Ψηφιακή Επεξεργασία Σημάτων – 1η Εργαστηριακή Άσκηση 2018

Όνομα: Βλαντισλάβ

Επώνυμο: Λέντελ

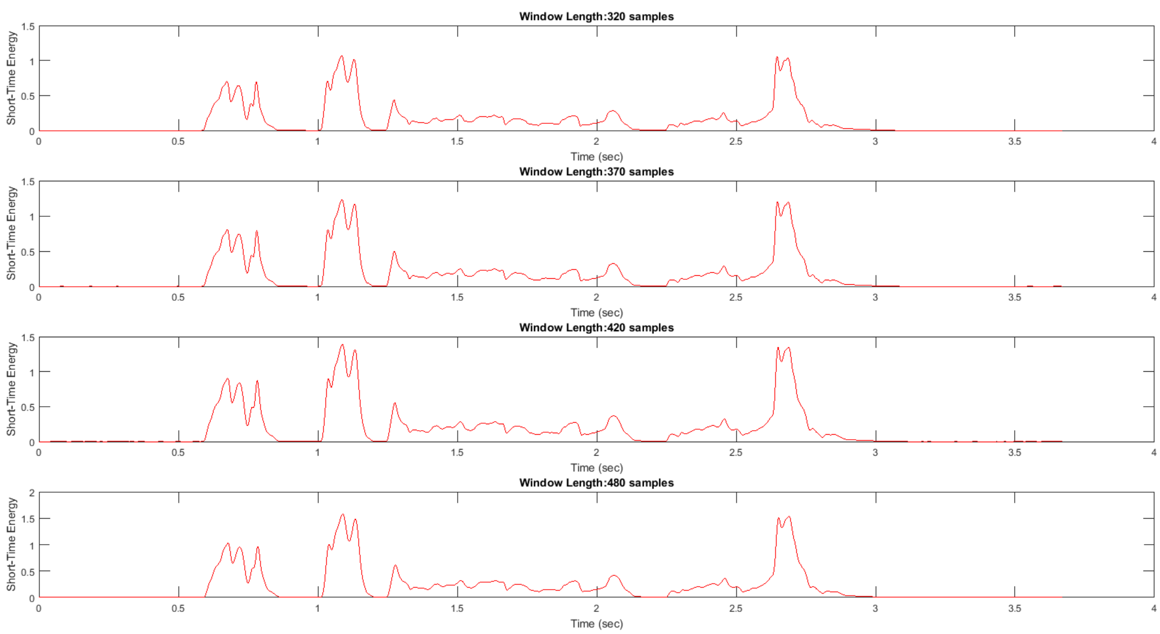
ΑΜ: 03114054

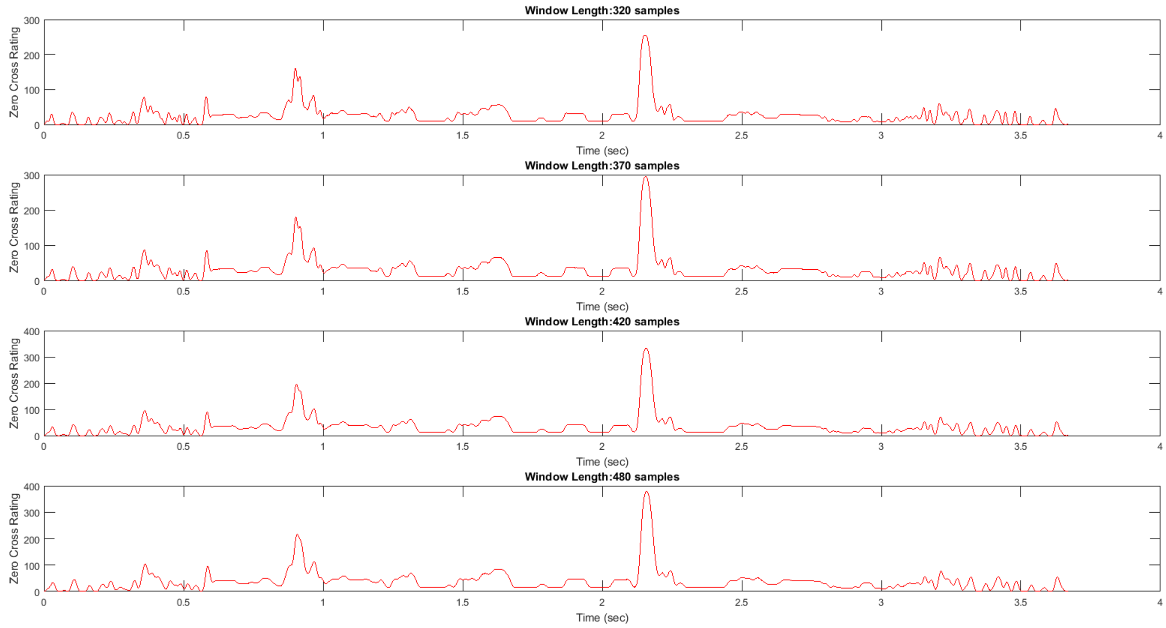
Μέρος 1ο

Ερώτημα 1.1 (ex1a.m)

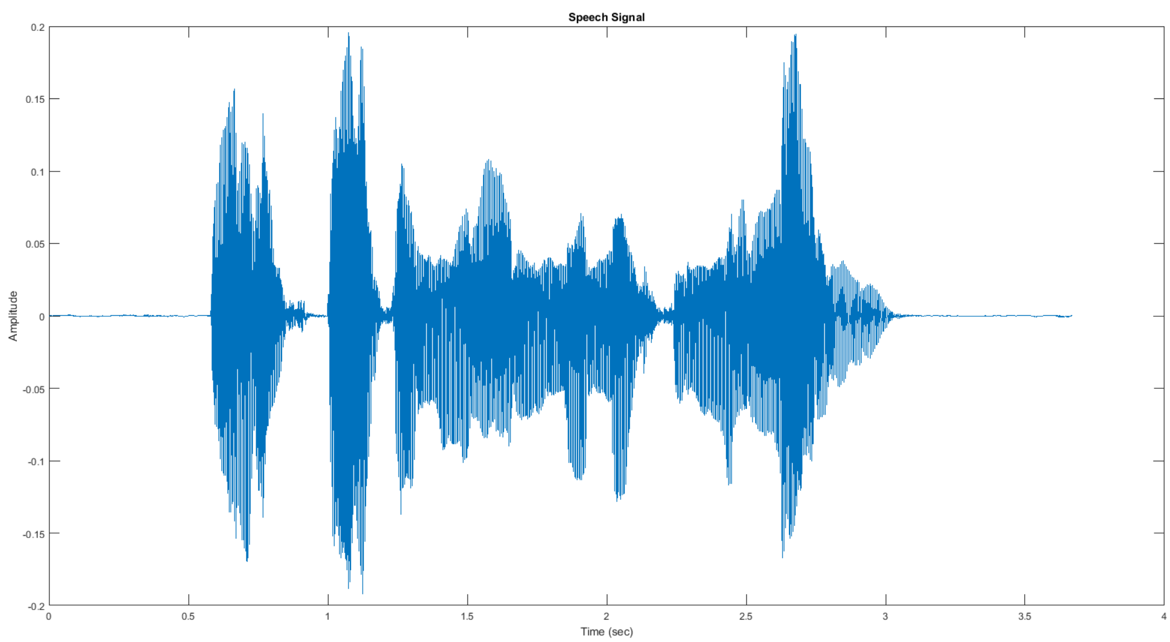
Σε αυτό το ερώτημα υπολογίζουμε την ενέργεια βραχέως χρόνου και τον ρυθμό εναλλαγής, αποθηκεύοντας το επιθυμητό μήκος παραθύρου στην μεταβλητή winLen, η οποία θα λαμβάνει σε κάθε επανάληψη του κύριου κώδικα αυξανόμενη τιμή της επιλογής μας, εντός των ζητούμενων ορίων, δηλαδή 20-30ms. Επειδή έχουμε συχνότητα δειγματοληψίας 16kHz, άρα και 16000 δείγματα του σήματος το δευτερόλεπτο, το πλήθος σε δείγματα του μήκους παραθύρου winLen θα είναι αντίστοιχα μεταξύ $20 \cdot 10^{-3} sec \cdot 16 \cdot 10^3 \frac{samples}{sec}$ και $30 \cdot 10^{-3} sec \cdot 16 \cdot 10^3 \frac{samples}{sec}$, ή $320 \leq winLen \leq 480$. Για τον υπολογισμό της ενέργειας χωρίζουμε το σήμα σε τμήματα με την συνάρτηση buffer, μήκους winLen το καθένα, φροντίζοντας να λάβει ως παράμετρο επικάλυψης των τμημάτων τέτοια ώστε το βήμα ολίσθησης παραθύρου να είναι 1 δείγμα, με βάση τον ορισμό και των δύο μεγεθών, οπότε η αντίστοιχη μεταβλητή θα είναι κάθε φορά winOverlap=winLen-1. Εφαρμόζουμε την ίδια διαδικασία και στις διαφορές: $|sgn[x[m]] - sgn[x[m-1]]|$, τις οποίες υπολογίσαμε με τις συναρτήσεις sign και diff. Έπειτα, πολλαπλασιάζουμε διαγώνιο πίνακα, που έχει τις τιμές παραθύρου Hamming, με πίνακα που έχει το σήμα σε τμήματα. Χρησιμοποιούμε τη συνάρτηση sparse για εξοικονόμηση μνήμης. Τέλος, υπολογίζουμε τα αθροίσματα των πινάκων κατά μήκος της κατάλληλης διάστασης, και ως οριζόντιο άξονα χρησιμοποιούμε διάνυσμα χρονικών στιγμών, το πλήθος των οποίων είναι ίσο με το πλήθος δειγμάτων του σήματος φωνής, ενώ η απόσταση μεταξύ τους θα είναι 1/Fs.

Τα αποτελέσματα που προέκυψαν είναι τα εξής:





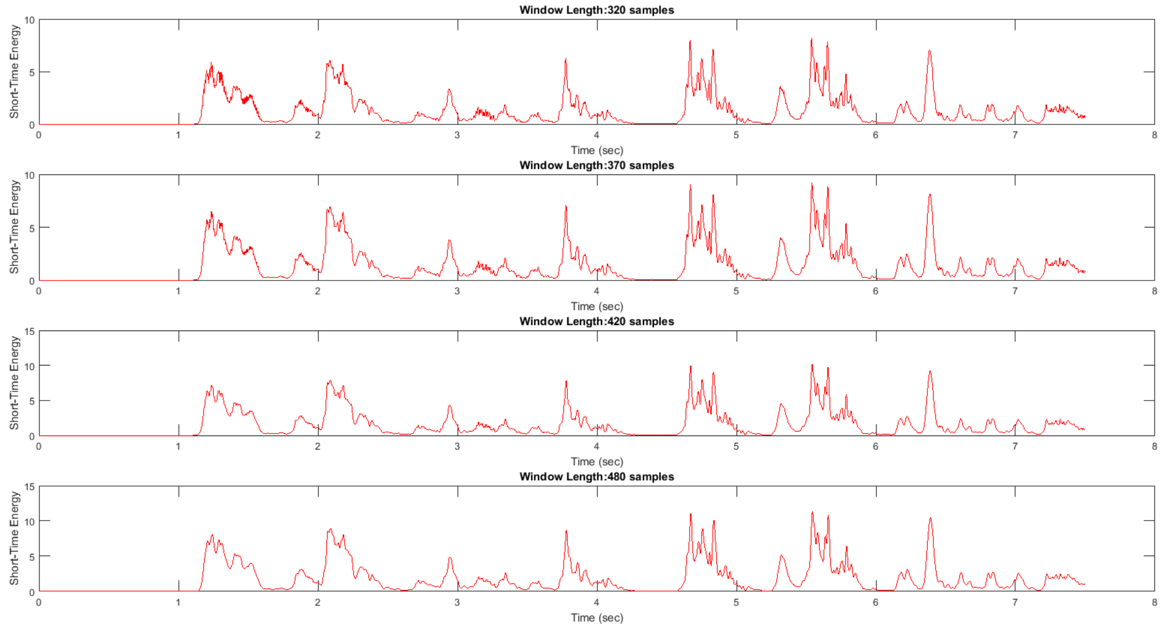
Παραθέτουμε και το πλάτος του σήματος συναρτήσει του χρόνου, καθώς μας διευκόλυνε να εντοπίσουμε χρονικό διάστημα με ήχους που μπορεί να μας ενδιαφέρουν, ακούγοντάς το παράλληλα με κάποιο πρόγραμμα που εμφανίζει τις χρονικές στιγμές με μεγαλύτερη ακρίβεια και προσφέρει ευελιξία στην χρήση, πχ Audacity το οποίο και χρησιμοποιήσαμε.

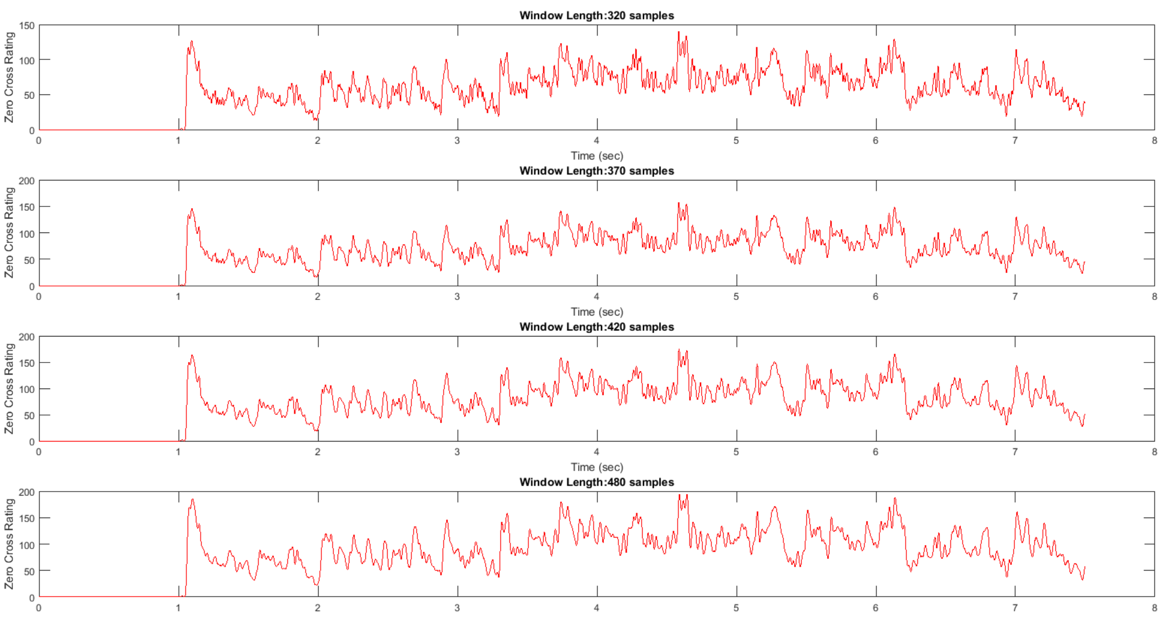


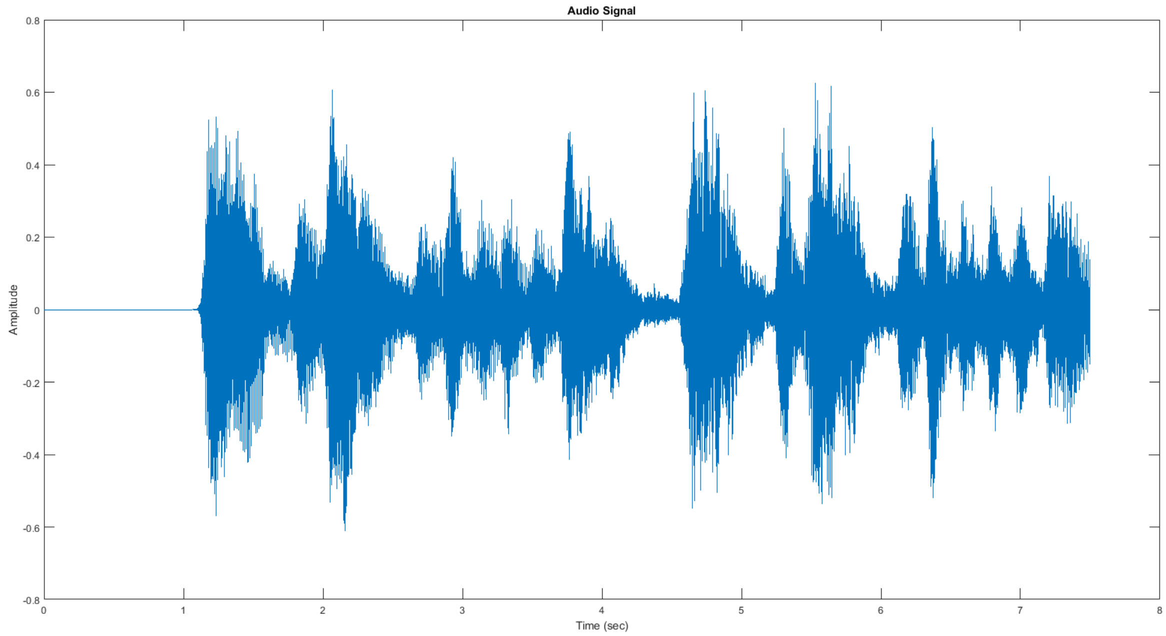
Παρατηρήσαμε ότι η ενέργεια βραχέως χρόνου παρουσιάζει μεγάλες τιμές σε διαστήματα που έχουμε φωνήεντα, για παράδειγμα με βάση το παραπάνω διάγραμμα μπορούμε να εντοπίσουμε το πρώτο και τελευταίο /ο/, καθώς και το πρώτο /a/ ανάμεσά τους, με σχετικά μεγάλη ακρίβεια. Γενικά παρουσιάζονται μικρότερες, αλλά υπαρκτές τιμές σε σημεία που δεν έχουμε σιωπή, ενώ διαστήματα με σύμφωνα που δεν ακούγονται ιδιαίτερα εμφανίζουν μηδενικές τιμές. Το δεύτερο διάγραμμα, από την άλλη, παρουσίασε μεγάλες τιμές σε διαστήματα που ακούγονται σύμφωνα, αλλά εμφάνιζε χαμηλότερες τιμές σε σημεία με φωνήεντα. Στην δική μας περίπτωση διακρίνουμε περισσότερο το /σ/ στο διάστημα 2.100-2.200, που παρουσιάζει τη μεγαλύτερη τιμή, αλλά και το /φ/ στο διάστημα 0.850-0.990. Αντίστοιχα, διαστήματα στα οποία εντοπίσαμε φωνήεντα προηγουμένως εμφανίζουν πολύ μικρό zero cross rating. Καταλήγουμε στο συμπέρασμα ότι με την ενέργεια βραχέως χρόνου μπορούμε να διακρίνουμε έμφωνους ήχους από σιωπή, και ταυτόχρονα με το ρυθμό εναλλαγής προσήμου διακρίνουμε άφωνους από έμφωνους ήχους.

Ερώτημα 1.2 (ex1b.m)

Εφαρμόζουμε τον ίδιο κώδικα και στο 2º σήμα ήχου. Τα διαγράμματα που προέκυψαν είναι:







Παρατηρούμε ότι το διάγραμμα με την ενέργεια παρουσιάζει μεγαλύτερες τιμές σε διαστήματα που ακούμε τους ήχους που προκύπτουν από το μουσικό όργανο, δηλαδή κάθε φορά που έχουμε «τριβή» με το βιολί. Το zero cross rating παρουσιάζει κορυφές σε αρκετά σημεία που η ενέργεια έχει τιμή που μηδενίζεται, ωστόσο σε αυτήν την περίπτωση ήχου δυσκολεύει η ανάλυση του σήματος, καθώς είναι αρκετά πιο δύσκολο να ξεχωρίσουμε διαστήματα στο συγκεκριμένο διάγραμμα.

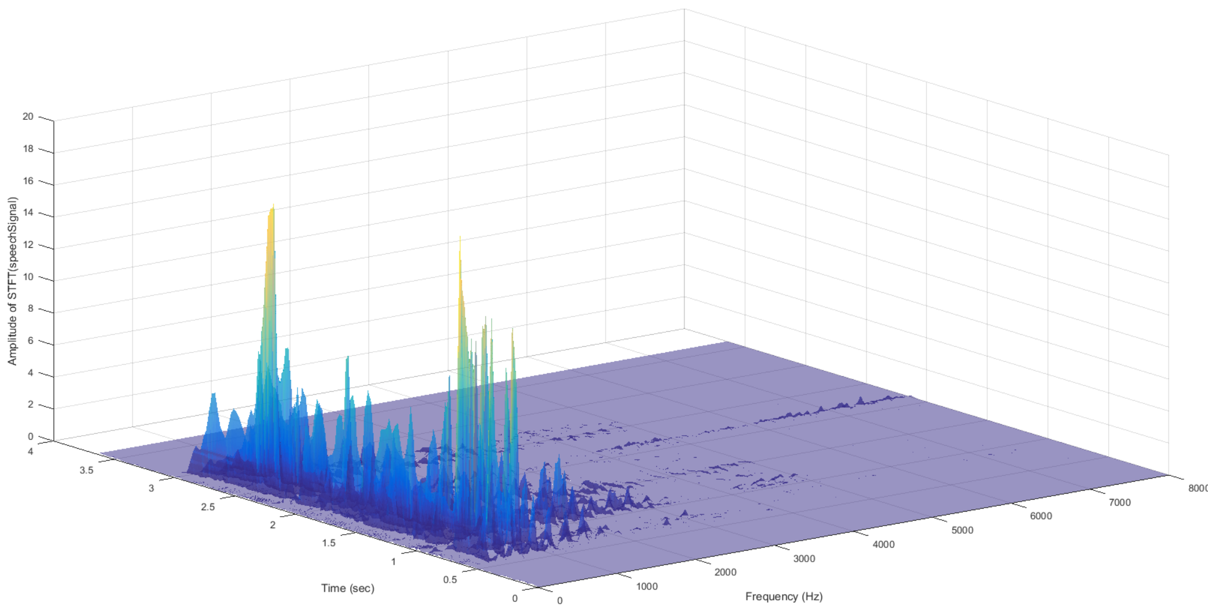
Μέρος 2ο

Ερώτημα 2.1 (ex2_1.m & mySTFT.m)

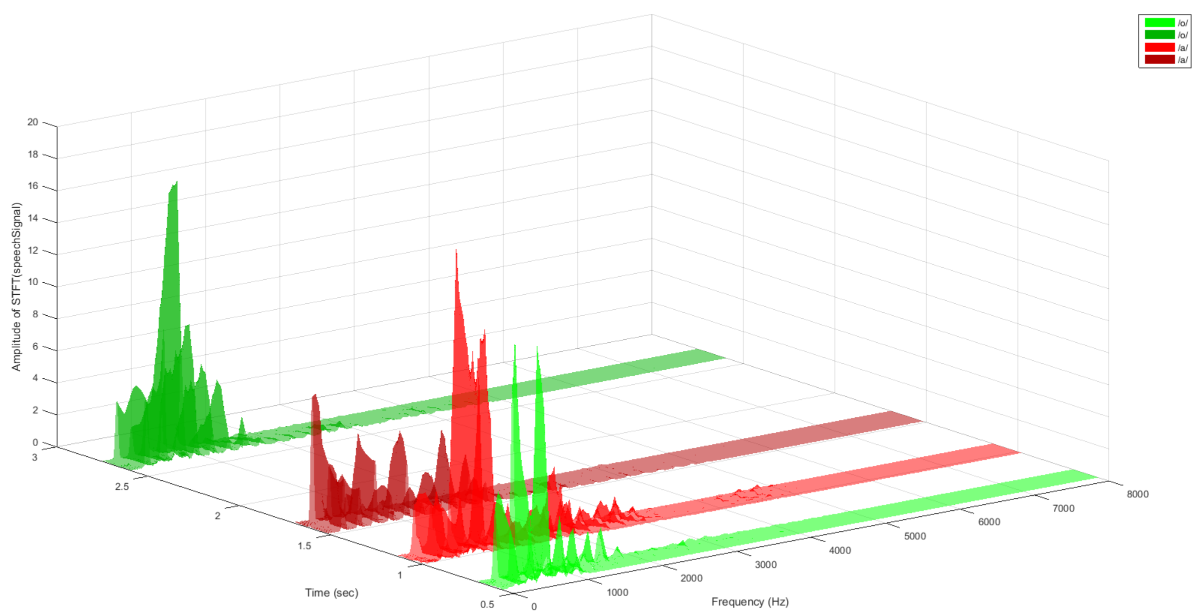
Σε αυτό το ερώτημα υπολογίζουμε και πάλι το μήκος του παραθύρου αλλά και του βήματος όπως προηγουμένως, $L = 40msec \cdot 16k\frac{samples}{sec}$ και $R = 20msec \cdot 16k\frac{samples}{sec}$. Για να υπολογίσουμε τον ζητούμενο διδιάστατο πίνακα που θα περιέχει τον STFT του σήματος, υλοποιούμε την συνάρτηση mySTFT.m και την καλούμε με τις κατάλληλες παραμέτρους. Στο σώμα της συνάρτησης, αρχικά, υπολογίζουμε το διάνυσμα με τις τιμές του παραθύρου Hamming, τις οποίες θα εφαρμόσουμε σε κάθε τμήμα του σήματος μήκους L. Η διαδικασία αυτή πραγματοποιείται μέσα σε βρόχο επανάληψης μήκους $\left\lceil\frac{length(signal)}{R}\right\rceil$ (το οποίο υπολογίζουμε με την συνάρτηση ceil), κατά την οποία θεωρούμε διάταξη με αρχή το πρώτο δείγμα το σήματος φωνής και υπολογίζουμε το δείγμα με θέση A, που θα είναι η αρχή του πλαισίου, και το δείγμα θέσης B, που θα είναι το τέλος του τρέχοντος πλαισίου i . Ο υπολογισμός γίνεται ολισθαίνοντας με βήμα R το ιδεατό παράθυρο μήκους $A - B + 1 = L$, οπότε βρίσκουμε τα επιθυμητά όρια του πλαισίου : $A = (\mathrm{i} - 1) * \mathrm{R} + 1, \mathrm{B} = (\mathrm{i} - 1) * \mathrm{R} + \mathrm{L}$. Έπειτα, εφαρμόζουμε στο κάθε πλαίσιο τον fft και αποθηκεύουμε το αποτέλεσμα στην γραμμή i του ζητούμενου πίνακα. Σημειώνουμε ότι η θέση B μπορεί να ξεπεράσει τα όρια του σήματος φωνής, οπότε σε αυτήν την περίπτωση θεωρούμε ότι το πλαίσιο φτάνει μέχρι και το τελευταίο δείγμα, και για να παραμείνει μήκους L το πλαίσιο προσθέτουμε μηδενικά στο τέλος του, χωρίς αυτό να επηρεάσει το αποτέλεσμα.

Ερώτημα 2.2 (ex2_2.m)

Για να κάνουμε αναπαράσταση του πλάτους του προηγούμενου πίνακα, φροντίσαμε να πάρουμε μόνο το πρώτο μισό του, καθώς το Matlab επιστρέφει μέσω του fft, στο δεύτερο μισό του πίνακα, τα πλάτη για αρνητικές συχνότητες. Ύστερα, υπολογίζουμε τον άξονα των συχνοτήτων: $fs = \left(0{:}\frac{N}{2} - 1\right) \cdot \frac{F_s}{N}$ και τον άξονα του χρόνου: $t = \left(0{:}\left\lceil\frac{length(signal)}{R}\right\rceil - 1\right) \cdot \frac{duration}{\left\lceil\frac{length(signal)}{R}\right\rceil}$, δηλαδή ως ποσοστό της διάρκειας του σήματος, βάσει της πρώτης διάστασης του πίνακα. Τέλος, αφού πρώτα πάρουμε το πλάτος του σήματος με τη συνάρτηση abs, αναπαριστούμε το αποτέλεσμα με την συνάρτηση surf:



/α/ και /ο/ εντοπίζουμε στα διαστήματα(sec): [0.587-0.700] (/ο/), [2.625-2.754] (/ο/), [1.010-1.154](/α/) και [1.550-1.700](/α/) με τη χρήση του audacity για καλύτερη ακρίβεια. Απομονώνουμε τα διαστήματα αυτά του παραπάνω διαγράμματος και τα εμφανίζουμε παρακάτω:



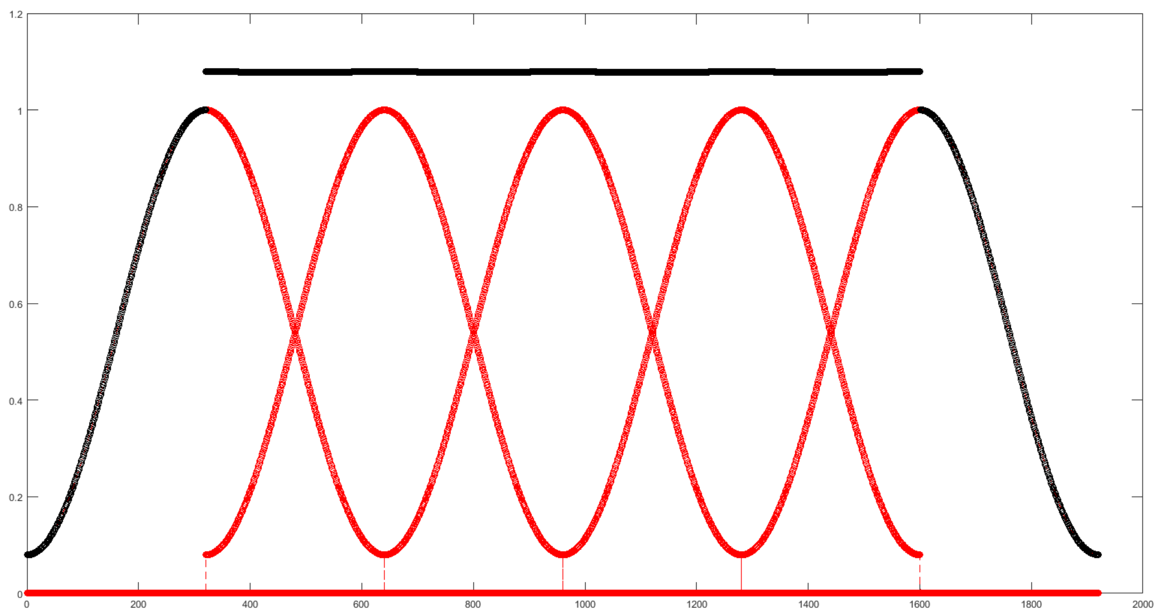
Παρατηρούμε ότι στα φωνήεντα αυτά ο STFT παρουσιάζει συνιστώσες με μεγαλύτερα πλάτη.

Ερώτημα 2.3 (ex2_3.m)

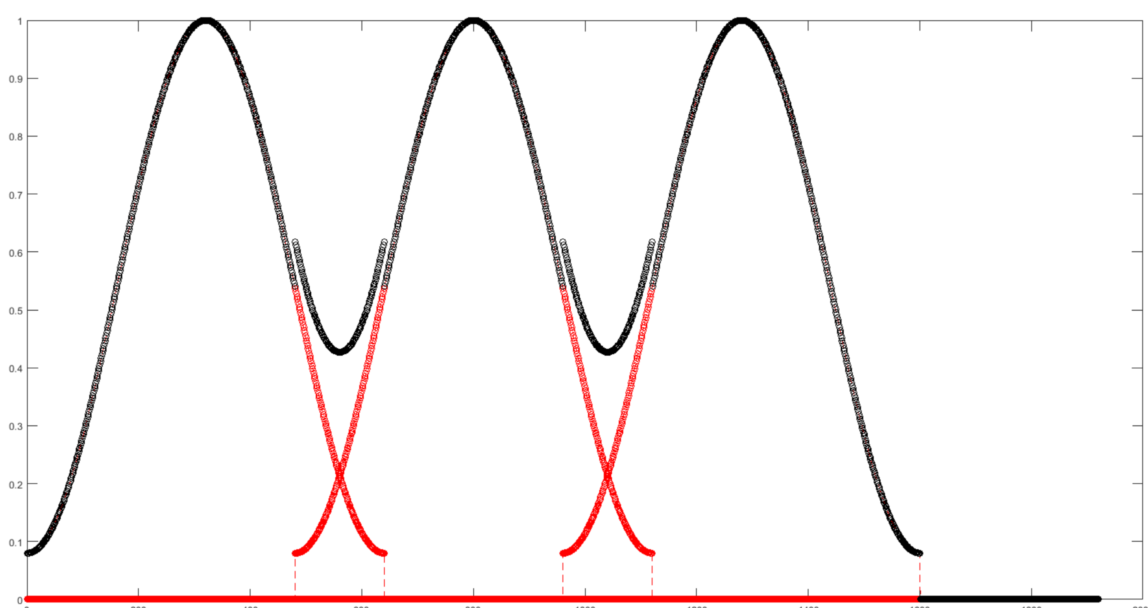
Για να ανακατασκευάσουμε το σήμα ακολουθούμε την αντίστροφη διαδικασία από αυτήν της συνάρτησης mySTFT. Πιο συγκεκριμένα, κατασκευάζουμε ένα διάνυσμα με μηδενικά στοιχεία και εντός βρόχου, στην i-οστή επανάληψη, προσθέτουμε στις θέσεις [Ai-Bi] του διανύσματος, από τον αντίστροφο DFT της i γραμμής του διδιάστατου πίνακα των προηγούμενων ερωτημάτων, τις στήλες από Ai εώς Bi, όπου Ai, Bi όπως είχαν υπολογιστεί σε προηγούμενο ερώτημα, δηλαδή: $Ai = (\mathrm{i}-1)\cdot \mathrm{R}+1, Bi = (\mathrm{i}-1)\cdot \mathrm{R}+\mathrm{L}$, με R, L τιμές ίσες με τις τιμές που χρησιμοποιήθηκαν κατά την κατασκευή του STFT του αρχικού σήματος. Το αποτέλεσμα το ακούμε με την χρήση της συνάρτησης sound, και καταγράφουμε το αποτέλεσμα στο ζητούμενο αρχείο wav .

Ερώτημα 2.4 (ex2_4.m)

Με την χρήση της συνάρτησης ola.m για μήκος παραθύρου 40ms και βήμα 20ms έχουμε το εξής αποτέλεσμα:



ενώ για βήμα 30ms:

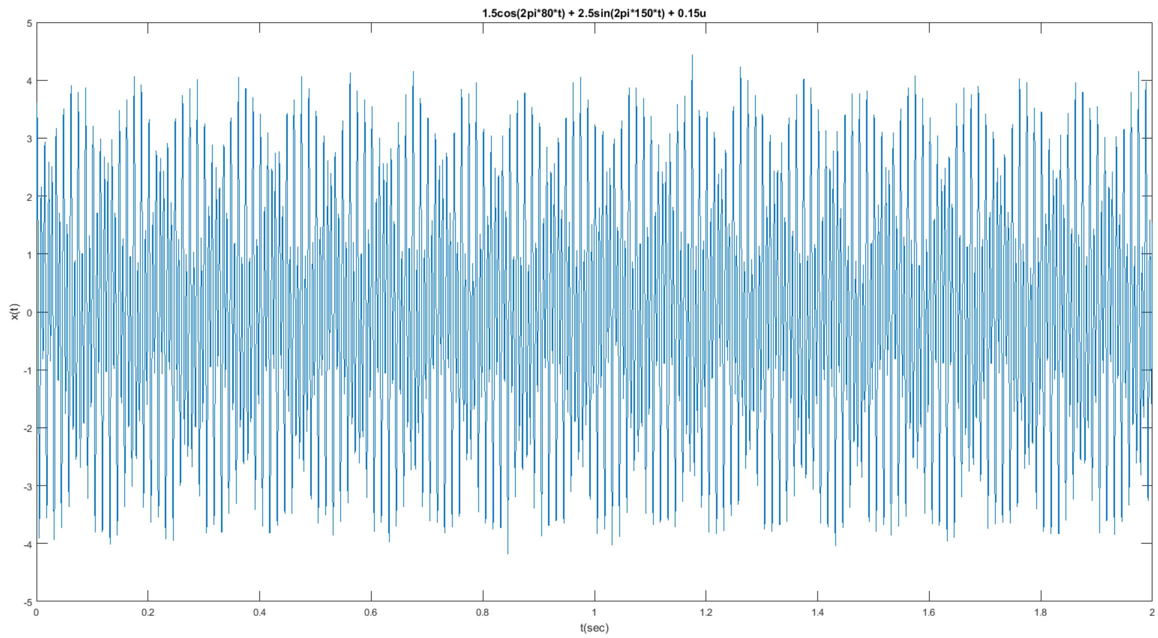


Παρατηρούμε ότι στην 1η περίπτωση το άθροισμα $\sum w[n - m \cdot R]$ ισούται με 1, κάτι που δεν ισχύει για το δεύτερο διάγραμμα. Επομένως, οι παράμετροι που χρησιμοποιήσαμε για τον υπολογισμό του STFT ικανοποιούν την συνθήκη OLA. Αντιθέτως, για βήμα ανάλυσης 30ms προέκυψε, μετά την ανακατασκευή, μια ελαφρά παραμορφωμένη έκδοση του αρχικού σήματος ήχου. Παραθέτουμε το αποτέλεσμα στο αρχείο: speech_utterance_R_30ms.wav .

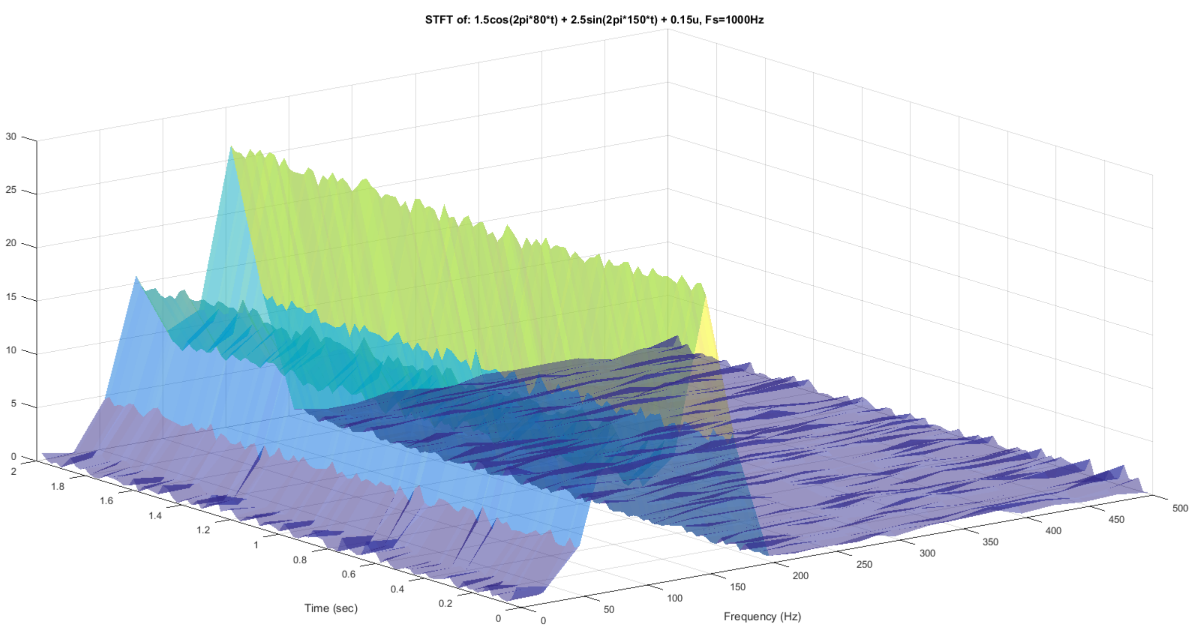
Μέρος 3ο

Ερώτημα 3.1 (ex3_1a.m , ex3_1b.m και ex3_1c.m)

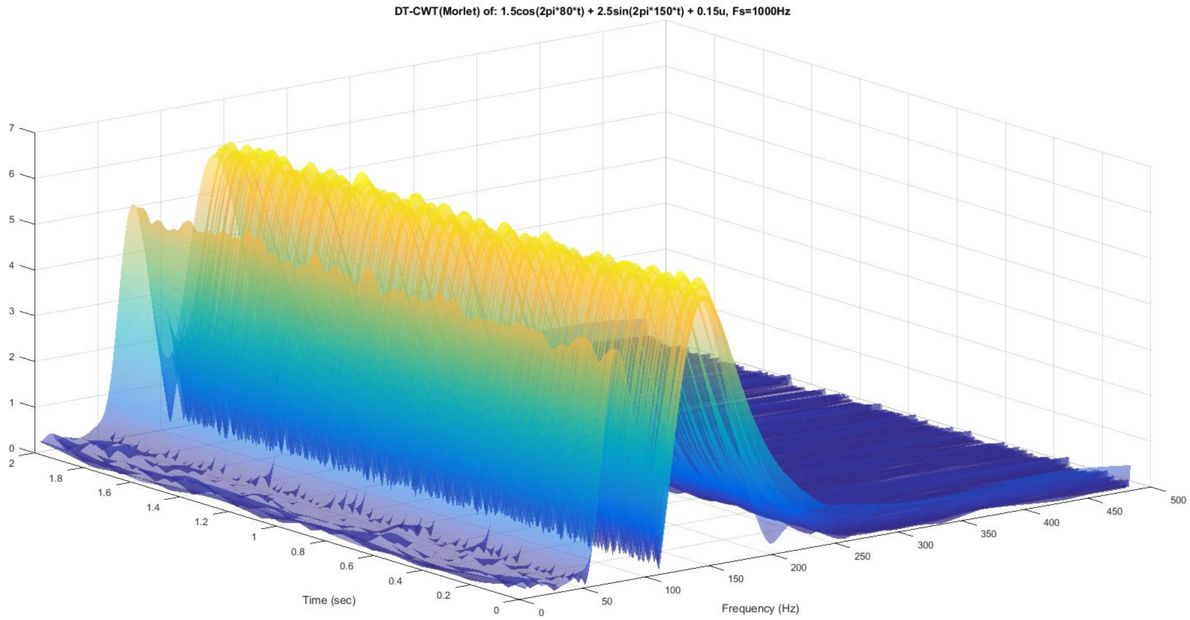
(α) Η αναπαράσταση του σήματος x(t) είναι:



(β) Το πλάτος του STFT με τη χρήση της ρουτίνας spectrogram είναι:



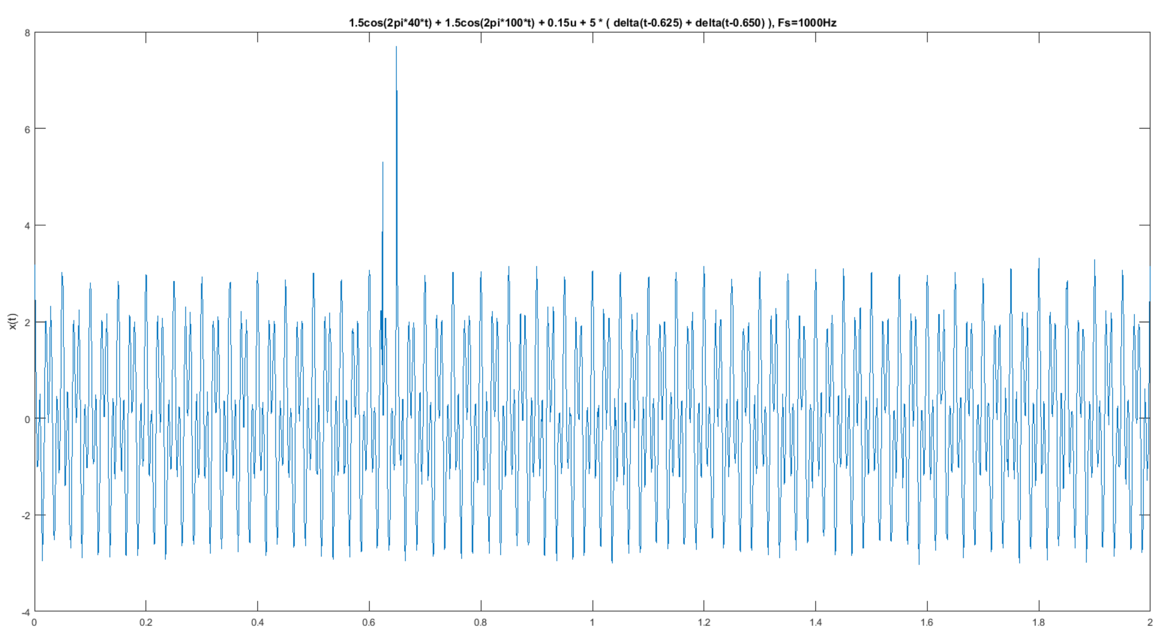
(γ) Με τη χρήση Wavelet αντίστοιχα προκύπτει:



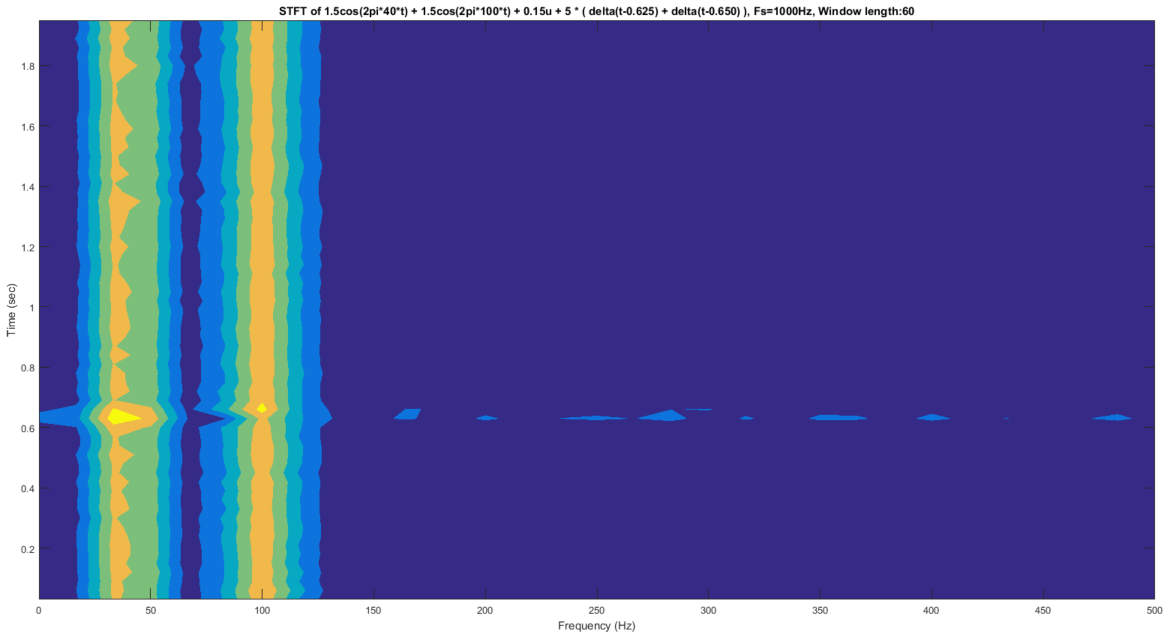
(δ) Παρατηρούμε ότι με τη χρήση Wavelet έχουμε εμφανώς καλύτερο αποτέλεσμα καθώς είναι πιο διακριτές οι αρμονικές των κύριων συχνοτήτων στον STFT του $2^{ου}$ διαγράμματος, εφόσον σε αυτές η άνοδος είναι πολύ πιο απότομη, δηλαδή το εύρος στον οποίο εμφανίζεται ο λοβός είναι μικρότερο. Επίσης, οι κορυφές στα πλάτη είναι πιο ομαλές. Από την άλλη, στο $1^{ο}$ διάγραμμα παρατηρούμε ότι στις συχνότητες μεταξύ των δύο κύριων συνιστωσών έχουμε επικάλυψη, διότι το εύρος των λοβών είναι μεγαλύτερο, κάτι που σημαίνει ότι για πιο κοντινές συχνότητες (πχ 110Hz και 140Hz) θα ήταν πιο δύσκολο να ξεχωρίσουμε την μία αρμονική από την άλλη.

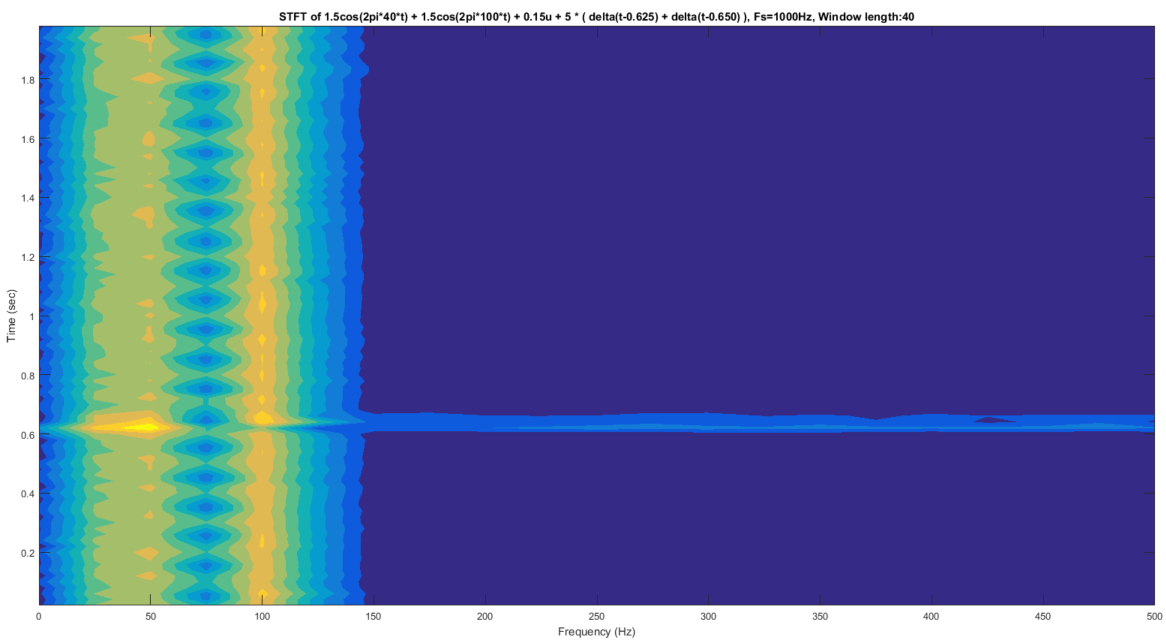
Ερώτημα 3.2 (ex3_2a.m , ex3_2b.m και ex3_2c.m)

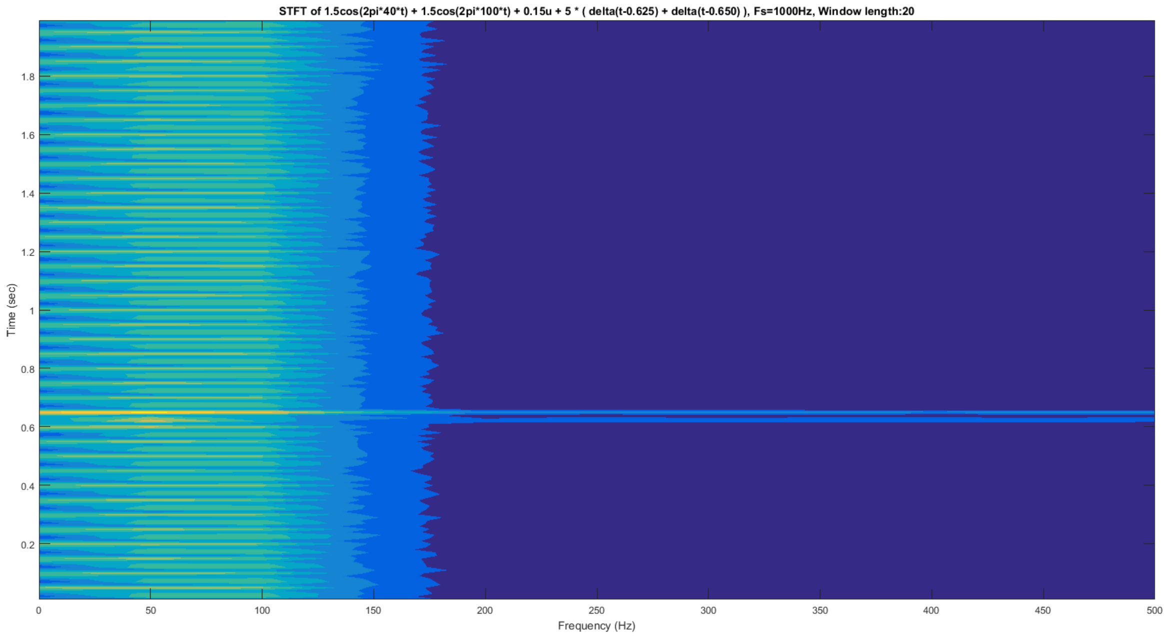
(α) Για το σήμα αυτό, προκειμένου να αναπαραστήσουμε την συνάρτηση δέλτα του συνεχούς στον διακριτό χρόνο, απλά προσομοιώσαμε το δέλτα του Kronecker, προσθέτοντας στις κατάλληλες θέσεις του διακριτού σήματος x την τιμή του πλάτους της δ (δηλαδή 5). Το διάγραμμα που προέκυψε είναι:



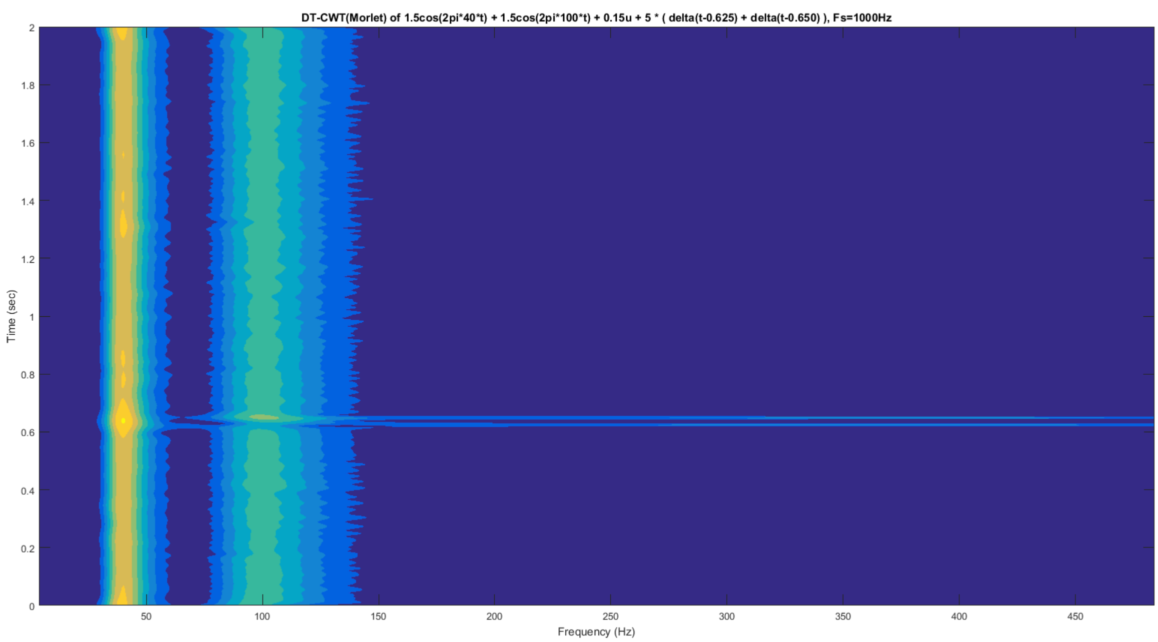
(β) Προσθέσαμε χρώμα με κατάλληλο όρισμα στην συνάρτηση contour για την διευκόλυνσή μας. Επίσης, για το κάθε μήκος παραθύρου χρησιμοποιήσαμε διαφορετικό figure του Matlab . Τα διαγράμματα είναι:







(γ) Αντίστοιχα, με την χρήση Wavelet προέκυψε το εξής:



(δ) Παρατηρούμε ότι με τον STFT μετασχηματισμό, προκειμένου να εντοπίσουμε τις δύο κύριες συνιστώσες, χρειαστήκαμε παράθυρο μήκους 60. Σε αυτό το διάγραμμα, όμως, δεν είναι εμφανές σε ποιο σημείο έχουμε απότομη διακριτή μεταβολή. Αυτή φαίνεται πολύ περισσότερο στο διάγραμμα με παράθυρο μήκους 20, στο οποίο, από την άλλη, δεν μπορούμε να ξεχωρίσουμε τις κύριες συνιστώσες. Επομένως, με αυτόν τον μετασχηματισμό δεν μπορούμε να πετύχουμε ταυτόχρονα τα 2 ζητούμενα. Εν αντιθέσει, με τον μετασχηματισμό DT-CWT, όπως βλέπουμε από το διάγραμμα, καταφέρνουμε να εντοπίσουμε τόσο τις διακριτές μεταβολές όσο και τις βασικές συχνότητες του σήματος.