

Ψηφιακή Επεξεργασία Σήματος

1η Εργαστηριακή Άσκηση

Ομάδα εργασίας:

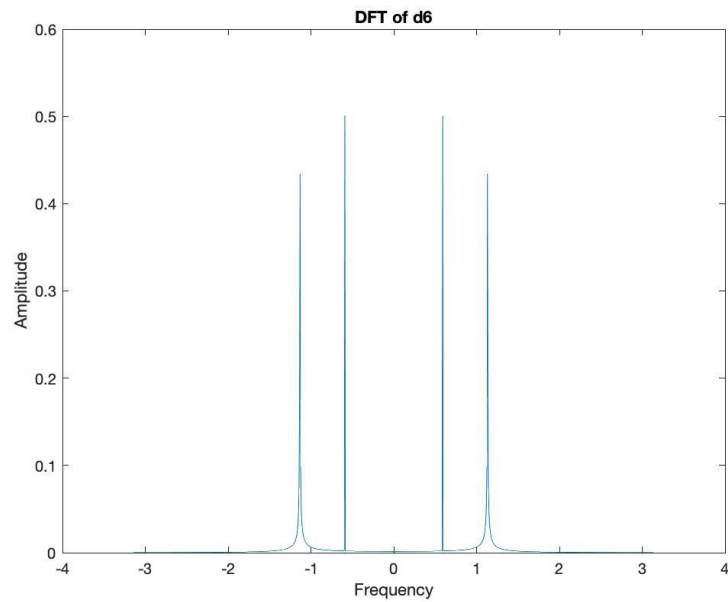
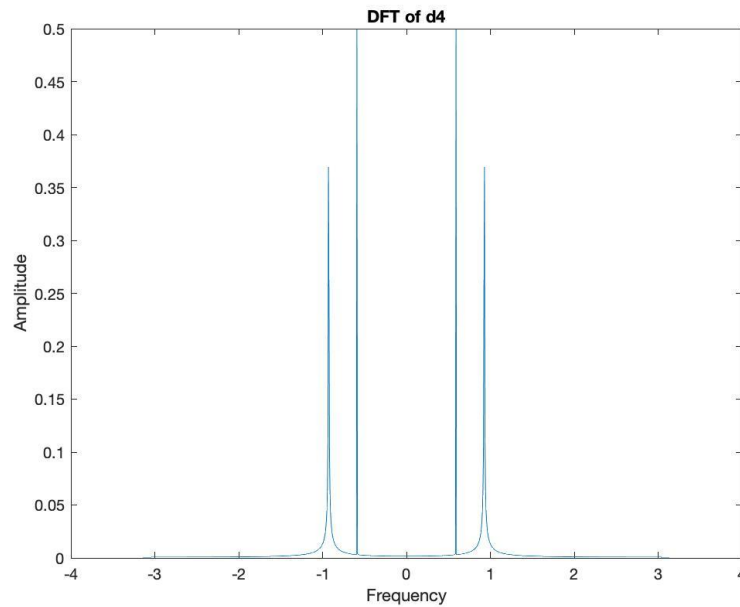
Σιφναίος Σάββας (ΑΜ: 03116080)

Στούμπου Βασιλική (ΑΜ: 03116026)

6ο εξάμηνο, ΗΜΜΥ

Μέρος 1ο

- 1.1) Δημιουργήσαμε τους 10 τόνους με βάση τον δοθέντα πίνακα συχνοτήτων-τόνων. Ο κάθε τόνος έχει μήκος 1000 δείγματα.
- 1.2) Υπολογίσαμε τον DFT των σημάτων $d_4[n]$ και $d_6[n]$ χρησιμοποιώντας τη συνάρτηση `fft()`. Επίσης, μέσω της συνάρτησης `fftshift()` μετατοπίσαμε το φάσμα του κάθε τόνου, ώστε οι κορυφές του φάσματος να αντιστοιχούν στις πραγματικές τιμές των συχνοτήτων ω .



- 1.3) Το σήμα που δημιουργήσαμε βάσει του αθροίσματος των AM μας αποτελείται από τους τόνους που αντιστοιχούν στο εξής νούμερο: 0 6 2 3 2 1 0 6. Στο διάνυσμα του σήματος, παρεμβάλαμε μεταξύ δύο τόνων ένα μηδενικό διάνυσμα 100 θέσεων, έτσι ώστε δύο διαδοχικοί τόνοι να διαχωρίζονται με 100 μηδενικά. Χρησιμοποιώντας τη συνάρτηση `audiowrite()`, ηχογραφήσαμε το παραπάνω σήμα και το αποθηκεύσαμε σε ένα αρχείο τύπου `.wav` με το όνομα `"tone_sequence.wav"`.
- 1.4) Για αυτό το ερώτημα ορίσαμε ένα τετραγωνικό και ένα Hamming παράθυρο μήκους 1000 δειγμάτων (με τις συναρτήσεις `rectwin()` και `hamming()` αντίστοιχα), τα οποία χρησιμοποιήσαμε στη συνέχεια προκειμένου να παραθυροποιήσουμε το παραπάνω σήμα. Λάβαμε 8 παραθυροποιημένα σήματα σε κάθε περίπτωση (όσοι δηλαδή και οι τόνοι του σήματος). Στη συνέχεια, με χρήση της `fft()` και `fftshift()` εφαρμόσαμε Μετασχηματισμό DFT σε κάθε παραθυροποιημένο σήμα αποθηκεύοντας τα αποτελέσματα σε έναν πίνακα 8 x 1000 (`WindowedSignal1` για το τετραγωνικό παράθυρο και `WindowedSignal2` για το παράθυρο Hamming). Κάθε γραμμή του πίνακα αντιστοιχεί στον DFT κάθε παραθυροποιημένου σήματος.
- 1.5) Για τις ανάγκες αυτού του ερωτήματος ορίσαμε τη συνάρτηση `mypeaks`, η οποία δέχεται ως ορίσματα ένα μετασχηματισμένο κατά DFT σήμα και τον αριθμό των τόνων που αυτό περιέχει. Για κάθε έναν από τους μετασχηματισμένους τόνους, βρίσκει τους δείκτες στο διάνυσμά του όπου το φάσμα του παρουσιάζει τις δύο κορυφές, με τη χρήση της `findpeaks()` και της `maxk()`. Ως έξοδο δίνει τους δείκτες αυτούς που αντιστοιχούν μόνο στις θετικές τιμές συχνοτήτων. Χρησιμοποιώντας αυτή τη συνάρτηση εντοπίσαμε τους δείκτες `k` για τους οποίους εμφανίζονται οι κορυφές στο φάσμα κάθε μετασχηματισμένου τόνου και τους αντιστοιχίσαμε στις πλησιέστερες στις `touch tone` συχνότητες.
- Επίσης, δημιουργήσαμε τον 10x3 πίνακα `"points"`, ο οποίος στις δύο πρώτες στήλες του περιέχει τους δείκτες στο διάνυσμα κάθε μετασχηματισμένου κατά DFT τόνου όπου εμφανίζονται οι κορυφές και στην τρίτη περιέχει τον αριθμό του αντίστοιχου τόνου.
- Σε αντιστοιχία, δημιουργούμε και τον πίνακα `frequencies`, ο οποίος περιέχει τις τιμές των πλησιέστερων στις `touch tone` συχνότητες που αντιστοιχούν σε κάθε αριθμό.
- 1.6) Δημιουργήσαμε τη συνάρτηση `ttdecode`, η οποία δέχεται ως όρισμα ένα τονικό σήμα. Αρχικά, διαχωρίζουμε τους άγνωστους τόνους από τα μηδενικά στο σήμα εισόδου και τους αποθηκεύουμε σε έναν πίνακα με πλήθος γραμμών ίσο με τον αριθμό των τόνων και πλήθος στηλών ίσο με 1000, όσα και τα δείγματα κάθε τόνου. Στη συνέχεια, εφαρμόζουμε μετασχηματισμό DFT σε κάθε τόνο και υπολογίζουμε την ενέργειά του. Χρησιμοποιούμε τη συνάρτηση `mypeaks` για να εντοπίσουμε τις θέσεις όπου η ενέργεια εμφανίζει κορυφές και τις συγκρίνουμε με

τις δύο πρώτες στήλες του πίνακα points. Έτσι, επιτυγχάνεται η αναγνώριση του νούμερου που αντιστοιχεί σε κάθε τόνο.

Χρησιμοποιήσαμε αυτή την συνάρτηση θέτοντας ως είσοδο το σήμα του ερωτήματος 1.3 και λάβαμε το αναμενόμενο αποτέλεσμα:

0 6 2 3 2 1 0 6

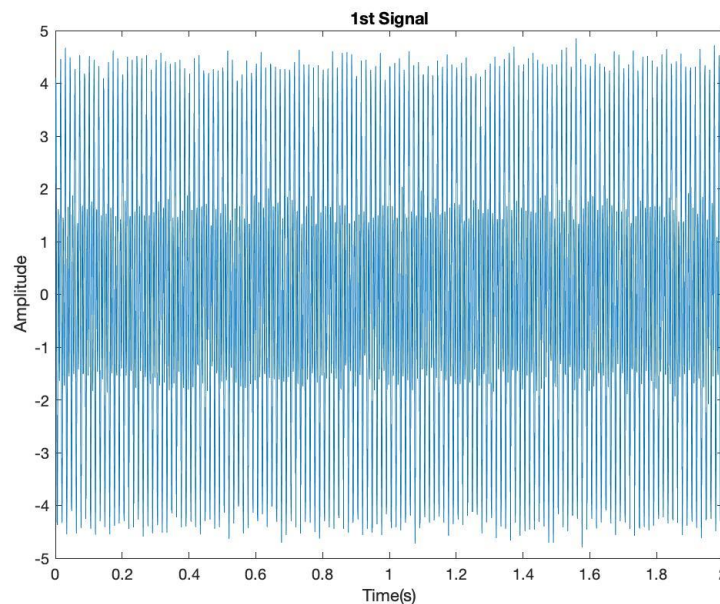
- 1.7) Τέλος, χρησιμοποιήσαμε τη συνάρτηση ttdecode() για να αναγνωρίσουμε το νούμερο που αντιστοιχεί στα σήματα easySig και hardSig. Τα αποτελέσματα ήταν τα εξής:

Για το easySig: 1 3 2 6 3 9 0 0

Για το hardSig: 9 0 9 6 3 2 1 1 9 1

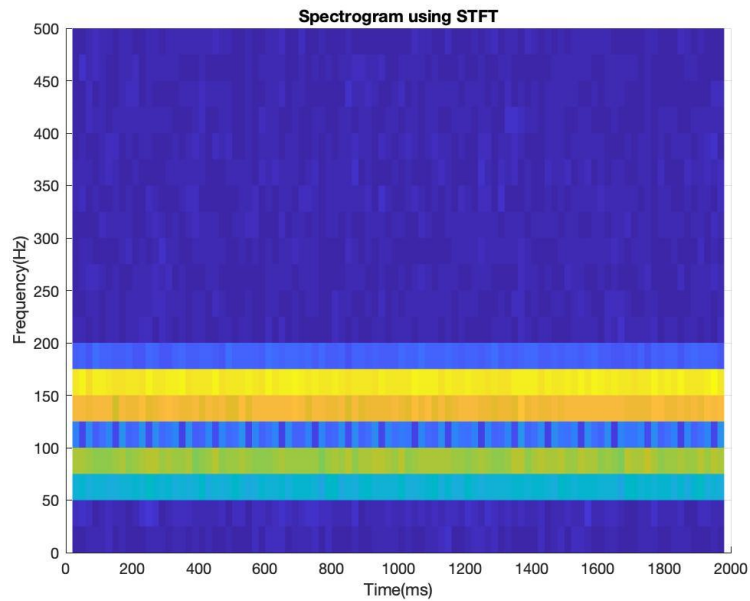
Μέρος 2ο

2.1) α] Δημιουργήσαμε το σήμα $x[n]$ από το σήμα $x(t) = 2 \cos(2\pi 70t) + 3 \sin(2\pi 140t) + 0.15v(t)$ στο οποίο εφαρμόσαμε δειγματοληψία με ρυθμό 1000 Hz, όπου $v(t)$ είναι λευκός Gaussian θόρυβος που υλοποιείται με τη συνάρτηση randn(). Παρακάτω φαίνεται η γραφική παράσταση του σήματος:

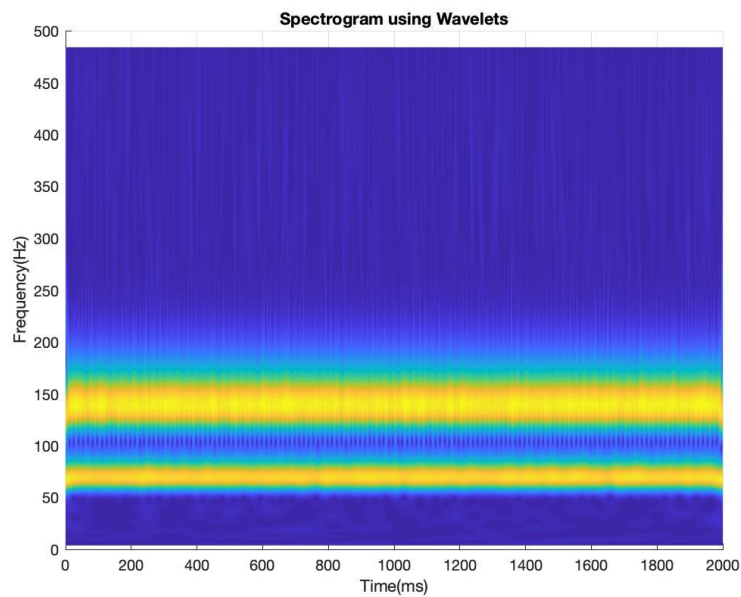


β] Στη συνέχεια, υπολογίσαμε τον STFT του σήματος με χρήση της συνάρτησης spectrogram(), με παράθυρο μεγέθους 40 msec και επικάλυψη 20 msec. Για να υπολογιστεί ο διακριτός Fourier μετασχηματισμός χρησιμοποιήθηκαν 40 δείγματα σε κάθε παράθυρο. Παρακάτω φαίνεται το σπεκτρογράφημα του μετασχηματισμένου κατά STFT σήματος. Για την απεικόνιση χρησιμοποιήσαμε τη συνάρτηση surf, η οποία

επιστρέφει ένα τρισδιάστατο γράφημα, το οποίο μετατρέψαμε σε δισδιάστατο μέσω της `view()`.

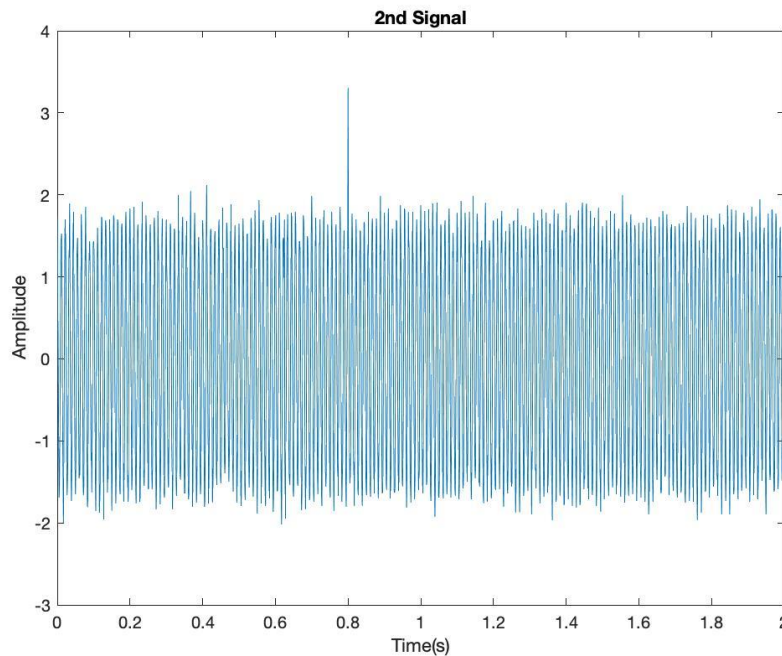


γ] Στο σημείο αυτό, υπολογίσαμε τον DT-CWT μέσω της συνάρτησης `cwtft`. Για τον ορισμό της κλίμακας και της συχνότητας χρησιμοποιήσαμε τη συνάρτηση `wavelets` του συμπληρωματικού υλικού. Σε όλες τις περιπτώσεις χρησιμοποιήθηκε το κυματίδιο Morlet. Ομοίως με το προηγούμενο ερώτημα, χρησιμοποιήθηκαν οι συναρτήσεις `surf` και `view` ώστε να προκύψει το ακόλουθο σπεκτρογράφημα:

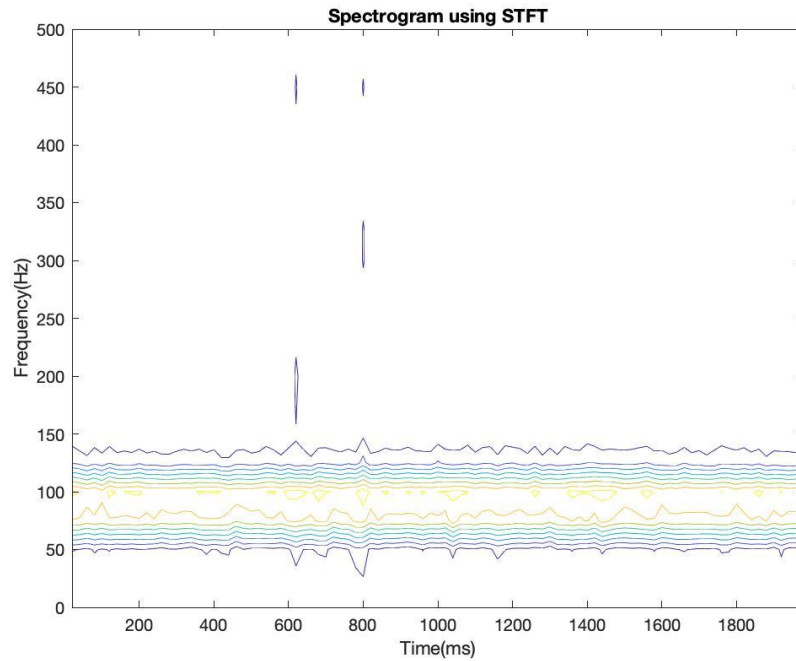


δ] Παρατηρούμε ότι στην περίπτωση του μετασχηματισμού κυματιδίων, προκύπτει ένα σπεκτρογράφημα με πολύ καλύτερη ανάλυση και σαφή διαχωρισμό του συχνοτικού του περιεχομένου (συχνότητες του συνημιτόνου και του ημιτόνου 70 και 140 Hz αντίστοιχα). Αντίθετα, στην περίπτωση του STFT οι συχνότητες δεν είναι ευδιάκριτες. Αυτό συμβαίνει γιατί κατά τον μετασχηματισμό κυματιδίων, σε αντίθεση με τον STFT, το παράθυρο που χρησιμοποιείται έχει μεταβλητό εύρος ανάλογα με την περιοχή συχνοτήτων. Σε χαμηλές συχνότητες, επιλέγεται παράθυρο μεγαλύτερης χρονικής διάρκειας, ενώ στις υψηλές συχνότητες επιλέγεται παράθυρο μικρότερης χρονικής διάρκειας. Αυτό έχει ως αποτέλεσμα στις υψηλές συχνότητες να έχουμε καλή ανάλυση στον χρόνο, ενώ στις χαμηλές συχνότητες να έχουμε καλή ανάλυση στη συχνότητα.

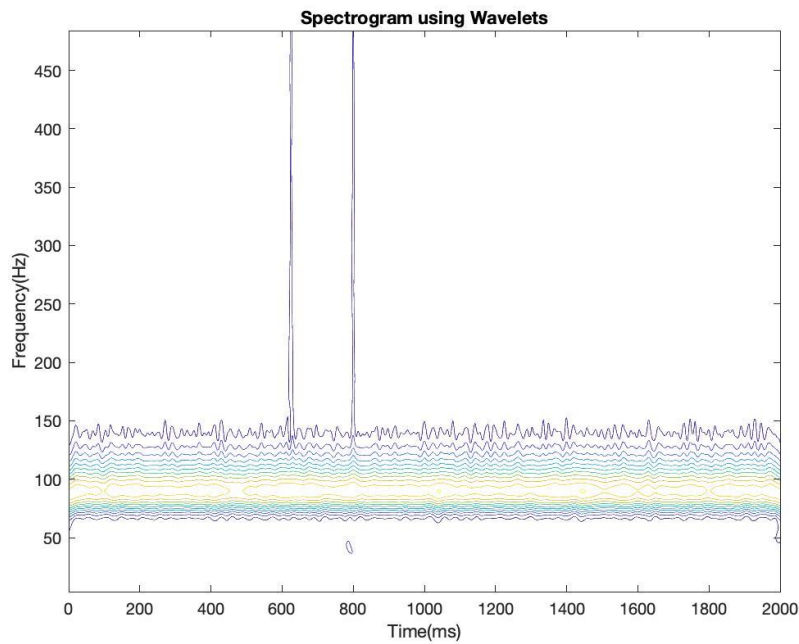
2.2) α] Δημιουργήσαμε το σήμα $y[n]$ από το σήμα $y(t) = 1.7 \cos(2\pi 90t) + 0.15v(t) + 1.7\delta(t - 0.625) + 1.7\delta(t - 0.800)$, στο οποίο εφαρμόσαμε δειγματοληψία με ρυθμό 1000 Hz, όπου $v(t)$ είναι λευκός Gaussian θόρυβος που υλοποιείται με τη συνάρτηση `randn()`. Παρακάτω φαίνεται η γραφική παράσταση του σήματος:



β] Στη συνέχεια, υπολογίσαμε τον STFT του σήματος με χρήση της συνάρτησης `spectrogram()`, με παράθυρο μεγέθους 40 msec και επικάλυψη 20 msec. Για να υπολογιστεί ο διακριτός Fourier μετασχηματισμός χρησιμοποιήθηκαν 40 δείγματα σε κάθε παράθυρο. Παρακάτω φαίνεται το σπεκτρογράφημα του μετασχηματισμένου κατά STFT σήματος. Για την απεικόνιση χρησιμοποιήσαμε τη συνάρτηση `contour`, η οποία επιστρέφει τις ισοδυναμικές γραμμές του σήματος αυτού.



γ] Στο σημείο αυτό, υπολογίσαμε τον DT-CWT μέσω της συνάρτησης `cwtft`. Για τον ορισμό της κλίμακας και της συχνότητας χρησιμοποιήσαμε τη συνάρτηση `wavelets` του συμπληρωματικού υλικού. Σε όλες τις περιπτώσεις χρησιμοποιήθηκε το κυματίδιο Morlet. Ομοίως με το προηγούμενο ερώτημα, χρησιμοποιήθηκε η συνάρτηση `contour` για την απεικόνιση των ισοδυναμικών γραμμών στο σπεκτρογράφημα.



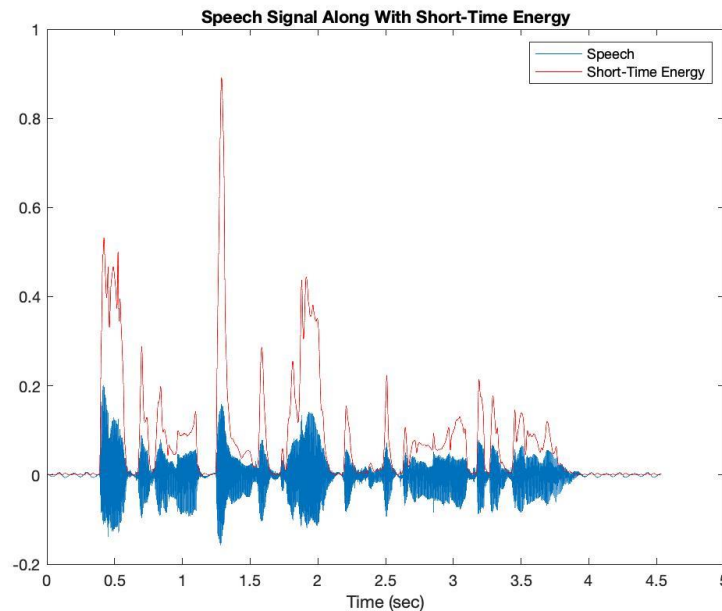
δ] Παρατηρούμε ότι οι ισοδυναμικές γραμμές στο διάγραμμα στην περίπτωση των κυματιδίων είναι πολύ πιο λεπτομερείς σε σχέση με την περίπτωση του STFT. Επίσης, οι δύο κρουστικές στις χρονικές στιγμές 0.625 sec και 0.800 sec αποτυπώνονται πλήρως μόνο στο διάγραμμα στην περίπτωση των κυματιδίων. Αυτό σημαίνει ότι στην περίπτωση του STFT είναι πιο δύσκολο να εντοπιστούν οι μεταβολές.

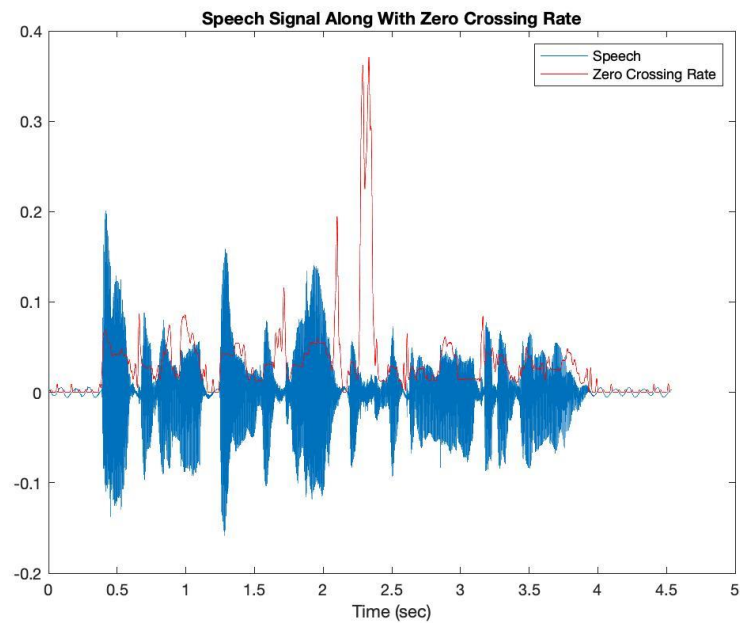
Μέρος 3^ο

3.1) Στην άσκηση αυτή υπολογίσαμε την ενέργεια βραχέος χρόνου (STE) και τον ρυθμό εναλλαγής προσήμου (ZCR) για το δοθέν σήμα φωνής `speech_utterance.wav`, το οποίο παραθυροποιήσαμε χρησιμοποιώντας σταθερού μήκους παράθυρα. Για λόγους σύγκρισης, εφαρμόσαμε τους υπολογισμούς μας για διάφορα μήκη παραθύρων.

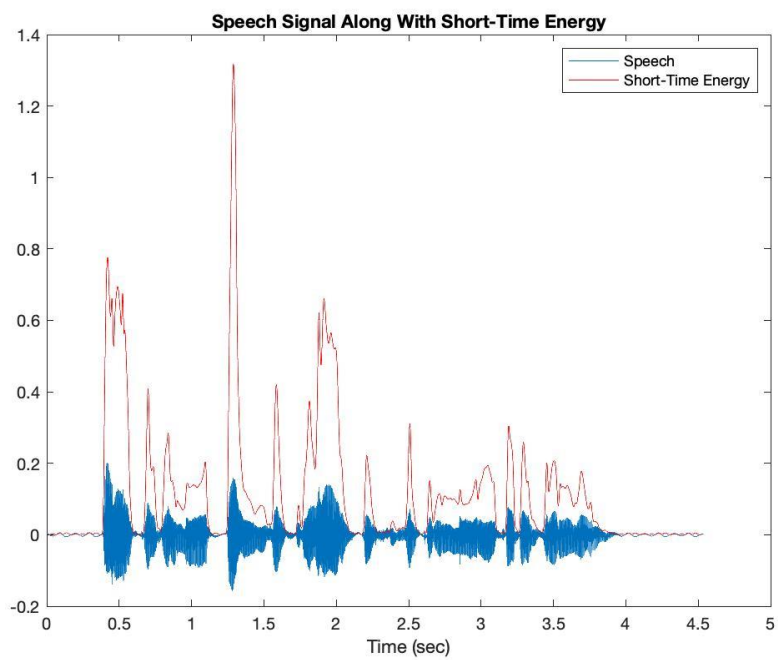
Παρακάτω παρουσιάζονται οι γραφικές παραστάσεις ενέργειας βραχέος χρόνου – σήματος φωνής και ρυθμού εναλλαγής προσήμου – σήματος φωνής σε κοινά διαγράμματα για μήκη παραθύρων 20 ms και 30 ms.

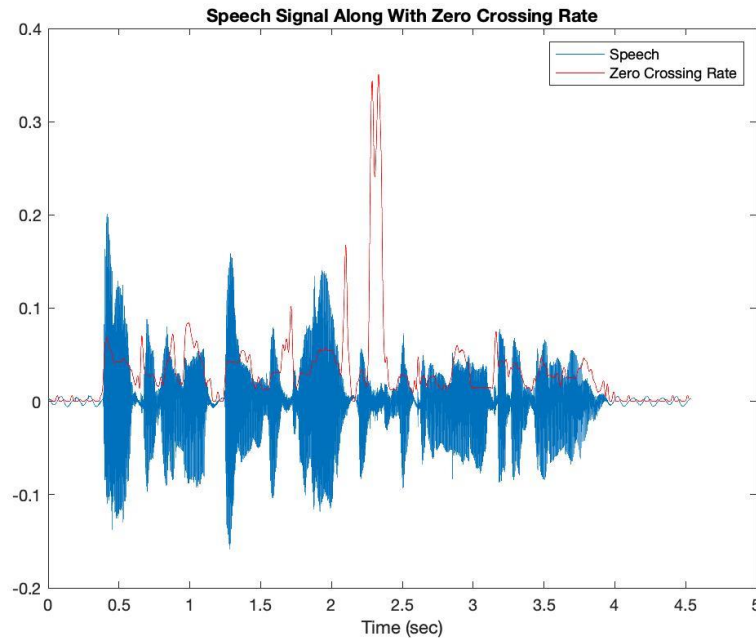
Μήκος παραθύρου: 20 ms





Μήκος παραθύρου: 30 ms





Γενικώς, παρατηρούμε ότι η ενέργεια βραχέος χρόνου συμπεριφέρεται αντίστροφα σε σχέση με τον ρυθμό εναλλαγής προσήμου: μεγάλη τιμή ενέργειας συνήθως αντιστοιχεί σε μικρή τιμή ρυθμού εναλλαγής προσήμου και το αντίθετο.

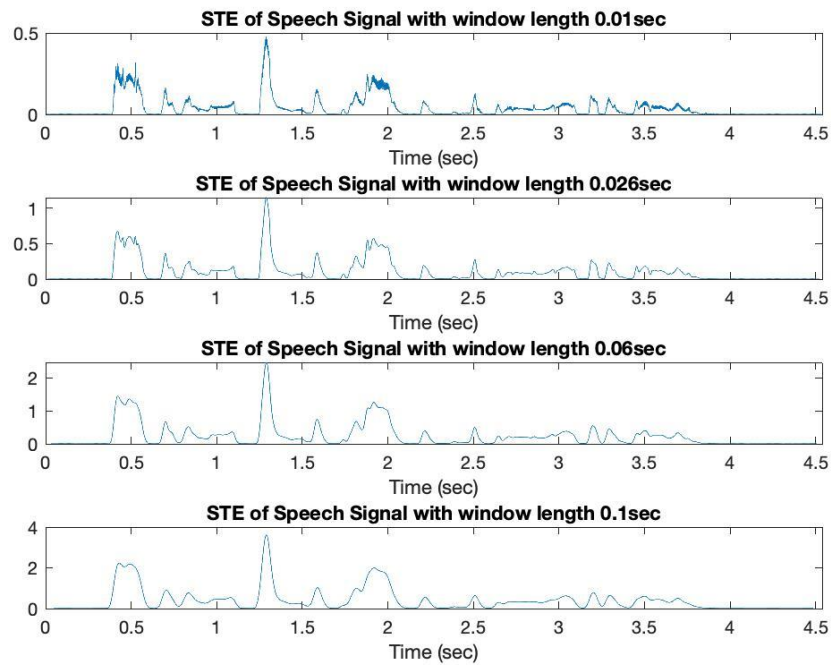
Σημειώνεται ότι από τα διαγράμματα αυτά μπορούμε να διακρίνουμε έμφωνους από άφωνους ήχους καθώς και σιωπή από ομιλία.

Για παράδειγμα στα χρονικά διαστήματα 0.00-0.38 sec και 3.960-4.528 sec παρατηρούμε ότι έχουμε σιωπή.

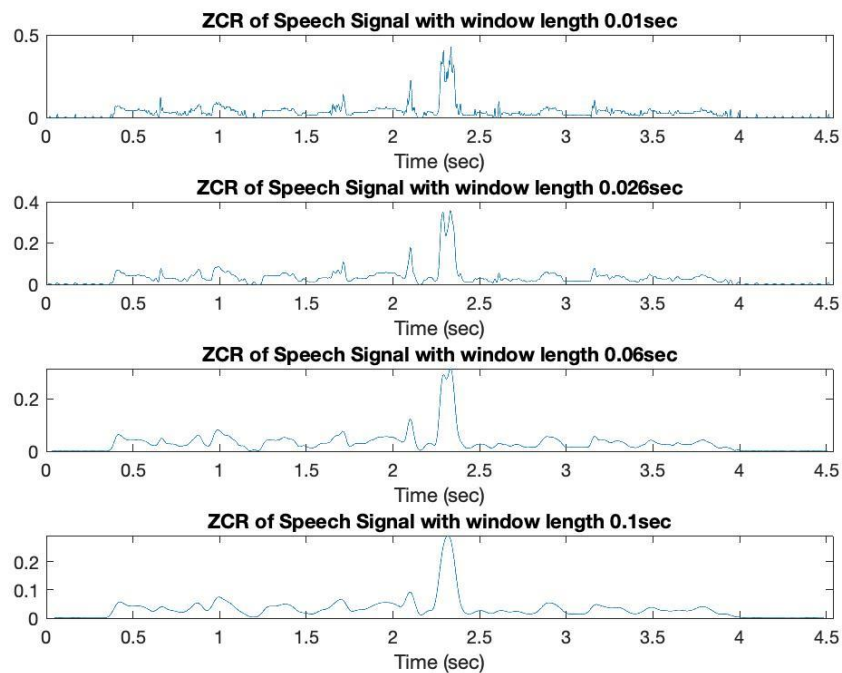
Κατά κανόνα, τα διαστήματα όπου παρατηρείται υψηλή ενέργεια βραχέος χρόνου και χαμηλός ρυθμός εναλλαγής προσήμου αντιστοιχούν σε έμφωνους ήχους, ενώ το αντίθετο συμβαίνει στην περίπτωση των άφωνων ήχων.

Συγκεκριμένα, το διάστημα 0.38-0.60 sec αντιστοιχεί στη λέξη «αλλά». Σε αυτό το διάστημα παρατηρούμε υψηλή ενέργεια βραχέος χρόνου και χαμηλό ρυθμό εναλλαγής προσήμου, επομένως η λέξη αυτή κατατάσσεται στους έμφωνους ήχους. Επιπλέον, στο διάστημα 1.25-1.40 sec, που αντιστοιχεί στο φωνήεν ε, παρατηρούμε ότι έχουμε και πάλι το αντίστοιχο φαινόμενο. Αντίθετα, στο διάστημα 2.27-2.35 sec, το οποίο αντιστοιχεί στο σύμφωνο τ, παρατηρούμε ιδιαίτερα χαμηλή τιμή της ενέργειας βραχέος χρόνου και υψηλό ρυθμό εναλλαγής προσήμου.

Παρακάτω παρατίθενται οι γραφικές απεικονίσεις των ενεργειών βραχέος χρόνου (STE) του σήματος, για παράθυρα διαφορετικού μήκους, σε κοινό διάγραμμα.



Επίσης, παραθέτουμε τις γραφικές απεικονίσεις του ρυθμού εναλλαγής προσήμου του σήματος για παράθυρα διαφορετικού μήκους.



Στο σημείο αυτό, αξίζει να σημειωθεί πως χρονική μεταβολή του μήκους του παραθύρου επιφέρει αισθητές αλλαγές στη γραφική απεικόνιση των δύο μεγεθών. Συγκεκριμένα, όπως φαίνεται και από τα παραπάνω δύο διαγράμματα, χρονική αύξηση του μήκους του παραθύρου συνεπάγεται απώλεια λεπτομερειών στα γραφήματα των παραπάνω μεγεθών. Παρατηρείται, δηλαδή, εξομάλυνση των γραφημάτων και συνεπώς μικρές αλλαγές στις τιμές της ενέργειας βραχέος χρόνου και του ρυθμού εναλλαγής προσήμου. Αντίθετα, μικρότερα χρονικά παράθυρα έχουν ως αποτέλεσμα πολύ απότομες μεταβολές στις τιμές των δύο μεγεθών, με αποτέλεσμα την υπερβολική κυμάτωση που παρατηρούμε στο διάγραμμα. Γι' αυτό και το ιδανικό μέγεθος παραθύρου κυμαίνεται από 20 έως 30 ms, που αποτελεί μια ενδιάμεση λύση.

Επίσης, παρατηρούμε ότι με την αύξηση του μήκους του παραθύρου οι τιμές της ενέργειας βραχέος χρόνου αυξάνονται, ενώ του ρυθμού εναλλαγής προσήμου μειώνονται. Τέλος, αξίζει να επισημάνουμε ότι καθώς αυξάνουμε το μήκος του παραθύρου, οι διαφοροποιήσεις μεταξύ των διαγραμμάτων είναι λιγότερες. Η μεγάλη διαφοροποίηση από διάγραμμα σε διάγραμμα γίνεται κυρίως για μικρές τιμές διάρκειας παραθύρου.

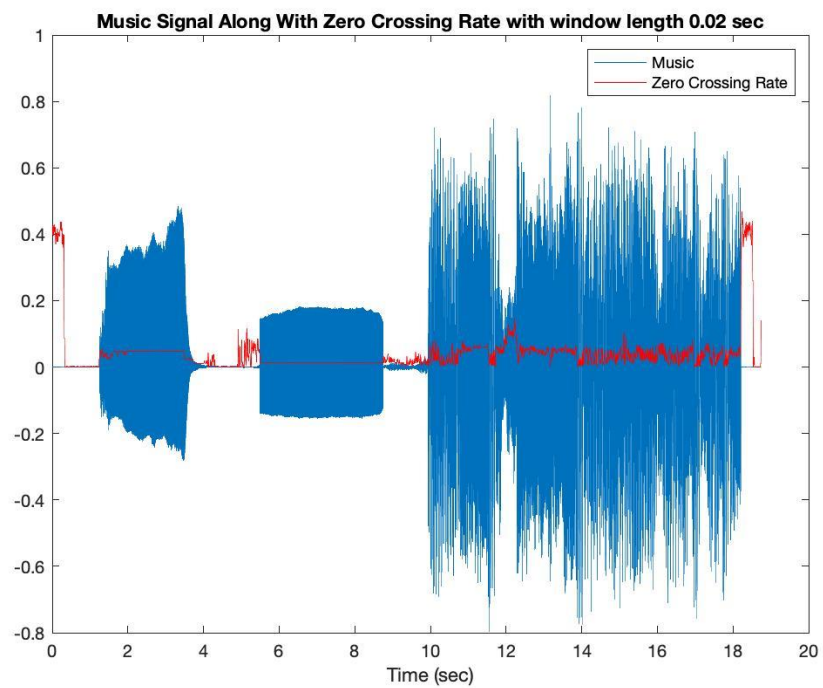
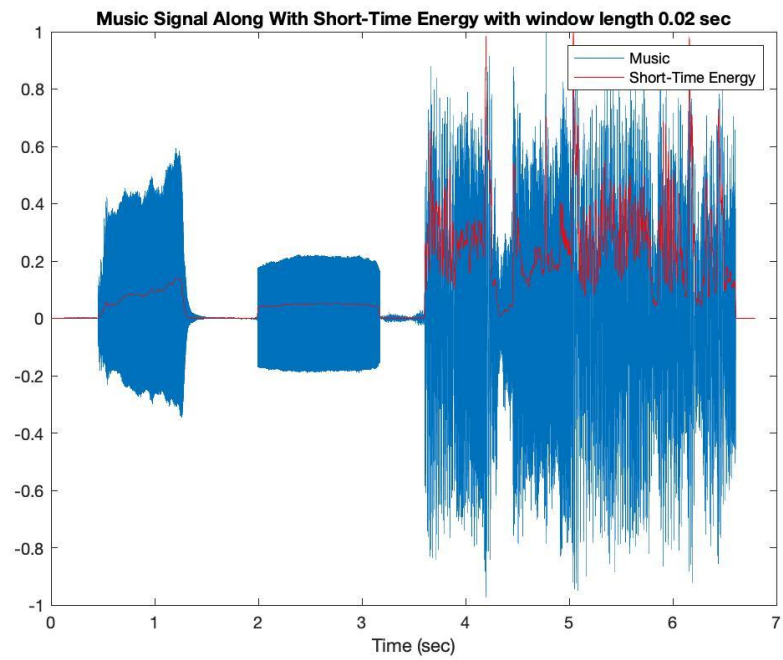
3.2) Στη συνέχεια, εξετάσαμε το σήμα μουσικής «music_cut.wav».

Παρακάτω παρουσιάζονται οι γραφικές παραστάσεις ενέργειας βραχέος χρόνου – σήματος μουσικής και ρυθμού εναλλαγής προσήμου – σήματος μουσικής σε κοινά διαγράμματα για μήκη παραθύρων 20 ms και 30 ms.

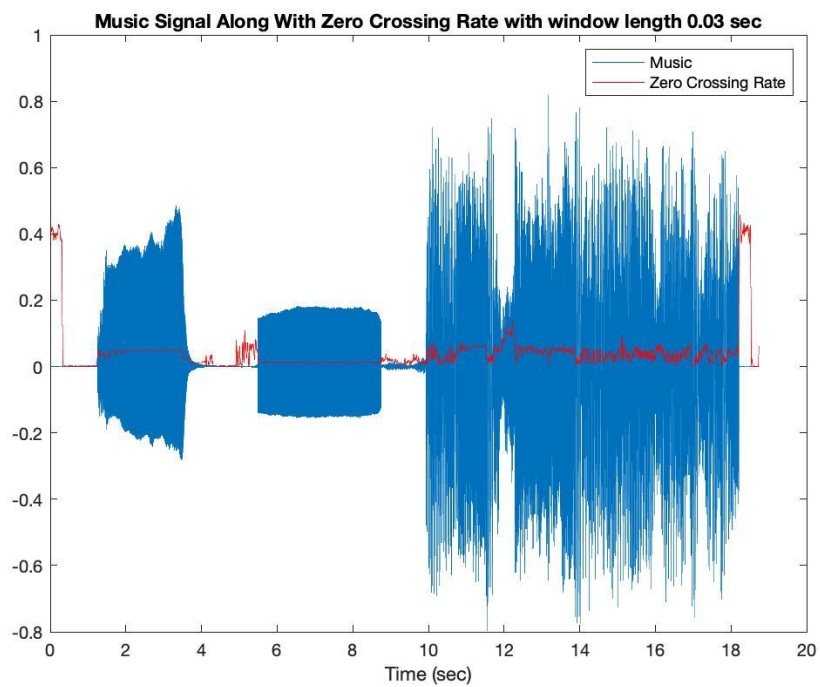
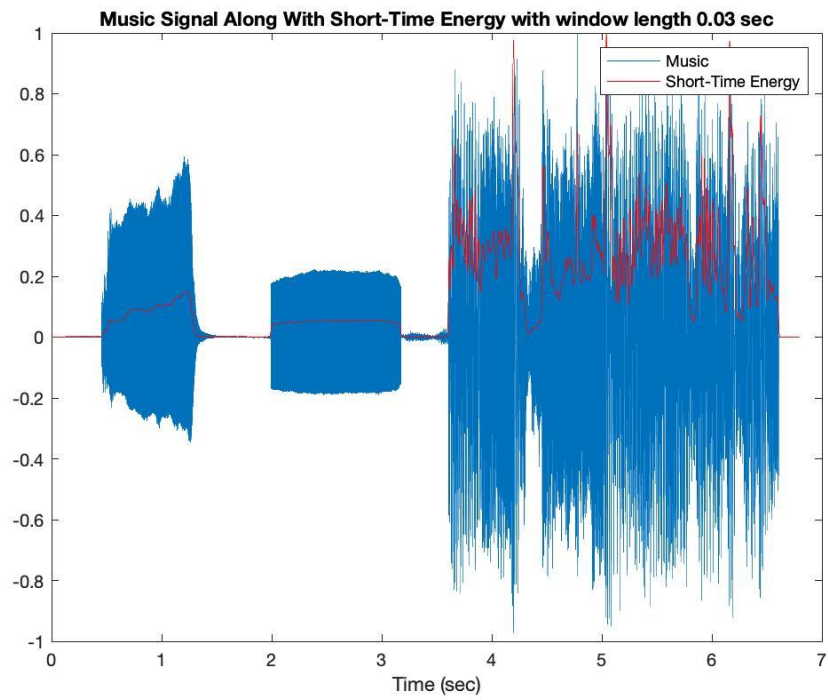
Να σημειωθεί ότι στην περίπτωση του σήματος μουσικής, επιλέξαμε να κανονικοποιήσουμε τις τιμές του σήματος, της ενέργειας βραχέος χρόνου και του ρυθμού εναλλαγής προσήμου για τις ανάγκες του διαγράμματος, ώστε να έχουμε καλύτερη εποπτεία των μεγεθών συγκριτικά. Ειδικά στην περίπτωση ενέργειας βραχέος χρόνου, επειδή οι τιμές της ήταν πολύ μεγάλες συγκριτικά με το σήμα, θα ήταν πιο δύσκολη η από κοινού εξέταση των διαγραμμάτων.

Το εκ πρώτης όψεως συγκρίσιμο πλάτος των μεγεθών των παρακάτω γραφημάτων οφείλεται στο γεγονός ότι και τα δύο μεγέθη έχουν κανονικοποιηθεί. Στην πραγματικότητα, όμως, η ενέργεια βραχέος χρόνου είναι δύο τάξεις μεγαλύτερη περίπου σε σχέση με τον ρυθμό εναλλαγής προσήμου.

Μήκος παραθύρου: 20 ms



Μήκος παραθύρου: 30 ms



Και εδώ επιβεβαιώνεται η αντίστροφη συμπεριφορά της ενέργειας βραχέος χρόνου και του ρυθμού εναλλαγής προσήμου, όπως φαίνεται από τα διαγράμματα.

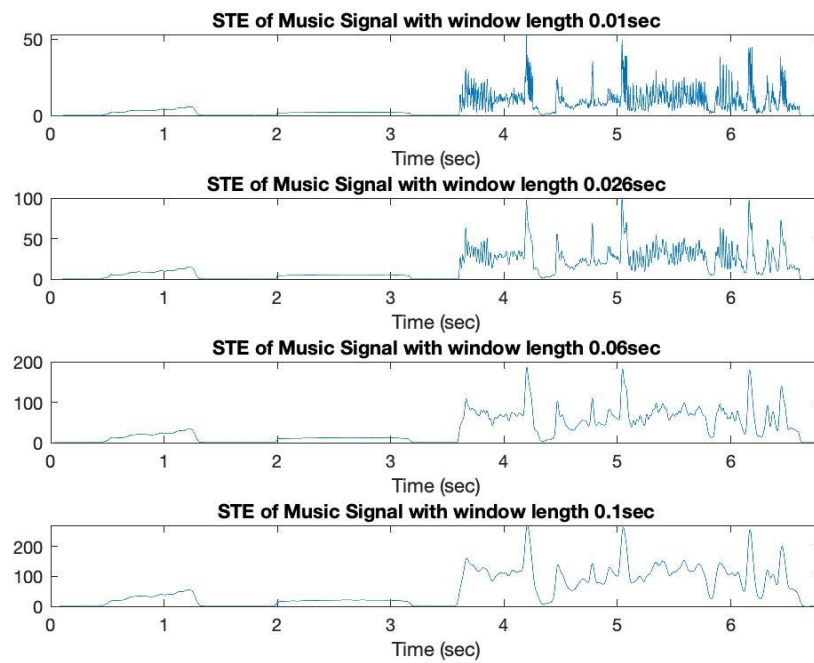
Από τα διαγράμματα αυτά μπορούμε να διακρίνουμε και πάλι έμφωνους από άφωνους ήχους καθώς και σιωπή από ήχο.

Για παράδειγμα στα χρονικά διαστήματα 0.00-0.47 sec και 1.77-1.99 sec παρατηρούμε ότι έχουμε σιωπή.

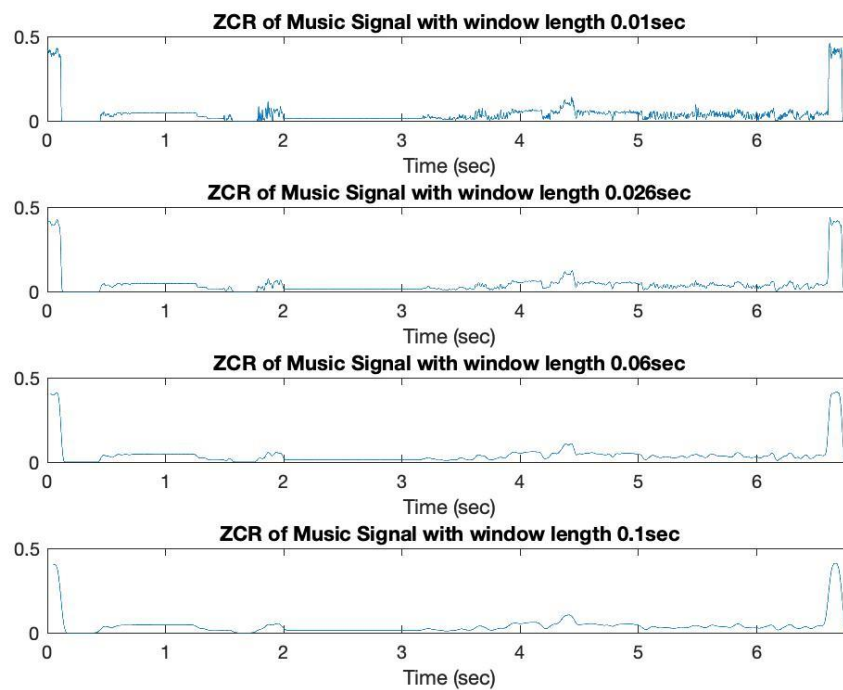
Κατά κανόνα, τα διαστήματα όπου παρατηρείται υψηλή ενέργεια βραχέος χρόνου και χαμηλός ρυθμός εναλλαγής προσήμου αντιστοιχούν σε έμφωνους ήχους, ενώ το αντίθετο συμβαίνει στην περίπτωση των άφωνων ήχων.

Συγκεκριμένα, το διάστημα 0.49-1.16 sec αντιστοιχεί σε ήχο βιολιού. Σε αυτό το διάστημα παρατηρούμε σχετικά υψηλή ενέργεια βραχέος χρόνου και χαμηλό ρυθμό εναλλαγής προσήμου, επομένως ο ήχος αυτός κατατάσσεται στους έμφωνους ήχους. Επιπλέον, στο διάστημα 1.99-3.18 sec, που αντιστοιχεί στον ήχο μίας κόρνας, παρατηρούμε ότι έχουμε και πάλι το αντίστοιχο φαινόμενο. Αντίθετα, στο διάστημα 3.19-3.6 sec, το οποίο αντιστοιχεί σε ήχο από κονσέρτο, παρατηρούμε ιδιαίτερα χαμηλή τιμή της ενέργειας βραχέος χρόνου και υψηλό ρυθμό εναλλαγής προσήμου, επομένως ο ήχος αυτός κατατάσσεται στους άφωνους ήχους. Τέλος, στο διάστημα 3.6-6.6 sec ακούμε ένα pop κομμάτι που χαρακτηρίζεται από πολύ υψηλή ενέργεια βραχέος χρόνου και πολύ χαμηλό ρυθμό εναλλαγής προσήμου, άρα αποτελεί έμφωνο ήχο.

Παρακάτω παρατίθενται οι γραφικές απεικονίσεις των ενεργειών βραχέος χρόνου (STE) του σήματος, για παράθυρα διαφορετικού μήκους, σε κοινό διάγραμμα.



Επίσης, παραθέτουμε τις γραφικές απεικονίσεις του ρυθμού εναλλαγής προσήμου του σήματος για παράθυρα διαφορετικού μήκους.



Και σε αυτή την περίπτωση παρατηρούμε πως χρονική μεταβολή του μήκους του παραθύρου επιφέρει αισθητές αλλαγές στη γραφική απεικόνιση των δύο μεγεθών. Συγκεκριμένα, όπως φαίνεται και από τα παραπάνω δύο διαγράμματα, χρονική αύξηση του μήκους του παραθύρου συνεπάγεται απώλεια λεπτομερειών στα γραφήματα των παραπάνω μεγεθών. Παρατηρείται, δηλαδή, εξομάλυνση των γραφημάτων και συνεπώς μικρές αλλαγές στις τιμές της ενέργειας βραχέος χρόνου και του ρυθμού εναλλαγής προσήμου. Αντίθετα, μικρότερα χρονικά παράθυρα έχουν ως αποτέλεσμα πολύ απότομες μεταβολές στις τιμές των δύο μεγεθών, με αποτέλεσμα την υπερβολική κυμάτωση που παρατηρούμε στο διάγραμμα. Γι' αυτό και το ιδανικό μέγεθος παραθύρου κυμαίνεται από 20 έως 30 ms, που αποτελεί μια ενδιάμεση λύση.

Επίσης, παρατηρούμε ότι με την αύξηση του μήκους του παραθύρου οι τιμές της ενέργειας βραχέος χρόνου αυξάνονται. Τέλος, αξίζει να επισημάνουμε ότι καθώς αυξάνουμε το μήκος του παραθύρου, οι διαφοροποιήσεις μεταξύ των διαγραμμάτων είναι λιγότερες. Η μεγάλη διαφοροποίηση από διάγραμμα σε διάγραμμα γίνεται κυρίως για μικρές τιμές διάρκειας παραθύρου.