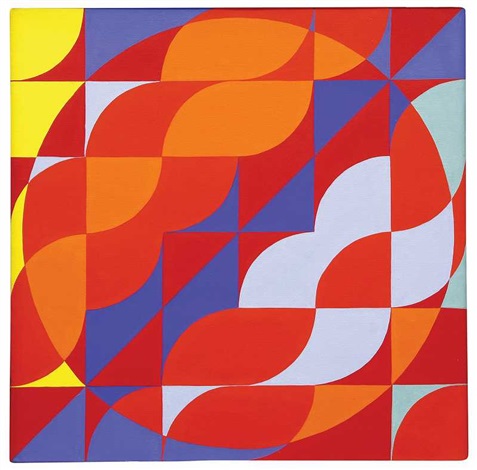
1. **Θεοδωρίδης Αριστομένης | 03115632**



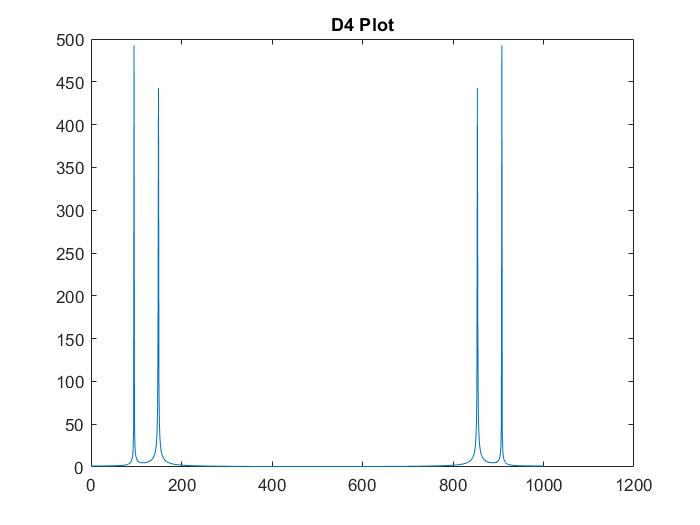
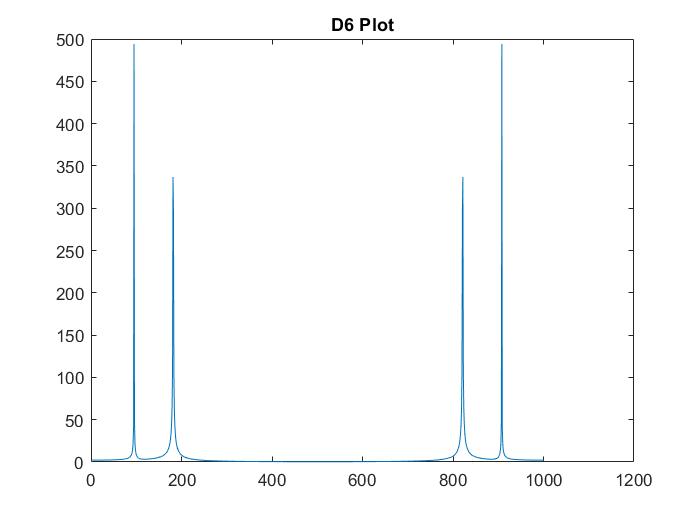
1. **Αβδελάς Λεωνίδας | 03113182**

Εισαγωγή στην Ψηφιακή Επεξεργασία Σημάτων με MATLAB και Εφαρμογές σε Ακουστικά Σήματα

1η εργαστηριακη ασκηση

# Μέρος 1- Σύστημα Εντοπισμού Τηλεφωνικών Τόνων (Telephone Touch – Tones)

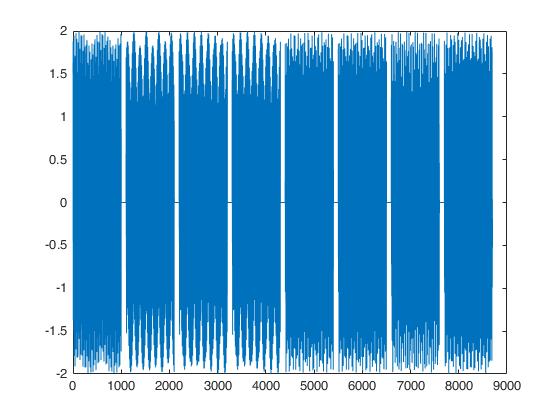
## 1.2

Από τα σήματα d4[n] και d6[n] δημιουργήσαμε τις εξής γραφικές παραστάσεις:

Παρατηρούμε ότι μπορούμε εύκολα να διακρίνουμε τα δύο σήματα μεταξύ τους, λόγω των peaks στις χαρακτηριστικές τους συχνότητες. Ακόμα, λόγω του αλγορίθμου της συνάρτησης **fft()** που χρησιμοποιούμε, οι αρνητικές συχνότητες έχουν μετατοπιστεί κυκλικά στις συχνότητες μεγαλύτερες των 500Hz. Οι θετικές συχνότητες, οι οποίες μας ενδιαφέρουν εδώ, είναι αυτές από 0-500Hz.

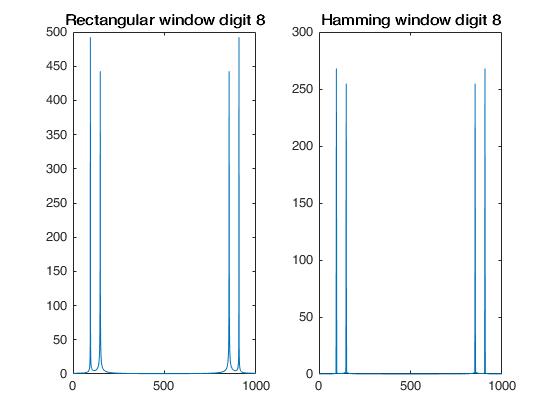
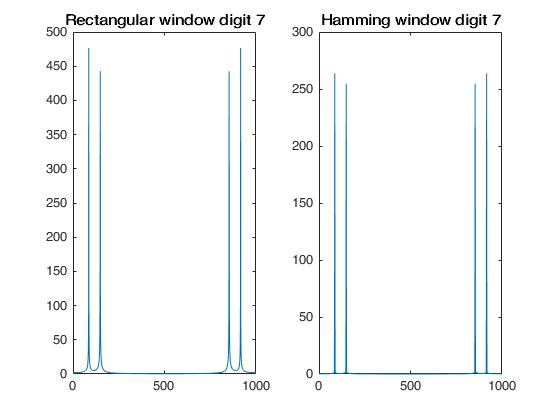
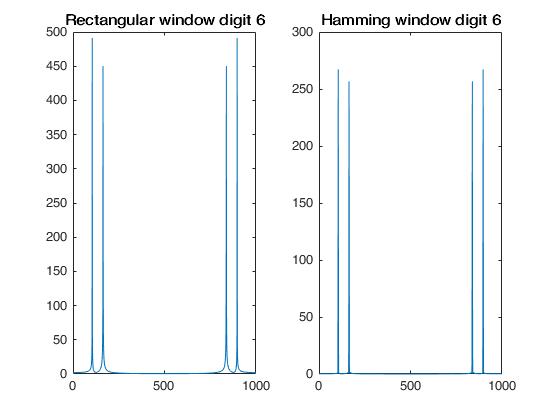
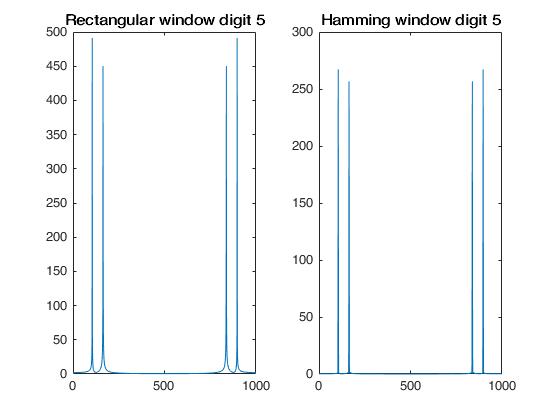
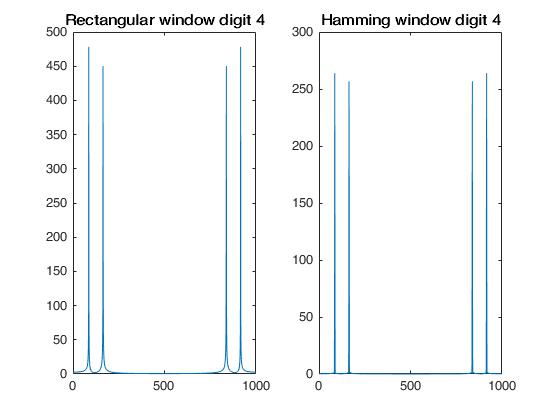
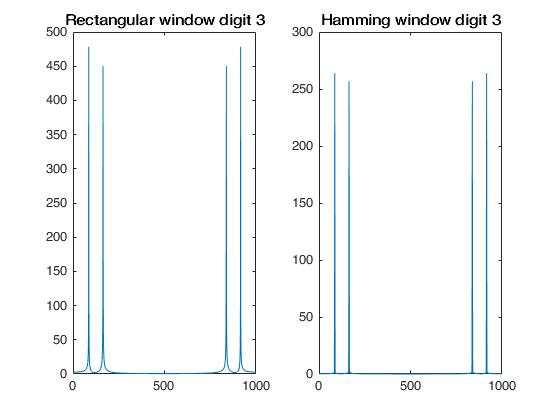
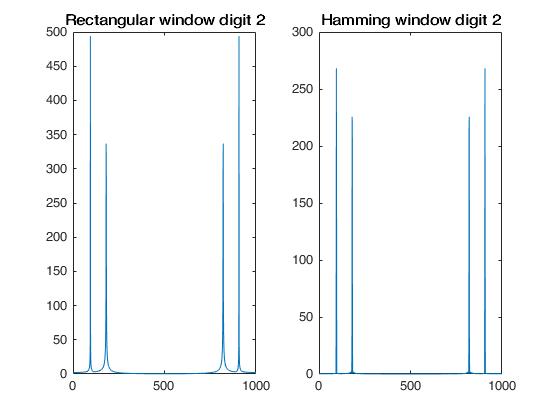
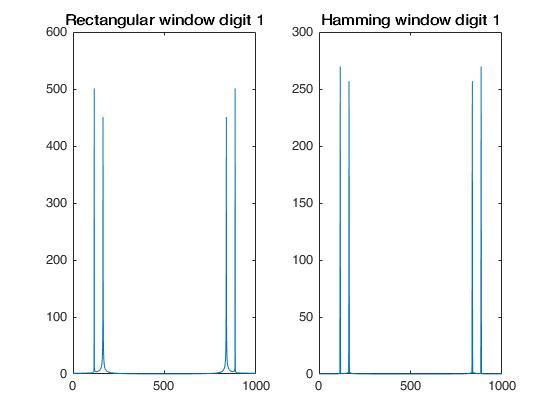
## 1.3

Το συνδυασμένο ΑΜ μας είναι 06228814. Για να δημιουργήσουμε το αρχείο ‘tone\_sequence.wav’, δημιουργήσαμε έναν πίνακα με τα tone sequences που δημιουργήσαμε στο ερώτημα 1.1 και 100 μηδενικά μεταξύ τους.



## 1.4

Παρακάτω βλέπουμε τις γραφικές που παίρνουμε από την παραθυροποίηση και εφαρμογή της συνάρτησης **fft()** στα παράθυρα του σήματος AM\_sound.



Όπως βλέπουμε, χρησιμοποιώντας παράθυρα Hamming, τα peaks συγκεντρώνονται γύρω από συγκεκριμένες συχνότητες και η διάκριση των κύριων συχνοτήτων είναι πιο ξεκάθαρη.

### 1.5

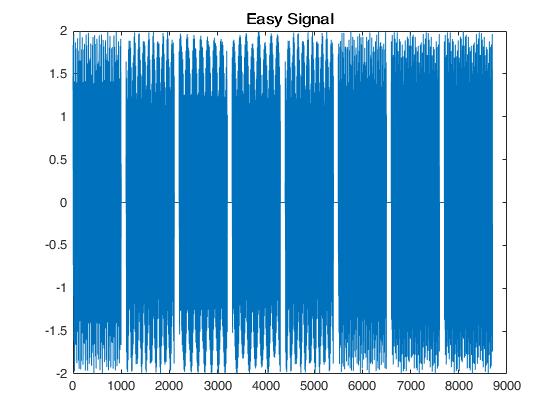
Για να υπολογίσουμε την λίστα με τις touch-tone συχνότητες, οπτικά αυτό που κάνουμε είναι να βρούμε τα δείγματα στα οποία αντιστοιχούν τα μέγιστα των συχνοτήτων. Λόγω του ότι δεν υπάρχουν απόλυτες συναρτήσεις Dirac στον πραγματικό κόσμο, οι μέγιστες τιμές δεν είναι σε ένα σημείο, αλλά σε ένα μικρό πεδίο γύρω από την κύρια συχνότητα. Για να αντιμετωπίσουμε αυτό το πρόβλημα, ελέγχουμε για το δεύτερο peak, σε περιοχή που είναι τουλάχιστον 10 δείγματα μακριά από το πρώτο.

## 1.6

Η συνάρτηση **ttdecode()** δέχεται ως παραδοχή, ότι όλοι οι τόνοι είναι 1000 δείγματα και ότι ανάμεσα σε κάθε τόνο υπάρχει ένα κενό 100 δειγμάτων. Έτσι, με αυτό το μοτίβο σπάει το σήμα σε παράθυρα μέχρι να φτάσει στο τέλος του και μετά χρησιμοποιεί μετασχηματισμό fft σε κάθε ένα από αυτά. Χρησιμοποιώντας παράθυρα Hamming, έχουμε καλύτερη διακριτική ικανότητα σχετικά με τις κορυφές των σημάτων μας, όπως αναφέρθηκε και παραπάνω. Έπειτα, αποσπούμε την πληροφορία σχετικά με το που βρίσκονται οι κορυφές (δηλαδή τα σημεία με μέγιστη ενέργεια) με τον ίδιο τρόπο που αναφέρθηκε στο βήμα 1.5. Για να βρούμε σε ποιο ψηφίο αντιστοιχούν αυτές οι κορυφές, συγκρίνουμε τα σημεία που πήραμε με τον πίνακα που δημιουργήσαμε στον βήμα 1.5, έχοντας ένα περιθώριο λάθους 3 δειγμάτων. Η συνάρτηση αυτή προσδιορίζει τέλεια τους τόνους που δημιουργήσαμε στο βήμα 1.3.

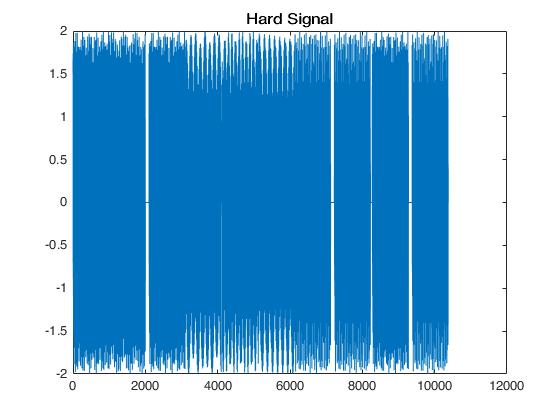
## 1.7

Το εύκολο σήμα, προσδιορίζεται από την **ttdecode()** που έχουμε δημιουργήσει χωρίς πρόβλημα, καθώς ακολουθεί τις παραδοχές που είχαμε ορίσει νωρίτερα. Αυτό φαίνεται εύκολα και από την γραφική του σήματος που ακολουθεί.



Η ακολουθία ψηφιών στην οποία καταλήγουμε είναι η 1 3 2 6 3 9 0 0.

Το hardSig, βλέπουμε εύκολα ότι δεν ακολουθεί τις παραδοχές που είχαμε ορίσει για την συνάρτηση μας στο βήμα 1.6. Αυτό που θα κάνουμε λοιπόν είναι να την αλλάξουμε ώστε να μπορεί να αντιμετωπίσει μεταβλητό μήκος κενού μεταξύ των τόνων, δεδομένου ότι οι τόνοι συνεχίζουν να είναι 1000 δειγμάτων.

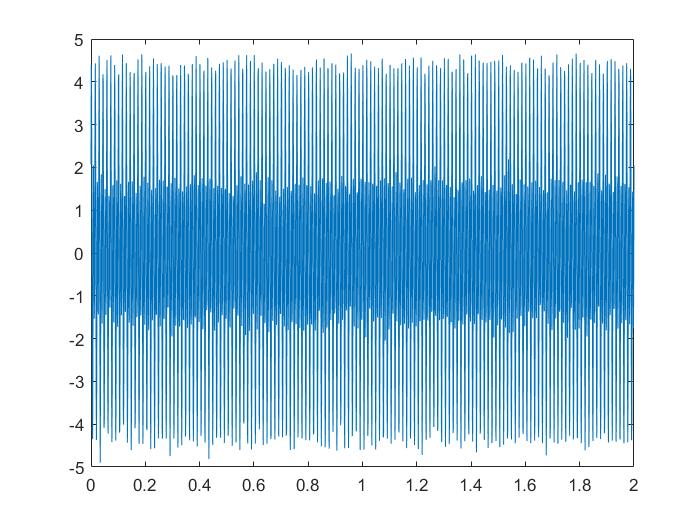


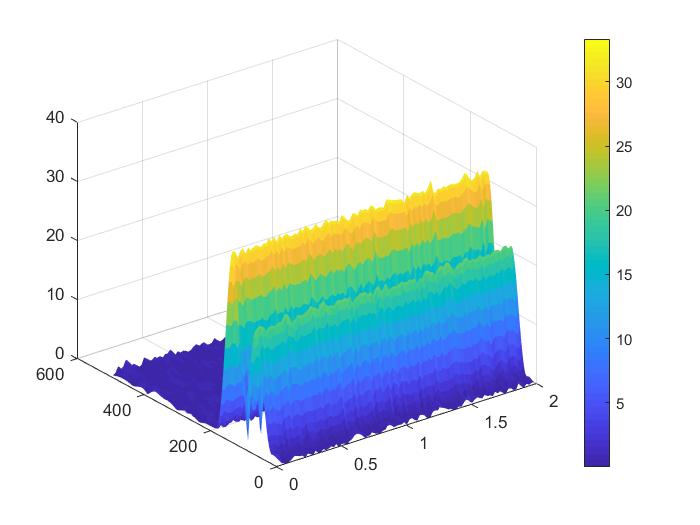
Ο τρόπος που το κάνουμε αυτό είναι να βρίσκουμε κάθε φορά που τελειώνει ένας τόνος, το πρώτο δείγμα που δεν είναι 0 στο σήμα μας και να παίρνουμε ένα παράθυρο 1000 δειγμάτων, το οποίο ξεκινάει από εκείνο το δείγμα, ως τον επόμενο τόνο. Δοκιμάζοντας την νέα υλοποίηση της **ttdecode()** στο σήμα του 1.3 και στο easySig παίρνουμε τα ίδια αποτελέσματα με πριν. Δοκιμάζοντας το στο hardSig, παίρνουμε την ακολουθία ψηφίων 9 0 9 6 3 2 1 1 9 1.

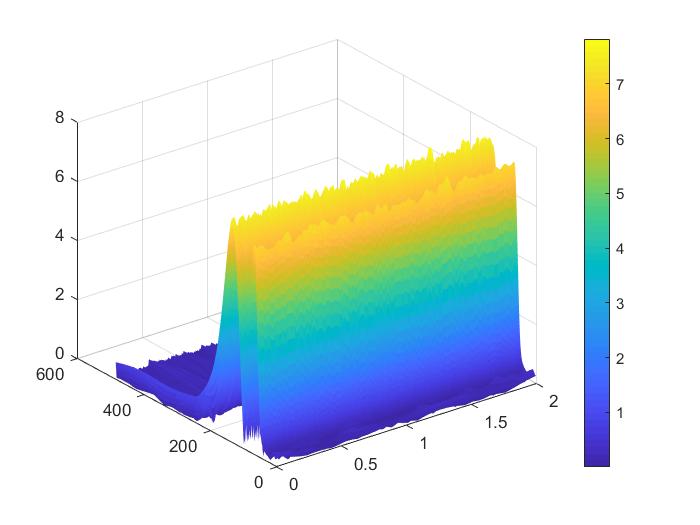
Μέρος 2ο - Φασματική Ανάλυση Ημιτονοειδών και Ανίχνευση Απότομων Μεταβάσεων με τον Μετ/σμό Fourier Βραχέος Χρόνου (STFT) και τον Μετ/σμό Wavelets (διακριτοποιημένο DT-CWT)

2.1

Αρχικά υπολογίστηκε το ακόλουθο σήμα:

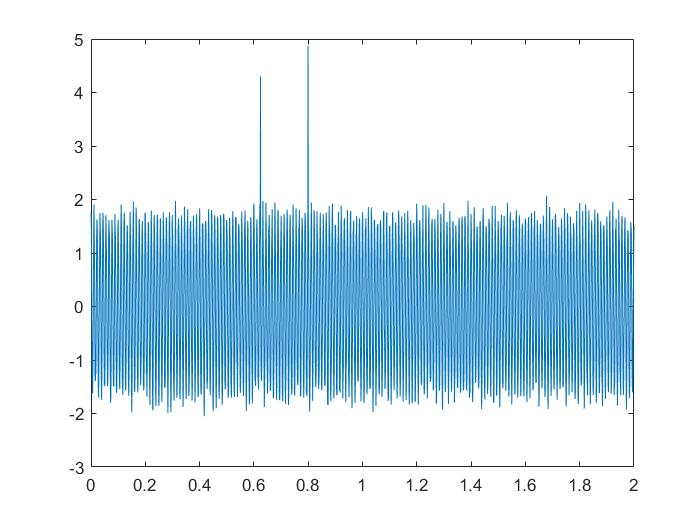
Στη συνέχεια, με χρήση της ρουτίνας **spectrogram()** υπολογίσαμε τον STFT μετασχηματισμό του σήματος με μήκος παραθύρου ίσο με 0.04sec και επικάλυψη ίση με 0.02sec. Με χρήση της συνάρτησης **surf()** πήραμε το ακόλουθο διάγραμμα για το πλάτος του STFT.

Ακολούθως, με τη χρήση της ρουτίνας **cwtft()**, επιλέγοντας “Morlet” wavelet, υπολογίσαμε το DT-CWT και με τη συνάρτηση **surf()** πήραμε την ακόλουθη αναπαράσταση του πλάτους του:

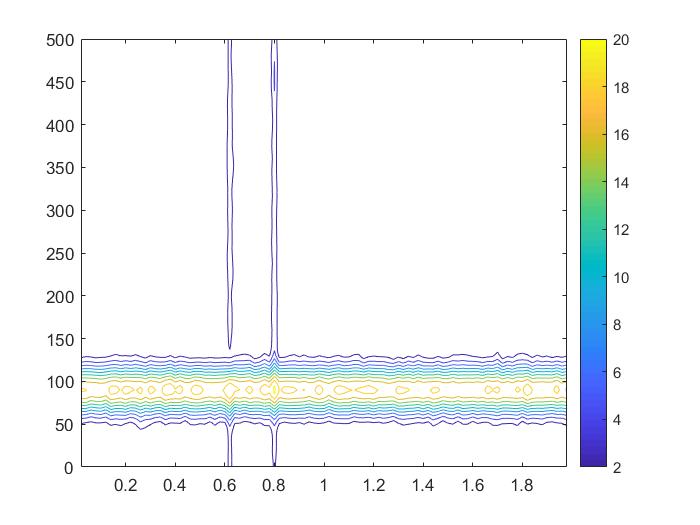
Παρατηρούμε ότι με τη χρήση των wavelets έχουμε καλύτερη ευκρίνεια και μεγαλύτερη ακρίβεια στον άξονα της συχνότητας, όμως και οι δυο μετασχηματισμοί καταφέρνουν να εντοπίσουν τις αλλαγές που έχουμε στη συχνότητα λόγω των ημιτονοειδών συναρτήσεων και του θορύβου.

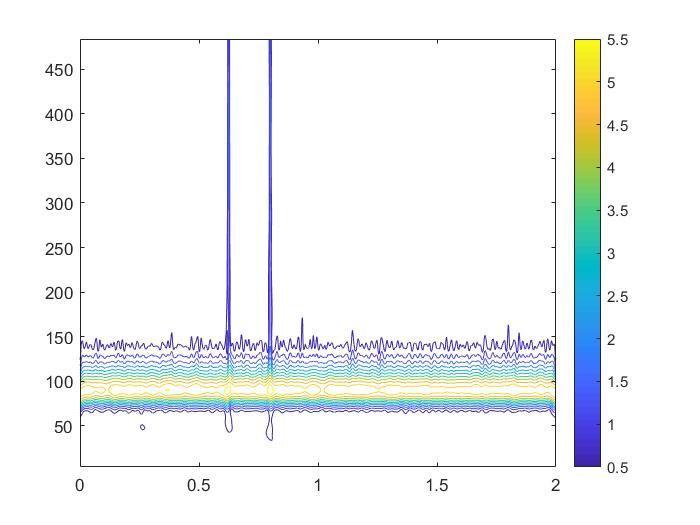
2.2

Αρχικά υπολογίστηκε το ακόλουθο σήμα:



Στη συνέχεια, με χρήση της ρουτίνας **spectrogram()** υπολογίσαμε τον STFT μετασχηματισμό του σήματος με μήκος παραθύρου ίσο με 0.04sec και επικάλυψη ίση με 0.02sec. Με χρήση της συνάρτησης **contour()** πήραμε το ακόλουθο διάγραμμα για το πλάτος του STFT.



Ακολούθως, με τη χρήση της ρουτίνας **cwtft()**, επιλέγοντας “Morlet” wavelet, υπολογίσαμε το DT-CWT και με τη συνάρτηση **contour()** πήραμε την ακόλουθη αναπαράσταση του πλάτους του:

Παρατηρούμε ότι με τη χρήση του STFT περιορίζεται ο θόρυβος καθώς δεν αντιλαμβάνεται τις απότομες αλλαγές στη συχνότητα, ενώ ο DT-CWT για μικρές συχνότητες είναι περίπου ίδιος με τον STFT και όσο μεγαλώνουν αρχίζει να παίρνει παράθυρα για καλύτερη ακρίβεια στη συχνότητα και έτσι αντιλαμβάνεται καλύτερα το θόρυβο και τις απότομες αλλαγές στη συχνότητα(για t=0.625 και t=0.8).

Μέρος 3ο - Χαρακτηριστικά Βραχέος Χρόνου Σημάτων Φωνής και Μουσικής (Ενέργεια και Ρυθμός Εναλλαγής Προσήμου)

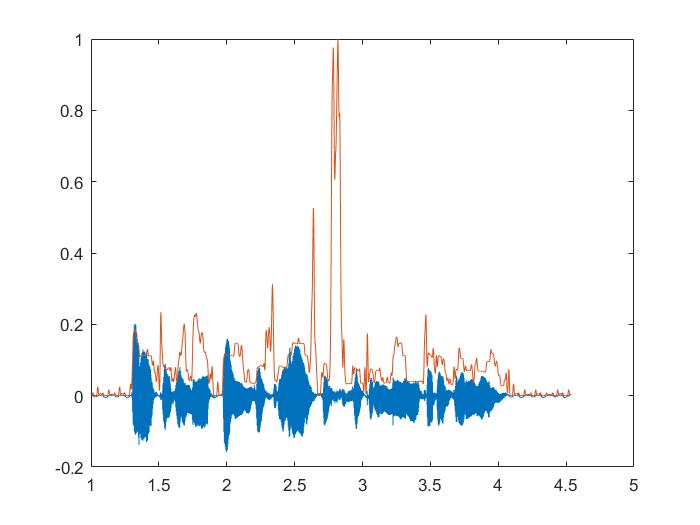
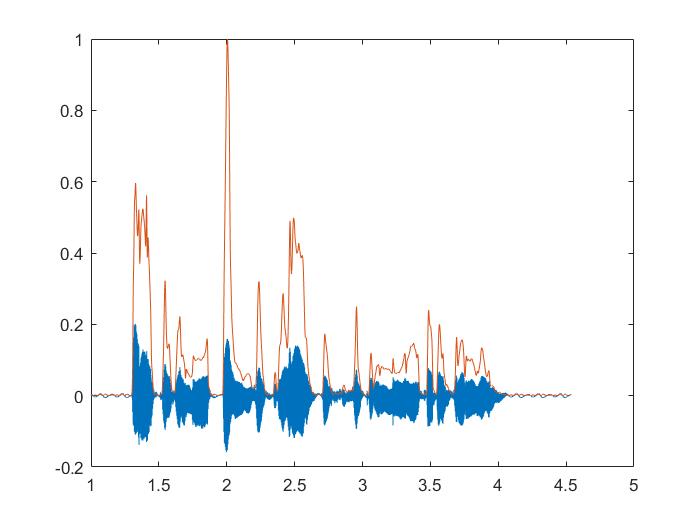
Για τον υπολογισμό της ενέργειας βραχέως χρόνου κατασκευάσαμε μια συνάρτηση που αρχικά σπάει το σήμα σε κομμάτια όσο το μέγεθος του παραθύρου, πολλαπλασιάζει κάθε κομμάτι με το παράθυρο και στη συνέχεια προσθέτει τα προσθέτει τα δείγματα που αντιστοιχούν στην ίδια χρονική στιγμή.

Για το ρυθμό εναλλαγής προσήμου χρησιμοποιείται αντίστοιχη συνάρτηση που πριν πολλαπλασιάσει με το παράθυρο περνάει το σήμα από την συνάρτηση **sign()** για να πάρει το πρόσημο μόνο και στη συνέχεια από την **abs()**. Στη συνέχεια η συνάρτηση κάνει ότι και η προηγούμενη.

Και οι δυο συναρτήσεις κανονικοποιούν τα αποτελέσματα για καλύτερη αναπαράσταση στα διαγράμματα.

3.1

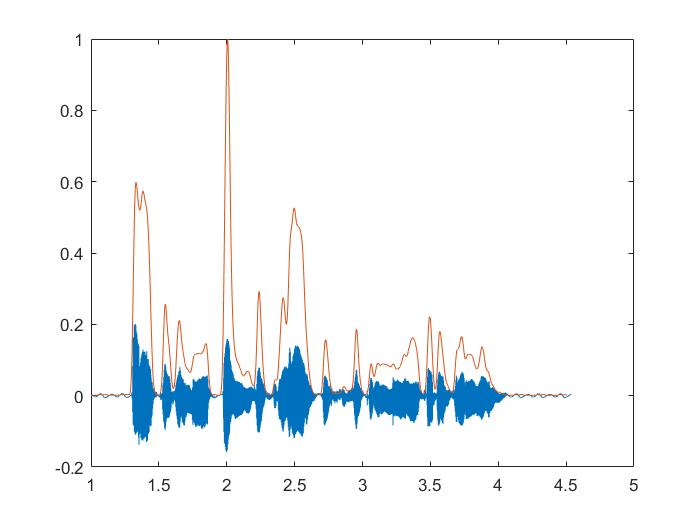
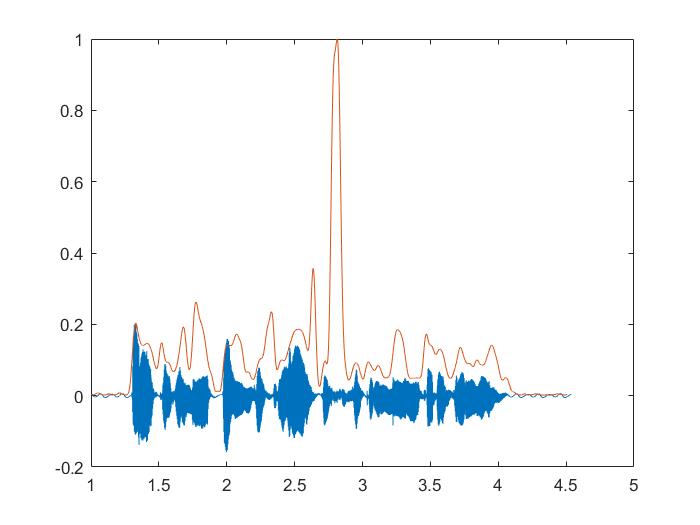
Στο ερώτημα αυτό χρησιμοποιήθηκε το σήμα φωνής που δόθηκε στα βοηθητικά αρχεία. Ακολουθούν τα διαγράμματα βραχέος χρόνου και ρυθμού εναλλαγής προσήμου αντίστοιχα.



*Μήκος παραθύρου 20ms.*

Παρατηρούμε ότι όσο μεγαλώνει η ένταση του σήματος, η ενέργεια βραχέος χρόνου αυξάνεται ενώ ο ρυθμός εναλλαγής προσήμου αυξάνεται όπου η ένταση του σήματος είναι μικρή.

Αν αυξήσουμε το παράθυρο:



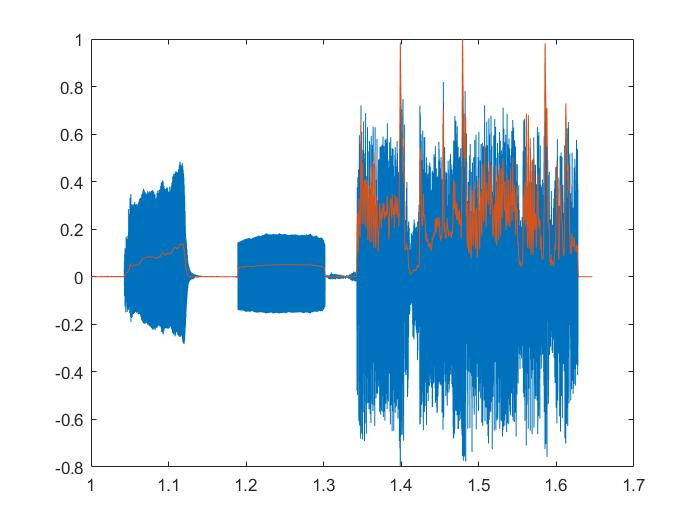
*Μήκος παραθύρου 80ms.*

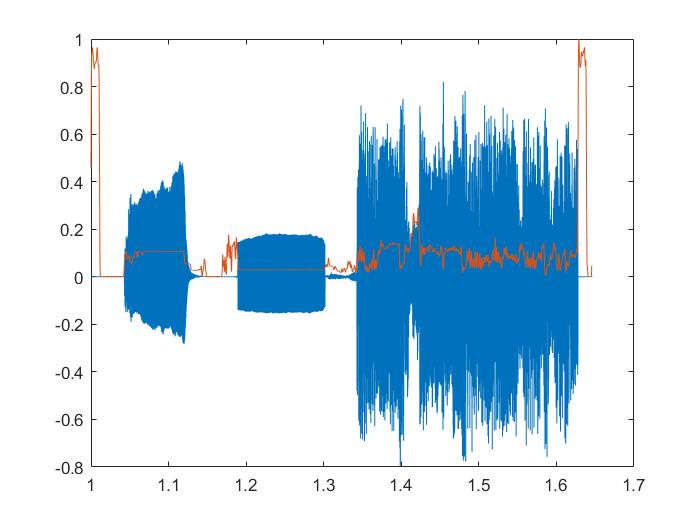
Παρατηρούμε ότι όταν μεγαλώνουμε το μήκος του παραθύρου η διακριτική μας ικανότητα στις απότομες αλλαγές μειώνεται.

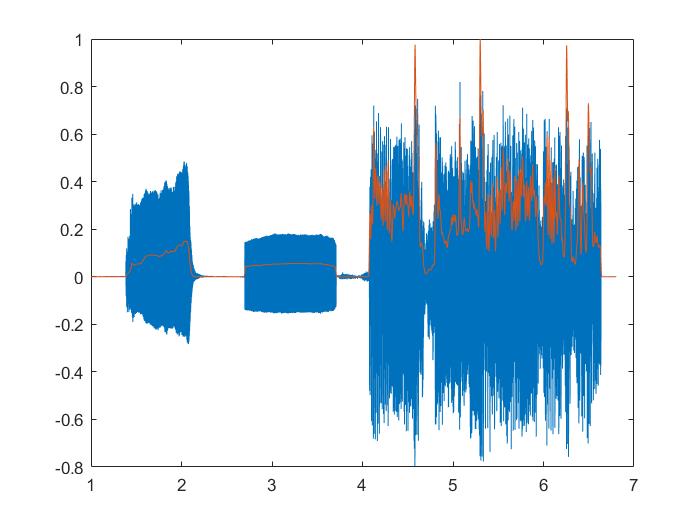
Με τη χρήση αυτών των μετρήσεων μπορούμε να διαχωρίσουμε τη φωνή από τη σιωπή αλλά και τους άφωνους από τους έμφωνους ήχους. Οι έμφωνοι ήχοι έχουν υψηλή ενέργεια βραχέος χρόνου σε σχέση με τους άφωνους και χαμηλό ρυθμό εναλλαγής προσήμου συγκριτικά με τους άφωνους. Αντίστοιχα η φωνή έχει υψηλή ενέργεια βραχέος χρόνου και χαμηλό ρυθμό εναλλαγής προσήμου συγκριτικά με τη σιωπή, η οποία μάλιστα έχει μηδενική ενέργεια.

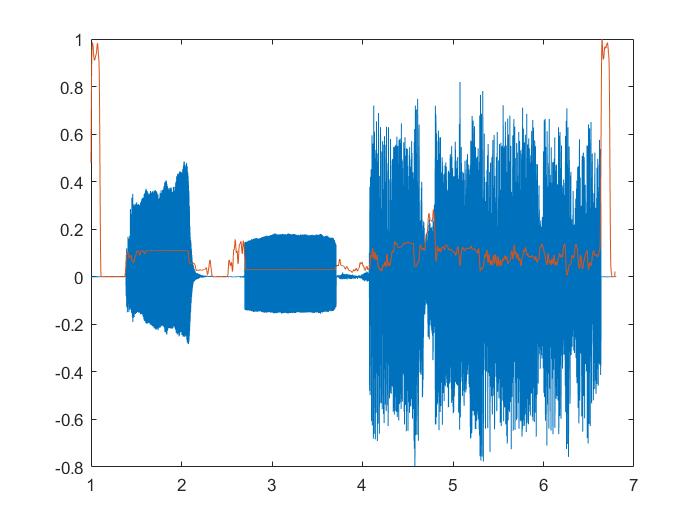
3.2

Στο ερώτημα αυτό χρησιμοποιήθηκε το σήμα μουσικής που δόθηκε στα βοηθητικά αρχεία. Ακολουθούν τα διαγράμματα βραχέος χρόνου και ρυθμού εναλλαγής προσήμου αντίστοιχα.



*Μήκος παραθύρου 20ms.*

Αν αυξήσουμε το παράθυρο:



*Μήκος παραθύρου 30ms*

Στο σήμα μουσικής μπορούμε εύκολα να διακρίνουμε, μέσω των μετρήσεων μας, πού ο ήχος είναι σταθερός – μια νότα για παράδειγμα - καθώς η ενέργεια αλλά και ο ρυθμός εναλλαγής προσήμου είναι σταθερός, ενώ όταν έχουμε μουσική με λόγια η ενέργεια μεταβάλλεται συνεχώς. Φυσικά και εδώ είμαστε ικανοί να ξεχωρίσουμε τον ήχο από τη σιωπή, αλλά είναι αρκετά πιο δύσκολο να ξεχωρίσουμε τους έμφωνους από τους άφωνους ήχους διότι ο έντονος θόρυβος έχει ως αποτέλεσμα ο ρυθμός εναλλαγής προσήμου να μην έχει σημαντικές διαφορές και στο διάγραμμα της ενέργειας μπορούμε να διακρίνουμε μόνο τρεις από τους τέσσερις έμφωνους ήχους του τραγουδιού καθώς υπάρχουν μονάχα τρία ξεσπάσματα στην ενέργεια που ξεπερνούν την ενέργεια του θορύβου.

# References

iitg.vlab.co.in. (2011). *Short Term Time Domain Processing of Speech* . Retrieved from http://vlab.amrita.edu/index.php?sub=59&brch=164&sim=857&cnt=1

Katsamanis Nassos, P. V. (n.d.). Retrieved from http://cvsp.cs.ntua.gr: http://cvsp.cs.ntua.gr/~nassos/resources/speech\_course\_2004/OnlineSpeechDemos/speechDemo\_2004\_Part1.html#2

Sharma, N. S. (n.d.). Retrieved from Short-time Energy and Zero Crossing Rate: https://www.mathworks.com/matlabcentral/fileexchange/23571-short-time-energy-and-zero-crossing-rate