheim, Schaffer and Buck

Όσον αφορά τη δυνατότητα φασματικής διάκρισης παρατηρούμε ότι οι τιμές βρίσκονται αρκετά κοντά μεταξύ τους. Φυσικά, οι ακριβείς τιμές των διαφορών εξαρτώνται από το τι κριτήρια λαμβάνει υπόψιν ο καθένας ώστε να θεωρήσει δύο κορυφές καθαρά διαχωρισμένες μεταξύ τους.

Για την ευκρίνεια του φάσματος, φαίνεται ότι για μεγαλύτερα N , υπάρχει μεγαλύτερη ευκρίνεια, ενώ όσο μειώνεται το N , η ευκρίνεια μειώνεται. Μπορεί να έχουμε μεγαλύτερη ευκρίνεια (τα περισσότερα δείγματα μας δίνουν περισσότερη πληροφορία για το τι γίνεται στο πεδίο συχνοτήτων) αλλά ο dft είναι πιο περίπλοκος

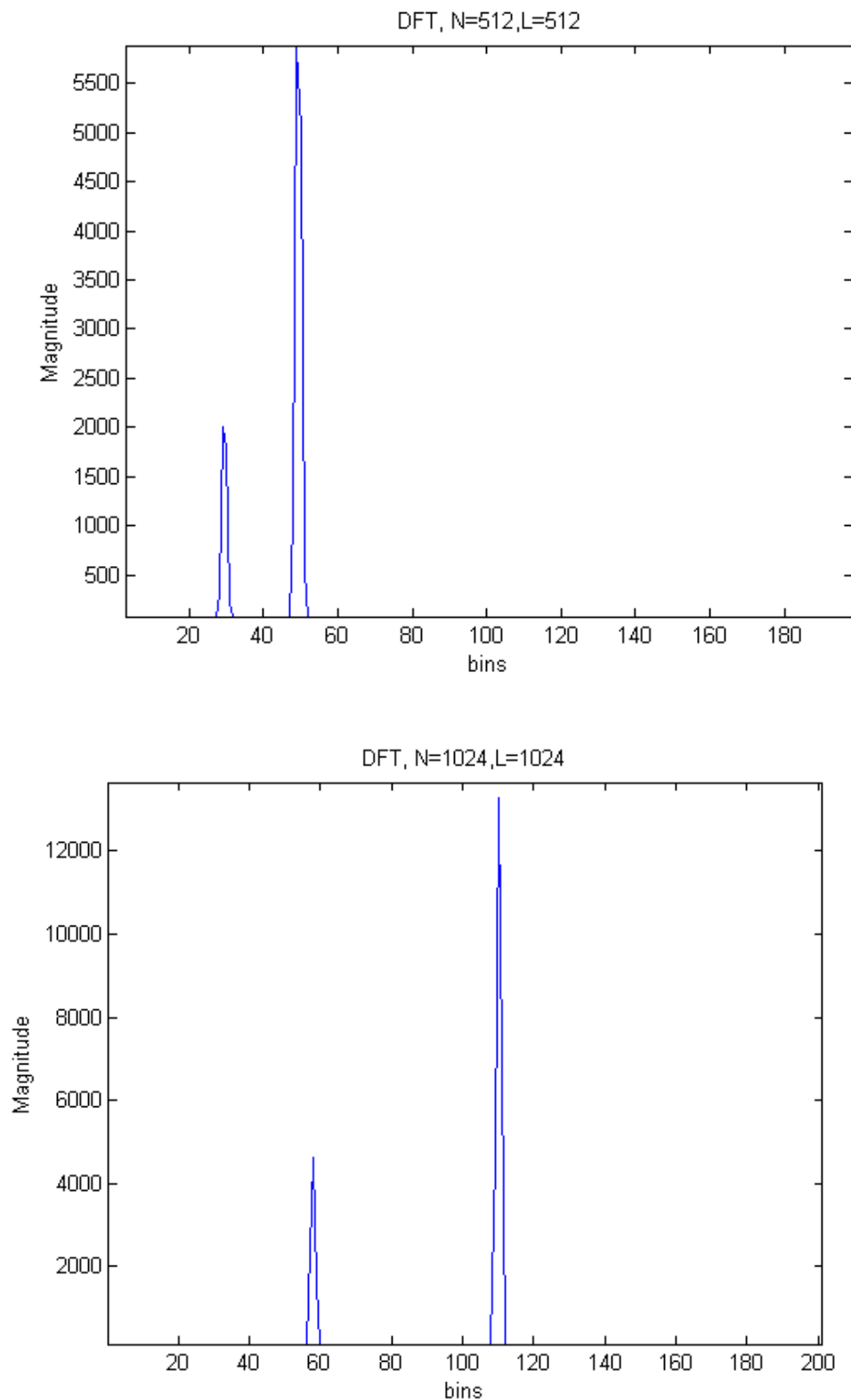


Σχ.2.2.α (πάνω αριστερά), Σχ.2.2.β (πάνω δεξιά)

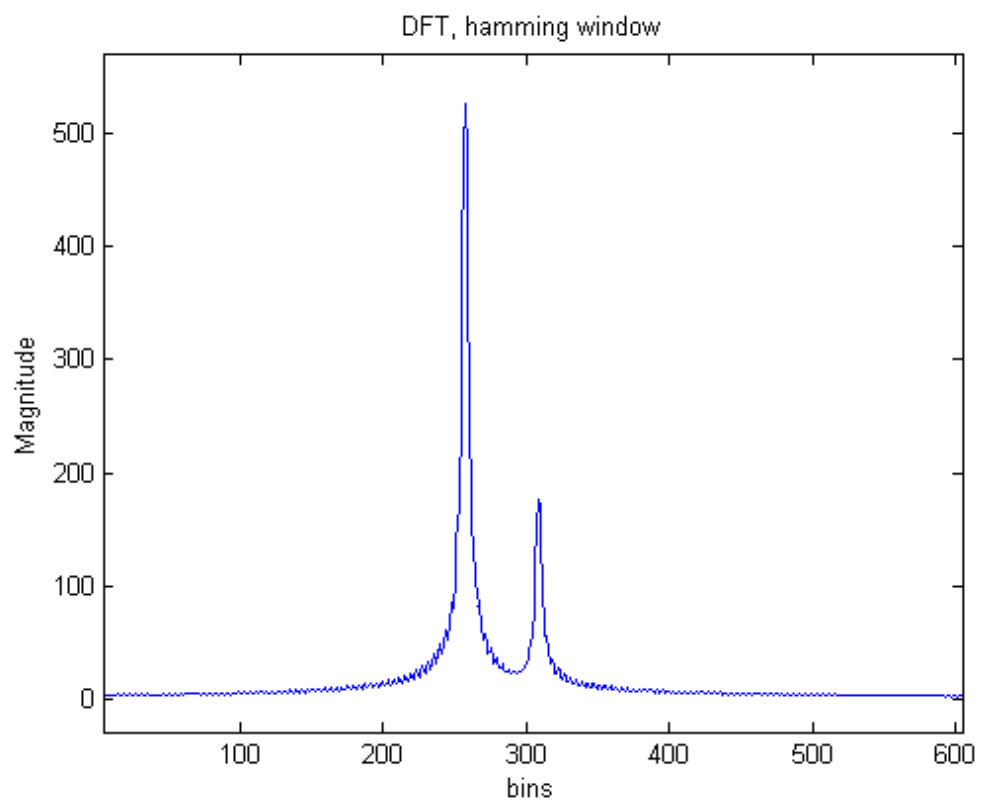
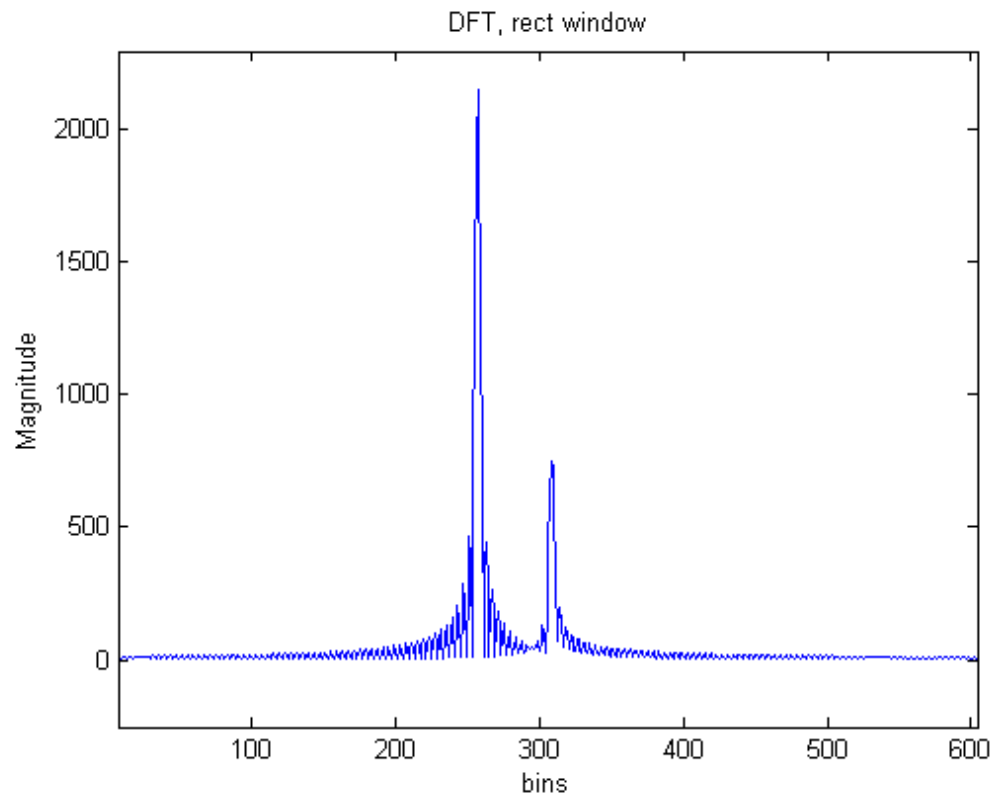
Σχ.2.2.γ (κάτω αριστερά)

Στα διπλανά σχήματα ισχύει: $w1=\pi/9$ και $w2=\pi/2$. Τα N, L διακρίνονται σε κάθε σχήμα

2.4. Στα επόμενα 2 σχήματα φαίνονται οι γραφικές παραστάσεις του σήματος $y[n]$ με διαφορετικά N, L . Στα σχήματα απεικονίζονται τα πρώτα 200 bins. Στις επόμενες περιπτώσεις δεν έχει γίνει zero-padding όπως ακριβώς και στο 2.2. όπου $N=256, L=256$. Επομένως η ευκρίνεια φάσματος και η δυνατότητα φασματικής διάκρισης συμπίπτουν. Για μεγαλύτερο N η ευκρίνεια, και άρα και η δυνατότητα διάκρισης, αυξάνεται, οπότε οι δύο κορυφές μπορούν να διακριθούν για μικρότερη διαφορά $\Delta\omega$.

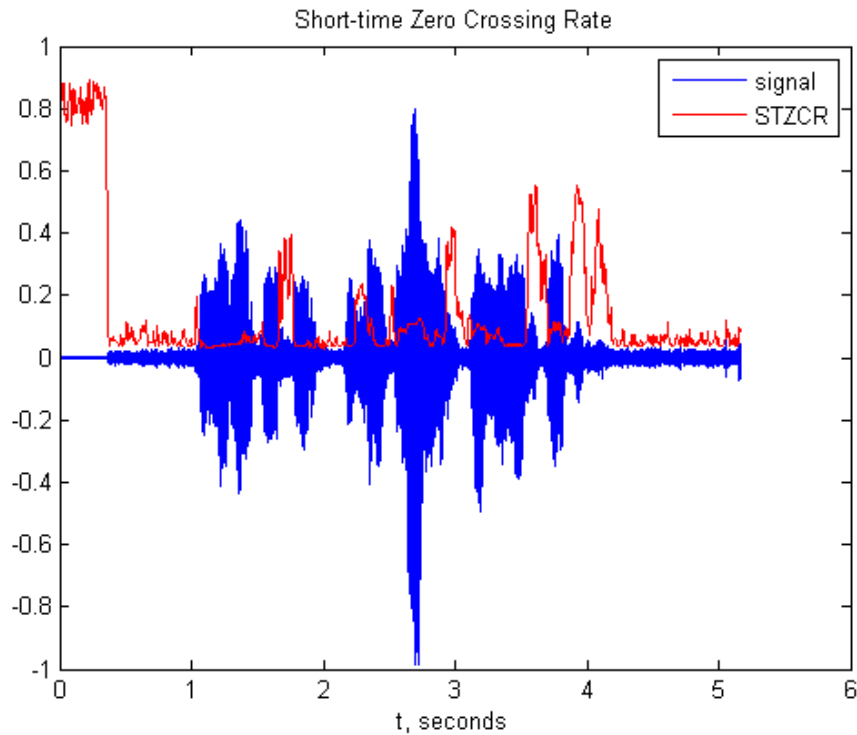


2.5. Για τα δεδομένα που μας δίνονται προκύπτουν τα ακόλουθα σχήματα. Φαίνεται ο DFT για τα πρώτα 600 bins. Ισχύουν οι ίδιες παρατηρήσεις για τις διαφορές *hamming* και *rectangular* που έγιναν και στο 1^ο μέρος.

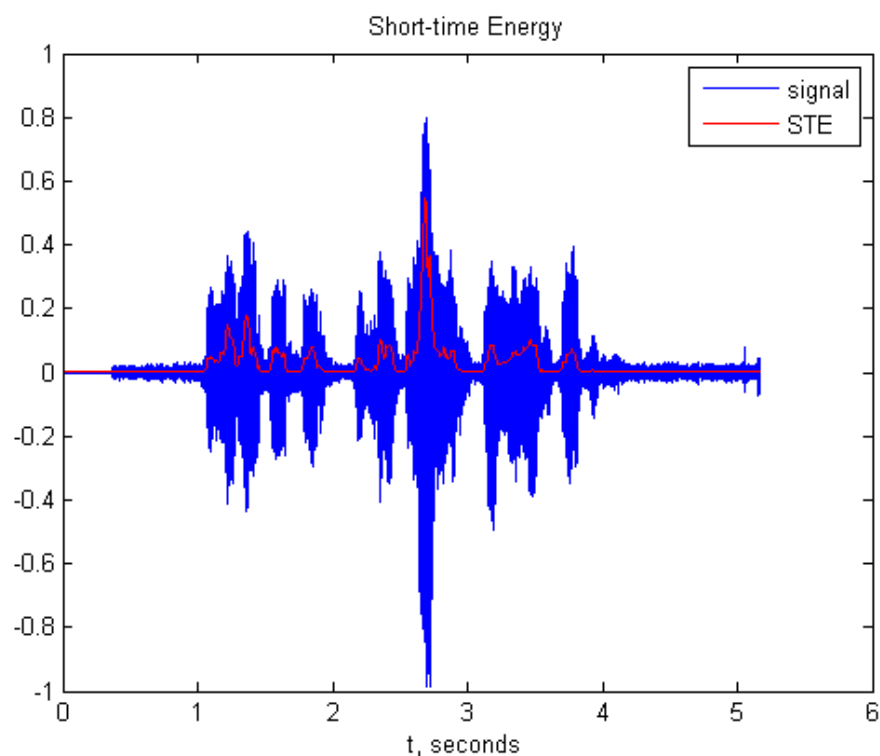


Part3. Χαρακτηριστικά βραχέος χρόνου σημάτων φωνής και μουσικής (ενέργεια και ρυθμός εναλλαγής προσήμου –zero crossing rate).

Η **ενέργεια βραχέος χρόνου** και το **zero-crossing rate** του σήματος του αρχείου `'speech_utterance.wav'` φαίνεται παρακάτω.

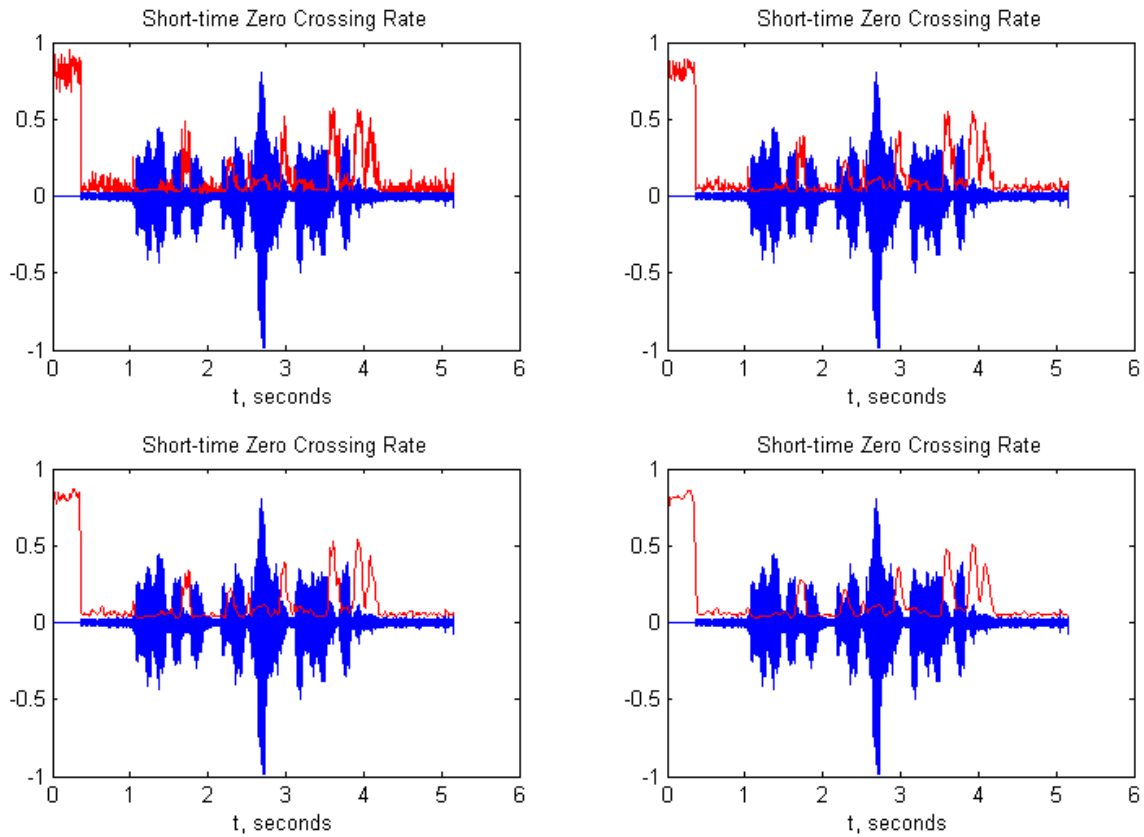


Σχ.3.1 Σήμα εισόδου και ZCR σε κλίμακα χρόνου.

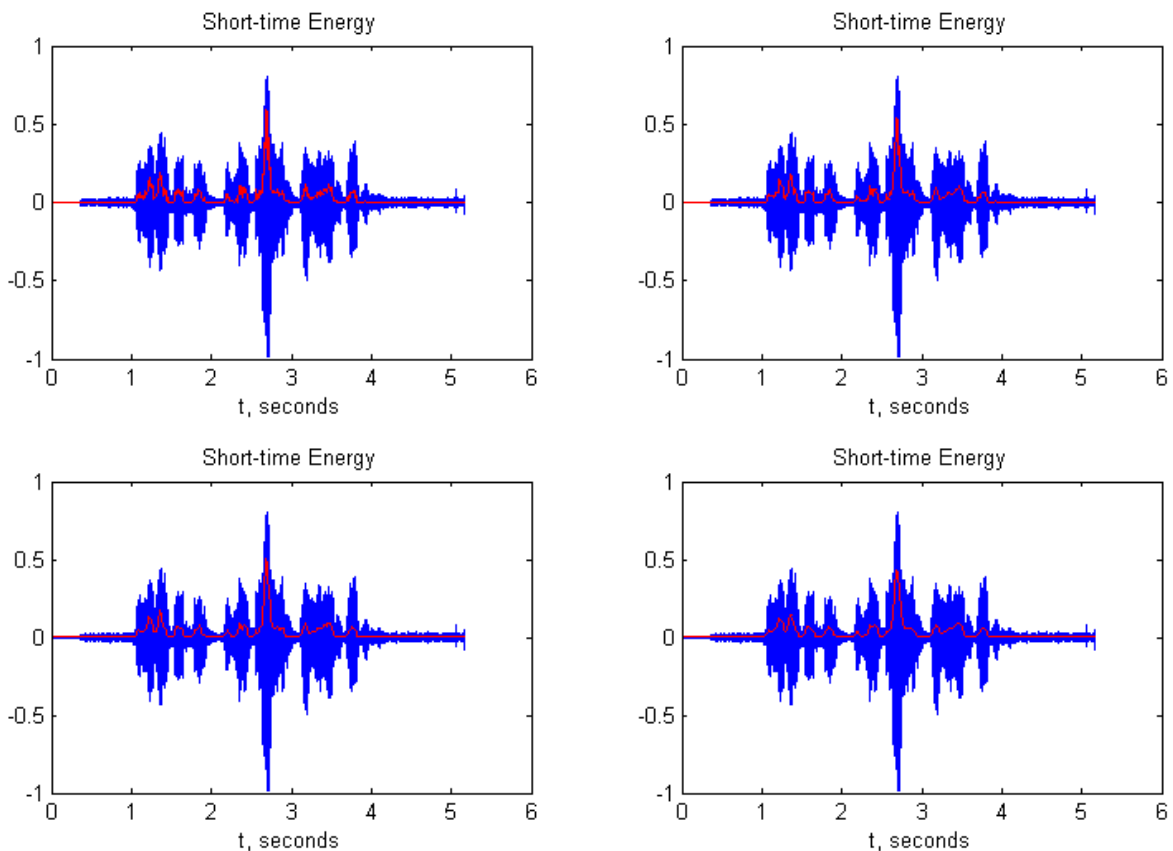


Σχ.3.1 Σήμα εισόδου και Ενέργεια βραχέος χρόνου σε κλίμακα χρόνου.

Έπειτα μεγαλώνουμε το μήκος του παραθύρου ώστε να δούμε τι επιπτώσεις θα έχει στις ζητούμενες κυματομορφές. Τα αποτελέσματα φαίνονται παρακάτω.



Σχ.3.2.1 Στο σχήμα βλέπουμε τη μορφή του ZCR με διαφορετικά μήκη παραθύρου L . Ξεκινώντας από πάνω αριστερά και πηγαίνοντας προς τα αριστερά είναι $L_1=10\text{ms}$, $L_2=20\text{ms}$, $L_3=40\text{ms}$, $L_4=80\text{ms}$



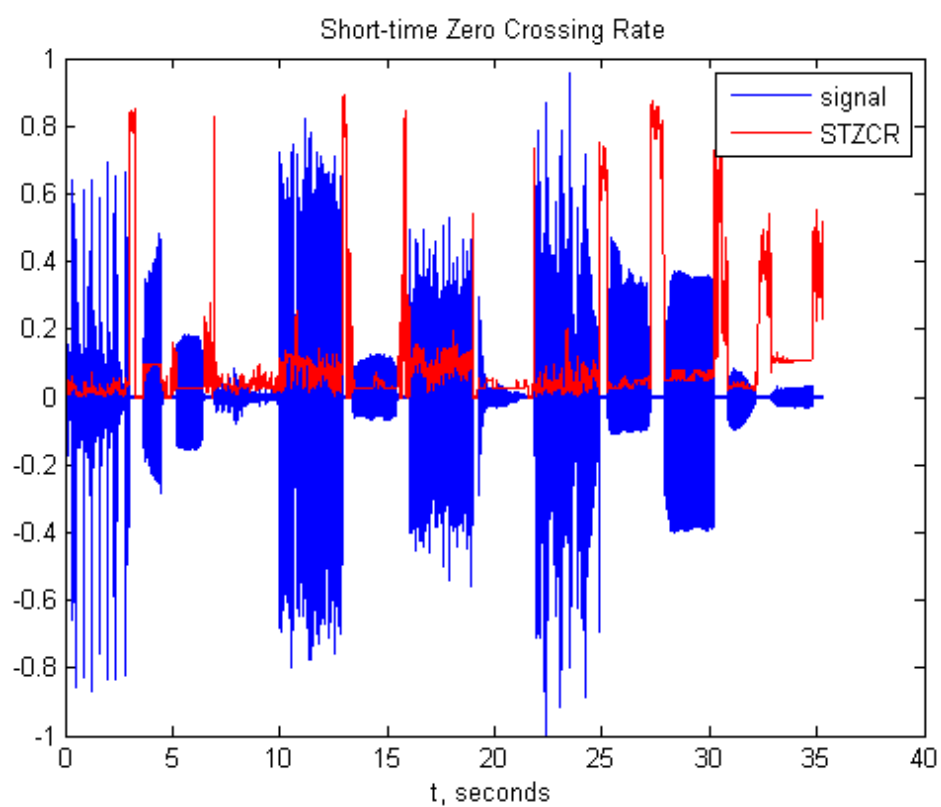
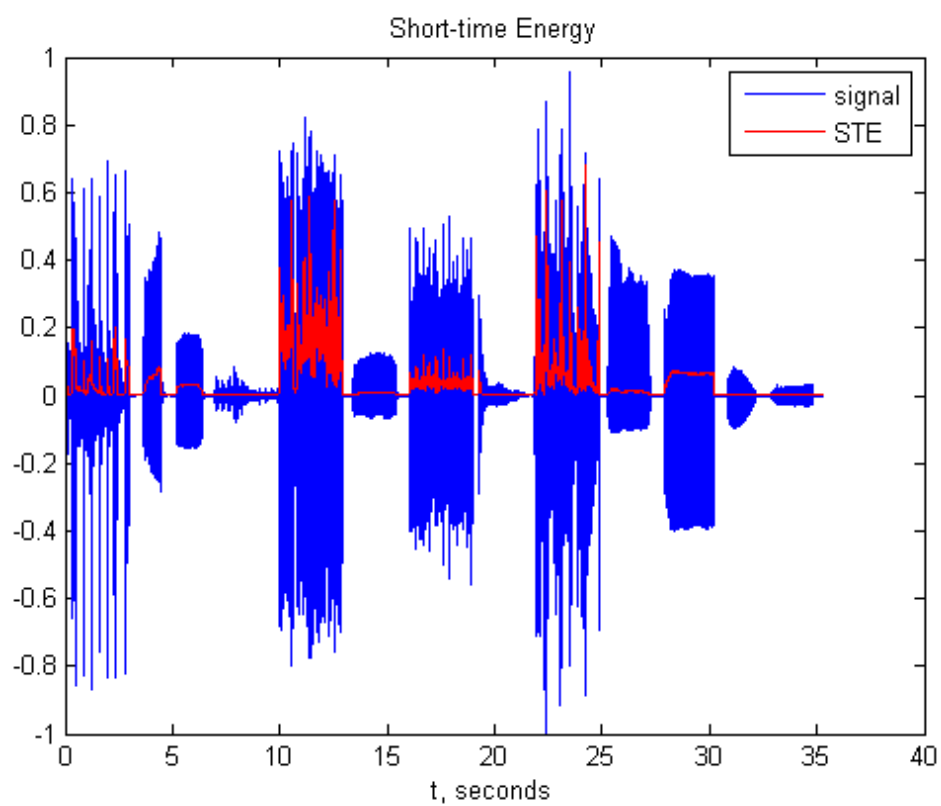
Σχ.3.2.2 Στο σχήμα βλέπουμε τη μορφή της ενέργειας με διαφορετικά μήκη παραθύρου L . Ξεκινώντας από πάνω αριστερά και πηγαίνοντας προς τα αριστερά είναι $L_1=10\text{ms}$, $L_2=20\text{ms}$, $L_3=40\text{ms}$, $L_4=80\text{ms}$

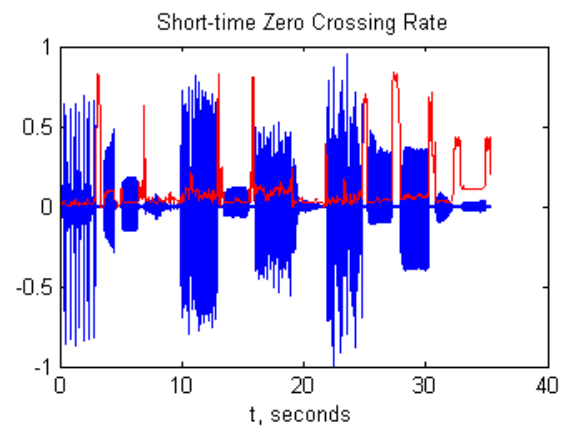
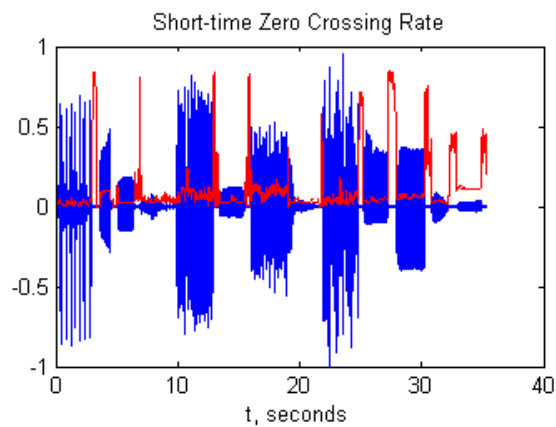
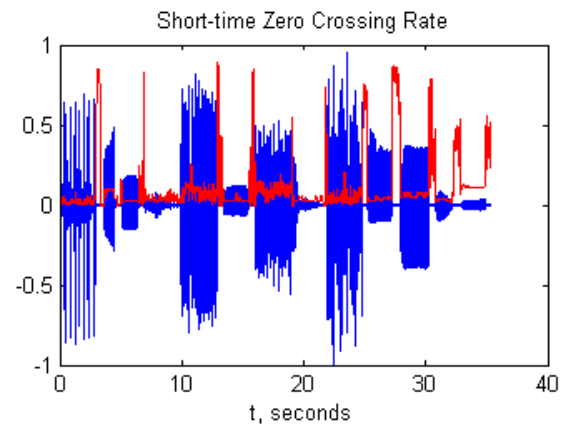
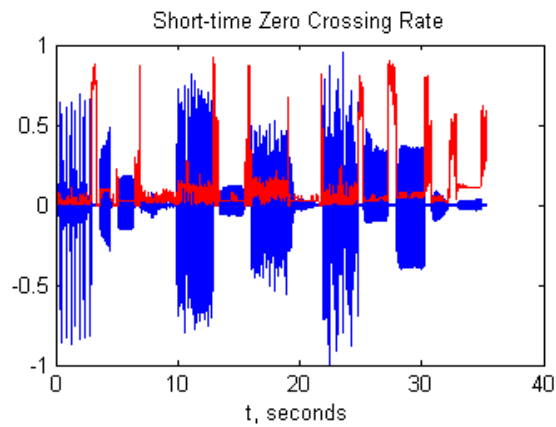
Τα αποτελέσματα για διαφορετικά μήκη παραθύρων είναι εμφανή. Όσο αυξάνεται το μήκος του παραθύρου, τόσο πιο λεία φαίνεται η κυματομορφή της ενέργειας βραχέος χρόνου, όπως αναμενόταν. Οπότε σε μεγάλα μήκη παραθύρων πολλές κορυφές της κυματομορφής της ενέργειας χάνονται. Άμεσο αποτέλεσμα είναι ότι όσο πιο μεγάλο είναι το μήκος του παραθύρου, τόσο πιο δύσκολη είναι η διάκριση μεταξύ **φωνής(voiced)** ή **σιωπής (unvoiced)**. Οι ήχοι φωνής έχουν την περισσότερη ενέργειά τους συγκεντρωμένη σε χαμηλές συχνότητες, ενώ αντίθετα η σιωπή έχει ενέργεια σε υψηλότερες συχνότητες.

Όμοια για το zcr όσο μεγαλώνει το παράθυρο, τόσο πιο λεία γίνεται η κυματομορφή (*smoothing and spreading of ZCR*). Έτσι μπορούν να χαθούν μικρές περιοχές που περιέχουν κορυφές με μεγάλο πλάτος και να παραμορφωθεί σημαντικά η κυματομορφή.

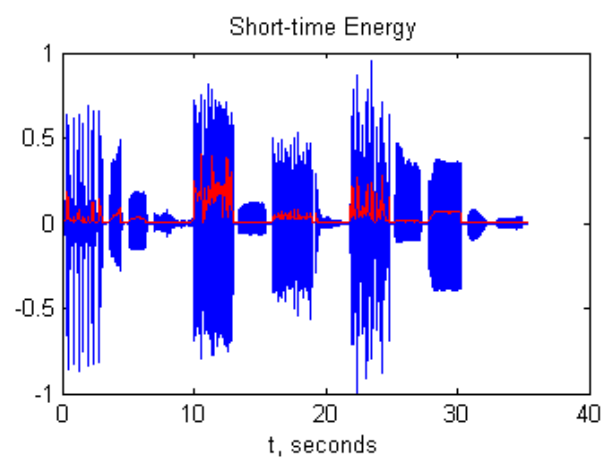
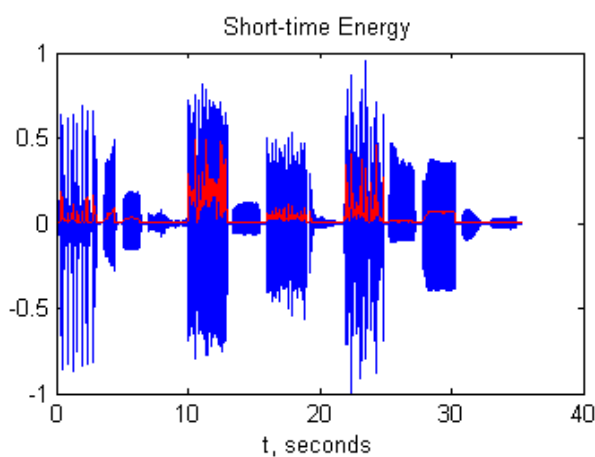
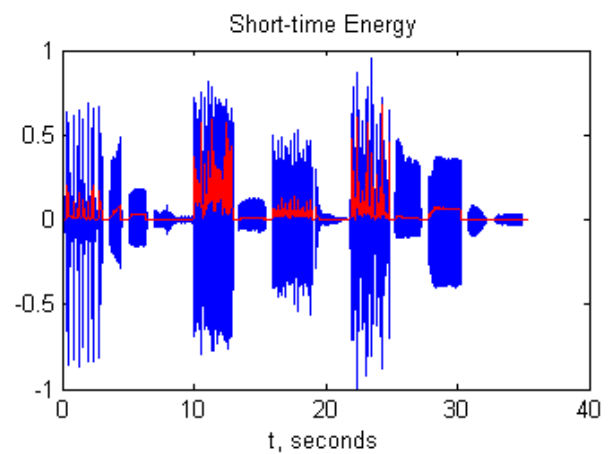
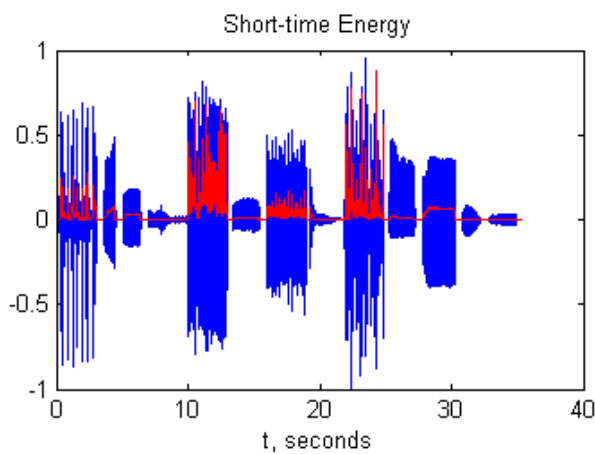
Η διάκριση μεταξύ άφωνων και έμφωνων ήχων μπορεί να γίνει λαμβάνοντας υπόψιν και την ενέργεια και το zcr. Οι έμφωνοι ήχοι βρίσκονται σε σημεία με πολύ υψηλή ενέργεια και αμελητέο zcr, ενώ οι άφωνοι ήχοι έχουν πολύ χαμηλή ενέργεια και αρκετά υψηλό zcr. Η σιωπή έχει μηδενική ενέργεια και μέγιστο zcr.

Παρακάτω φαίνεται και η ανάλυση του αρχείου `'music.wav'`.





Σχ.3.3.1 Στο σχήμα βλέπουμε τη μορφή του ZCR με διαφορετικά μήκη παραθύρου L . Ξεκινώντας από πάνω αριστερά και πηγαίνοντας προς τα αριστερά είναι $L_1=10\text{ms}$, $L_2=20\text{ms}$, $L_3=40\text{ms}$, $L_4=80\text{ms}$



Σχ.3.3.2 Στο σχήμα βλέπουμε τη μορφή της ενέργειας με διαφορετικά μήκη παραθύρου L . Ξεκινώντας από πάνω αριστερά και πηγαίνοντας προς τα αριστερά είναι $L_1=10\text{ms}$, $L_2=20\text{ms}$, $L_3=40\text{ms}$, $L_4=80\text{ms}$