Ψηφιακή Επεξεργασία Σημάτων
1η Εργαστηριακή Άσκηση
Θέμα: Εισαγωγή στην Ψηφιακή Επεξεργασία Σημάτων με MATLAB και Εφαρμογές σε Ακουστικά Σήματα

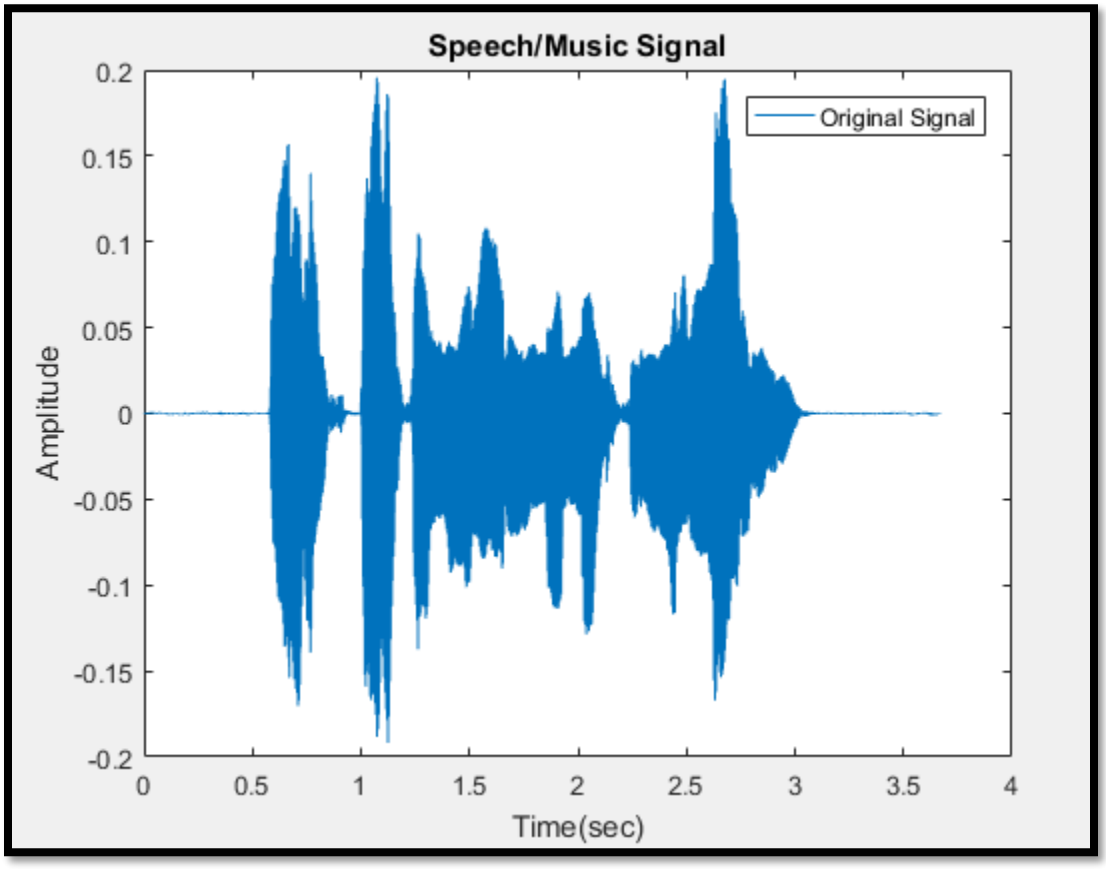
ΠΑΝΑΓΙΩΤΑΡΑΣ ΗΛΙΑΣ ΑΜ: 03115746
ΑΘΑΝΑΣΙΟΣ ΚΟΥΡΑΝΤΟΣ ΑΜ: 03115167

Ακαδ. Έτος 2017-18 | Ημ. Παράδοσης 30/03/18 | 6ο Εξάμηνο

Μέρος 1ο - Χαρακτηριστικά Βραχέος Χρόνου Σημάτων Φωνής και Μουσικής (Ενέργεια και Ρυθμός Εναλλαγής Προσήμου)

1.1)

Για την υλοποίηση αυτού του ερωτήματος θα χρησιμοποιήσουμε το σήμα φωνής της πρότασης ‘’ Όλα αυτά ήταν η άμυνα μες στο μυαλό μου’’ που περιέχεται στο αρχείο **speech_utterance.wav. (**Η συχνότητα δειγματοληψίας είναι 16 kHz). Το σήμα σε συνάρτηση με τον χρόνο παρουσιάζεται ακολούθως:



Στόχος μας είναι, χρησιμοποιώντας τις μετρήσεις βραχέος χρόνου που έχουν παρουσιαστεί, να καταφέρουμε να παρατηρήσουμε τοπικά χαρακτηριστικά του δοθέντος σήματος.

Θα χρησιμοποιήσουμε την ενέργεια βραχέος χρόνου(Short Time Energy) η οποία ορίζεται ως :

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

, όπου w[n] είναι ένα μετακινούμενο παράθυρο (τύπου Hamming στην δεδομένη άσκηση)

Η ενέργεια βραχέως χρόνου (STE - Short Time Energy) είναι η ενέργεια που περιέχει ένα σήμα, υπολογισμένη ανά μικρά, παραθυρομένα τμήματα του αρχικού σήματος πληροφορίας. Η μέτρηση της STE ενός ηχητικού σήματος μπορεί να χρησιμοποιηθεί για την διάκριση μεταξύ έμφωνης και άφωνης ομιλίας καθώς και για την ανίχνευση της μετάβασης από τη μία στην άλλη κατηγορία, αφού η ενέργεια της έμφωνης ομιλίας είναι πολύ μεγαλύτερη από αυτή της άφωνης.

Επίσης, θα κάνουμε χρήση του ρυθμού εναλλαγής προσήμου (Zero Crossing Rate) που ορίζεται ως:

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

Ο ρυθμός διελεύσεων γύρω από το μηδέν (ZCR - Zero Crossing Rate) υποδεικνύει τη συχνότητα εναλλαγής πρόσημου του πλάτους ενός σήματος. Σε μαθηματικούς όρους ένα σημείο διέλευσης από το μηδέν είναι ένα σημείο όπου το πρόσημο μιας συνάρτησης μεταβάλλεται (από θετικό σε αρνητικό ή αντίστροφα) και αναπαριστάται με την διέλευση από τον άξονα στη γραφική παράσταση της συνάρτησης.

Με αυτές τις μέτρηση θα προσπαθήσουμε να κάνουμε τον διαχωρισμό μεταξύ έμφωνων και άφωνων τμημάτων του σήματος φωνής, καθώς επίσης και της φωνής από την σιωπή. Επίσης, θα διερευνήσουμε την επίδραση του μήκους του παραθύρου στις μετρήσεις αυτές.

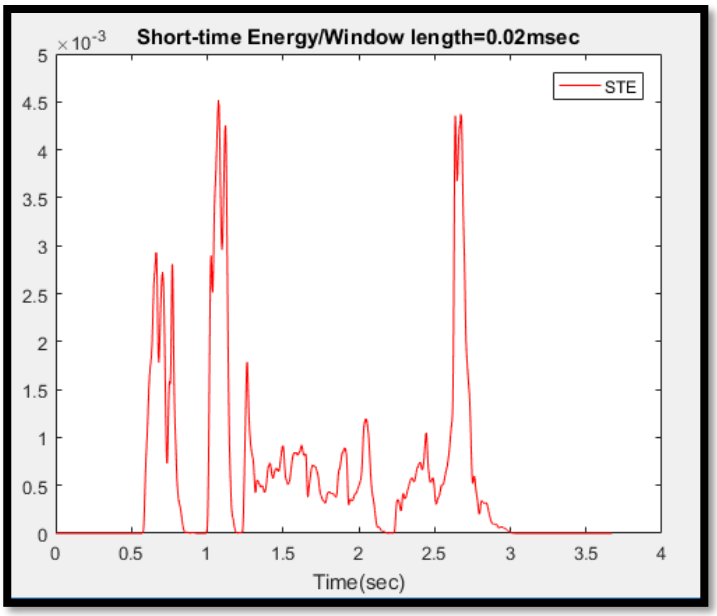
Υλοποίηση

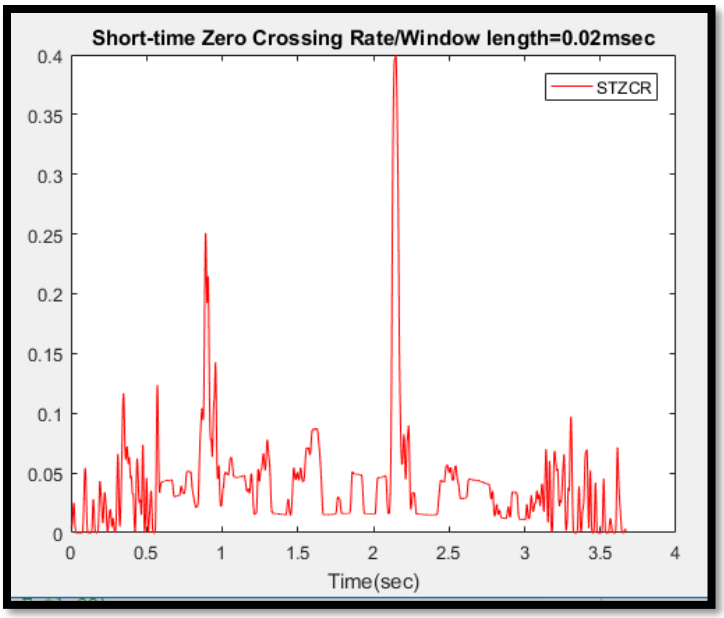
Αρχικά, θα υπολογίσουμε την ενέργεια βραχέος χρόνου, καθώς και τον ρυθμό εναλλαγής προσήμου. Σημειώνεται πως γίνεται, αρχικά, χρήση παραθύρου τύπου Hamming μήκους 20 ms.

Παρατηρούμε από τον τύπο της ενέργειας βραχέος χρόνου πως αθροίζουμε διαδοχικά τον σήμα μας πολλαπλασιασμένο με το μετακινούμενο παράθυρο (στο τετράγωνο), κάτι το οποίο περιλαμβάνει την πράξη της συνέλιξης. Γνωρίζουμε πως συνέλιξη στον χρόνο αντιστοιχεί σε πολλαπλασιασμό στο πεδίο της συχνότητας. Αυτός είναι, λοιπόν, και ο τρόπος που υλοποιείται η συνάρτηση winconv.m (window convolution), την οποία καλεί η energy.m για να υπολογίσει την ενέργεια βραχέος χρόνου.

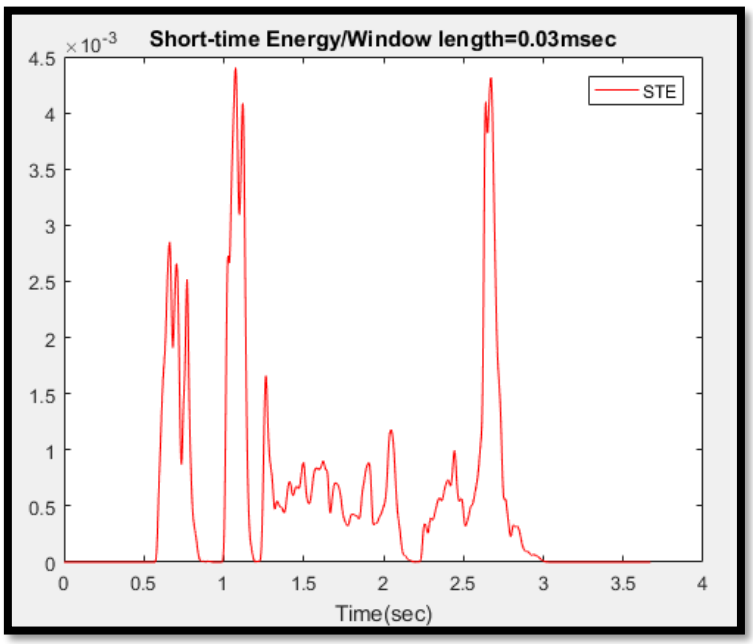
Αντίστοιχα, στον τύπο του ρυθμού εναλλαγής προσήμου αθροίζουμε διαδοχικά την διαφορά sgn[x[m]] – sgn[x[m-1]] πολλαπλασιασμένη με το μετακινούμενο παράθυρο, κάτι το οποίο εμπεριέχει την πράξη της συνέλιξης, η οποία υπολογίζεται όπως πριν. Έτσι, η συνάρτηση zerocross.m καλεί πρώτα την sgn.m (που υπολογίζει το sgn(x)) και έπειτα την winconv ,για να υπολογίζει το Zero Crossing Rate.

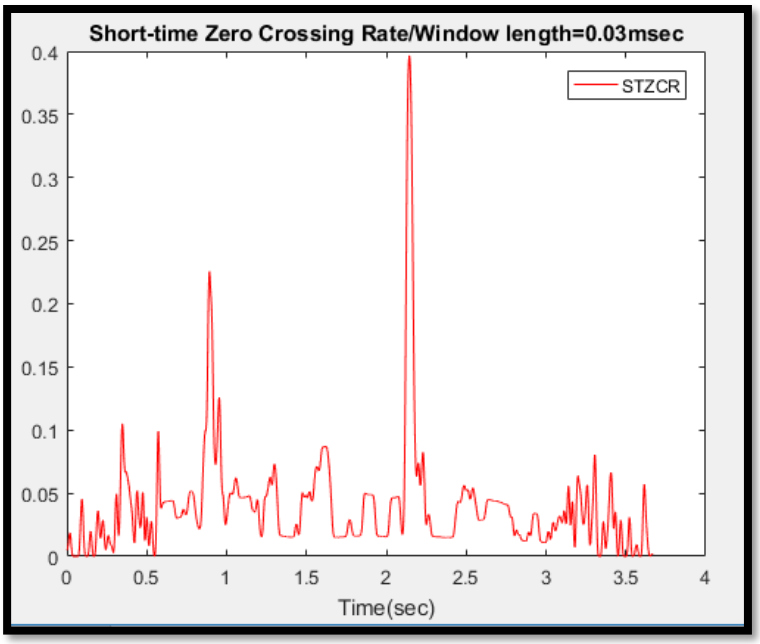
Τα διαγράμματα που προκύπτουν είναι τα ακόλουθα:





Μεγαλώνουμε το παράθυρο στα 30 ms και επαναλαμβάνουμε την παραπάνω διαδικασία:

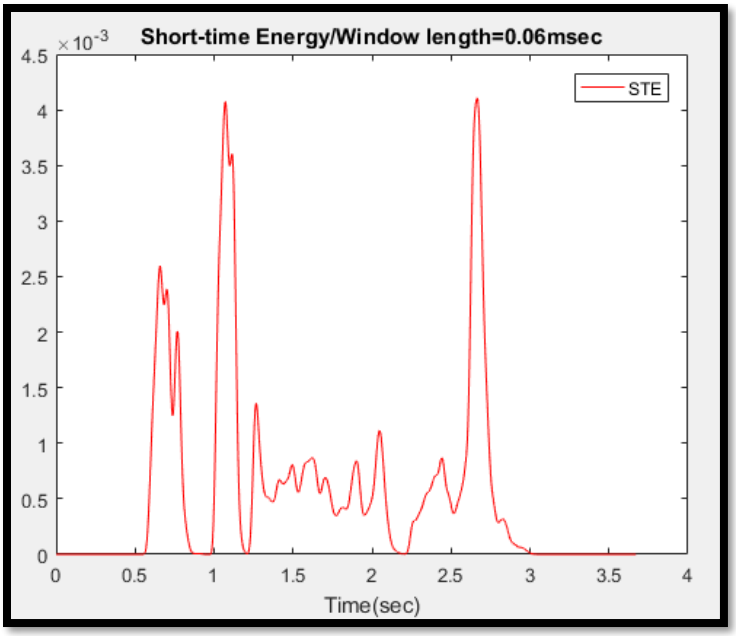


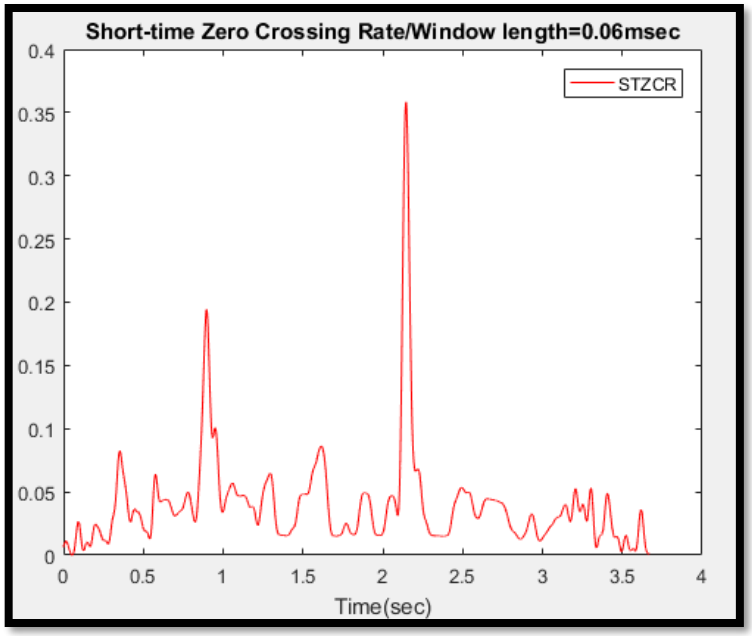


Ήδη παρατηρούμε την επίπτωση που έχει η αύξηση του μήκους του παραθύρου στις μετρήσεις μας. Βλέπουμε πως ορισμένες λεπτομέρειες (απότομες μεταβάσεις) που υπήρχαν στο προηγούμενο παράθυρο εδώ αρχίζουν να ‘’χάνονται’’, δηλαδή οι μεταβάσεις γίνονται πιο

ομαλές. Αυτό συμβαίνει, καθώς η κάθε μέτρηση (βραχέος χρόνου ή ρυθμού εναλλαγής προσήμου) δεν λαμβάνεται για κάθε δείγμα του αρχικού μας σήματος, αντιθέτως λόγω του βαθυπερατού χαρακτήρα του παραθύρου, η STE και ο ZCR είναι συχνοτικά περιορισμένοι στο εύρος ζώνης του παραθύρου (το οποίο είναι φυσικά μικρότερο από 16 kHz).

Το φαινόμενο γίνεται πιο έντονο όταν διπλασιάσουμε το μήκος του παραθύρου στα 60ms.





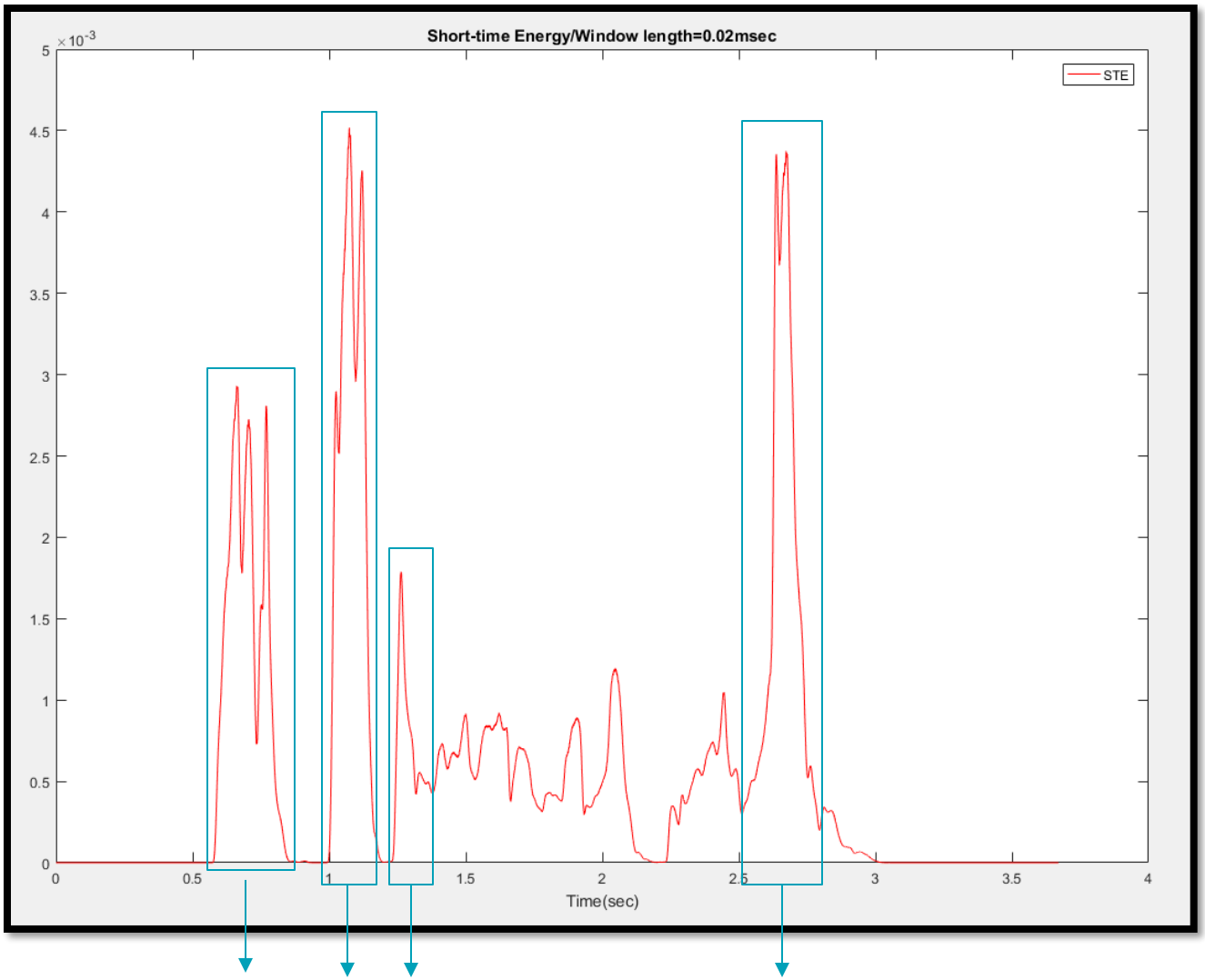
Έπειτα, θα προσπαθήσουμε να διαχωρίσουμε την φωνή από την σιωπή και τους έμφωνους από τους άφωνους ήχους κάνοντας χρήση των παραπάνω μετρήσεων.

Γνωρίζουμε πως μπορούμε να χρησιμοποιήσουμε την ενέργεια βραχέος χρόνου για να διακρίνουμε τους έμφωνους από τους άφωνους ήχους (ή τμημάτα φωνής), αφού οι άφωνοι ήχοι έχουν σημαντικά χαμηλότερη ενέργεια. Αντίθετα οι έμφωνοι ήχοι έχουν αρκετά υψηλότερη ενέργεια.

Αντίστοιχα, η μέτρηση του ρυθμού εναλλαγής προσήμου μας επιτρέπει να διαχωρίσουμε, επίσης, τους έμφωνους από τους άφωνους ήχους, καθώς επίσης και την φωνή από την σιωπή, όπως θα δούμε στην συνέχεια. Εν γένει, οι άφωνοι ήχοι έχουν σημαντικά υψηλότερο ρυθμό, ενώ το αντίθετο ισχύει για τους έμφωνους ήχους.

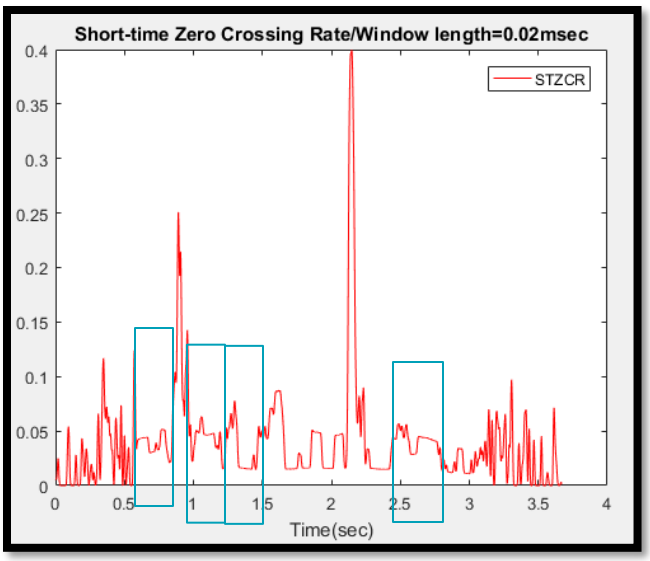
Σημειώνεται πως γίνεται χρήση παραθύρου τύπου Hamming μήκους 20 ms για καλύτερη ευκρίνεια στα διαγράμματα. Στις παρακάτω κυματομορφές έχουν απομονωθεί συγκεκριμένα τμήματα των μετρήσεων που αντιστοιχούν σε κάποιους άφωνους ή έμφωνους ήχους. Στο αρχείο Part1.m βρίσκονται σε σχόλια τα αντίστοιχα τμήματα, μαζί με μία εντολή sound ώστε να μπορούν (ξεχωριστά και όχι όλα μαζί) να ακουστούν.

Έμφωνοι ήχοι (κάθε επιλεγμένη περιοχή αντιστοιχεί σε έναν ή περισσότερους μαζί):



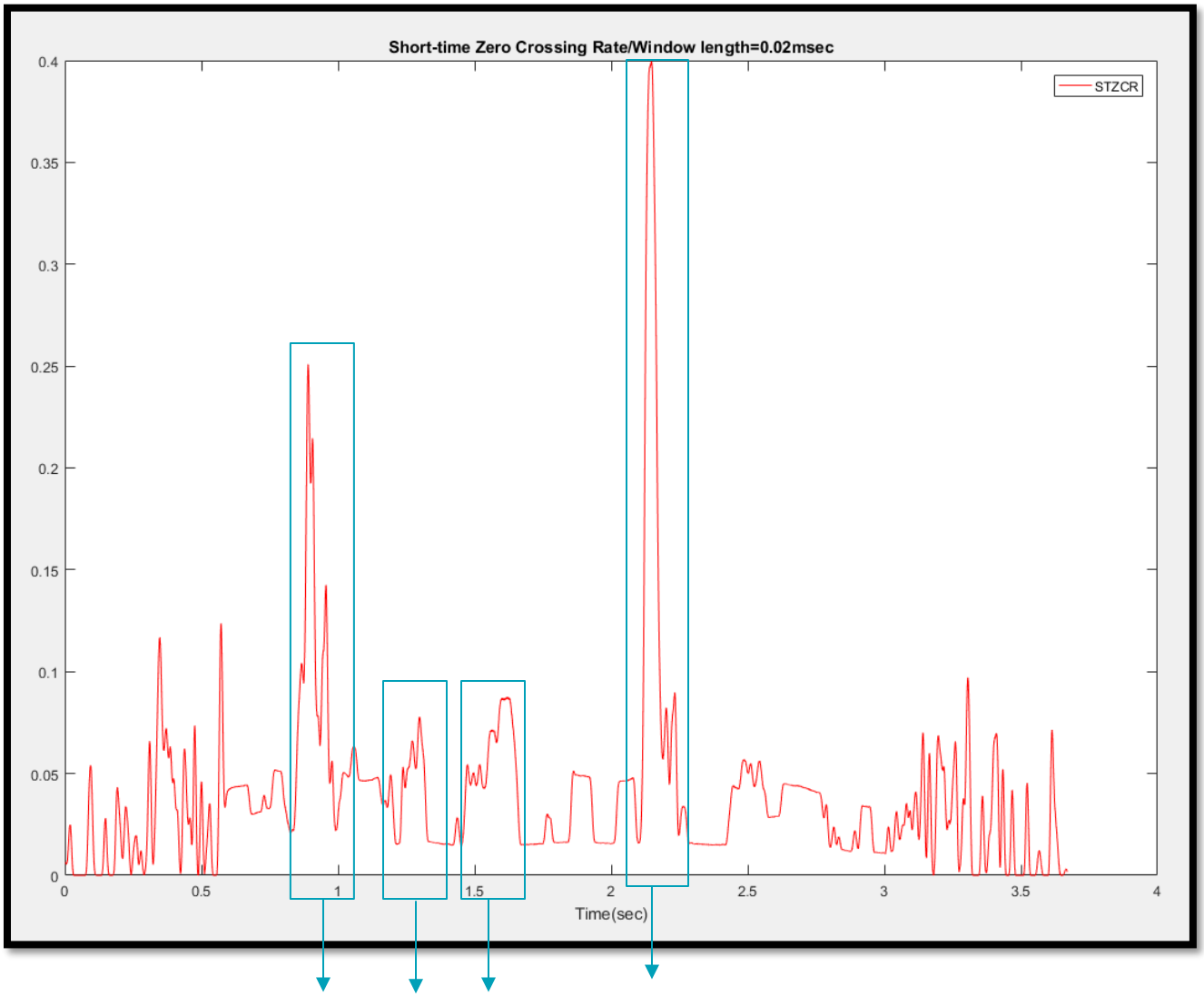
/o/,/l/,/a/ /a/,/h/ /a/,/h/ /a/,/l/,/o/

‘‘ Όλα αυτά ήτ αν η άμυνα μες στο μυαλό μου ’’



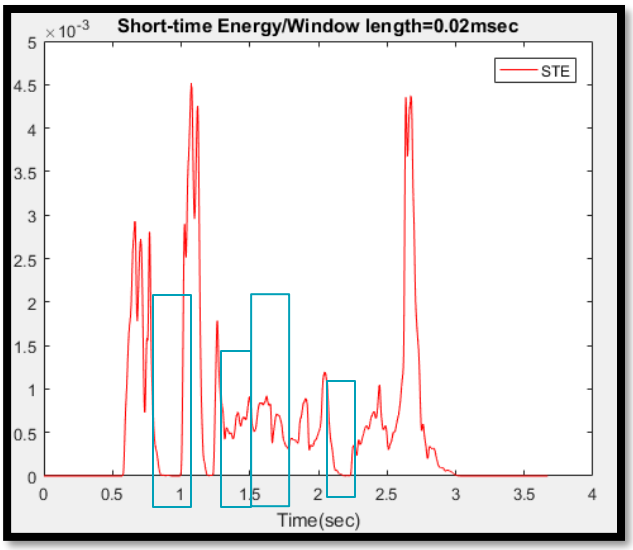
- Είναι άξιο να σημειωθεί πως παρατηρώντας το διάγραμμα ZCR θα δούμε πως στις αντίστοιχες περιοχές οι έμφωνοι ήχοι έχουν μικρό (εάν όχι σχεδόν μηδενικό) ρυθμό εναλλαγής προσήμου. Αντίθετα, στις περιπτώσεις όπου το STE ήταν χαμηλό, το ZCR είναι αρκετά υψηλότερο.

Άφωνοι ήχοι (κάθε επιλεγμένη περιοχή αντιστοιχεί σε έναν ή περισσότερους μαζί):



/f/,/t/ /t/ /m/ /s/,/t/

‘‘ Όλα αυτά ήταν η άμυνα μες στο μυαλό μου ’’

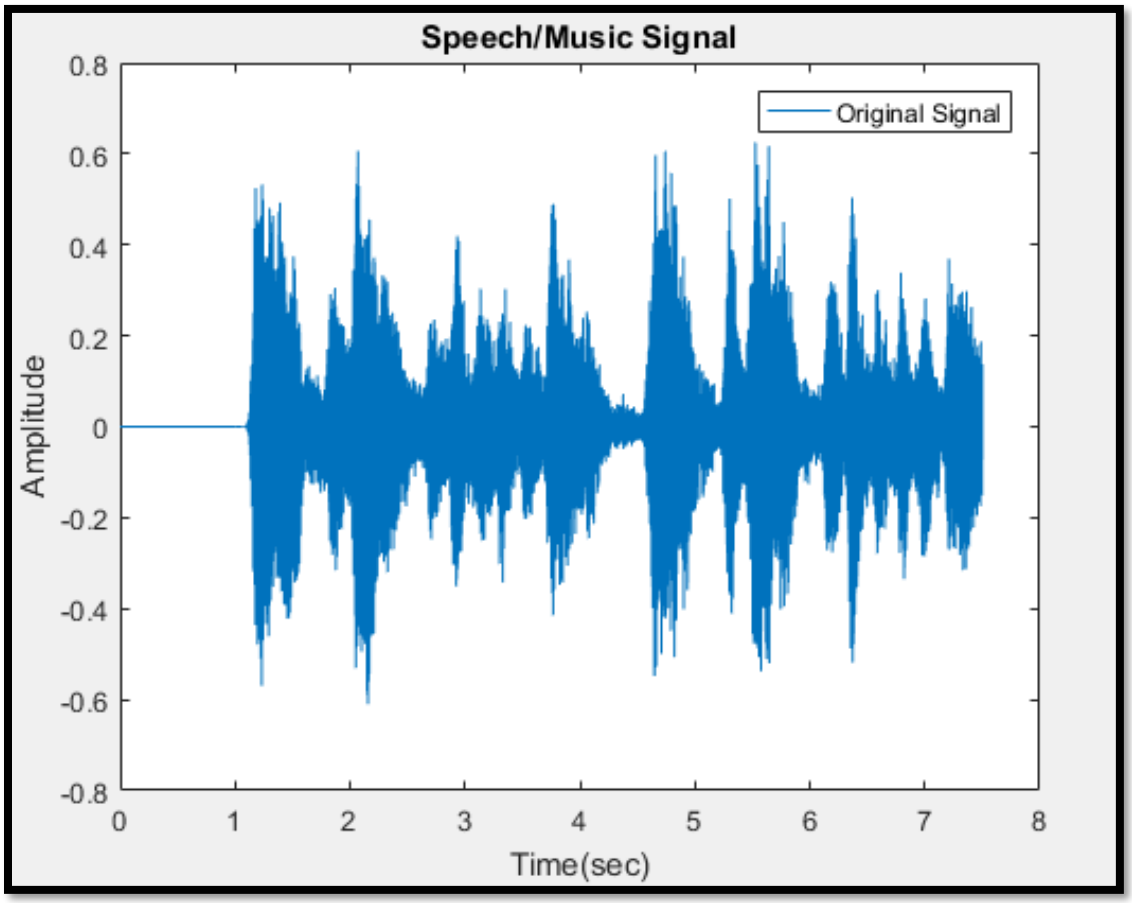


- Είναι εξίσου άξιο να προστεθεί πως παρατηρώντας το διάγραμμα STE θα δούμε πως στις αντίστοιχες περιοχές οι άφωνοι ήχοι έχουν μικρή (εάν όχι σχεδόν μηδενική) ενέργεια. Αντίθετα, στις περιπτώσεις όπου το ZCR ήταν χαμηλό, το STE είναι αρκετά υψηλότερο.

Αναφορικά με την φωνή και την σιωπή ισχύει πως πριν τα 0.6 sec παρατηρούμε (από το αρχικό σήμα) πως δεν υπάρχει φωνή, κάτι το οποίο φαίνεται και στο διάγραμμα ενέργειας όπου έχουμε μηδενική STE. Ωστόσο, βλέπουμε πως το ZCR κάνει πολλές, απότομες και μικρές διακυμανσεις, που σημαίνει πως ενώ δεν έχουμε φωνή στο συγκεκριμένο τμήμα υπάρχει θόρυβος στην συγκεκριμένη ηχογράφηση. Ο θόρυβος δεν φαίνεται στην ενέργεια, αλλά οι περιελίξεις που κάνει γλυρω από το μηδέν ανιχνεύονται από το ZCR και φαίνονται στο διάγραμμα. Το ίδιο ακριβώς ισχύει και για το τμήμα μετά τα 3 sec. Η ενέργεια είναι μηδενική, καθώς και το αρχικό σήμα φωνής σταματάει, ενώ το ZCR έχει τις γνωστές απότομες μεταβολές λόγω θορύβου.

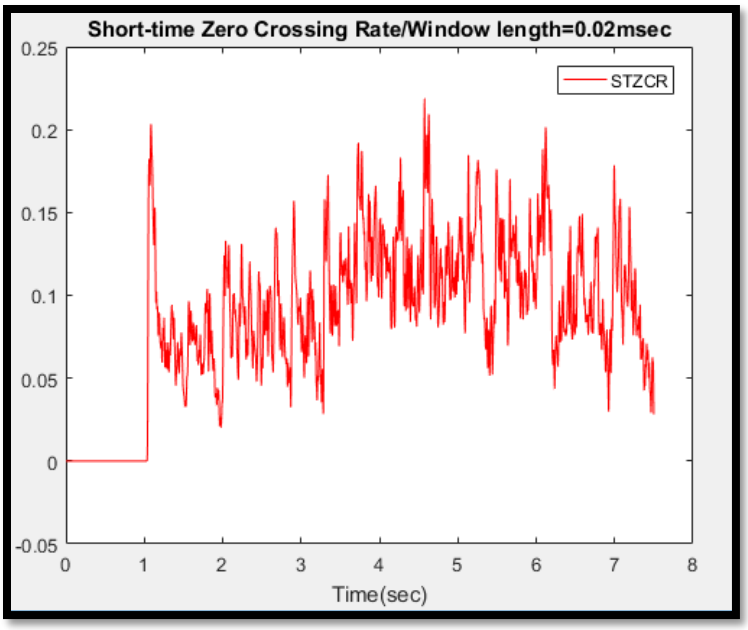
1.2)

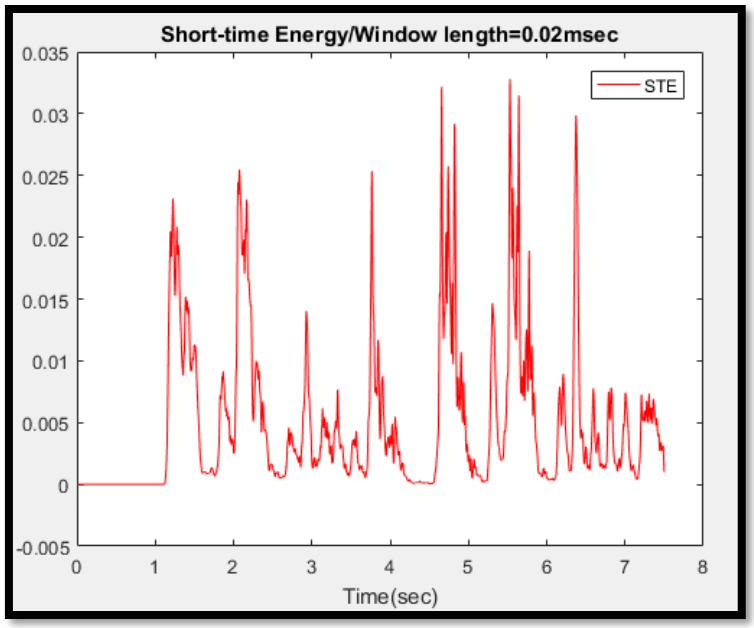
Ακολουθώντας την ίδια διαδικασία πρόκειται να αναλύσουμε με παρόμοιο τρόπο το σήμα μουσικής που βρίσκεται στο αρχείο music.wav. Το σήμα σε συνάρτηση με το χρόνο είναι το ακόλουθο:



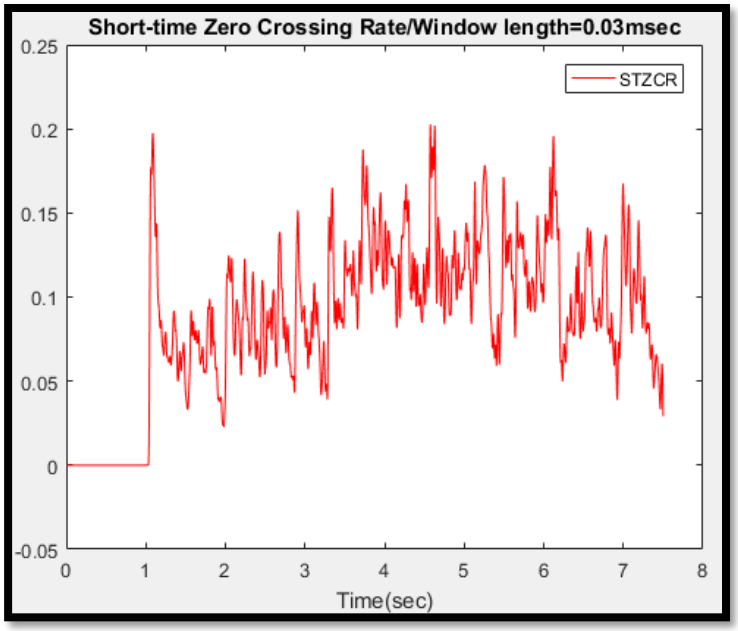
Αρχικά, θα υπολογίσουμε την ενέργεια βραχέος χρόνου, καθώς και τον ρυθμό εναλλαγής προσήμου. Σημειώνεται πως γίνεται, αρχικά, χρήση παραθύρου τύπου Hamming μήκους 20 ms.

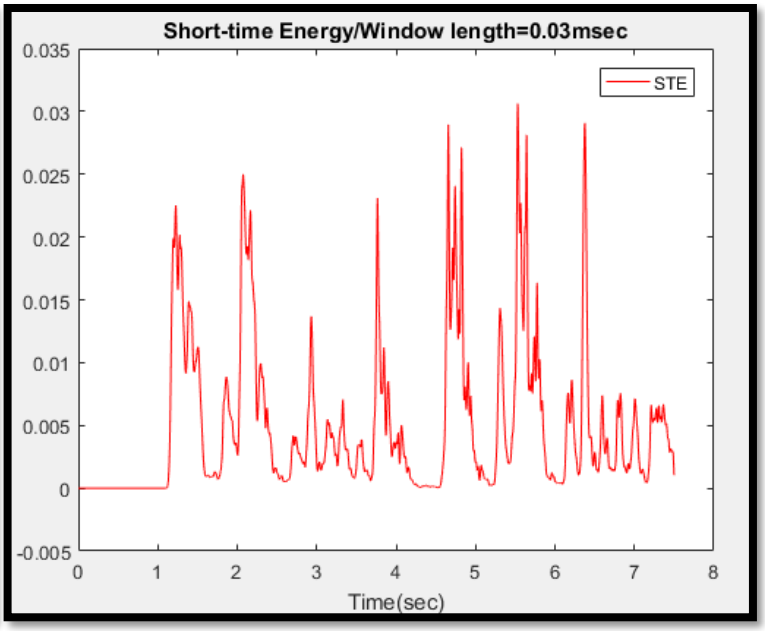
Τα διαγράμματα που προκύπτουν είναι τα ακόλουθα:



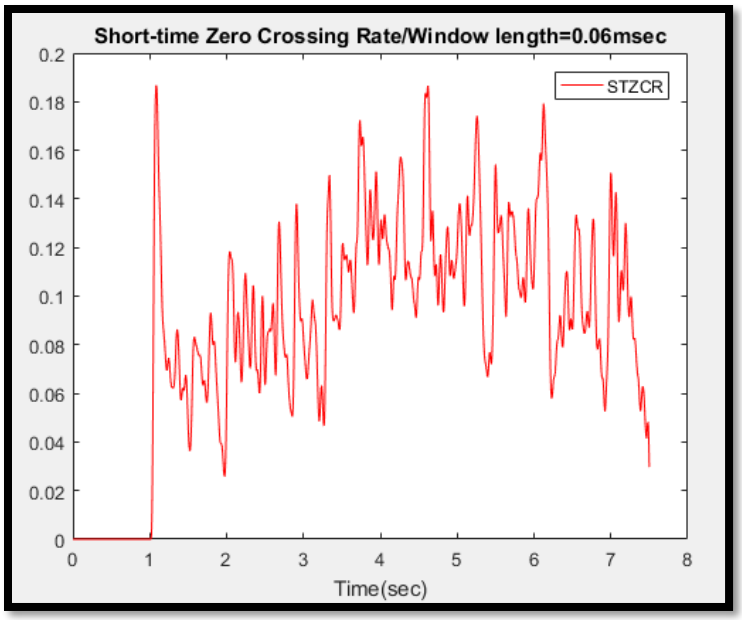


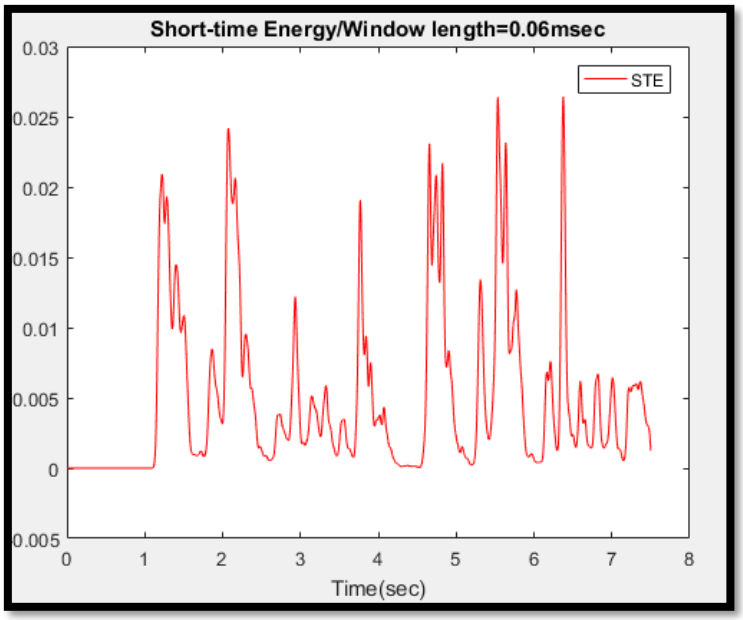
Μεγαλώνουμε το παράθυρο στα 30 ms και επαναλαμβάνουμε την παραπάνω διαδικασία:





Όπως έχει ήδη εξηγηθεί οι μεταβολές στις μετρήσεις γίνονται ελαφρώς ομαλότερες με την αύξηση του μήκους του παραθύρου. Το φαινόμενο γίνεται πιο έντονο με μήκος παραθύρου 60ms:





Στην περίπτωση του σήματος μουσικής θα προσπαθήσουμε να ερμηνεύσουμε τα δύο διαγράμματα, μιας και οι νότες/ήχοι του βιολιού και εν γένει των εγχόρδων δεν μπορούν να κατηγοριοποιηθούν σε έμφωνους και άφωνους ήχους. Σημειώνεται πως θα γίνει παραθύρου μήκους 20 msec για την ακόλουθη ανάλυση.

Αρχικά, παρατηρούμε πως το βιολί δεν εμφανίζεται μέχρι το 1 sec, ενώ στο διάστημα που μεσολαβεί η ενέργεια βραχέος χρόνου είναι μηδενική (σιωπή), καθώς και ο ρυθμός εναλλαγής προσήμου, κάτι που δηλώνει την απουσία (ή την μικρή παρουσία) θορύβου.

Όταν οι νότες αρχίζουν να εμφανίζονται και μέχρι την λήξη του σήματος θα παρατηρήσουμε πως η ενέργεια αυξάνεται μόνο όταν έχουμε έναν ήχο του εγχόρδου. Στα διαστήματα που εμφανίζεται μια νότα, η ενέργεια είναι μη μηδενική και υψηλή, ενώ στα ενδιάμεσα διαστήματα σχεδόν μηδενική ή πολύ μικρή.

Αναφορικά με τον ρυθμό εναλλαγής προσήμου, βλέπουμε πως από την στιγμή που οι ήχοι του βιολιού εμφανιστούν ο ZCR είναι διαρκώς μη μηδενικός και παράλληλα παρουσιάζει απότομες μεταβολές, όπως μας δείχνουν και τα σχήματα παραπάνω. Αυτό οφείλεται στο γεγονός ότι στα έγχορδα (και άρα και το βιολί), από την στιγμή που μία νότα θα παιχτεί, οι χορδές του βρίσκονται σε μία διαρκή ταλάντωση για να την κάνουν να ηχήσει. Έτσι, ουσιαστικά, στο σήμα μουσικής δεν υπάρχει κανένα τμήμα του που να μην έχει κάποια χορδή/νότα να ταλαντώνεται/παίζει , οπότε και το ZCR που ανιχνεύει τις περιελίξεις γύρω από το μηδέν δεν θα είναι ποτέ μηδενικό.

Μέρος 2ο - Ανάλυση και Σύνθεση Σήματος με τον Μετ/σμό Fourier Βραχέος Χρόνου (STFT)

2.1)

Ο μετασχηματισμός Fourier βραχέος χρόνου (Short – Time Fourier Transform--STFT) είναι ένας μετασχηματισμός που χρησιμοποιείται για τον προσδιορισμό της ημιτονοειδούς συχνότητας ενός σήματος, καθώς αλλάζει με την πάροδο του χρόνου. Εν γένει, μας βοηθάει να μελετήσουμε το συχνοτικό περιεχόμενο ενός χρονικά μεταβαλλόμενου σήματος, αναλύοντάς το σε μικρά χρονικά διαστήματα ίσης διάρκειας και υπολογίζοντας χωριστά το μετασχηματισμό Fourier Διακριτού Χρόνου (Discrete-Time Fourier Transform) σε καθένα από αυτά.

Σε αυτήν την άσκηση θα υλοποιήσουμε τον STFT με την βοήθεια του οποίου θα παρατηρήσουμε πως μεταβάλλεται το συχνοτικό περιεχόμενο του φωνητικού σήματος **speech utterance.wav** (η συχνότητα δειγματοληψίας είναι 16 kHz) με το πέρασμα του χρόνου.
Η μαθηματική έκφραση του STFT είναι η ακόλουθη:

$$STFT(\tau, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-\tau]e^{-j\omega n},$$
$$\text{όπου } \tau \in Z \text{ και } w[n] \text{ το παράθυρο της επιλογής μας}$$

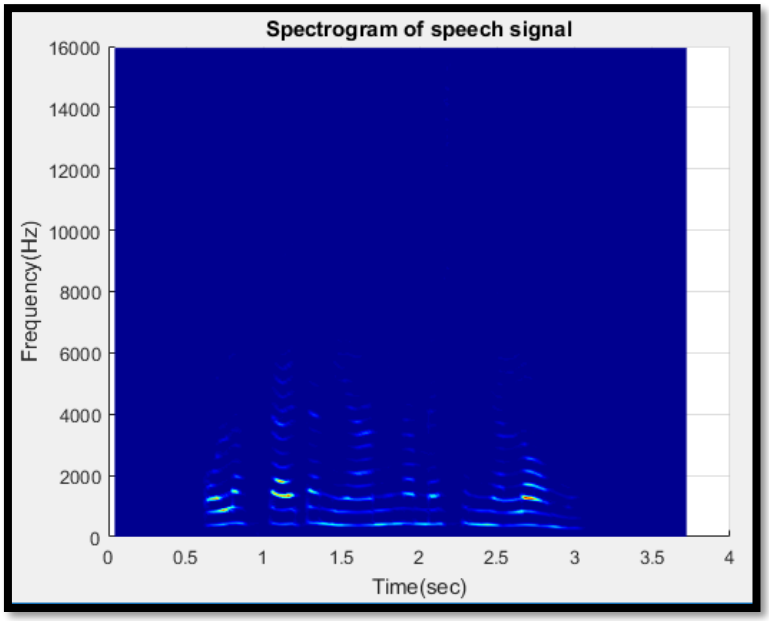
Υλοποίηση

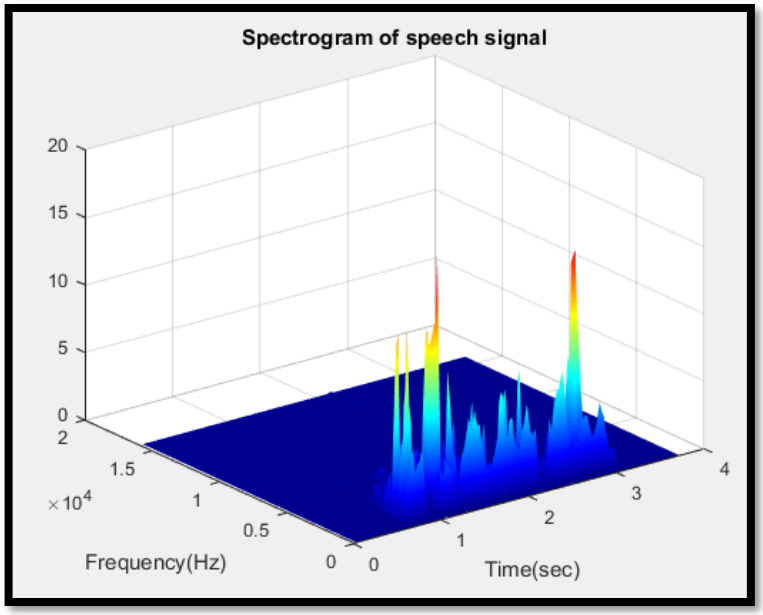
Στην MATLAB, για την υλοποίηση του STFT, χρησιμοποιείται ο DFT , οπότε η συχνότητα ω του αρχικού σήματος φωνής δειγματοληπτείται στα $\omega_k = \frac{2\pi k}{N}$, όπου N το μήκος του DFT μεγαλύτερο ή ίσο από το μήκος L του παραθύρου που θα χρησιμοποιήσουμε. Για την υλοποίηση του STFT,επίσης, επιλέγουμε το είδος του κυλιόμενου παραθύρου να είναι Hamming, το μήκος του παραθύρου L στα 40msec, καθώς και το βήμα ολίσθησης του παραθύρου R στα 20msec (σημαίνει επικάλυψη παραθύρου 50%).

Υλοποιούμε την συνάρτηση με όνομα mySTFT.m η οποία παίρνει σαν είσοδο το σήμα φωνής και τις μεταβλητές του παραθύρου και επιστρέφει τον Μετ/σμό STFT. Βρίσκουμε τον αριθμό των χρονικών παραθύρων και κάθε ένα από αυτά το πολλαπλασιάζουμε με το αρχικό μας σήμα, αφού πρώτα σιγουρευτούμε πως τα παράθυρα είναι κατάλληλα κεντραρισμένα μεταξύ τους. Τέλος, υπολογίζουμε τον DFT του παραθυροποιημένου σήματος κάνοντας χρήση της συνάρτησης fft. Ο διδιάστατος πίνακας που προκύπτει στο τέλος, αντιστοιχεί στον Μετ/σμό STFT(τ ,ω_k).

2.2)

Αφότου υπολογίσουμε τις συχνότητες συνεχούς χρόνου $f_k = \frac{\omega_k f_s}{2\pi}$, αναπαραστούμε το πλάτος |STFT(τ, f)| (σπεκτρογράφημα) με τις κατάλληλες τιμές στους άξονες του χρόνου. Σημειώνεται πως γίνεται χρήση της συνάρτησης **surf** , ο οποία μας επιτρέπει να εμφανίσουμε το σπεκτρογράφημα σε δυσδιάστατη ή τρισδιάστατη μορφή με την προσθήκη ή την αφαίρεση της εντολής **view(0, 90).** Τα διαγράμματα που προκύπτουν είναι τα παρακάτω:



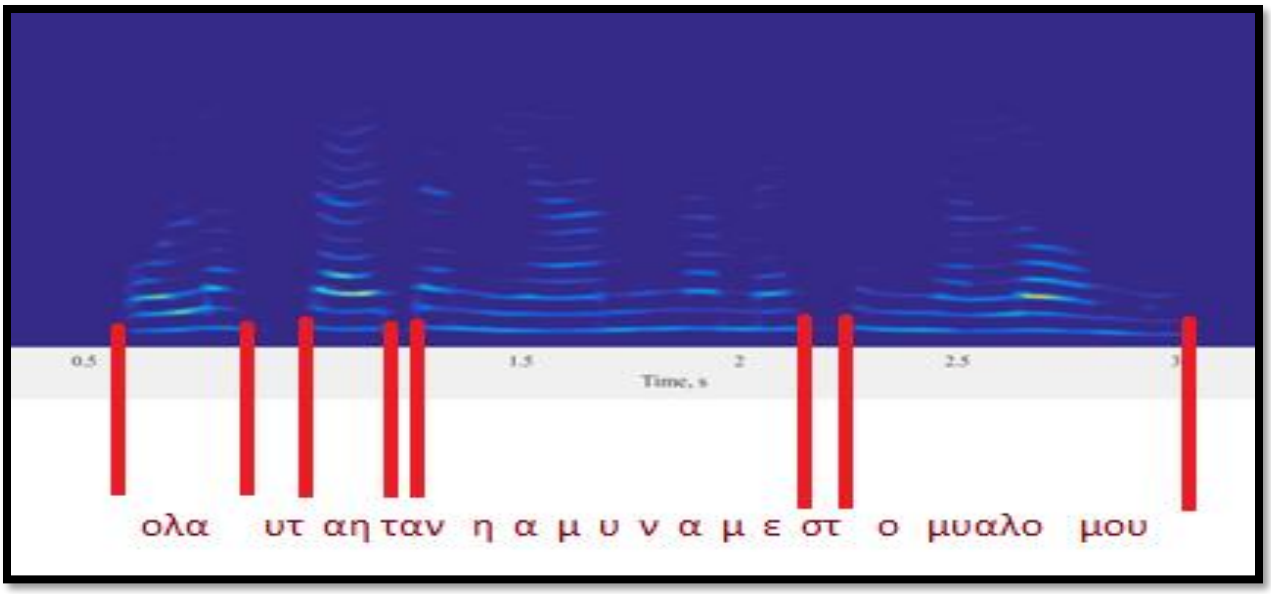


Στην ομιλία, η πηγή του ήχου, παρέχεται κυρίως από τη δόνηση των φωνητικών πτυχών. Από μαθηματική άποψη,η δόνηση φωνητικών πτυχών είναι πολύπλοκη και αποτελείται από την θεμελιώδη συχνότητα και τις αρμονικές.(Οι αρμονικές προκύπτουν πάντοτε ως ακέραια πολλαπλάσια της θεμελιώδους (x1, x2, x3, κλπ).

Στο διάγραμμα του πλάτους οι γραμμές που βλέπουμε ονομάζονται διαμορφωτές (formants) δηλαδή η συγκέντωση ακουστικής ενέργειας γύρω από μια συγκεκριμένη συχνότητα στο ηχητικό κύμα της ομιλίας. Από τεχνική άποψη, αντιπροσωπεύει ένα σύνολο παρακείμενων αρμονικών που ενισχύονται από έναν συντονισμό σε κάποιο μέρος της φωνητικής οδού.

Η διαφορά,λοιπόν, των φωνηέντων από τα σύμφωνα παρατηρείται στο γεγονός ότι τα σύμφωνα δεν έχουν διαμορφωτές. Οπότε από το διάγραμμα τα διαστήματα που υπάρχουν οι διαμορφωτές αντιστοιχούν σε φωνήεντα ενώ στα κενά βρίσκονται σύμφωνα.

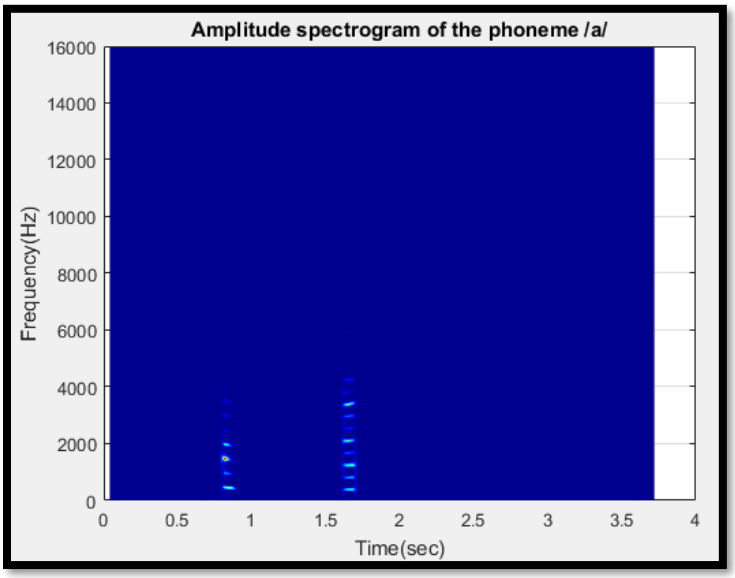
Η διαφορά φαίνεται καλύτερα στο ακόλουθο διάγραμμα:

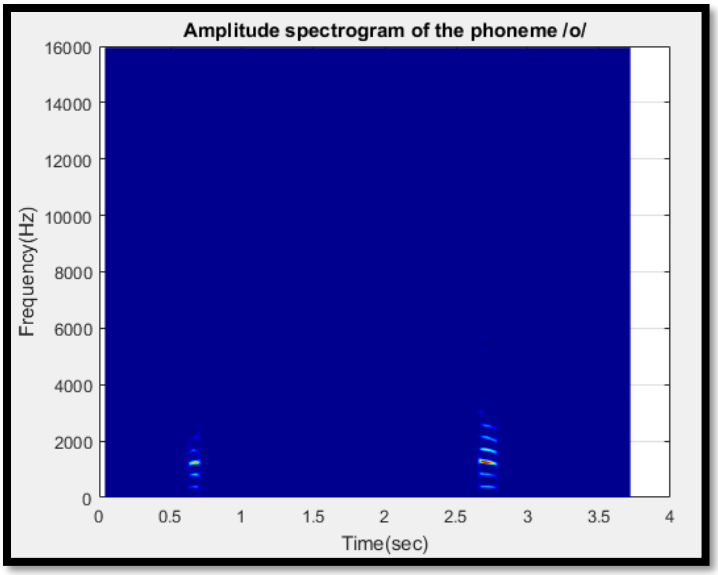


Παρατηρούμε πως τα διαστήματα που αποτελούνται από δύο συνεχόμενα σύμφωνα (τα οποία καταλαμβάνουν περισσότερο χρόνο) δεν διαθέτουν διαμορφωτές. Στα υπόλοιπα διαστήματα αν υπάρχουν σύφωνα, αυτά δεν έχουν μεγάλη διάρκια, οπότε τα φωνήεντα υπερισχύουν και για αυτό εντοπίζονται διαμορφωτές . Όμως επιρεάζονται σε έναν βαθμό αφού βλέπουμε διακυμάνσεις στην ένταση και στο πλάτος.

Σημειώνεται πως η στοματική κοιλότητα, ανάλογα με το σχήμα που έχει μια συγκεκριμένη στιγμή, λειτουργεί σαν φίλτρο του οποίου η απόκριση διαφέρει στις αρμονικές συχνότητες που παράγονται στις φωνητικές χορδές. Κάποιες από αυτές τις ενισχύει και κάποιες τις εξασθενεί. Οι αρμονικές συχνότητες που ενισχύονται είναι και οι διαμορφωτές (formants) που αναφέρθηκαν, επειδή ακριβώς διαμορφώνουν την ποιότητα του φωνήεντος.

Ακολούθως απομονώνουμε δύο χρονικά τμήματα τα οποία αντιστοιχούν στο φωνήεν /α/ και άλλα δύο για το φωνήεν /ο/ της δοσμένης πρότασης. Κάνοντας χρήση και της συνάρτησης sound είμασταν σε θέση να εντοπίσουμε ακριβώς τα χρονικά διαστήματα των ζητούμενων φωνηέντων. Έπειτα, υπολογίζουμε τον STFT αυτών των τμημάτων, μηδενίζοντας το υπόλοιπο σήμα, με αποτέλεσμα να προκύψουν τα ακόλουθα διαγράμματα:

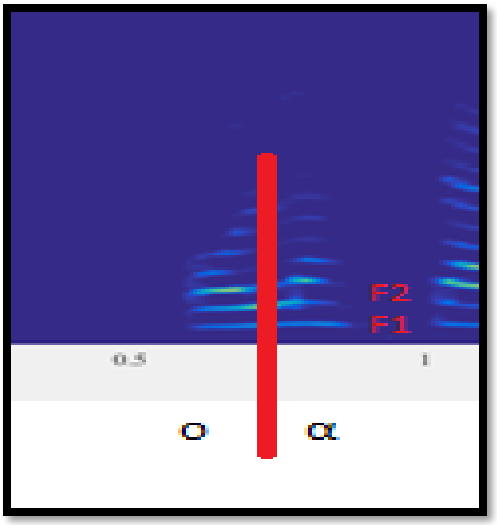




Από τα διαγράμματα παρατηρούμε ότι τα δύο φωνήεντα ως προς το πλάτος δεν διαφέρουν πολύ ,κάτι που το περιμέναμε αφού είναι και τα δύο στην ίδια κατηγορία έμφωνων ήχων. Οι διαφορές στη συχνότητα που παρατηρούμε, οφείλονται στις κινήσεις της γλώσσας κυρίως, γιατί αλλάζουν το σχήμα της φωνητικής οδού και, κατά συνέπεια τις αντηχήσεις της, γεγονός που έχει ως αποτέλεσμα τις διαφορετικές συχνότητες των διαμορφωτών κάθε φωνήεντος.

Ο πρώτος διαμορφωτής έστω F1 σχετίζεται αντιστρόφως με το ύψος της γλώσσας δηλαδή όσο ψηλότερα βρίσκεται η γλώσσα στη στοματική κοιλότητα, τόσο χαμηλότερα είναι ο F1. Αντίθετα ο δεύτερος διαμορφωτής F2 σχετίζεται με το πόσο οπίσθια είναι η θέση της γλώσσας δηλαδή όσο πιο πίσω βρίσκεται η γλώσσα τόσο χαμηλότερα βρίσκεται ο F2.

Οπότε με βάσει τα παραπάνω περιμένουμε ότι το F1 στο /α/ και στο /ο/ θα βρίσκεται στην ίδια συχνότητα και το F2 του /α/ θα είναι πιο ψηλά από του /ο/. Αυτό φαίνεται και από τον παρακάτω απόκομμα του πρώτου διαγράμματος :



- Δύο ακόμα διαφορές παρατηρούνται στην διάρκεια και στην ένταση των φωνηέντων. Το /ο/ ως πιο χαμηλό φωνήεν έχει μεγαλύτερη διάρκεια αλλά μικρότερη ένταση από το /α/ .

2.3)

Μια χρήσιμη ιδιότητα του Μετ/σμού STFT, είναι ότι μπορούμε να ανασυνθέσουμε το αρχικό σήμα, δεδομένου ότι τηρούνται κάποιες συνθήκες μεταξύ του L και του R, όπως θα δούμε στη συνέχεια, για το συγκεκριμένο είδος παραθύρου. Σε αυτό το ερώτημα θα υλοποιήσουμε τον αντίστροφο Μετ/σμό ISTFT (Inverse STFT), για να ανακτήσουμε πίσω στο χρόνο το αρχικό σήμα και να το ακούσουμε.

Δημιουργούμε την συνάρτηση με όνομα myISTFT.m η οποία παίρνει σαν είσοδο το μετασχηματισμένο σήμα φωνής και τις μεταβλητές του παραθύρου και επιστρέφει το ανακατασκευασμένο σήμα. Αρχικά, υπολογίζουμε τον αριθμό των δειγμάτων του ανακατασκευασμένου σήματος σύμφωνα με το μήκος και την επικάλυψη του παραθύρου.

Έπειτα, για την ανακατασκευή του αρχικού σήματος, ακολουθείται η αντίστροφη διαδιακασία. Για την ανακατασκευή του κάθε παραθύρου, εφαρμόζουμε αντίστροφο IDFT μετ/σμό με χρήση της συνάρτησης ifft. Για να είναι επιτυχής η ανακατασκευή του αρχικού σήματος από τα επιμέρους ανακατασκευασμένα γειτονικά παράθυρα, χρησιμοποιείται η τεχνική OverLap-Add (OLA) (εξηγείται στο ερώτημα 2.4), που θεωρεί πως τα ανακατασκευασμένα πλαίσια εμφανίζουν την ίδια επικάλυψη με αυτή που θεωρήθηκε κατά τον υπολογισμό του STFT, οπότε το τελικό σήμα προκύπτει με την κατάλληλη (αφού το κάθε πλαίσιο τοποθετηθεί σωστά στο χρόνο) πρόσθεση των επικαλυπτόμενων πλαισίων.

Αφού ανακατασκευάσουμε το αρχικό σήμα φωνής, μπορούμε να το ακούσουμε με τη βοήθεια της συνάρτησης sound και το αποθηκεύουμε σαν αρχείο ήχου με το όνομα **speech_utterance_rec.wav**, κάνοντας χρήση της συνάρτησης audiowrite. Παρατηρούμε πως το ανακτασκευασμένο σήμα ακούγεται ίδιο με το αρχικό (το γιατί αναλύεται στο 2.4).

2.4)

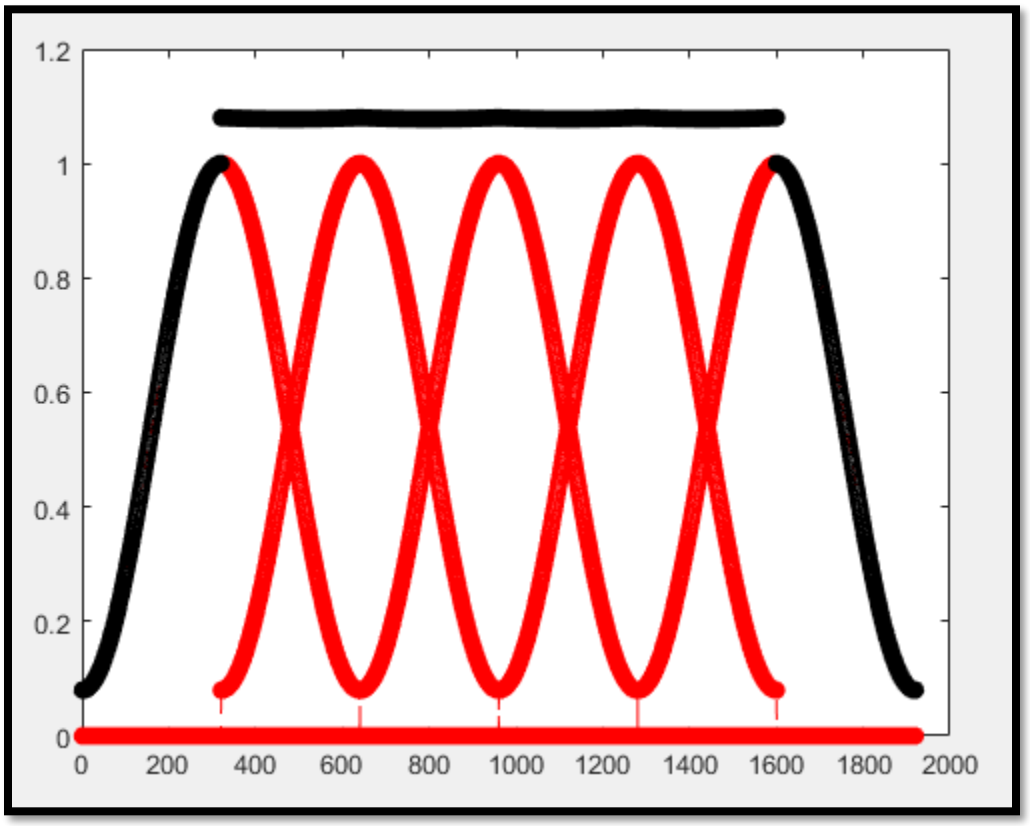
Γνωρίζουμε πως μπορούμε να αναλύσουμε ένα σήμα x[n] σε πλαίσια χρησιμοποιώντας ένα παράθυρο, όπως συμβαίνει και στην δεδομένη άσκηση με ένα παράθυρο τύπου Hamming. Μία σημαντική παράμετρος σε αυτή την ανάλυση είναι το χρονικό βήμα ανάλυσης (R), που είναι ο αριθμός των δειγμάτων μεταξύ των χρόνων έναρξης διαδοχικών πλαισίων, δηλαδή ο αριθμός των δειγμάτων κατά τον οποίο μετακινούμε κάθε επόμενο παράθυρο. Μία δεύτερη σημαντική παράμετρος είναι το μήκος του παραθύρου (L).

Για να δουλέψει η ανάλυση σε πλαίσια θα πρέπει να μπορούμε να ανακατασκευάσουμε το σήμα x[n] από τα επιμέρους επικαλυπτόμενα παράθυρα, ιδανικά με απλή πρόσθεσή τους στις αρχικές χρονικές τους θέσεις. Καταλήγουμε, λοιπόν, στην εξής συνθήκη:

$$x[n] = \sum_{m=-\infty}^{\infty} x_m[n] \ \ \alpha\nu \ \kappa\alpha\iota \ \mu\acute{o}\nu o \ \alpha\nu \sum_{m \in Z} w[n - mR] = 1, \forall n \in Z$$

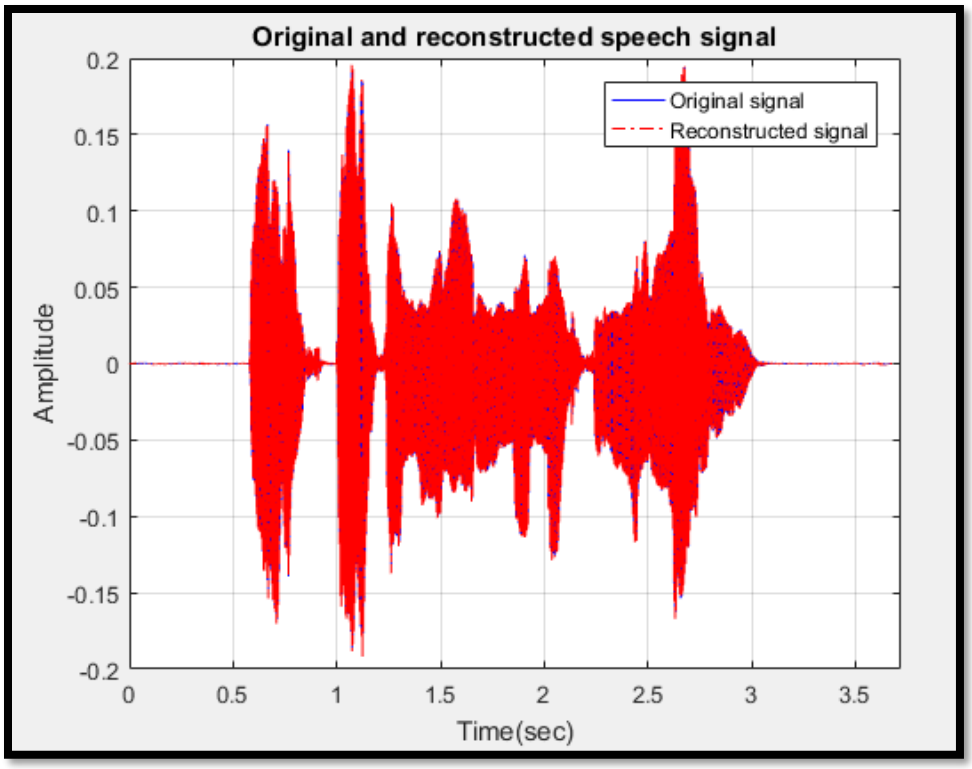
Συνεπώς, θα πρέπει να ελέγχουμε αν το άθροισμα των παραθύρων ισούται με 1 (ή κάποια σταθερά) με βάση το δεδομένο μήκος παραθύρου και overlap. Η συνθήκη αυτή θα πρέπει να ελέγχεται πριν την ανάλυση σε παραθυρωμένα επικαλυπτόμενα πλαίσια για να συμπεράνουμε αν θα είναι συνατή η ανασύνθεση του συνολικού σήματος.

Σημειώνεται πως για τον έλεγχο της συνθήκης χρησιμοποιούμε μήκος παραθύρου 40ms που αντιστοιχεί σε 640 δείγματα για συχνότητα δειγματοληψίας 16kHz και overlap 50%. Κάνουμε, λοιπόν, χρήση της συνάρτησης ola.m και αυτό είναι το αποτέλεσμα που παίρνουμε:

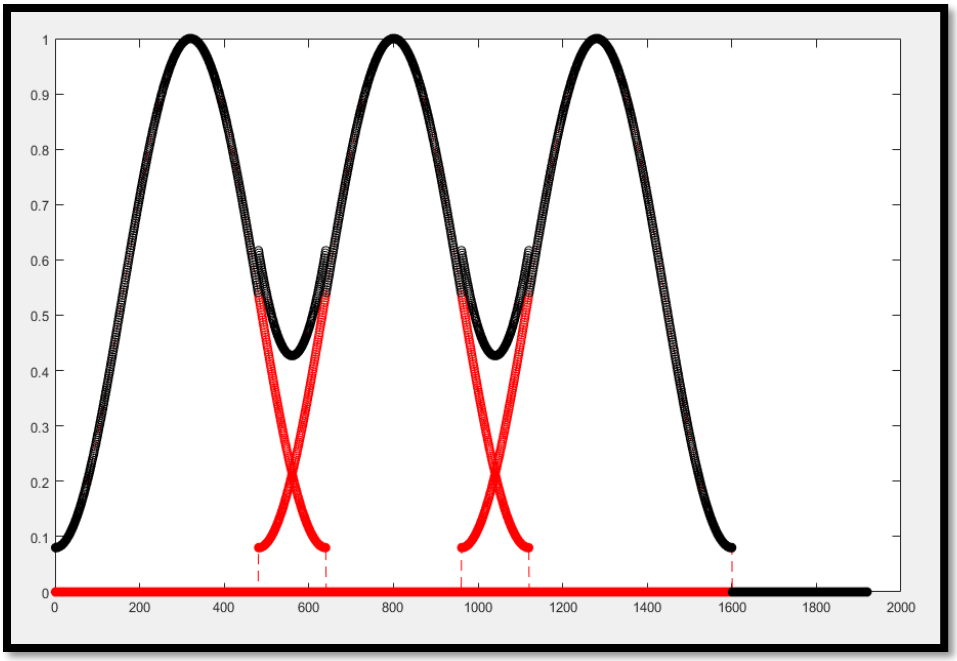


Παρατηρούμε πως το άθροισμα των επικαλυπτόμενων (στην αρχή και το τέλος δεν έχουμε επικάλυψη) παραθύρων ισούται με μία σταθερά (περίπου 1.1), οπότε η συνθήκη OLA ικανοποιείται και το σήμα μπορεί να ανακατασκευαστεί.

Παράλληλα, επιβεβαιώνουμε πως η ανακατασκευή είναι επιτυχής με το ακόλουθο διάγραμμα του αρχικού με το ανακατασκευασμένο σήμα:



Ελέγχουμε την συνθήκη OLA με παράθυρο 40msec και βήμα ανάλυσης 30 msec και παίρνουμε το εξής αποτέλεσμα:



- Παρατηρούμε πως το άθροισμα των επικαλυπτόμενων παραθύρων δεν ισούται με μία σταθερά, κάτι το οποίο επιβεβαιώνει το γεγονός πως το ανακατασκευασμένο σήμα σε αυτή την περίπτωση δεν είναι ίδιο με το αρχικό. Αντίθετα εμφανίζει μεταβολές, οι οποίες με την σειρά τους θα επηρεάζουν το τελικό σήμα. Ακούγοντάς το, βλέπουμε πως η διαφορά είναι αρκετά μεγάλη και αισθητή.

Μέρος 3ο - Φασματική Ανάλυση Ημιτονοειδών και Ανίχνευση Απότομων Μεταβάσεων με τον Μετ/σμό Fourier Βραχέος Χρόνου (STFT) και τον Μετ/σμό Wavelets (διακριτοποιημένο DT-CWT)

3.1)

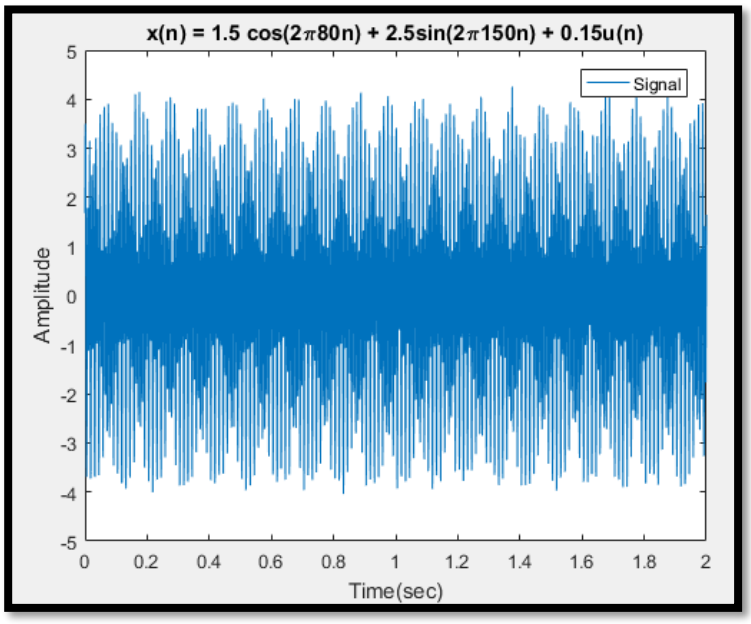
Γνωρίζουμε πως ένα από τα χαρακτηριστικά του STFT είναι ότι το μήκος L του παραθύρου w είναι σταθερό και επιλέγεται εξαρχής. Το μήκος L του παραθύρου w[n] που θα χρησιμοποιηθεί καθορίζει την σχέση μεταξύ της διακριτικής ικανότητας στη συχνότητα και της ανάλυσης στο χρόνο. Μικρό παράθυρο πετυχαίνει καλή ανάλυση στο χρόνο με όμως χειρότερη διακριτική ικανότητα στην συχνότητα, ενώ αντίστροφα, μεγάλο παράθυρο στο χρόνο πετυχαίνει καλή διακριτική ικανότητα στο πεδίο συχνοτήτων χάνοντας σε ανάλυση στον χρόνο. Συνήθεις τιμές για εφαρμογές narrow-band επεξεργασίας φωνής είναι L=30-50msec.

Αντίστοιχα, στον μετασχηματισμό των Wavelets χρησιμοποιούμε μία βασική συνάρτηση ψ(t) η οποία μπορεί να μετατοπιστεί κατά τ και να σμικρυνθεί κατά s. Στην περίπτωση του συνεχούς χρόνου ορίζεται ως:

$$CWT(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t)\psi^*(\frac{t-\tau}{s})dt$$

Αντίθετα με τον STFT, στην περίπτωση των Wavelets δεν υπάρχει ο περιορισμός του σταθερού μήκους παραθύρου όπως συνέβαινε στον STFT, όπως θα δούμε στην συνέχεια. Αυτό σημαίνει πως ο Wavelet Transform μπορεί να επιλέξει αν θα αφιερώσει την διακριτική του ικανότητα σε ορισμένες συχνότητες, ενώ σε άλλες να έχει καλύτερη ανάλυση στον χρόνο. Στην παρούσα άσκηση, εφόσον θα δουλέψουμε με ένα διακριτό σήμα, θα εφαρμοστεί η διακριτοποιημένη μορφή του DT-CWT (Discrete-time Continuous Wavelet Transform), η οποία αποτελεί παραλλαγή του γνωστου διακριτού μετ/σμού DWT.

α) Παρατίθεται το σήμα που θα αναλύσουμε:



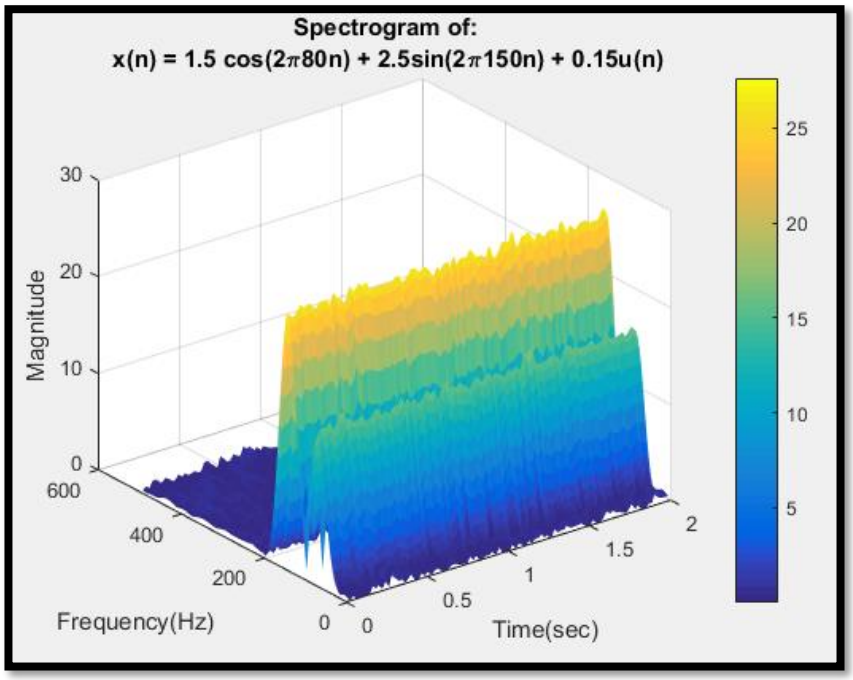
$$x(t) = 1.5\cos(2\pi 80t) + 2.5\sin(2\pi 150t) + 0.15\upsilon(t)$$

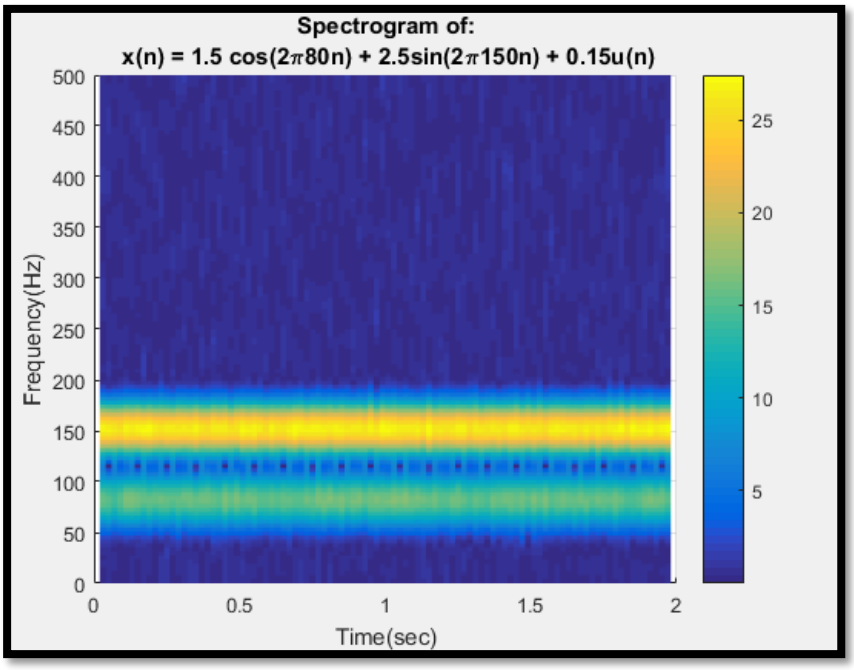
,όπου υ(t) λευκός Gaussian θόρυβος μηδενικής μέσης τιμής (χρήση συνάρτησης randn)

Η συχνότητα δειγματοληψίας ρυθμίζεται στο 1kHz και η δειγματοληψία θα γίνει στο διάστημα [0,2]sec. Το διακριτό σήμα x[n] σε συνάρτηση με τον χρόνο φαίνεται στο διπλανό διάγραμμα.

β) Για τον υπολογισμό του STFT του διακριτού σήματος x[n] επιλέγουμε μήκος παραθύρου ίσο με 40msec και επικάλυψη 20msec. Με την χρήση της συνάρτησης spectogram παίρνουμε τον STFT του σήματος (S), καθώς και τις κατάλληλες τιμές χρόνου (T) και συχνότητας (F) για τον σχεδιασμό του. Τέλος, με την συνάρτηση surf αναπαριστούμε την απόλυτη τιμή του S είτε σε 3-διάστατο χώρο είτε σε 2-διάστατο με την επιπλέον χρήση της εντολής view(0,90).

Τα διαγράμματα που προκύπτουν είναι τα ακόλουθα:

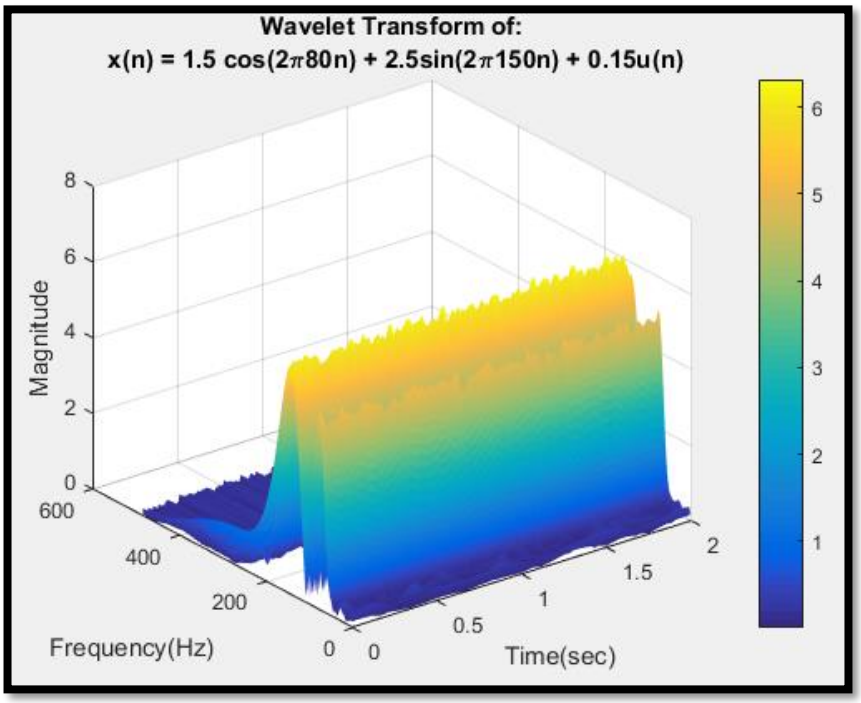


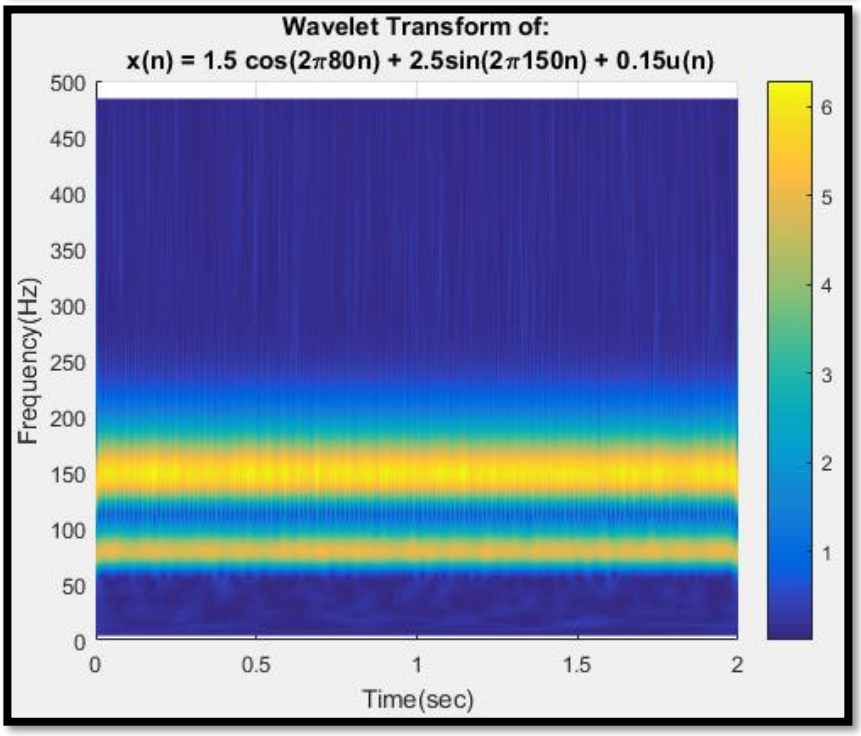


- Είναι έυκολο να παρατηρήσει κανείς την χαρακτηριστική ιδιότητα του STFT: το σταθερό παράθυρο. Αυτό φαίνεται κυρίως στο δεύτερο διάγραμμα, όπου η διακριτική του ικανότητα στην συχνότητα και η ανάλυση στο χρόνο παραμένει εξίσου σταθερή.
- Κύριος στόχος μας είναι ο εντοπισμός των βασικών συχνοτήτων του σήματός μας (80 και 150 Hz) με την προσθήκη θορύβου. Όπως θα παρατηρήσουμε και στα δύο διαγράμματα ο STFT καταφέρνει να εντοπίσει τις βασικές συχνότητες. Ταυτόχρονα, γίνεται εμφανές το tradeoff μεταξύ της διακριτής ικανότητας στο πεδίο της συχνότητας και της ανάλυσης στον χρόνο, αφού δεν έχουμε τόσο καλή ανάλυση στον χρόνο σε όλο το φάσμα συχνοτήτων.
- Γίνεται,παράλληλα, αντιληπτό πως η διακριτική ικανότητα στη συχνότητα, ενός μετ/σμού, σε συνδιασμό με την ανάλυση στον χρόνο είναι περιορισμένα από την Αρχή Αβεβαιότητας του Heisenberg, η οποία θα αναλυθεί στη συνέχεια.

γ) Για τον υπολογισμό του DT-CWT κάνουμε αρχικά χρήση της δοθείσας συνάρτησης wavescales για τον υπολογισμό των κλιμάκων s και των αντίστοιχων ψευδο-συχνοτήτων f, οι οποίες αποσκοπούν στην καλύτερη σύγκριση των δύο μετασ/μων. Έπειτα, γίνεται χρήση της συνάρτησης cwft, στην οποία περνάμε, επίσης, ως όρισμα 'Wavelet','morl' , έτσι ώστε να χρησιμοποιήσει το κυματίδιο Morlet για τον υπολογισμό του DT-CWT. Τέλος, αναπαριστούμε το πλάτος του μετασ/μού που μόλις υπολογίσαμε με την συνάρτηση surf, η οποία όπως έχει ήδη αναφερθεί μας επιτρέπει να δούμε το διάγραμμα σε μορφή 3D ή 2D.

Τα διαγράμματα που προκύπτουν είναι τα ακόλουθα:





- Όπως προηγουμένως, έτσι και τώρα εύκολα διακρίνουμε την χαρακτηριστική ιδιότητα του Wavelet Transform: την εναλλαγή μεταξύ καλύτερης ανάλυσης στο χρόνο και της καλύτερης διακριτικής ικανότητας στο πεδίο των συχνοτήτων. Κοιτάζοντας το δεύτερο διάγραμμα, παρατηρούμε πως για χαμηλές συχνότητες (κάτω από 200 Hz) υπερτερεί η διακριτική ικανότητα στη συχνότητα, κάτι που επιτρέπει στον μετασ/μό να εντοπίσει τις βασικές συχνότητες του σήματος (80 και 150 Hz) (όπως φαίνεται και στο πρώτο διάγραμμα). Αντίθετα σε υψηλότερες μη βασικές συχνότητες υπερτερεί η ανάλυση στο χρόνο, όπως φαίνεται από το δεύτερο διάγραμμα.
- Ο DT-CWT βλέπουμε πως ενώ εντοπίζει τις βασικές συχνότητες, καταφέρνει να μην ''χαραμίσει'' την διακριτική του ικανότητα για συχνότητες που δεν μας ενδιαφέρουν. Αντ'αυτού επιλέγει να έχει καλύτερη ανάλυση στο χρόνο σε αυτές.

δ) Στα πλαίσια αυτού του ερωτήματος, και οι δύο μετασχηματισμοί καταφέρνουν να εντοπίσουν τις βασικές συχνότητες, κάτι που είναι και ο πρωταρχικός στόχος μας. Άρα, στη δεδομένη περίπτωση μπορούμε να τους θεωρήσουμε ισοδύναμους/αποτελεσματικούς. Ωστόσο, παρατηρήσαμε πως ο Wavelet Transform δεν περιορίζεται από το σταθερό παράθυρο του STFT και μπορεί να επιλέξει ανάμεσα σε διακριτική ικανότητα και ανάλυση στον χρόνο ανάλογα με τη συχνότητα.

Σημειώνεται πως αυτό που περιορίζει έναν μετασχηματισμό από το να ικανοποιεί και τις δύο απαιτήσεις ταυτόχρονα είναι η Αρχή της Απροσδιοριστίας του Heisenberg που υποστηρίζει πως δεν μπορούμε να έχουμε καλό εντοπισμό στη συχνότητα και τον χρόνο την ίδια στιγμή. Συγκεκριμένα, αν Δf είναι το διάστημα στη συχνότητα και Δt το αντίστοιχο διάστημα στον χρονο πρέπει να ισχύει πάντοτε:

$$\Delta t \cdot \Delta f \geq \frac{1}{2}$$

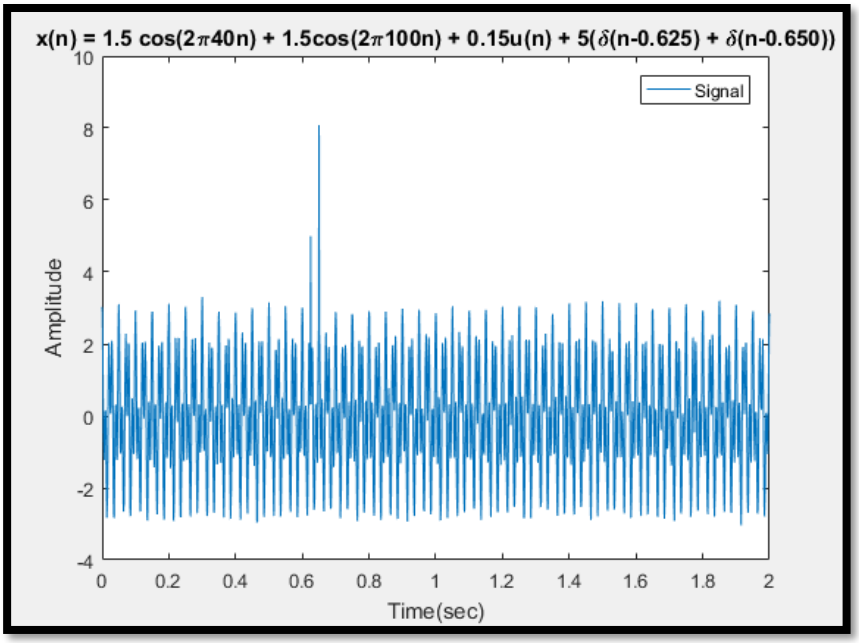
3.2)

α) Το επόμενο σήμα που θα αναλύσουμε είναι το ακόλουθο:

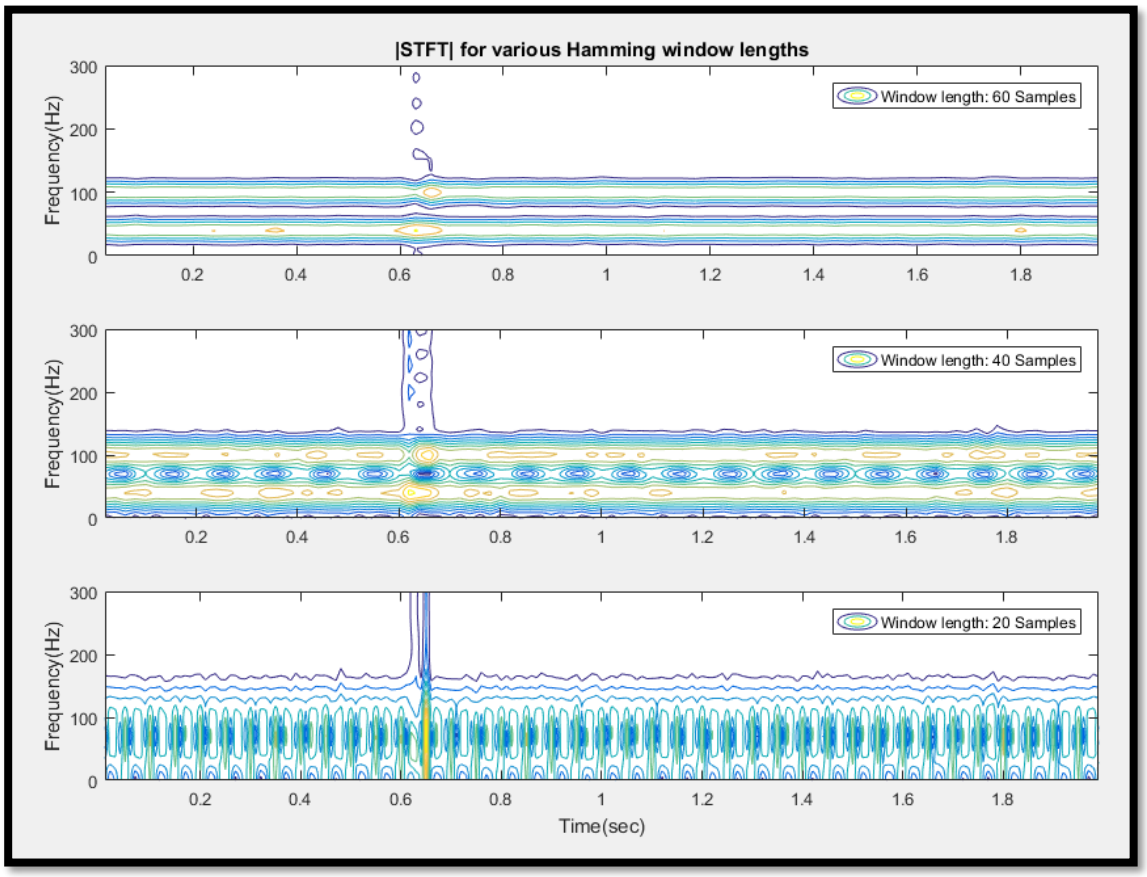
$$x(t) = 1.5\cos(2\pi 40t) + 1.5\cos(2\pi 100t) + 0.15\upsilon(t) + 5(\delta(t-0.625) + \delta(t-0.650))$$

,όπου υ(t) λευκός θόρυβος μηδενικής μέσης τιμής

Δειγματοληπτούμε με συχνότητα 1 kHz το σήμα x(t) στο διάστημα [0,2]sec και παίρνουμε το διακριτό σήμα x[n]. Σε αυτό το ερώτημα θέλουμε στην ανάλυσή μας τόσο να μελετήσουμε το συχνότικό περιεχόμενο του σήματος (2 κύριες ημιτονικές συχνότητες, όπως στο προηγούμενο ερώτημα), όσο και να εντοπίσουμε τις απότομες μεταβολές σε σύντομο χρονικό διάστημα, τις οποίες έχει το σήμα. Σημειώνεται πως επειδή οι συναρτήσεις dirac (δ(t)) τείνουν στο άπειρο, πρέπει να τις περιορίσουμε στο 0. Η γραφική παράσταση του σήματος x[n] σε συνάρτηση με το χρόνο απεικονίζεται ακολούθως:

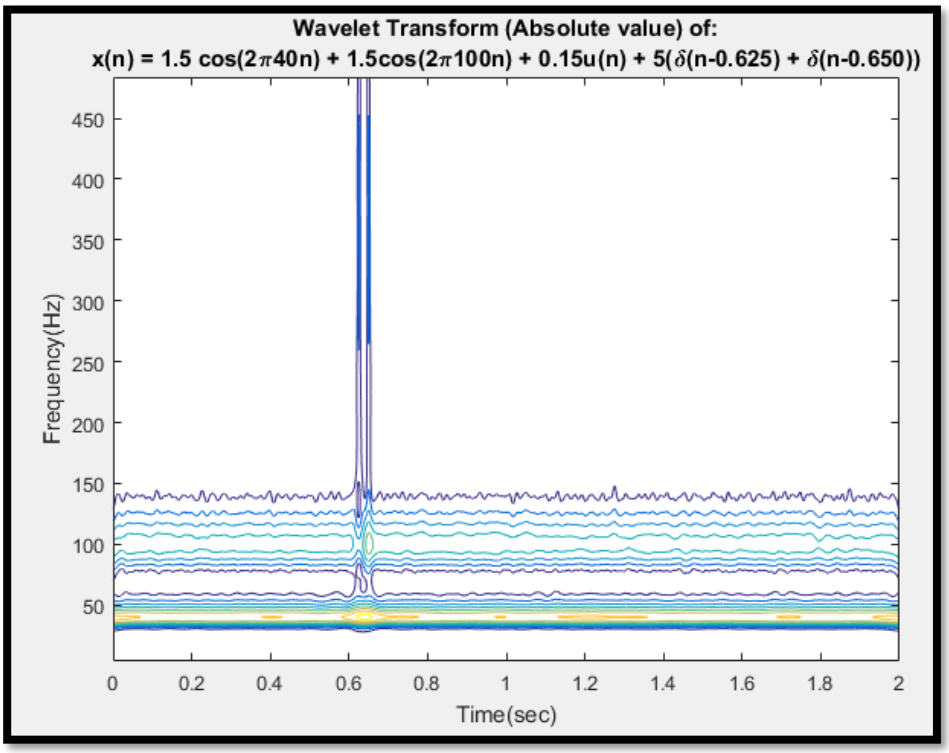


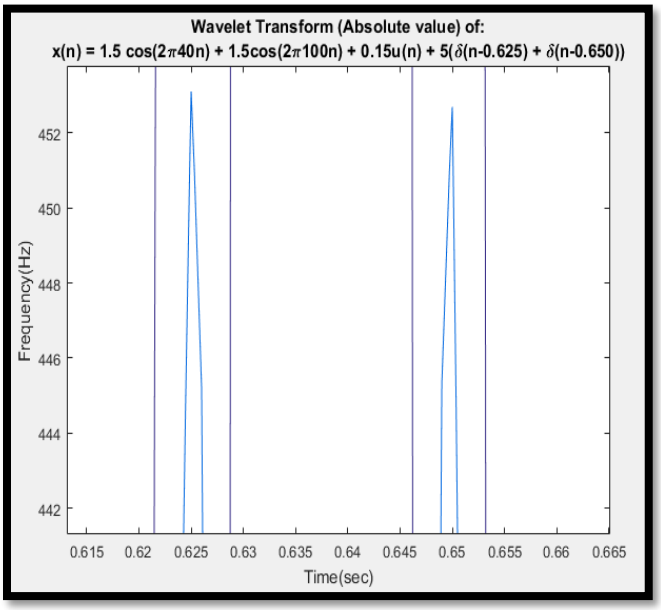
β) Για τον υπολογισμό του STFT θα χρησιμοποιήσουμε μήκοι παραθύρου: 60, 40, 20 msec και επικάλυψη 50% κάθε φορά. Όπως και στο προηγούμενο ερώτημα, αφού δημιουργήσουμε το παράθυρο τύπου Hamming με μήκος παραθύρου ένα από τα παραπάνω, κάνουμε χρήση της συνάρτησης spectogram και παίρνουμε τον STFT, τις αντίστοιχες κυκλικές συχνότητες εκφρασμένες στα πλαίσια της συχνότητας δειγματοληψίας, καθώς και τις κατάλληλες τιμές χρόνου. Τέλος, κάνοντας χρήση της συνάρτησης contour αναπαριστούμε το πλάτος του STFT σε δυσδιάστατο χώρο για τα διάφορα μήκη παραθύρου:



Παρατηρούμε, αρχικά, πως όσο αυξάνουμε το μήκος παραθύρου, τόσο ο STFT τείνει να επιλέγει καλύτερη διακριτική ικανότητα στο πεδίο συχνότητας, αλλά χειρότερη ανάλυση στο χρόνο. Όπως βλέπουμε στο πρώτο σχήμα, έχουμε αρκετά καλό εντοπισμό των βασικών συχνοτήτων του σήματός μας (40 και 100 Hz), ενώ αρκετά κακό εντοπισμό των απότομων διακριτών μεταβολών. Προχωράμε στο σχήμα με μήκος παραθύρου 40 msec, όπου βλέπουμε πως ενώ ο εντοπισμός των βασικών συχνοτήτων δυσχεραίνει, ο εντοπισμός των απότομων μεταβολών αρχίζει να βελτιώνει. Τέλος, στο τελευταίο διάγραμμα, με μήκος παραθύρου ίσο με 20 msec, γίνεται αισθητή η καλύτερη ανάλυση στο χρόνο, αφού μπορεί να εντοπίσει ελαφρώς τις απότομες μεταβολές. Αντιθέτως, η διακριτική του ικανότητα στη συχνότητα έχει χειροτερεύσει ακόμα περισσότερο με αποτέλεσμα να μην μπορεί ο STFT να εντοπίσει σωστά τις βασικές συχνότητες του σήματος.

γ) Παρόμοια με το ερώτημα 3.1, κάνουμε χρήση της συνάρτησης wavescales για να εξάγουμε τον πίνακα με τις κλίμακες s και τις ψευδο-συχνότητες f (για τον σχεδιασμό του DT-CWT), και έπειτα, καλούμε την cwft με επιλογή του Morlet κυματιδίου για να πάρουμε τον DT-CWT. Τελικά. Με την χρήση της συνάρτησης contour είμαστε σε θέση να εμφανίσουμε το πλάτος του DT-CWT σε δυσδιάστατο επίπεδο.





Θα παρατηρήσουμε πως η χαρακτηριστική ιδιότητα του Wavelet μετασχηματισμού να ενναλάσσει από καλύτερη διακριτική ικανότητα στη συχνότητα σε καλύτερη χρονική ανάλυση γίνεται εμφανής για ακόμα μία φορά. Θα παρατηρήσουμε πως σε χαμηλές συχνότητες (κάτω από 70 Hz περίπου) υπερτερεί η διακριτική ικανότητα στο πεδίο της συχνότητας, με αποτελέσμα να καταφέρνει να εντοπίσει την πρώτη κύρια συχνότητα (40Hz) με καλή ακρίβεια, ενώ δεν γίνονται εμφανής ακόμα οι απότομες μεταβολές. Όσο ανεβαίνουμε στη συχνότητα θα δούμε πως ενώ καταφέρνει να εντοπίσει με (σχεδόν) απόλυτη ακρίβεια τις διακριτές απότομες μεταβολές (ειδικά σε αρκετά υψηλότερες συχνότητες), η διακριτική του ικανότητα στην συχνότητα χάνεται με αποτέλεσμα να μην καταφέρνει να εντοπίσει (σχεδόν καθόλου) την δεύτερη βασική μας συχνότητα (100Hz).

δ) Στα πλαίσια αυτού του ερωτήματος αντιλαμβανόμαστε πως για να κρίνουμε κάποιον μετασχηματισμό θα πρέπει να κάνουμε μία επιλογή. Αν μας ενδιαφέρει κυρίως ο εντοπισμός των βασικών συχνοτήτων τότε ο μετασχηματισμός STFT θα αποτελέσει μία καλή επιλογή καθώς με ένα μεγάλο παράθυρο είμαστε σε θέση να αποκτήσουμε καλή ακρίβεια στο πεδίο της συχνότητας. Ωστόσο, πρέπει να είμαστε έτοιμοι να κάνουμε έναν συμβιβασμό με την ανάλυση στον χρόνο, όπου δεν θα είμαστε σε θέση να εντοπίσουμε τις απότομες μεταβολές των συναρτήσεων δέλτα. Ο μετασχηματισμός DT-CWT (με το επιλεγμένο κυματίδιο) ενώ είναι σε θέση να κάνει καλό εντοπισμό των χαμηλών βασικών συχνοτήτων, αδυνατεί να εντοπίσει την υψηλότερες.

Αντίστοιχα, αν μας ενδιαφέρει κυρίως ο εντοπισμός των διακριτών απότομων μεταβολών, μια καλή επιλογή θα αποτελούσε ο DT-CWT, αφού σε υψηλότερες συχνότητες επιλέγει καλή ανάλυση στον χρόνο με αποτέλεσμα να είναι σε θέση να τις εντοπίσει με πολύ καλή ακρίβεια. Ωστόσο, δεν είναι σε θέση να εντοπίσει για αυτόν ακριβώς τον λόγο όλες τις βασικές συχνότητες. Ένας μετασχηματισμός STFT με μικρό παράθυρο ενδεχομένως να μας αρκούσε σε αυτή την περίπτωση αλλά λόγω του σταθερού του παραθύρου το πιο πιθανό είναι να μην είναι σε θέση να εντοπίσει καμία άλλη βασική συχνότητα.

Εφόσον, παρουσιάστηκαν τα υπερ και τα κατά του κάθε μετασχηματισμού είναι στην διακριτική ευχαίρεια του καθενός η επιλογή κατάλληλου μετασχηματισμού, ανάλογα με τις απαιτήσεις του.

Λόγω της Αρχής της Αβεβαιότητας, που αναφέρθηκε σε προηγούμενο ερώτημα αναλυτικά, σημειώνουμε πως δεν είναι δυνατό για έναν μετασχηματισμό να πετυχαίνει τον εντοπισμό και των διακριτών απότομων μεταβολών και των βασικών συχνοτήτων, καθώς το ένα απαιτεί καλή ανάλυση στον χρόνο και το άλλο καλή διακριτική ικανότητα. Υπενθυμίζουμε πως δεν είναι δυνατό να έχουμε ταυτόχρονα καλό εντοπισμό στη συχνότητα και τον χρόνο, αφού ο Heisenberg μας περιορίζει στο ½ του γινομένου Δt*Δf.

Βιβλιογραφία

1. Oppenheim & Schafer-Discrete Time Signal Processing
2. MATLAB Documentation – MathWorks
3. Εκπαιδευτικό υλικό μαθήματος (Slides)