



Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

**Σχολή Τεχνολογικών Εφαρμογών
Τμήμα Μηχανικών Πληροφορικής**

Πτυχιακή εργασία

**Εξαγωγή Σημασιολογικών και Μορφολογικών
Χαρακτηριστικών σε Μουσικά Αρχεία**

Καλαϊτζάκη Ευαγγελία (ΑΜ:2202)

**Επιβλέπων καθηγητής : Δρ. Αθανάσιος Μαλάμος
Αναπληρωτής Καθηγητής**



Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

**Σχολή Τεχνολογικών Εφαρμογών
Τμήμα Μηχανικών Πληροφορικής**

Πτυχιακή εργασία

**Εξαγωγή Σημασιολογικών και Μορφολογικών
Χαρακτηριστικών σε Μουσικά Αρχεία**

Καλαϊτζάκη Ευαγγελία (ΑΜ:2202)

**Επιβλέπων καθηγητής : Δρ. Αθανάσιος Μαλάμος
Αναπληρωτής Καθηγητής**

Επιτροπή Αξιολόγησης :

**Μαλάμος
Αθανάσιος**

**Παναγιωτάκης
Σπυρίδων**

**Καπετανάκης
Κωνσταντίνος**

.....

.....

.....

Μαρτιος, 2014

Ευχαριστίες

Με την ολοκλήρωση της πτυχιακής μου εργασίας, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή κ. Μαλάμο Αθανάσιο για την ευκαιρία που μου έδωσε να ασχοληθώ με το θέμα της πτυχιακής εργασίας που του πρότεινα και επέλεξα προσωπικά.

Επίσης, θα ήθελα εγκάρδια να ευχαριστήσω τον Ζάμπογλου Μάρκο, για τις χρήσιμες συμβουλές του και την πολύτιμη καθοδήγηση του καθ' όλη τη διάρκεια της εκπόνησης αυτής της πτυχιακής εργασίας καθώς και για την βοήθειά του στην αντιμετώπιση των προβλημάτων που συνάντησα.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου για την υπομονή και την υποστήριξη που μου έδειξαν καθώς επίσης και τους αγαπημένους μου φίλους Γιώργο Α., Τζούλια Ζ. και Ελένη Β. για την συμπαράστασή τους στις δύσκολες στιγμές που πέρασα στην διάρκεια των σπουδών μου.

Περίληψη

Με τις τρέχουσες τεχνολογίες αποθήκευσης και μετάδοσης αρχείων ψηφιακής μουσικής, έχει καταστεί πια εφικτή η συγκρότηση τεράστιων μουσικών συλλογών, όχι μόνο σε επίπεδο φορέων και επιχειρήσεων, αλλά και στο επίπεδο του απλού οικιακού χρήστη. Σ' ένα τέτοιο περιβάλλον, όπου η πρόσβαση στη μουσική είναι πλέον μια διαδικασία απλή, ταχεία και οικονομική, το ερευνητικό ενδιαφέρον στρέφεται στη διαχείριση της μουσικής πληροφορίας, και στην παροχή προηγμένων δυνατοτήτων αναζήτησης στο χρήστη. Ο κλάδος της Ανάκτησης Μουσικής Πληροφορίας (Music Information Retrieval) στοχεύει στην οργάνωση μουσικών συλλογών με βάση τα πραγματικά ηχητικά χαρακτηριστικά κάθε μουσικού κομματιού/τραγουδιού (όπως ο ρυθμός ή το ηχόχρωμα), και όχι την πληροφορία κειμένου (όπως το όνομα του καλλιτέχνη).

Θεμελιώδες βήμα για κάθε εγχείρημα Ανάκτησης Μουσικής Πληροφορίας είναι η εξαγωγή περιγραφικών χαρακτηριστικών από τα μουσικά αρχεία μιας συλλογής. Ο γενικός όρος αυτός αναφέρεται σε ένα διάνυσμα τιμών, που αποσκοπεί στην αναπαράσταση κάποιας πτυχής της μουσικής πληροφορίας. Μπορούμε έτσι, για παράδειγμα, να σχεδιάσουμε και να εξάγουμε περιγραφείς για τη μελωδία ενός μουσικού κομματιού, τα χαρακτηριστικά της ενορχήστρωσής του, το ρυθμό του, ή το μέτρο του. Έχοντας την περιγραφική πληροφορία για όλα τα κομμάτια μιας συλλογής, μπορούμε στη συνέχεια να προχωρήσουμε σε ένα πλήθος εγχειρημάτων, από την αυτόματη ταξινόμηση βάσει μουσικού είδους έως την αυτόματη συγκρότηση λίστας αναπαραγωγής με βάση τα μουσικά κριτήρια ενός συγκεκριμένου χρήστη.

Στην παρούσα πτυχιακή εργασία θα ασχοληθούμε συγκεκριμένα με την αυτόματη εύρεση της ρυθμικής πληροφορίας ενός μουσικού κομματιού που βρίσκεται σε μορφή Mp3. Θα γίνει μελέτη του όρου «Beat Tracking» και θα ερευνήσουμε τον τρόπο με τον οποίο μπορούμε να εξάγουμε τέτοιου είδους πληροφορία από μουσικά αρχεία με την βοήθεια του δυναμικού προγραμματισμού. Στόχος μας, είναι να γίνει συγκριτική αξιολόγηση των αποτελεσμάτων που εξάγουμε για την ανίχνευση του ρυθμού, από διάφορες συναρτήσεις/μεθόδους και παραμετροποιήσεις που υπάρχουν στην βιβλιογραφία του συγκεκριμένου ερευνητικού πεδίου. Δημιουργούμε συνδυασμούς όλων αυτών των διαφορετικών αλγορίθμων και παραμετροποιήσεων πάνω σε μία μεγάλη βάση μουσικών κομματιών (Mp3) που υπάρχει διαθέσιμη στο σύνολο των αρχείων του Magnatagatune (σύνολο δεδομένων για την εξαγωγή μουσικών χαρακτηριστικών) και στην συνέχεια συγκρίνουμε τις τιμές των αποτελεσμάτων τους με τις ακριβείς τιμές που λαμβάνουμε από τα δεδομένα των αντίστοιχων XML αρχείων των μουσικών κομματιών αυτών, που επίσης υπάρχουν στα αρχεία του Magnatagatune.

Η υλοποίηση του πειραματικού μέρους θα γίνει στο προγραμματιστικό περιβάλλον του Matlab, χρησιμοποιώντας κυρίως έτοιμα Toolboxes που μας επιτρέπουν την επεξεργασία των μουσικών χαρακτηριστικών των ψηφιακών σημάτων.

Λέξεις –Κλειδιά:

Ανάκτηση Μουσικής Πληροφορίας, Ανίχνευση Ρυθμού, Συνάρτηση Ανίχνευσης του Onset μιας νότας, MagnaTagATune.

Abstract

With the current storage and transmission technology of digital music files ,the constitution of large music collections is feasible now ,not only in the level of institutions and enterprises ,but also in the level of the simple home user. In such an environment ,where the access to the music is just a typical ,fast and economic procedure ,the research interest is turning around to the music information management and to the provision of advanced search capabilities to the user. The branch of Music Information retrieval aims to the organization of music collections regarding to the real acoustic characteristics of each music track/song (such as the rhythm or the timbre), and not the text information (like the artist's name).

The fundamental step for each trial of music information retrieval is the export of descriptive characteristics from the music files of a collection. This general term refers to one values vector that intends to represent some aspect of the music information .We can thus ,for example, design and export descriptions about the melody of a music track , the instrumentation characteristics ,the rhythm or the measure .Having the descriptive information about all the collection tracks ,we can continue subsequently to a large amount of projects from the automatic classification based on genre to the automatic composition of a playlist based on the musical criteria of a certain user.

In this thesis we are going to engage particularly with the automatic finding of rhythmic information of a music track in mp3 format. There will be a study of the term "Beat Tracking" and we are going to search the way we can automatically export such information from music files with the assistance of a dynamic programming.

Our target is the benchmarking of results obtained from various functions/methods (Onset Detection Functions) and parameterizations that there are in the bibliography of this certain research area for the detection of rhythm.

We create combinations between all these different algorithms and parameterizations on a big base of tracks for recording which are represented by audio-type mp3 files that are available to all Magnatagatune files (data set for musical features extraction). Then, we compare the result values with the accurate values obtained from the data of the corresponding .xml files of music tracks that there are also in Magnatagatune files.

The implementation of the experimental part will be held in the programming environment of Matlab, using mostly ready Toolboxes that allow us to process the musical characteristics of digital signals.

Key Words:

Music Information Retrieval, Beat Tracking, Onset Detection Function, MagnaTagATune.

Περιεχόμενα

Ευχαριστίες.....	3
Περίληψη.....	4
Abstract.....	5
Πίνακας Περιεχομένων.....	6
Κατάλογος Σχημάτων.....	8
1 Εισαγωγή	9
1.1. Διάρθρωση Εργασίας	10
1.2. MIR (Music Information Retrieval)	11
1.3. Εφαρμογές MIR (MIR Research Tasks)	12
1.4. Πληροφορίες Μουσικού Περιεχομένου	15
1.5 Ρυθμός (Rhythm).....	24
2 Αυτόματη Εύρεση Ρυθμικής Πληροφορίας	26
2.1 Beat Tracking	26
2.2 Στάδια /Μεθοδολογία Υλοποίησης Beat Tracking.....	29
2.2.1 Περιγραφή Ασυμπέιστου Αρχείου	30
2.2.2 Περιγραφή Spectrogram.....	32
2.2.2.1 Φασματογράφημα του σήματος	32
2.2.3 Περιγραφή Onset Detection Function	34
2.2.3.1. Σχόλια πάνω στις μεθόδους ανίχνευσης των onsets.....	Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.
2.2.4 Περιγραφή Correlation	50
2.2.5 Περιγραφή Peak-Picking.....	53
3 Πειραματικό Μέρος-Υλοποίηση	53
3.1. Ανάπτυξη Λογισμικού-Υλοποίηση Αλγορίθμων	53
3.2. Βάση/Πειραματικά Δεδομένα	54
3.3. Αλγόριθμοι	56
4 Αποτελέσματα	63

5	Συμπεράσματα-Κατακλείδα	75
5.2	Δεν υλοποιήθηκαν	78
5.3	Μελλοντικές Επεκτάσεις	79
6	Βιβλιογραφία.....	80
7.	Λεξιλόγιο-Συνοτμεύσεις	Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.

Κατάλογος Σχημάτων

1.1 Περιοχές Έρευνας του MIR	σελ. 13
1.2 Περιοχές Έρευνας του MIREX, τα έτη 2005,2009,2014	σελ. 14
1.3 Κατηγοριοποίηση στόχων MIR με βάση τον ρυθμό εξειδίκευσής τους	σελ. 15
1.4 Πολλαπλάσια και Υποπολλαπλάσια του τετάρτου(60 sec) σελ24.....	σελ. 24
2.1 Στάδια Υλοποίησης Beat Tracking	σελ. 29
2.2 Διάγραμμα Ροής Αλγορίθμου Εύρεσης Onset	σελ. 30
2.3 Spectrogram-Masri(dB(1)-mel(1)-nmel(40)).....	σελ. 33
2.4 Spectrogram-Masri(dB(0)-mel(0)-nmel(0)).....	σελ. 33
2.5 Onset, Attack, Transient και Delay μιας νότας	σελ. 34
2.6 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (CD).....	σελ. 36
2.7 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (NWPD).....	σελ. 36
2.8 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (PD).....	σελ. 37
2.9 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (RCD).....	σελ. 37
2.10 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (Duxbury).....	σελ.38
2.11 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (E).....	σελ.38
2.12 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (HFC).....	σελ.39
2.13 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (Kullback).....	σελ.39
2.14 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (Masri).....	σελ.40
2.15 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (MoT).....	σελ.40
2.16 OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0 (WPD).....	σελ.41
2.17 OnsetCorrscollection-1-1-40-0-0-xcorr-Ellis-Ellis-Ellis-0 (Duxbury).....	σελ.41
2.18 OnsetCorrscollection-1-1-40-0-0-xcorr-Ellis-Ellis-Ellis-0 (E).....	σελ.42
2.19 OnsetCorrscollection-1-1-40-0-0-xcorr-Ellis-Ellis-Ellis-0 (HFC).....	σελ.42
2.20 CorrelationsCorrscollection-1-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0-(Masri)	σελ. 52
2.21 CorrelationsCorrscollection-1-1-40-0-0-xcorr-Ellis-Ellis-Ellis-0-(Duxbury)...	σελ. 52

1 Εισαγωγή

Με την ολοένα αυξανόμενη επέκταση του διαδικτύου, ο όγκος των ελεύθερων προς τους χρήστες μουσικών αρχείων, συνεχώς μεγεθύνεται. Ο τρόπος με τον οποίο οι άνθρωποι αλληλεπιδρούν πάνω στην μουσική, εντοπίζουν και επεξεργάζονται την μουσική, αλλάζει συνεχώς καθώς επεκτείνεται και ο χώρος των διαθέσιμων εφαρμογών και ψηφιακών συσκευών (π.χ: mp3 players, tablets, smart phones κ.ά) που παρέχουν την δυνατότητα ακρόασης μουσικών κομματιών έως και την επεξεργασία αυτών (π.χ: αποκοπή μέρους ενός μουσικού κομματιού, συνένωση 2 ή περισσότερων μουσικών καταγραφών κλπ).

Αντιλαμβανόμαστε, λοιπόν, πως καθώς υπάρχει εξέλιξη της τεχνολογίας γενικότερα αλλά και πιο συγκεκριμένα στο πεδίο της επεξεργασίας του ήχου και της μουσικής, είναι λογικό επακόλουθο να παρατηρείται η ανάγκη για μετατροπή του αναλογικού σήματος μουσικών αρχείων σε ψηφιακό σήμα, με στόχο την επεξεργασία του, την μελέτη και την εξαγωγή μορφολογικών χαρακτηριστικών του μουσικού κομματιού.

Στην συγκεκριμένη πτυχιακή εργασία θα ασχοληθούμε με την **αυτόματη εύρεση της ρυθμικής πληροφορίας** πάνω σε μία μεγάλη βάση μουσικών κομματιών που η αρχική τους μορφή είναι τύπου .mp3. Θα χρησιμοποιήσουμε 11 διαφορετικές, ήδη υπάρχουσες συναρτήσεις ανίχνευσης έναρξης μιας νότας (**Onset Detection Functions**) και σε συνδυασμό με αλλαγές διαφορετικών τιμών και εκτελέσεων ορισμένων παραμέτρων (π.χ: τιμή dB=0 (απλό spectrogram) ή dB=1 (spectrogram εκφρασμένο σε λογαριθμικούς άξονες.)) που λειτουργούν πάνω στα διάφορα στάδια του αλγορίθμου για την ανίχνευση του Tempo. Από τα αποτελέσματα όλων αυτών των δυνατών συνδυασμών που έχουμε δημιουργήσει, λαμβάνουμε μία τιμή, το **BeatPeriod**, η οποία εκφράζει το χρονικό διάστημα που χρειάζεται για να δημιουργηθεί ένας χτύπος, που σε επαναλαμβανόμενη αναπαραγωγή του χτύπου αυτού, δημιουργείται ο ρυθμός του μουσικού κομματιού. Στη συνέχεια, συγκρίνουμε την τιμή αυτή, με την ακριβή τιμή του ρυθμού του κάθε μουσικού κομματιού που παίρνουμε από τα xml αρχεία που υπάρχουν στο σύνολο των αρχείων της Magnatagatune. Δημιουργούμε με τον τρόπο αυτό μια νέα τιμή που εκφράζει τον δείκτη απόκλισης (ή βέλτιστη απόκλιση) του Beat Period με την ακριβή τιμή που έχουμε εξάγει από τα xml αρχεία. Όσο μικρότερη είναι η τιμή του δείκτη απόκλισης που προκύπτει (κοντά στο μηδέν), τόσο πιο επιτυχής ήταν η ανίχνευση της χρονικής στιγμής που εκφράζει τον ρυθμό.

1.1. Διάρθρωση Εργασίας

Στο **1^ο κεφάλαιο** γίνεται μία εισαγωγική αναφορά στο θέμα και στον στόχο της παρούσας πτυχιακής εργασίας. Στην συνέχεια παρουσιάζουμε το ερευνητικό πεδίο MIR. Αναφέρουμε την εξέλιξη της Ανάκτησης Μουσικής Πληροφορίας (Music Information Retrieval-MIR) καθώς επίσης και τις εφαρμογές του πεδίου αυτού. Τέλος, περιγράφονται οι πληροφορίες και τα χαρακτηριστικά των Μουσικών Περιεχομένων.

Στο **2^ο κεφάλαιο**, ασχολούμαστε με τις μεθόδους εύρεσης ρυθμικής πληροφορίας και αναλύουμε τους τρόπους/στάδια με τα οποία μπορούμε να καταφέρουμε με δυναμικό τρόπο, να εξάγουμε από μουσικά κομμάτια το χαρακτηριστικό στοιχείο του ρυθμού. Συγκεκριμένα, ασχολούμαστε με θέματα : εκτίμηση του tempo , Beat Tracking , και Onset Detection Functions.

Στο **3^ο κεφάλαιο**, παρουσιάζεται όλο το πειραματικό κομμάτι της συγκεκριμένης πτυχιακής εργασίας και ο τρόπος υλοποίησης του.

Στο **4^ο κεφάλαιο**, βρίσκονται οι πίνακες των αποτελεσμάτων από τους αλγορίθμους που δουλέψαμε όλων των συνδυασμών των Onset Det. Function και των παραμέτρων τους.

1.2. MIR (Music Information Retrieval)

Οι προσπάθειες για ψηφιοποίηση της μουσικής που βρίσκεται ηχογραφημένη σε κάποιο αναλογικό μέσο, παρατηρούνται κυρίως από δισκογραφικές εταιρίες, από μουσικούς/μουσικόφιλους ή από πολιτιστικούς οργανισμούς. Επίσης, με την ανάπτυξη των διαδικτυακών εφαρμογών ψηφιακής σύνθεσης ή μίξης των μουσικών καταγραφών, σε συσχέτιση με την εκτεταμένη χρήση των σελίδων κοινωνικής δικτύωσης, έχει δώσει την δυνατότητα σε πολλούς καλλιτέχνες να αναπτύξουν και να κοινοποιήσουν δημόσια τα μουσικά έργα τους, δίχως να έρχονται αντιμέτωποι με τις οικονομικές και τις διαδικαστικές δυσκολίες που σίγουρα θα αντιμετώπιζαν στο παρελθόν.

Όπως πολύ εύκολα μπορεί να αντιληφθεί κάποιος, οι σημερινοί χρήστες διαδικτύου έχουν την δυνατότητα προσπέλασης σε εξαιρετικά μεγάλου μεγέθους όγκο μουσικής πληροφορίας, ο οποίος αναπτύσσεται συνεχώς με ραγδαίους ρυθμούς. Παλαιότερα, το διαθέσιμο μουσικό υλικό ήταν αρκετά περιορισμένο προς τους χρήστες για την δημιουργία προσωπικής συλλογής, που αποτελούταν συνήθως από εκατοντάδες μουσικά κομμάτια. Στις μέρες μας όμως, οι υπηρεσίες/εφαρμογές που υπάρχουν σε απευθείας σύνδεση (online) με τον χρήστη όπως είναι το YouTube, το iTunes, το 8Tracks και άλλες πολλές, δίνουν την δυνατότητα της άμεσης πρόσβασης σε έναν τεράστιο όγκο μουσικών συλλογών με το πάτημα ενός κουμπιού.

Ενώ στο παρελθόν ο μόνος τρόπος για να αποκτήσει ένας άνθρωπος νέα ακούσματα πάνω στην μουσική περιοριζόταν στην αγορά δίσκων μουσικών αρχείων από δισκοπωλεία, στην ακρόαση μουσικών κομματιών από ραδιοφωνικές εκπομπές, στην αναστροφή με άλλα άτομα με πλούσιο υλικό και γνώσεις μουσικών συλλογών, σήμερα με την επέκταση της χρήσης του διαδικτύου και των διαφορετικών ειδών σελίδων κοινωνικής δικτύωσης, αυτή η διαδικασία μπορεί να πραγματοποιηθεί σε χρονικό διάστημα μόλις λίγων λεπτών, με οποιαδήποτε θετική ή αρνητική συνέπεια μπορεί να εμφανίσει η ευκολία αυτή.

Ο υπέρογκος αυτός χώρος μουσικής πληροφορίας που έχει αναπτυχθεί, με τον οποίο έρχεται σε επαφή ένας χρήστης διαδικτύου σχεδόν καθημερινά, είναι αρκετά δύσκολο να τον διαχειριστεί και να τον ερευνήσει κάποιος χωρίς την βοήθεια κάποιων αυτόματων μηχανισμών. Είναι πολύ εύκολο να οδηγηθεί ο χρήστης σε άσκοπη υπερπληροφόρηση κατά την διάρκεια της αναζήτησης συγκεκριμένων μουσικών πεδίων από την στιγμή που δεν υπάρχουν συγκεκριμένα φίλτρα αναζήτησης. Η μεγάλη σε όγκο συλλογή μουσικών επιλογών που διατίθεται σε έναν χρήστη ως προς την ακρόαση αυτών των επιλογών, έχει φέρει ως αποτέλεσμα το φαινόμενο «Τυραννία Επιλογής» ('Tyrranny of Choice') κατά την οποία ο χρήστης δυσκολεύεται πάρα πολύ να αποφασίσει τι να επιλέξει.

Συνεπώς, για την λύση αυτού του προβλήματος ήταν αναγκαία η ανάπτυξη στρατηγικών και εφαρμογών, οι οποίες θα διευκολύνουν την πρόσβαση και την επιλογή μουσικών συλλογών, τόσο σε νέες όσο και σε παλαιότερες συλλογές μουσικού περιεχομένου. Ο κλάδος που παρουσιάστηκε για την αντιμετώπιση του θέματος αυτού ονομάζεται **Music Information Retrieval (Ανάκτηση Μουσικών Πληροφοριών)**.

«Το MIR αποτελεί μια αναπτυσσόμενη περιοχή έρευνας αφιερωμένη κυρίως στο να

καλύψει τις ανάγκες των χρηστών για πληροφόρηση και αλληλεπίδραση με τη μουσική πληροφορία.» [7]

Οι κυριότερες ομάδες χρηστών που επωφελούνται από τα ερευνητικά πεδία του MIR είναι οι τρεις παρακάτω:

- **Επιχειρήσεις**, οι οποίες δραστηριοποιούνται στον τομέα των ηχογραφήσεων, συγκέντρωσης αλλά και διάδοσης μουσικών κομματιών.
- **Τελικοί Χρήστες**, που έχουν ως στόχο να ανακαλύψουν νέα ακούσματα στα διαφορετικά είδη μουσικής που υπάρχουν.
- **Επαγγελματίες**, όπως είναι οι μουσικοί/μουσικολόγοι, οι μουσικοί παραγωγοί, οι προγραμματιστές (για την επεξεργασία του ήχου από ψηφιακά σήματα) καθώς και οι δικηγόροι (ασχολούνται με τον τομέα των δισκογραφικών δικαιωμάτων).

1.3. Εφαρμογές MIR (MIR Research Tasks)

Περιοχές Έρευνας του MIR

Το MIR ασχολείται με την ανάλυση, την εξαγωγή και την χρήση των πληροφοριών που σχετίζονται από μία μουσική οντότητα οποιουδήποτε είδους (καλλιτέχνες, είδος μουσικού κομματιού κτλ).

Παρακάτω, ακολουθεί η παρουσίαση των περιοχών έρευνας και εφαρμογών του MIR.

Εξαγωγή Μουσικών Χαρακτηριστικών (Feature Extraction)
Ακουστική Ομοιότητα (Acoustic Similarity)
Δομική Ανάλυση (Structural Analysis)
Ανάλυση Μουσικής Τονικότητας (Music Tonality Analysis)
Audio Fingerprinting
Μουσική Αναγνώριση (Music Identification)
Αναζήτηση και Ανακάλυψη Μουσικής (Music Search and Discovery)
Αυτόματη Σήμανση (Automatic Tagging)

Παρακολούθηση Δικαιωμάτων (Copyright Monitoring)
Ταξινόμηση κατά είδος (Genre Classification)
Αναγνώριση Καλλιτέχνη (Artist Recognition)
Αναγνώριση Μουσικού Οργάνου (Instrument Identification)

Εικόνα 1.1 :Περιοχές Έρευνας του MIR.

Σε αυτό το σημείο, αξίζει να αναφέρουμε ότι καθ' όλη την διάρκεια της εξέλιξης των ερευνητικών εργασιών στον τομέα του MIR, τα πεδία στα οποία επικεντρώνονται οι έρευνες γίνονται όλο και πιο εξειδικευμένα.

Το **MIREX** (Music Information Retrieval Evaluation Exchange), αποτελεί μία εκδήλωση που οργανώνεται κάθε χρόνο (από το 2005), με σκοπό να εισαχθούν νέα πεδία έρευνας τα οποία προτείνονται από τους ερευνητές που ασχολούνται με την επεξεργασία μουσικών περιεχομένων, καθώς επίσης και να εξετασθεί η αποτελεσματικότητα των ήδη υπαρχόντων MIR αλγορίθμων που αναπτύχθηκαν κατά το προηγούμενο έτος, δημιουργώντας με αυτόν τον τρόπο το έναυσμα στους ερευνητές του κλάδου, προκειμένου να βελτιώσουν ως προς την ακρίβεια τους υπάρχοντες αλγορίθμους.

Για να αναδείξουμε τον ρυθμό και τον τρόπο με τον οποίο εξελίσσονται τα πεδία των ερευνών του MIREX, θα παρατηρήσουμε στον παρακάτω πίνακα τις περιοχές έρευνας που προτάθηκαν από το MIREX το 2005 (την 1^η χρονιά που δημιουργήθηκε η εκδήλωση), το 2009 και το 2014.

2005	2009	2014
Audio Artist Identification	<u>Audio Onset detection</u>	<u>Audio Classification (Train/Test) Tasks</u>
Audio Drum Detection	<u>Audio Artist Identification</u>	<u>Audio K-POP Genre Classification</u>
Audio Genre Classification	<u>Audio Genre Classification</u>	<u>Audio K-POP Mood Classification</u>
Audio Melody Extraction	<u>Audio Tag Classification</u>	<u>Audio Cover Song Identification</u>
Audio Onset Detection	<u>Audio Music Mood Classification</u>	<u>Audio Tag Classification</u>
Audio Tempo Extraction	<u>Audio Cover Song Identification</u>	<u>Audio Music Similarity and Retrieval</u>
Audio and Symbolic Key Finding	<u>Real-time Audio to Score Alignment (a.k.a Score Following)</u>	<u>Symbolic Melodic Similarity</u>
Symbolic Genre Classification	<u>Query by Singing/Humming</u>	<u>Audio Onset Detection</u>
	<u>Multiple Fundamental Frequency Estimation & Tracking</u>	<u>Audio Key Detection</u>

	<u>Audio Chord Detection</u>	<u>Real-time Audio to Score Alignment (a.k.a Score Following)</u>
	<u>Audio Melody Extraction</u>	<u>Query by Singing/Humming</u>
	<u>Query by Tapping</u>	<u>Audio Melody Extraction</u>
	<u>Audio Beat Tracking</u>	<u>Multiple Fundamental Frequency Estimation & Tracking</u>
		<u>Audio Chord Estimation</u>
		<u>Query by Tapping</u>
		<u>Audio Beat Tracking</u>
		<u>Structural Segmentation</u>
		<u>Audio Tempo Estimation</u>
		<u>Discovery of Repeated Themes & Sections</u>

Εικόνα 1.2: Περιοχές Έρευνας του MIREX τα έτη 2005,2009 και 2014.

Μπορούμε να διαχωρίσουμε τα MIR συστήματα σε κατηγορίες, με βάση τον στόχο με τον οποίο καλούνται να πετύχουν και ο διαχωρισμός αυτός αφορά τον βαθμό εξειδίκευσης τους (specificity).

- **Συστήματα Υψηλής Εξειδίκευσης**

Θεωρούνται τα συστήματα τα οποία καλούνται να αναγνωρίσουν κάποιο μεμονωμένο μουσικό κομμάτι από μία μεγάλη βάση δεδομένων/συλλογών, παίρνοντας ως δεδομένα (inputs) ένα μικρού μεγέθους μουσικό απόσπασμα το οποίο στη συνέχεια σχετίζεται με κάποια από τις εφαρμογές της μουσικής αναγνώρισης (Audio-FingerPrinting) και διαχείρισης των μουσικών δικαιωμάτων.

- **Συστήματα Χαμηλής Εξειδίκευσης**

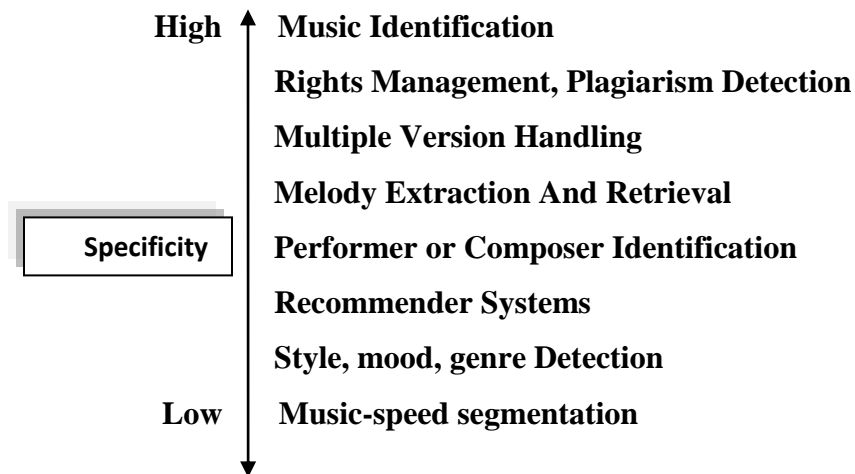
Είναι τα συστήματα τα οποία βασίζονται σε πληροφορίες εννοιολογικού περιεχομένου, όπως είναι η ταξινόμηση ανά είδος του μουσικού κομματιού ή η ταξινόμηση ανά ύφος.

- **Ενδιάμεσες Θέσεις εξειδίκευσης**

Στις ενδιάμεσες θέσεις εξειδίκευσης, θεωρούνται ότι ανήκουν όλοι οι υπόλοιποι στόχοι με βάση το κατά πόσο μπορεί να είναι αυστηρή η οριοθέτηση μεταξύ του σωστού ή του λάθους αποτελέσματος για την ανάκτηση του περιεχομένου της μουσικής.

Αξίζει να σημειωθεί, ότι ο συγκεκριμένος διαχωρισμός αφορά κυρίως τις εφαρμογές που

διαχειρίζονται και επεξεργάζονται πληροφορίες που προέρχονται από το ψηφιακό ηχητικό σήμα. Στην σχήμα που ακολουθεί, απεικονίζεται η κλίματα της εξειδίκευσης των MIR πεδίων που αναφέραμε παραπάνω.



Εικόνα 1.3: Κατηγοριοποίηση στόχων MIR με βάση τον βαθμό εξειδίκευσής τους.

1.4. Πληροφορίες Μουσικού Περιεχομένου

Τα χαρακτηριστικά τα οποία αποτελούν τις πληροφορίες μουσικών περιεχομένων, εξάγονται απευθείας από τον ηχητικό μουσικό ψηφιακό σήμα. Η συγκεκριμένη αυτή διαδικασία εξαγωγής των χαρακτηριστικών, αποτελεί μια απλή και κοινή ενέργεια για τον άνθρωπο. Η υπολογιστική ικανότητα του ανθρώπινου εγκεφάλου είναι τεράστια και λειτουργεί σε τέτοιο βαθμό ώστε να μπορεί να αξιοποιήσει αρκετά ικανοποιητικά την σημασιολογικής μορφής πληροφορία για την αναζήτηση διαφορών αλλά και ομοιοτήτων μεταξύ δύο ή περισσότερων ήχων καθώς επίσης και για την αναζήτηση ομαδοποίησης των ήχων αυτών. Αντίθετα, η ίδια διαδικασία γίνεται με πολύ πιο δύσκολο τρόπο, όταν υλοποιείται από τα αυτοματοποιημένα υπολογιστικά συστήματα που βασίζονται στα χαρακτηριστικά περιεχομένου που εξάγονται, εφόσον τα χαρακτηριστικά αυτά διαθέτουν ελάχιστη έως και καθόλου σημασιολογική σημασία. Το

πρόβλημα που προκύπτει στον τομέα της ανάκτησης μουσικών πληροφοριών, που βασίζεται στο περιεχόμενο, είναι ένα πρόβλημα του οποίου ο ορισμός δεν έχει ξεκαθαριστεί, εφόσον δημιουργεί ένα σημασιολογικό κενό μεταξύ των εννοιών υψηλού επιπέδου και σε περιγραφές χαμηλού επιπέδου, ανάμεσα δηλαδή στο ηχητικό σήμα και στα σημασιολογικά στοιχεία των περιεχομένων του.

Παραδείγματος χάριν, μια ηχογράφηση ενός μουσικού κομματιού, στον υπολογιστή μεταφράζεται ως μια σειρά αριθμών που αποτελούν τις τιμές των samples. Σε υψηλότερο σημασιολογικό επίπεδο, το ίδιο μουσικό κομμάτι εκφράζεται από νότες συγκεκριμένων χρονικών διαστημάτων (διάρκεια νότας). Για τον άνθρωπο, είναι αρκετά εύκολο να αντιληφθεί τις υψηλούς βαθμούς σημασιολογικές έννοιες ως μουσικές οντότητες (θέματα, περίοδος, μοτίβα κλπ) και ως ακούσματα που παράγουν κάποιο συγκεκριμένο συναίσθημα (εφορία, έκσταση, συγκίνηση κλπ).

Οι πληροφορίες μουσικού περιεχομένου χωρίζονται σε δύο κατηγορίες:

- **Υψηλού Επιπέδου Μουσικά Χαρακτηριστικά**
(High Level Music Features)

- Ρυθμός
- Κλίμακα
- Μελωδία
- Ηχόχρωμα κ.ά

- **Χαμηλού Επιπέδου Μουσικά Χαρακτηριστικά**
(Low Level Music Features)

- Φάσμα Μέτρου Βραχέως Χρόνου (Short-Time Magnitude Spectrum)
- Φάσμα MEL, το Χρωμάγραμμα κ.ά.

Στα χαμηλού επιπέδου μουσικά χαρακτηριστικά, η πληροφορία εξάγεται απευθείας από το ηχητικό σήμα και τις περισσότερες φορές δεν εμπεριέχουν κάποια σημασιολογική ερμηνεία που να μπορεί να εκλάβει ο τελικός χρήστης. Χρησιμοποιούνται κυρίως για να περιγράψουν το σήμα ώστε να συνθέσουν χαρακτηριστικά υψηλού επιπέδου.

Παρακάτω, παρουσιάζεται τμήμα από τα χαρακτηριστικά μουσικού περιεχομένου που χρησιμοποιούνται συχνά στο MIR.

- **Χρονικά χαρακτηριστικά** (Temporal Features)

Ο χρόνος αποτελεί το ενδογενές πεδίο ορισμού ενός ηχητικού σήματος. Όλα τα χαρακτηριστικά που εκφράζουν τον χρόνο, έχουν ως κοινό σημείο ότι εξάγονται απευθείας από το ηχητικό σήμα δίχως να είναι απαραίτητη η ενέργεια για μετασχηματισμό αυτού.

➤ **Χαρακτηριστικά Βασισμένα σε Μηδενισμούς (Zero Crossing-Based Features)**

• **Ρυθμός Μηδενισμού (Zero Crossing Rate- ZCR)**

Ο ρυθμός Μηδενισμού είναι ο αριθμός των μηδενισμών που δημιουργούνται στο πεδίο του χρόνου, σε χρονικό διάστημα ενός δευτερολέπτου και αποτελεί το μέτρο της κύριας συχνότητας του ηχητικού σήματος.

➤ **Χαρακτηριστικά βασισμένα σε Ενέργεια (Power-Based Features)**

- Η **Ενέργεια** του σήματος, ορίζεται ως το τετράγωνο του πλάτους μιας κυματομορφής. Είναι η ενέργεια, που μεταφέρεται σε χρονικά διαστήματα του ενός δευτερολέπτου και επομένως αποτελεί το μέσο τετράγωνο του σήματος.

• **Ενέργεια Βραχέως Χρόνου (Short-time Energy – STE or Root-Mean Square Energy - RMS).**

Η **Ενέργεια Βραχέως Χρόνου** περιγράφει την περιβάλλουσα ενός ηχητικού σήματος και χρησιμοποιείται πολύ συχνά από πολλούς τομείς του MIR. Εκφράζει την μέση Ενέργεια ανά πλαίσιο.

• **Χρονικό Κέντρο Βάρους (Temporal Centroid)**

Χαρακτηρίζει την μέση τιμή της περιβάλλουσας στον χρόνο του σήματος. Είναι το σημείο στο χρόνο που εντοπίζεται η μεγαλύτερη Ενέργεια του σήματος.

• **Λογαριθμικός Χρόνος Attack (Log Attack Time - LAT)**

Εκφράζει το attack ενός ήχου. Είναι ο λογάριθμος που περιγράφει τον χρόνο από το ξεκίνημα του ηχητικού σήματος έως και το χρονικό σημείο εκείνο στο οποίο το πλάτος παίρνει τη μέγιστη τιμή. Χρησιμοποιείται κυρίως για την κατάταξη μουσικών οργάνων.

➤ **Χαρακτηριστικά Φυσικής Συχνότητας(Physical Frequency Features):**

Τα χαρακτηριστικά Φυσικής Συχνότητας ορίζονται στο πεδίο των συχνοτήτων ή της αυτοσυσχέτισης.

Υπάρχουν πολλοί τρόποι ώστε να φθάσουμε στα πεδιά αυτά. Ο πιο συνηθισμένος τρόπος, είναι η εφαρμογή του μετασχηματισμού Fourier.

Άλλοι, εξίσου συνηθισμένοι τρόποι είναι οι εξής:

- Μετασχηματισμός Συνημιτόνου
- Μετασχηματισμός Κυματιδίων
- Μετασχηματισμός Constant-Q

- **Χαρακτηριστικά Βασισμένα στον Μετασχηματισμό Fourier Βραχέως Χρόνου (Short – Time Fourier Transform-Based Features):**

Ο μετασχηματισμός STFT καταλήγει σε πραγματικές και μιγαδικές τιμές. Οι πραγματικές τιμές αναλογούν στην κατανομή των συνιστωσών των συχνοτήτων (φασματική περιβάλλουσα), ενώ οι μιγαδικές μεταφέρουν πληροφορία της φάσης των συνιστωσών.

- **Φασματική Ροή (Spectral Flux - SF)**

Εκφράζει μια L2 – νόρμα διανύσματος της διαφοράς του φασματικού πλάτους μεταξύ κάποιων πλαίσιαων.

Ορίζεται ως η μέση διαφορά ανάμεσα σε δύο συνεχόμενα πλαίσια του μετασχηματισμού Fourier βραχέως χρόνου:

Χαμηλή Φασματική Ροή έχουν τα σταθερά σήματα ή εκείνα τα σήματα που εμφανίζουν αργή μεταβολή των φασματικών ιδιοτήτων τους. (πχ: ο θόρυβος)
Μπορούμε να την αντιληφθούμε ως την υποτυπώδη προσέγγιση της τραχύτητας ενός ήχου.

- **Φασματική Κλίση (Spectral Slope)**

Ως Φασματική Κλίση εννοούμε την προσέγγιση του φασματικού σχήματος δημιουργώντας γραμμική παλινδρόμηση. Εκφράζει την ελάττωση των φασματικών πλατών από τις τιμές των πιο χαμηλών συχνοτήτων προς τις υψηλότερες συχνοτήτες.

- Φασματικές Κορυφές (Spectral Peaks)
- Φασματική Εξάπλωση (Spectral Spread)
- Φασματική Ασυμμετρία (Spectral Skewness)
- Φασματική Κύρτωση (Spectral Kurtosis)

- **Χαρακτηριστικά Αντιληπτικών Συχνοτήτων(Perceptual Frequency Features)**

Στην συγκεκριμένη ενότητα, αναλύουμε τα χαρακτηριστικά εκείνα που έχουν σημασιολογικής φύσεως σημασία, στα πλαίσια των ανθρωπίνων ακουστικών

αντιλήψεων και κατηγοριοποιούνται με βάση την ακουστική ιδιότητα που περιγράφουν.

- **Φωτεινότητα (Brightness)**

Η Φωτεινότητα περιγράφει την φασματική κατανομή των συχνοτήτων και εκφράζει το κατά πόσο σε ένα ηχητικό σήμα κυριαρχούν υψηλές ή χαμηλές συχνότητες.

Ένας ήχος, είναι περισσότερος φωτεινός όταν κυριαρχεί το περιεχόμενο των υψηλών συχνοτήτων. Η Φωτεινότητα ενός ήχου, είναι στενά συνδεδεμένη με την οξύτητα που αντιλαμβανόμαστε στον ήχο αυτό με την απλή ακρόαση του.

- **Φασματικό Κεντροειδές (Spectral Centroid - SC)**

Το Φασματικό Κεντροειδές (SC), ορίζεται ως το κέντρο βάρους Φασματικού Πλάτους. Εμφανίζει το σημείο του φάσματος που είναι συγκεντρωμένη η περισσότερη Ενέργεια και αφορά την επικρατούσα συχνότητα του ηχητικού σήματος.

- **Οξύτητα (Sharpness)**

Η Οξύτητα, αποτελεί την διάσταση του ηχοχρώματος που επιδρά από την κύρια συχνότητα σημάτων της στενής ζώνης και αυξάνεται ανάλογα με την αύξηση της ισχύος των υψηλών συχνοτήτων του φάσματος.

Χρησιμοποιείται κυρίως σε εφαρμογές Μουσικής Ομοιότητας.

- **Φασματικό Κέντρο (Spectral Center)**

Είναι η συχνότητα εκείνη που η μισή ενέργεια του φάσματος υπάρχει κάτω από αυτήν την τιμή (κέντρο) και η άλλη μισή πάνω από αυτή. Εκφράζει την προσέγγιση της κατανομής Ενέργειας και συχνά χρησιμοποιείται σε εφαρμογές παρακολούθησης του ρυθμού.

- **Τονικότητα (Tonality):**

Η τονικότητα είναι το χαρακτηριστικό του ήχου που αποσυνδέει τους θορυβώδεις ήχους από τους τονικούς.

Οι θορυβώδεις ήχοι έχουν ένα φάσμα διάρκειας και επομένως η τονικότητα είναι ελάχιστη, αντίθετα με τους τονικούς ήχους που τα φάσματα τους είναι γραμμικά (line spectra) και έχουν υψηλή τονικότητα.

Η τονικότητα έχει να κάνει με την ένταση του τόνου (pitch strength), που περιγράφει την ένταση του αισθητού ήχου.

- **Φασματική Διασπορά (Spectral Dispersion)**

Εκφράζει το μέτρο της εξάπλωσης φάσματος γύρω από το φασματικό κέντρο.

- **Σημείο Φασματικού Roll off (Spectral Rolloff Point)**

Περιγράφεται ως η συχνότητα, στην οποία περιέχεται το 85% (ή 95%) της συνολικής ενέργειας. Η τιμή αυτής της συχνότητας αυξάνεται όταν αυξήσουμε το εύρος ζώνης του σήματος.

- **Φασματική Ομαλότητα (Spectral Flatness)**

Η Φασματική Ομαλότητα, υπολογίζει τον βαθμό στον οποίο οι συχνότητες του φάσματος είναι κατανεμημένες κανονικά. Χρησιμοποιείται πολύ σε συχνά σε εφαρμογές Audio FingerPrinting.

- **Παράγοντας Φασματικής Κορυφογραμμής (Spectral Crest Factor)**

Ο Παράγοντας Φασματικής Κορυφογραμμής αποτελεί το μέτρο της μη ομαλότητας του φάσματος. Χρησιμοποιείται για τη διαφοροποίηση των τονικών ήχων από τους θορύβους. Ορίζεται ως η αναλογία της μέγιστης και της μέσης φασματικής ενέργειας μιας υποζώνης. Χρησιμοποιείται συχνά σε εφαρμογές Audio FingerPrinting.

- **Ένταση (Loudness):**

Αποτελεί το χαρακτηριστικό της ακουστικής αίσθησης όπου οι ήχοι κατατάσσονται στη κλίμακα απαλοί ή δυνατοί.

- **Ενιαία Ένταση (Integral Loudness)**

- **Τόνος(Pitch):**

Αποτελεί την βασική διάσταση ενός ήχου σε συνδυασμό με την ένταση, το ηχόχρωμα και την διάρκεια. Ορίζεται ως το χαρακτηριστικό εκείνο το οποίο κατατάσσει τους ήχους στη κλίμακα χαμηλός/βαθύς ή υψηλός/οξύς.

- **Βασική Συχνότητα (Fundamental Frequency)**

Ορίζεται ως η χαμηλότερη συχνότητα της σειράς των αρμονικών και αποτελεί μια γενικότερη προσέγγιση του τόνου.

- **Ιστόγραμμα Τόνου(Pitch Histogram)**

Το ιστόγραμμα Τόνου περιγράφει το περιεχόμενο της τονικότητας ενός σήματος με αρκετά περιεκτικό τρόπο και χρησιμοποιείται αρκετά συχνά στην κατάταξη μουσικού είδους. Στη μουσική ανάλυση, ο τόνος αντιστοιχεί στις μουσικές νότες, και έτσι το pitch histogram παρέχει την αναπαράσταση της κατανομής των μουσικών νοτών στο μουσικό κομμάτι.

- **Χρώμα / Χροιά (Chroma):**

Σύμφωνα με τον Shepard (1964) η αντίληψη του μουσικού τόνου (νότας) μπορεί να προσδιοριστεί από δύο διαστάσεις:

- 1) Το ύψος του τόνου
- 2) Το χρώμα του τόνου.

Η κλίμακα/διάσταση του ύψους του τόνου διαιρείται από τις μουσικές οκτάβες.

Οι μουσικές νότες της ίδιας τονικής κατηγορίας (οκτάβας) έχουν το ίδιο χρώμα και δημιουργούν παρόμοια ηχητική αίσθηση.

- **Χρωματογράφημα (Chromagram)**

Το χρωματογράφημα ουσιαστικά αποτελεί ένα φασματογράφημα (spectrogram) που εκφράζει την φασματική ενέργεια για καθεμία από τις 12 διαφορετικές τονικές κλάσεις. Βασίζεται στο λογαριθμικό φάσμα βραχέως χρόνου Fourier. Οι συχνότητες κβαντίζονται (αντιστοιχίζονται) στις 12 τονικές κλάσεις με την βοήθεια μιας συσσωρευτικής συνάρτησης. Ουσιαστικά αντιστοιχίζει όλες τις συχνότητες σε μία και μόνο οκτάβα. Αυτή η συσσώρευση, οδηγεί στην συμπύκνωση του φάσματος, επιτρέποντας με αυτόν τον τρόπο μια πιο συγκροτημένη περιγραφή των αρμονικών σημάτων.

- **Τονικό Κεντροειδές (Tonal Centroid)**

Από το χρωματογράφημα που έχουμε εξάγει υπολογίζεται το 6-διάστατο διάστημα τονικού κεντροειδούς, που αντιστοιχεί σε προβολή συγχορδίων στους κύκλους Μεγάλης Τρίτης, Μικρής Τρίτης και Πέμπτης. (διαστήματα μεταξύ των τόνων.)

- **Προφίλ Τονικότητας (Pitch Profile)**

Το Προφίλ Τονικότητας, αποτελεί μια ακριβή αναπαράσταση του περιεχομένου που αναφέρεται στο ύψος του τόνου ενώ λαμβάνει υπόψη και τον κακό συντονισμό του ύψους που εισάγεται από τα μη-συντονισμένα μουσικά όργανα και είναι αρκετά έντονο

απέναντι στους θορυβώδεις ήχους των κρουστών μουσικών οργάνων.

- **Αρμονικότητα (Harmonicity)**

Αποτελεί την ιδιότητα που διαφοροποιεί και κατηγοροποιεί τα περιοδικά σήματα (αρμονικοί ήχοι) από τα μη περιοδικά (θορυβώδεις ήχοι).

Οι αρμονικές, αποτελούν τις συχνότητες και είναι ακέραια πολλαπλάσια της θεμελιώδους συχνότητας. Συνεπώς στο αρμονικό φάσμα, δημιουργούνται κορυφές στη κύρια συχνότητα και στα ακέραια πολλαπλάσια της. Τα χαρακτηριστικά αρμονικότητας χρησιμοποιούνται κυρίως στην αναγνώριση των μουσικών οργάνων (π.χ. τα έγχορδα όργανα εμφανίζουν πιο αρμονική συγκρότηση από εκείνη που παρουσιάζουν τα κρουστά).

- **Αρμονικός Λόγος (Harmonic Ratio)**

Αρμονικό Λόγο αναφέρουμε τον λόγο της ενέργειας της θεμελιώδους συχνότητας ως προς την συνολική ενέργεια του πλαισίου.

- **Μέτρα Μη- Αρμονικότητας (Inharmonicity Measures)**

- **Περιγραφείς Φασματικού Ηχοχρώματος (Spectral Timbral Descriptors)**

Αναφέρεται στην αρμονική δομή του μουσικού ήχου και βασίζεται στην εκτίμηση της θεμελιώδους συχνότητας και την ανίχνευση των αρμονικών κορυφών στο φάσμα.

Στατιστικές ροπές (moments) αρμονικών συχνοτήτων και των πλάτους τους:

- Harmonic Spectral Centroid - HSC
- Harmonic Spectral Deviation - HSD
- Harmonic Spectral Spread – HSS
- Harmonic Spectral Variation – HSV

- **Cepstral Χαρακτηριστικά (Cepstral Features)**

Τα cepstral χαρακτηριστικά χρησιμοποιούνται κυρίως ως εξομαλυμένες αναπαραστάσεις του λογαριθμικού πλάτους ενός φάσματος και τονίζουν τα χαρακτηριστικά ηχοχρώματος και του τόνου.

Χαρακτηριστικά Βασισμένα σε Συστοιχίες Αντιληπτικών Φίλτρων (Perceptual Filter Bank – Based Features):

Το **cepstrum** ορίζεται ως ο Μετασχηματισμός Fourier του λογαρίθμου πλάτους φάσματος από το αρχικό σήμα.

Συντελεστές Cepstral Συχνότητας Μελ (Mel – Frequency Cepstral Coefficients - MFCCs)

Οι Συντελεστές Cepstral Συχνότητας Mel, εκφράζουν την πληροφορία του ηχοχρώματος (δηλαδή την φασματική περιβάλλουσα) ενός σήματος.

1.5 Ρυθμός (Rhythm)

Ο ρυθμός και το tempo απαρτίζουν χαρακτηριστικά του ήχου που σχετίζονται με την διάρκεια της διαμόρφωσή του. Τις ρυθμικές δομές μπορούμε να τις εξάγουμε αναλύοντας διαμορφώσεις των χαμηλών συχνοτήτων στο χρόνο.

Ο ρυθμός αποτελεί μία από τις χαρακτηριστικές ιδιότητες του ηχητικού σήματος που εκφράζει ένα μοτίβο αλλαγής του ηχοχρώματος αλλά και της αλλαγής της ενέργειας στον χρόνο. Βάσει των καταγραφών των Zwicker και Fastl, η ακουστική αίσθηση του ρυθμού, στηρίζεται στην χρονική μεταβολή της έντασης. Τα ρυθμικά μοτίβα εξάγονται αναλύοντας τα πλάτη διαμόρφωσης των χαμηλών συχνοτήτων.

Ο Ρυθμός διακρίνεται από τα πρότυπα των μουσικών μονάδων, τα οποία εμφανίζονται σε διαφορετικά ιεραρχικά επίπεδα ρυθμού. Το διάστημα/μήκος που διαρκεί μία νότα, εναπόκειται από την χρονικό διάστημα που παίζεται η νότα αυτή όπως επίσης και από το μουσικό tempo αλλά και τον μετρικό οπλισμό. Ως **beats**, αποκαλούμε τις βασικές ρυθμικές μονάδες και ο ρυθμός επανάληψης αυτών των beats, οδηγεί στην εξαγωγή του Tempo. Το Tempo, είναι αντιστρόφως ανάλογο της περιόδου των beats. Έστω ότι η περίοδος ενός beat είναι $T_b(\text{sec})$, το Tempo υπολογίζεται ως εξής: $T=60/ T_b(\text{sec})$. Αυτό σημαίνει πως αν ο χρόνος του beat ανά δευτερόλεπτο είναι 60, τότε το Tempo του μουσικού κομματιού είναι 1 beat/sec.

Στο παρακάτω σχήμα, βλέπουμε τις διάρκειες των νοτών. (Αν, τέταρτο=60,Πολλαπλάσια και υποπολλαπλάσια ενός τετάρτου και οι μεταξύ τους σχέσεις).

Βασικές Αξίες	Παρεσιγμένα	Τρίηχα	Πεντάηχα
$\circ = 15 (60 \div 4)$	$\circ. = 10 (30 \div 3)$	$^3\circ = 22,5 (7,5 \times 3)$	
$\text{♩} = 30 (60 \div 2)$	$\text{♩}. = 20 (60 \div 3)$	$^3\text{♩} = 45 (15 \times 3)$	$^5\text{♩} = 37,5 (7,5 \times 5)$
$\text{♪} = 60$	$\text{♪}. = 40 (120 \div 3)$	$^3\text{♪} = 90 (30 \times 3)$	$^5\text{♪} = 75 (15 \times 5)$
$\text{♫} = 120 (60 \times 2)$	$\text{♫}. = 80 (240 \div 3)$	$^3\text{♫} = 180 (60 \times 3)$	$^5\text{♫} = 150 (30 \times 5)$
$\text{♬} = 240 (60 \times 4)$	$\text{♬}. = 160 (480 \div 3)$	$^3\text{♬} = 360 (120 \times 3)$	$^5\text{♬} = 300 (60 \times 5)$
$\text{♭} = 480 (60 \times 8)$	$\text{♭}. = 320 (960 \div 3)$	$^3\text{♭} = 720 (240 \times 3)$	$^5\text{♭} = 600 (120 \times 5)$
$\text{♮} = 960 (60 \times 16)$	$\text{♮}. = 640 (1920 \div 3)$	$^3\text{♮} = 1440 (480 \times 3)$	$^5\text{♮} = 1200 (240 \times 5)$

Εικόνα 1.4 : Πολλαπλάσια και Υποπολλαπλάσια του τετάρτου (60 sec.).

Πηγή: <http://3euk114.blogspot.gr/2007/10/blog-post.html>

➤ Χαρακτηριστικά Ρυθμού

- Μέτρο (Meter)

Είναι ένα τυπικό μέγεθος που επινόησε ο άνθρωπος με σκοπό να μπορεί να συγκρίνει όμοια πράγματα, όπως είναι οι αποστάσεις (μήκος/μέτρα) αλλά και οι χρονικές διάρκειες (χρόνος/δευτερόλεπτα), μάζες (κιλό). Με την χρήση αυτού του όρου, δημιουργούμε κατατάξεις ανάμεσα στα μεγέθη και με τον τρόπο αυτό βοηθούμε τη μνήμη, την αντίληψη, την σύγκριση και την δημιουργικότητα. Στην μουσική, μέτρο ονομάζεται η οργάνωση των κύριων μετρικών μονάδων/ beats σε ίσα χρονικά διαστήματα, ομάδες-σηματισμούς/ patterns. Η οργάνωση των μετρικών μονάδων σε ζεύγη, τριάδες ή τετράδες είναι πλέον κοινή. Μέτρο, είναι η απόσταση ανάμεσα από δύο διαστολές.

- Διαστολή / Barline

Τα μέτρα διαχωρίζονται από κάθετες γραμμές, οι οποίες εκτείνονται από την πρώτη μέχρι την πέμπτη γραμμή του πενταγράμμου. Οι γραμμές αυτές αποτελούν τις διαστολές.

- Μετρική Παλμού (Pulse Metric)

Μέτρο της ρυθμικότητας του ήχου. Όταν οι συσχετίσεις στις υποζώνες εμφανίζουν κορυφές, η μετρική παλμού έχει υψηλή τιμή, δηλαδή υπάρχει έντονη ρυθμική δομή στον ήχο.

- Περιοδικότητα Ζώνης (Band Periodicity)

Η περιοδικότητα ζώνης υπολογίζει την ύπαρξη της ρυθμικής δομής στο σήμα.

- Φάσμα Beat (Beat Spectrum)

Το Φάσμα Beat, επισημαίνει την ομοιότητα ενός σήματος σε συνάρτηση με των διαφορετικών χρονικών καθυστερήσεων. Οι κορυφές στο φάσμα beat, προσκομίζουν δυνατούς παλμούς που έχουν ένα συγκεκριμένο ρυθμό επανάληψης.

Χρησιμοποιείται κυρίως στην onset detection function αλλά και στον προσδιορισμό των ομοιοτήτων στο ρυθμικό μέρος μιας μουσικής ή και στην κατάτμηση κομματιού σε διαφορετικά ρυθμικά κομμάτια που το αποτελούν.

2 Αυτόματη Εύρεση Ρυθμικής Πληροφορίας

Ένα από τα κυριότερα χαρακτηριστικά στοιχεία της μουσικής είναι ο ρυθμός και είναι αντιληπτός ακόμη και από άτομα που δεν έχουν καμία μουσική παιδεία. Η ανάλυση του ρυθμού, είναι ένα καθοριστικό στάδιο για την κατανόηση αλλά και την περιγραφή ενός μουσικού αποσπάσματος. Ένα μουσικό κομμάτι καθορίζεται από μία βασική ρυθμική δομή: τα μουσικά γεγονότα περιγράφονται σε συγκεκριμένες χρονικές στιγμές, έχουν ορισμένη χρονική διάρκεια, αποτελούν ομάδες στο χρόνο και επαναλαμβάνονται περιοδικά. Όπως αναφέραμε και στην ενότητα 1.5, τα beats αποτελούν μία ακολουθία από παλμούς που ισαπέχουν μεταξύ τους και που παρουσιάζονται περιοδικά στη μουσική. Είναι οι χρονικές στιγμές εκείνες, στις οποίες καθώς ακούμε ένα μουσικό κομμάτι θα χτυπούσαμε παράλληλα με την μουσικό το πόδι μας. Το αντίστροφο της περιόδου του beat, εκφράζεται ως Tempo. Συχνά το εκφράζουμε σε beats/ λεπτό (Beats Per Minute) αντί για συχνότητα σε Hz. Η εκτίμηση της ρυθμικής πληροφορίας εντοπίζεται στη χρονική κατάτμηση της μουσικής, σε συνάρτηση με κάποια κριτήρια και επιτυγχάνεται ανακαλύπτοντας και ενισχύοντας τις ενδογενείς περιοδικότητες ενός μουσικού κομματιού.

2.1 Beat Tracking

Με τον όρο Beat Tracking (Ανίχνευση του Ρυθμού ενός Μουσικού Κομματιού) εννοούμε την εξαγωγή μιας αλληλουχίας στιγμιοτύπων παλμών/κτύπων από ένα μουσικό ηχητικό σήμα, η οποία ανταποκρίνεται στο σωστό ρυθμικό χτύπημα του ποδιού ενός ανθρώπου-ακροατή στο ίδιο αυτό μουσικό ηχητικό σήμα. Περιλαμβάνει δύο περιορισμούς τους οποίους και θα πρέπει να ικανοποιεί:

Αφ' ενός, τα επιλεγμένα στιγμιότυπα των παλμών θα πρέπει γενικότερα να αντιστοιχούν στις στιγμές του μουσικού κομματιού που υποδηλώνεται ότι υπάρχει κάποιο beat, όπως για παράδειγμα είναι το διακριτό σημείο της έναρξης μιας νότας από ένα μεμονωμένο όργανο μιας ορχήστρας. Αφ' ετέρου, η συλλογή των beats θα πρέπει να αντανakλά σε ένα τοπικό και συνεχές διάστημα ανάμεσα από δύο διαδοχικά beats τα οποία σε συνδυασμό με τα beats θα προσδιορίζουν τον μουσικό ρυθμό. Οι συγκεκριμένοι διπλοί περιορισμοί καθορίζουν την αξιοποίηση τους μέσω της εφαρμογής δυναμικού προγραμματισμού με σκοπό την εύρεση μιας σειράς χρονικών στιγμών από beats, τα οποία θα έχουν ως αποτέλεσμα την εξαγωγή του tempo.

Περιγράφουμε ένα σύστημα ανίχνευσης των beats, το οποίο αρχικά εκτιμά ένα γενικό tempo, χρησιμοποιώντας το με σκοπό να κατασκευάσει μία συνάρτηση μετάβασης και στη συνέχεια με την βοήθεια του δυναμικού προγραμματισμού να υπολογίζει την καλύτερη δυνατή συλλογή από χρονικές στιγμές των beats που αντικατοπτρίζουν το tempo. Επίσης, αντιστοιχίζονται με στιγμές υψηλού 'onset strength' σε μία συνάρτηση ανίχνευσης της έναρξης που εξάγεται από το ηχητικό μας σήμα.

Αυτή η απλή και υπολογιστικά αποτελεσματική διαδικασία φαίνεται να αποδίδεται πολύ καλά στο **MIREX-06 Beat Tracking Training Data**, επιτυγχάνοντας μία μέση ακρίβεια των beats λίγο χαμηλότερα από το 60% των δεδομένων ανάπτυξης. Εξετάζουμε επίσης τον αντίκτυπο της υπόθεσης ενός σταθερού ρυθμού-στόχου και στην συνέχεια

δείχνουμε ότι το σύστημα είναι συνήθως σε θέση να ακολουθεί τις αλλαγές του tempo σε ένα εύρος $\pm 10\%$ του στοχευόμενου tempo.

2.1.1. Αλγόριθμος δυναμικού προγραμματισμού

Θεωρούμε ότι έχουμε ένα συγκεκριμένο σταθερό tempo το οποίο δίνεται εκ των προτέρων. Στόχος του beat tracker είναι να αναπαράγει μία σειρά από χρονικές στιγμές των beats που αντιστοιχούν στα αντιλαμβανόμενα onsets ενός ηχητικού σήματος ενώ την παράλληλα συγκροτούν από μόνα τους ένα τακτικό ρυθμικό μοτίβο. Μπορούμε να προσδιορίσουμε μία αντικειμενική απλή συνάρτηση που συνδυάζει και τους δύο αυτούς κατευθύνσεις:

$$C(\{t_i\}) = \sum_{i=1}^N O(t_i) + \alpha \sum_{i=2}^N F(t_i - t_{i-1}, \tau_p)$$

Όπου: $\{t_i\}$, είναι η ακολουθία των N beats που βρέθηκαν από την συνάρτηση ανίχνευσης των Onset (Onset Detection Function), και $O(t)$ είναι η συνάρτηση ανίχνευσης των onsets που δίνει υψηλές τιμές στις χρονικές στιγμές που θα συνιστούσαν καλές επιλογές για τα beats, με κριτήριο τις τοπικές ακουστικές ιδιότητες. Το α είναι εκείνο που ρυθμίζει την ισορροπία της σπουδαιότητας των δύο όρων και το $F(\Delta t, \tau_p)$ είναι συνάρτηση η οποία μετράει τη αλληλουχία μεταξύ του διαστήματος Δt ανάμεσα στα beats και της επιδιωκόμενης απόστασης τ_p που ορίζεται από το Tempo-target.

Παραδείγματος χάριν, χρησιμοποιούμε μία απλή μέθοδο τετραγωνικού λάθους η οποία εφαρμόζεται πάνω στον λογαριθμικό λόγο του αληθινού ως προς το ιδανικό διάστημα, δηλαδή:

$$F(\Delta t, \tau_p) = -(\log \frac{\Delta t}{\tau})^2$$

Έχει μέγιστη τιμή $\max=0$, όταν $\Delta t = \tau$, γίνεται αυξανόμενα αρνητική για μεγαλύτερες αποκλίσεις, και είναι συμμετρική σε έναν λογαριθμικό ως προς το χρόνο άξονα, ώστε $F(k\tau, \tau) = F(\tau/k, \tau)$.

Εν συνεχεία, υποθέτουμε ότι ο χρόνος είναι κβαντισμένος σε κάποιο άλλο κατάλληλο πλέγμα.

Το χαρακτηριστικό-κλειδί της αντικειμενικής συνάρτησης, είναι ότι η χρονική ακολουθία με το βέλτιστο score μπορεί να υπολογισθεί αναδρομικά. Για να συλλέξουμε το μεγαλύτερο δυνατό score $C^*(t)$ από όλες τις ακολουθίες που ολοκληρώνουν στη χρονική στιγμή t , καθορίζουμε την αναδρομική σχέση:

$$C^*(t) \equiv O(t) + \max_{\tau=0 \dots t} \{ \alpha F(t - \tau, \tau_p) + C^*(\tau) \}$$

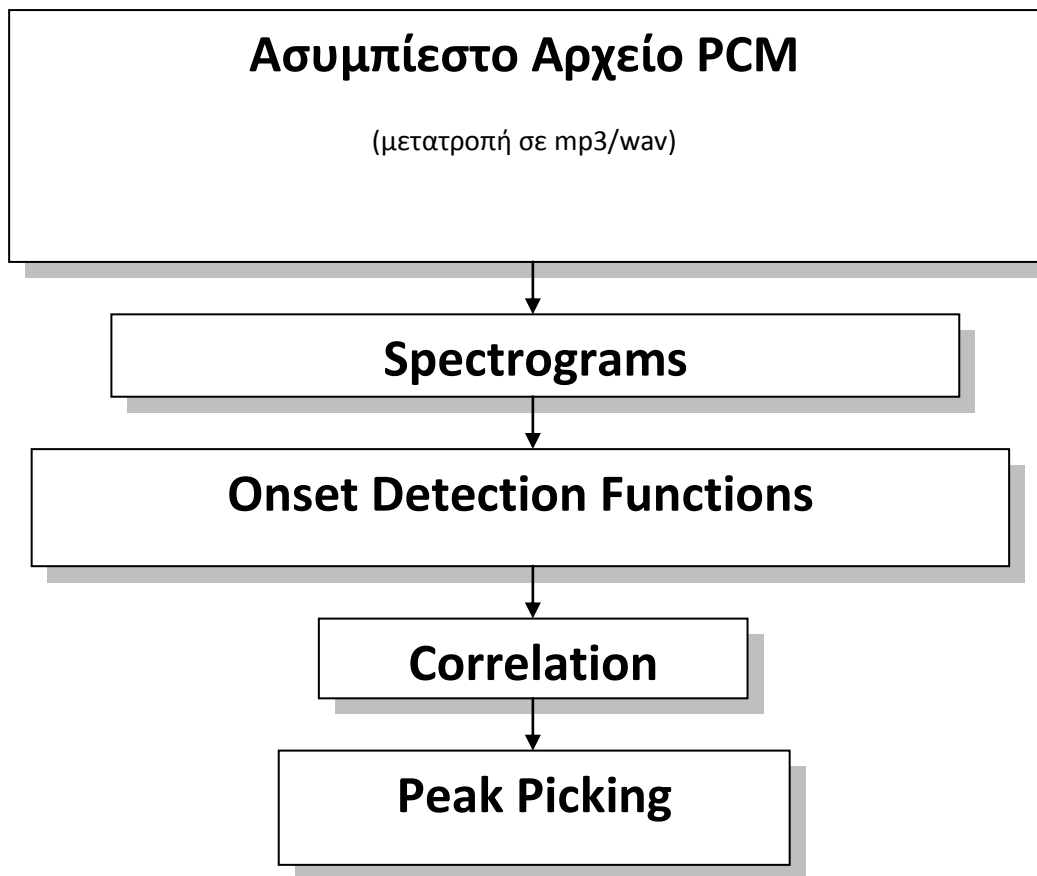
Η εξίσωση αυτή βασίζεται στην παρατήρηση πως το καλύτερο score κατά τη χρονική στιγμή t , αποτελεί το τοπικό Onset Strength, συν το καλύτερο score του προηγούμενου beat τ που αυξάνει το άθροισμα του βέλτιστου Score και του κόστους μετάβασης από την συγκεκριμένη χρονική στιγμή. Καθώς υπολογίζουμε το $C^*(t)$, καταγράφουμε επίσης και την χρονική στιγμή από το προηγούμενο beat, που απέδωσε το βέλτιστο score:

$$P^*(t) = \arg \max \{ \alpha F(t - \tau, \tau_p) + C^*(\tau) \}$$

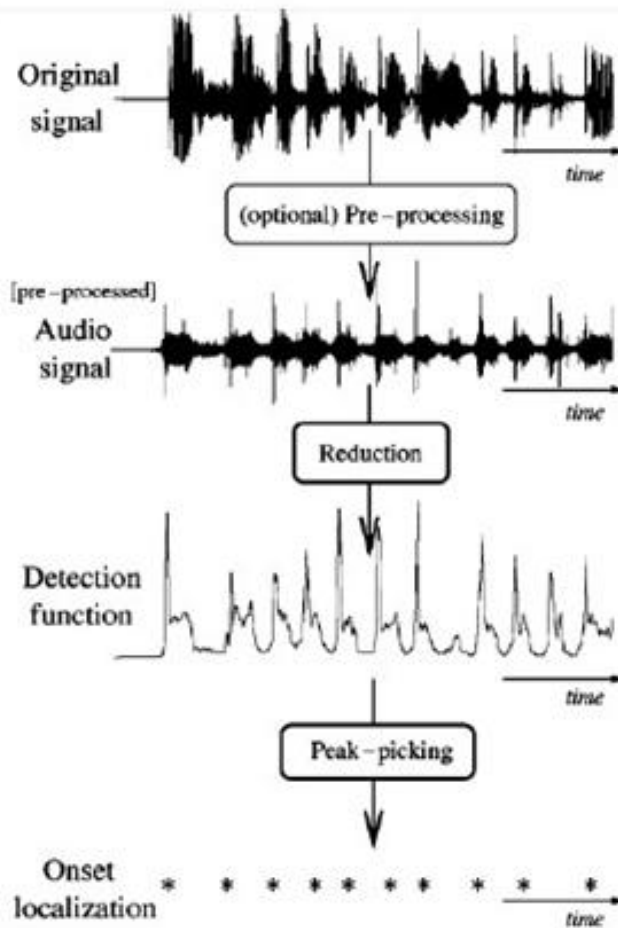
Στην πραγματικότητα είναι απαραίτητο να ψάξουμε μόνο ένα περιορισμένο εύρος από τ εφόσον ο γρήγορα αυξανόμενος όρος ποινής F καθιστά απίθανο η βέλτιστη στιγμή του προηγούμενου beat να είναι μακριά από το $t - \tau_p$. Έτσι η αναζήτηση γίνεται για $\tau = t - 2\tau_p \dots t - \tau_p / 2$.

Για να βρούμε το σύνολο των beats που βελτιστοποιούν την αντικειμενική συνάρτηση για μια δεδομένη συνάρτηση αντίχρευσσης, ξεκινάμε υπολογίζοντας τα C^* και P^* για κάθε στιγμή ξεκινώντας από το 0. Μόλις ολοκληρωθεί αυτή η διαδικασία, κοιτάμε για τη μεγαλύτερη τιμή του C^* ; έτσι έχουμε την τελική εμφάνιση beat t_N - όπου το N , ο συνολικός αριθμός των beats, δεν είναι ακόμα γνωστός σε αυτό το σημείο. Έπειτα με 'οπισθοδρόμηση' μέσω του P^* , βρίσκουμε το προηγούμενο beat $t_{N-1} = P^*(t_N)$, και σταδιακά δουλεύοντας προς τα πίσω, φτάνουμε στην αρχή του σήματος. Αυτό δίνει ολόκληρη τη βέλτιστη ακολουθία $\{t_i\}^*$. Χάρη στο δυναμικό προγραμματισμό, ερευνήθηκε αποδοτικά ολόκληρο το εκθετικού μεγέθους σύνολο από όλες τις δυνατές ακολουθίες beat σε γραμμικό χρόνο. Αυτό ήταν εφικτό διότι στην αντικειμενική συνάρτηση γεγονότα μεταγενέστερα του t_i δε μπορούν να επηρεάσουν τη συνεισφορά των προγενέστερων γεγονότων στο κόστος. [17]

2.2 Στάδια /Μεθοδολογία Υλοποίησης Beat Tracking



Εικόνα 2.1: Στάδια Υλοποίησης Beat Tracking



Εικόνα 2.2: Διάγραμμα Ροής Αλγορίθμου Εύρεσης Onset.

(Η εικόνα προέρχεται από το [14])

2.2.1 Περιγραφή Ασυμπίεστου Αρχείου

Οι μορφές ήχου στοχεύουν στην παρουσίαση των ψηφιακών ηχογραφήσεων των παραστάσεων. Η ψηφιακή απόκτηση βασίζεται στη δειγματοληψία του σήματος και στην κβαντοποίηση των τιμών του δείγματος. Οι περισσότερες μη συμπίεσμένες μορφές ήχου βασίζονται στην παρουσίαση του Pulse Code Modulation (PCM) όπου κάθε δείγμα παρουσιάζεται ως ένας παλμός με ένα δεδομένο εύρος. Στερεοφωνικές ή πολυκαναλικές πληροφορίες παρουσιάζονται συνήθως παρεμβάλλοντας τα δείγματα που έχουν παρθεί από κάθε κανάλι.

Οι δύο σχετικές παράμετροι των PCM μορφών ήχου είναι ο ρυθμός δειγματοληψίας που είναι ο ρυθμός των ισαπεχόντων δειγμάτων που λαμβάνονται στη μονάδα του χρόνου και η ανάλυση πλάτους που είναι ο αριθμός των bits που χρησιμοποιούνται για να παραστήσουν κάθε δείγμα. Και οι δύο παράμετροι

σχετίζονται με την αντιλαμβανόμενη ποιότητα του ήχου. Οι τιμές που έχουν επιλεγεί για audio CD- 44.1 kHz για το ρυθμό δειγματοληψίας και 16 bits για την ανάλυση πλάτους κάθε καναλιού- γίνονται το πρότυπο για υψηλής ποιότητας ήχο, ακόμα και αν η πραγματική τεχνολογία επιτρέπει ψηφιακό ήχο για υψηλότερο ρυθμό δειγματοληψίας και ανάλυσης πλάτους. Διαθέσιμες φόρμες ήχου- όπως η AIFF, η WAVE και η AU- μπορούν εύκολα να μετατραπούν από τη μία στην άλλη. Ως εκ τούτου, η επιλογή της φόρμας ήχου δεν είναι ένα σχετικό ζήτημα. Ένα παράδειγμα μιας PCM παρουσίασης ενός ηχητικού σήματος φαίνεται στην εικόνα 2.2, ο ήχος είναι μια παράσταση για πιάνο του σκορ που φαίνεται στην εικόνα 1.1.

Οι μορφές που βασίζονται στην PCM δεν είναι ιδιαίτερα αποτελεσματικές στην απασχόληση της μνήμης. 10 δευτερόλεπτα απόλυτης ησυχίας χρειάζονται το ίδιο διάστημα χρόνου όπως 10 δευτερόλεπτα ενός tutti, i.e κλπ, όλο το σύνολο παίζει ταυτόχρονα, ενώ εκτελείται από μια συμφωνική ορχήστρα. Οι συμπίεσμένες φόρμες έχουν αναπτυχθεί να ξεπερνούν αυτόν τον περιορισμό. Ανάμεσα στις διαφορετικές συμπίεσμένες φόρμες, η MPEG 1 και η Layer 3 (MP3) είναι οι πιο γνωστές. Μια πρόσθετη παράμετρος των συμπίεσμένων φορμών είναι το bitrate, που είναι ο αριθμός των bits που χρειάζονται για να κωδικοποιήσουμε ένα δευτερόλεπτο μουσικής. Άλλες ιδιότητες φόρμες έχουν προταθεί, πιο συγκεκριμένα η Windows media Audio, που επιτυγχάνει καλή συμπίεση με καλή ποιότητα ήχου. Ξεκάθαρα, η έλλειψη δημόσιων προδιαγραφών αυτών των φορμών τις κάνουν λιγότερο κατάλληλες για ΜIP εργασίες.

Υπάρχει ένας αριθμός φορμών ήχου που, αντί να παρουσιάζουν την εξέλιξη του ήχου, περιγράφουν την τεχνική σύνθεσης και τις παραμέτρους σύνθεσης που χρειάζονται για να παράγουμε ένα δεδομένο ήχο ή μια παράσταση. Μπορούν να θεωρηθούν το ακουστικό παράλληλο των διανυσματικών γραφικών και είναι ιδιαίτερα χρήσιμες στην παρουσίαση συνθετικών ήχων. Αυτές οι φόρμες επιτυγχάνουν μια μεγάλη τιμή συμπίεσης, επειδή οι παράμετροι ελέγχου αλλάζουν πολύ πιο σιγά από ότι οι παράμετροι ήχου και έχουν συμπεριληφθεί στο MPEG-4 πρότυπο. Στην κατάσταση της τέχνης, αυτές οι φόρμες δεν βρίσκουν εφαρμογή σε MIR, κυρίως επειδή είναι περιορισμένες σε μια τάξη ήχων που δεν έχουν ιδιαίτερο ενδιαφέρον για τους τελικούς χρήστες.

Μια άλλη εξέλιξη του MPEG προτύπου είναι το MPEG-7, το οποίο λαμβάνει υπόψη επίσης την κωδικοποίηση των περιγραφέων ήχου, συγκεκριμένα για τη διάσταση ηχοχρώματος που μπορεί να χρησιμοποιηθεί για MIR εργασίες. Μια προσέγγιση των περιγραφέων ταξινόμησης ήχου σε MPEG-7 έχει προταθεί (2).

Το MP3 είναι ένας τύπος ψηφιακού συμπίεσμένου αρχείου ήχου. Παρέχει τη δυνατότητα αναπαράστασης ήχου κωδικοποιημένου με μορφή Pulse Code Modulation (PCM) (διαμόρφωση με βάση κωδικούς παλμών), δεσμεύοντας όμως πολύ λιγότερο χώρο (για δεδομένα) σε σχέση με τις άμεσες μεθόδους. Αυτό γίνεται χρησιμοποιώντας ψυχοακουστικά μοντέλα για να απορρίψει τμήματα ή περιοχές του ηχητικού φάσματος που δεν ακούει το ανθρώπινο αυτί και καταγράφοντας την υπόλοιπη πληροφορία με αποτελεσματικό τρόπο. Παρόμοιες μέθοδοι χρησιμοποιούνται από το JPEG, ένα πρότυπο συμπίεσης εικόνων με απώλειες οπτικών λεπτομερειών μη αντιληπτών από το ανθρώπινο μάτι.

2.2.2 Περιγραφή Spectrogram

2.2.2.1 Φασματογράφημα του σήματος

Το **Φασματογράφημα** (Spectrogram) αποτελεί την χρονικά μεταβαλλόμενη αναπαράσταση του φάσματος ενός ηχητικού σήματος, που παρουσιάζει την μεταβολή της φασματικής πυκνότητας του σήματος σε συνάρτηση με τον χρόνο. Η συνηθέστερη μορφή ενός Spectrogram, είναι ένα γράφημα δύο διαστάσεων στο οποίο ο οριζόντιος άξονας εκφράζει τον χρόνο και ο κάθετος τις τιμές των συχνοτήτων.

Ο υπολογισμός ενός Spectrogram γίνεται κυρίως με την βοήθεια του Μετασχηματισμού Fourier. Πιο συγκεκριμένα, χρησιμοποιείται ο Μετασχηματισμός Fourier Βραχέως Χρόνου (STFT: Short Time Fourier Transform).

Ο υπολογισμός του **STFT**, ενός διακριτού σήματος $X(n)$, δίνεται από την παρακάτω σχέση:

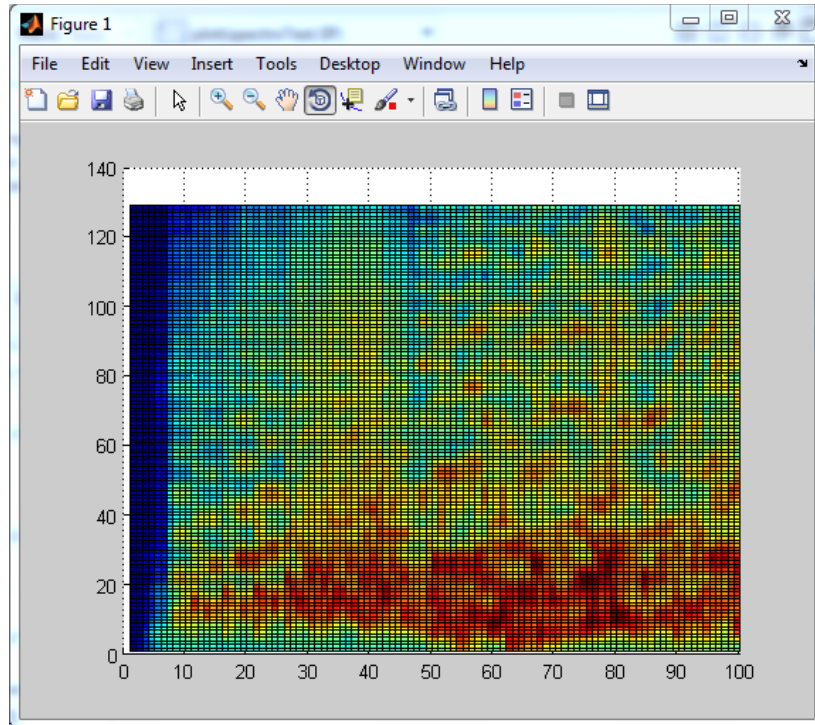
$$STFT\{X(n)\}(m, \omega) \equiv X(m, \omega) = \sum_{n=-\infty}^{+\infty} X(n)w[n - m]e^{-nj\omega}$$

όπου $w[n]$: συνάρτηση παραθύρου. (Συνήθως χρησιμοποιούνται τα παράθυρα Hann /Gaussian)

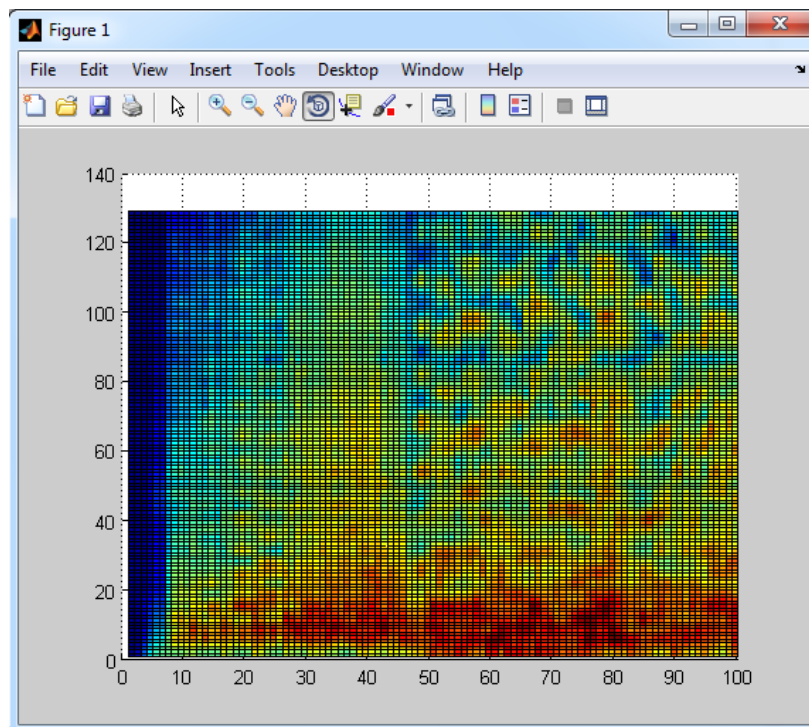
Το Spectrogram στη συνέχεια, ορίζεται ως το τετράγωνο του πλάτους του STFT.

$$spect\{x(t)\}(\tau, \omega) \equiv |X(\tau, \omega)|^2$$

Στις εικόνες 2.3 και 2.4 που ακολουθούν βλέπουμε τα φασματογραφήματα της Onset Detection Function Masri για τις τιμές των παραμετροποιήσεων dB(1)-mel(1)-nmel(40) και dB(0)-mel(0)-nmel(0) αντίστοιχα.



Εικόνα 2.3: Spectrogram-Masri-(dB(1)-mel(1)-nmel(40))



Εικόνα2.4: Spectrogram-Masri-(dB(0)-mel(0)-nmel(0))

2.2.3 Περιγραφή Onset Detection Function

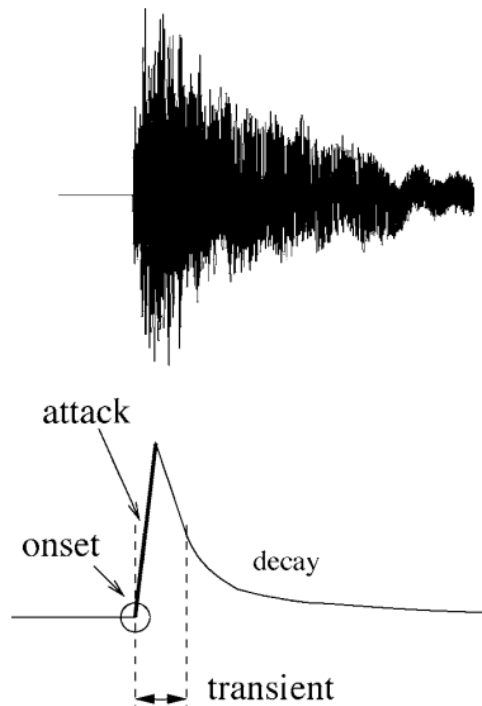
Εξαγωγή συνάρτησης ανίχνευσης

Η συνάρτηση ανίχνευσης αποτελεί τον μετασχηματισμό ενός ηχητικού σήματος που παρουσιάζει την εμφάνιση των transients στο αρχικό μας σήμα. Είναι η κύρια διαδικασία, αν όχι σε όλους, τότε στους περισσότερους αλγορίθμους ανίχνευσης των onsets.

Υπάρχουν δύο κατηγορίες τεχνικών για την εξαγωγή της συνάρτησης ανίχνευσης:

- Οι ντετερμινιστικές τεχνικές που βασίζονται στη χρήση χαρακτηριστικών του σήματος, χρονικών ή φασματικών (με ή χωρίς χρήση πληροφορίας για τη φάση)
- Οι στατιστικές τεχνικές που βασίζονται στην υπόθεση ότι το σήμα μπορεί να περιγραφεί από ένα στατιστικό μοντέλο.

Κάποια παραδείγματα συναρτήσεων ανίχνευσης είναι η περιβάλλουσα του σήματος στο πεδίο του χρόνου, η ενέργεια βραχέος χρόνου, η φασματική ροή και η μεταβολή της φάσης. [17]



Εικόνα 2.5: Onset, Attack, Transient και Delay μιας Νότας. [14]

Επιλογή Κορυφών

Εάν η συνάρτηση ανίχνευσης έχει σχεδιαστεί κατάλληλα, τα onsets και άλλα απότομα γεγονότα θα οδηγήσουν σε καλά εντοπισμένα αναγνωρίσιμα χαρακτηριστικά στη συνάρτηση ανίχνευσης. Συνήθως αυτά τα χαρακτηριστικά είναι τοπικά μέγιστα, συχνά αντικείμενα κάποιου βαθμού μεταβλητότητας στο μέγεθος και το σχήμα, και επισκιάζονται από 'θόρυβο' που είτε είναι πραγματικός θόρυβος που υπάρχει στο σήμα, ή άλλες πτυχές του σήματος που δεν σχετίζονται με τα onsets. Επομένως, ένας εύρωστος αλγόριθμος επιλογής κορυφών είναι απαραίτητος για να εκτιμήσει τους χρόνους των onsets του σήματος. Η διαδικασία της επιλογής κορυφών από μια συνάρτηση ανίχνευσης μπορεί να χωριστεί σε 3 στάδια: μετα-επεξεργασία, κατωφλίωση, και μια τελική διαδικασία απόφασης.[17]

Μετα-επεξεργασία:

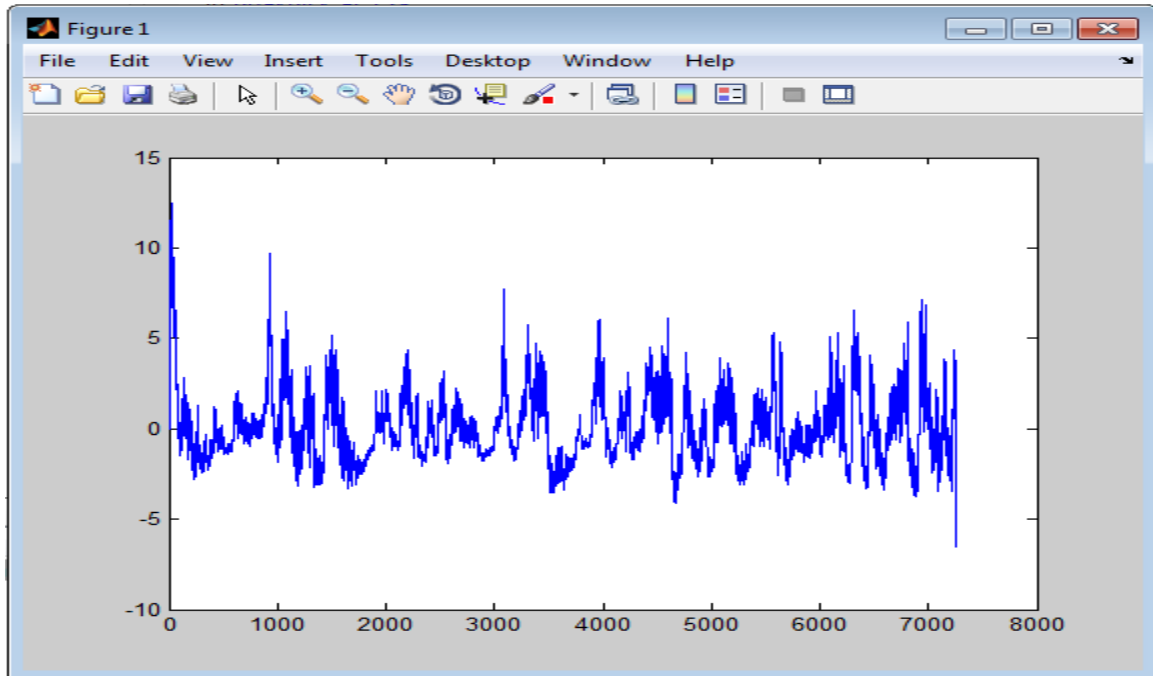
Όπως και η προ-επεξεργασία, η μετα-επεξεργασία είναι ένα προαιρετικό βήμα που εξαρτάται και από τη μέθοδο που ακολουθήθηκε για την εύρεση της συνάρτησης ανίχνευσης. Ο σκοπός της μετα-επεξεργασίας είναι να διευκολύνει τα προβλήματα της κατωφλίωσης και της επιλογής κορυφών αυξάνοντας την ομοιομορφία και τη συνεκτικότητα των χαρακτηριστικών της συνάρτησης ανίχνευσης που σχετίζονται με onsets, ιδανικά μετασχηματίζοντάς τα σε απομονωμένα, εύκολα ανιχνεύσιμα τοπικά μέγιστα. Μέθοδοι μετα-επεξεργασίας είναι το smoothing για την αφαίρεση του θορύβου, και διαδικασίες που χρειάζονται για την επιτυχή επιλογή των παραμέτρων κατωφλίωσης (κανονικοποίηση, αφαίρεση της DC συνιστώσας).[17]

Κατωφλίωση:

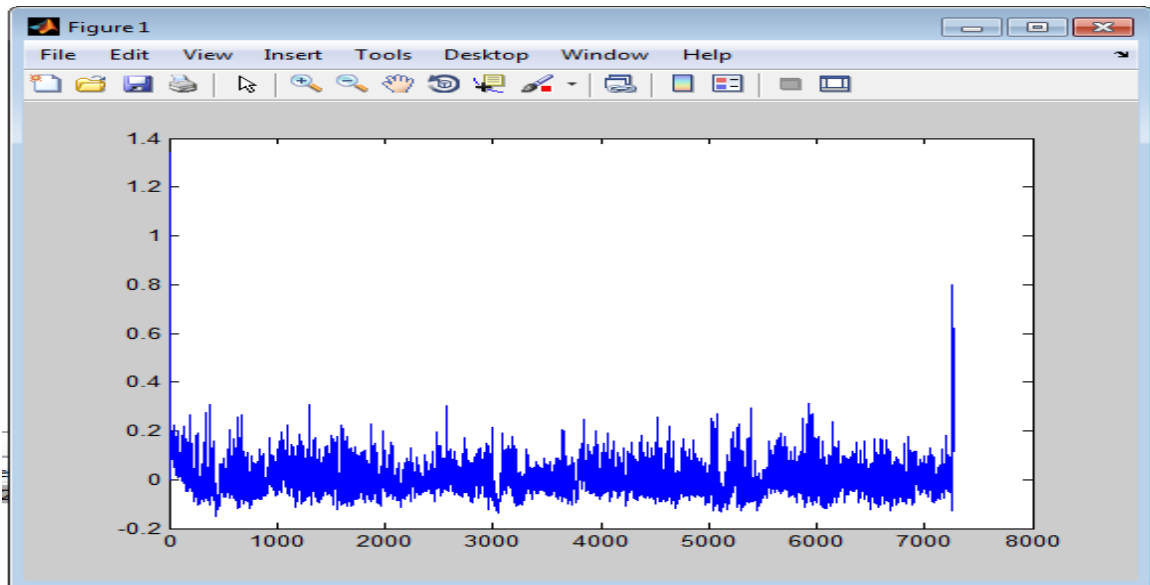
Για κάθε τύπο συνάρτησης ανίχνευσης και ακόμα και μετά τη μετα-επεξεργασία, θα υπάρχει ένας αριθμός κορυφών που δε θα σχετίζονται με τα onsets. Συνεπώς, είναι απαραίτητο να οριστεί ένα κατώφλι που θα χωρίζει τα γεγονότα σε αυτά που σχετίζονται με ένα onset και σε αυτά που δεν σχετίζονται. Το κατώφλι μπορεί να είναι σταθερό ή προσαρμοστικό. [17]

Τελική επιλογή κορυφών: Μετά τη μετα-επεξεργασία και την κατωφλίωση της συνάρτησης ανίχνευσης, το μόνο που μένει για την επιλογή κορυφών είναι ο προσδιορισμός των τοπικών μεγίστων που υπερβαίνουν το καθορισμένο κατώφλι.[17]

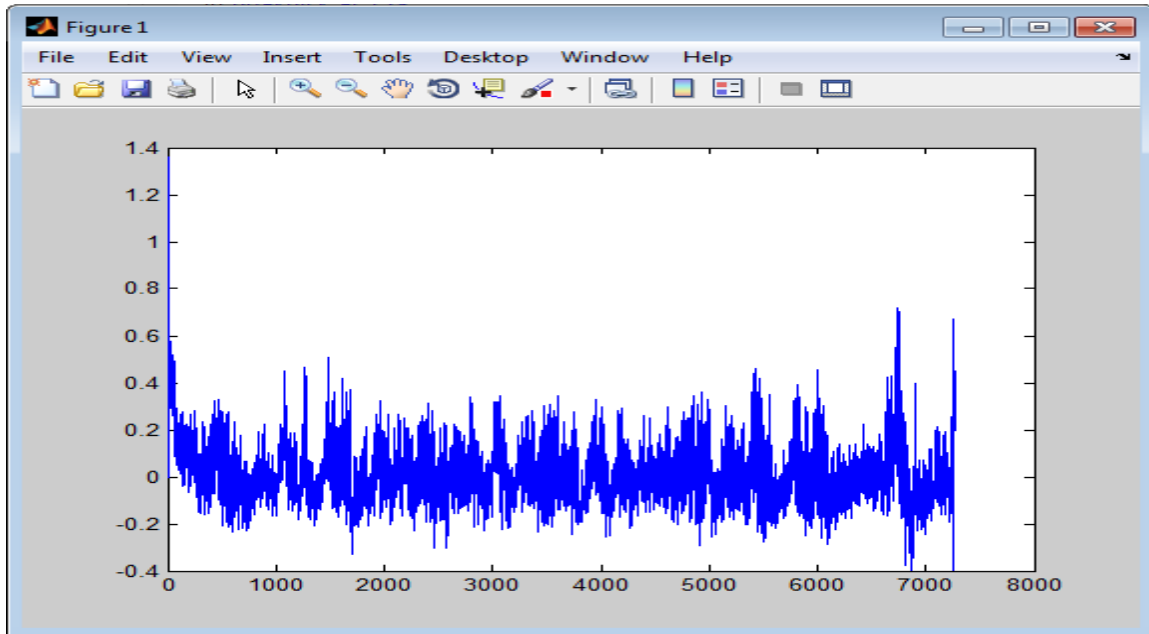
Παρακάτω παρουσιάζονται οι εικόνες διάφορων Onset Detection Function και των τιμών των παραμέτρων που τις υπολοιοποιούν. Τα ονόματα των Onset Det. Function βρίσκονται μέσα στις παρενθέσεις. Παρακάτω θα μελετήσουμε με λεπτομέρεια τι σημαίνουν οι παραμετροποιήσεις 0-0-0-0-xcorr-Ellis-Ellis-Ellis-0.



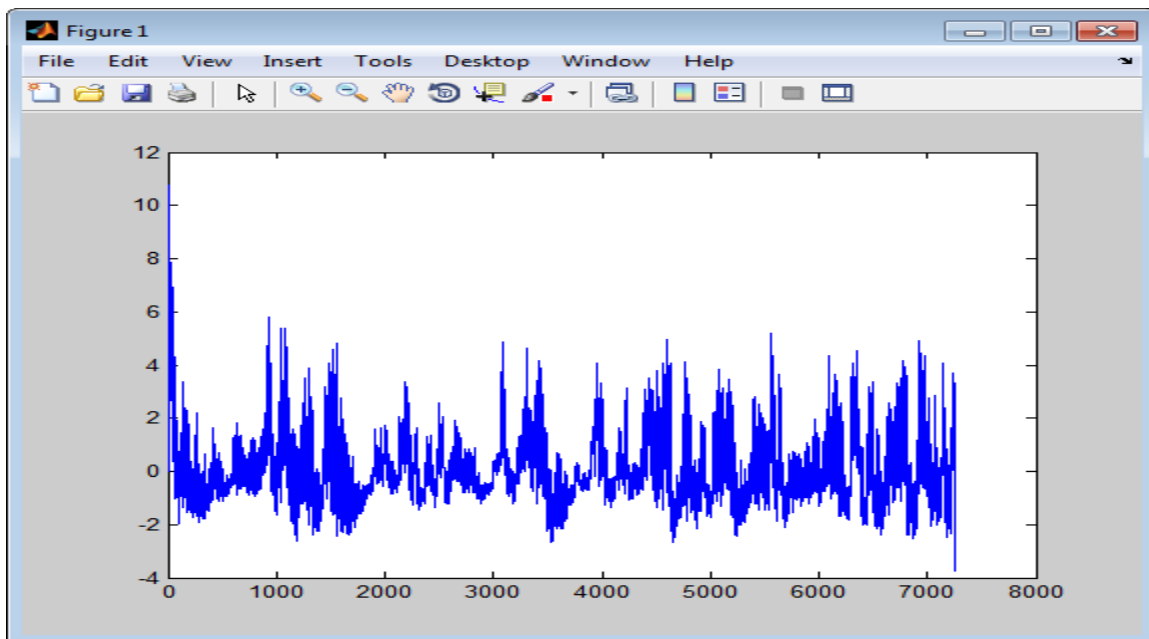
Εικόνα 2.6: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(CD)-3ο κομμάτι



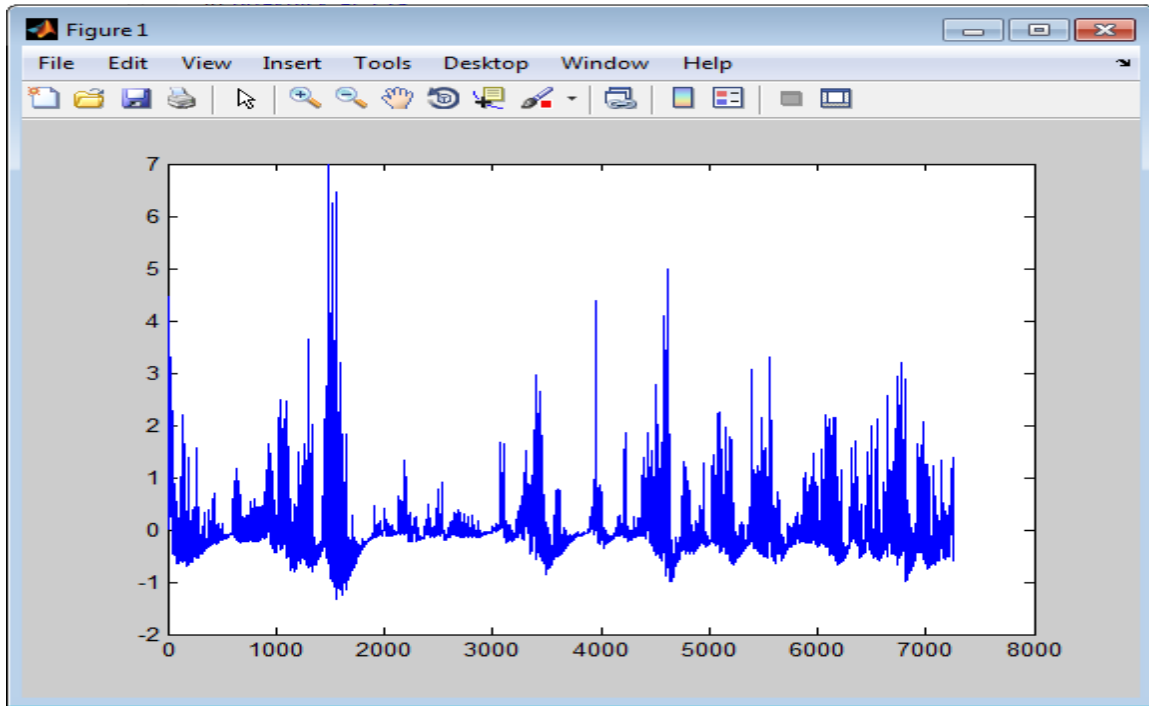
Εικόνα 2.7: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(NWPD)-3ο κομμάτι



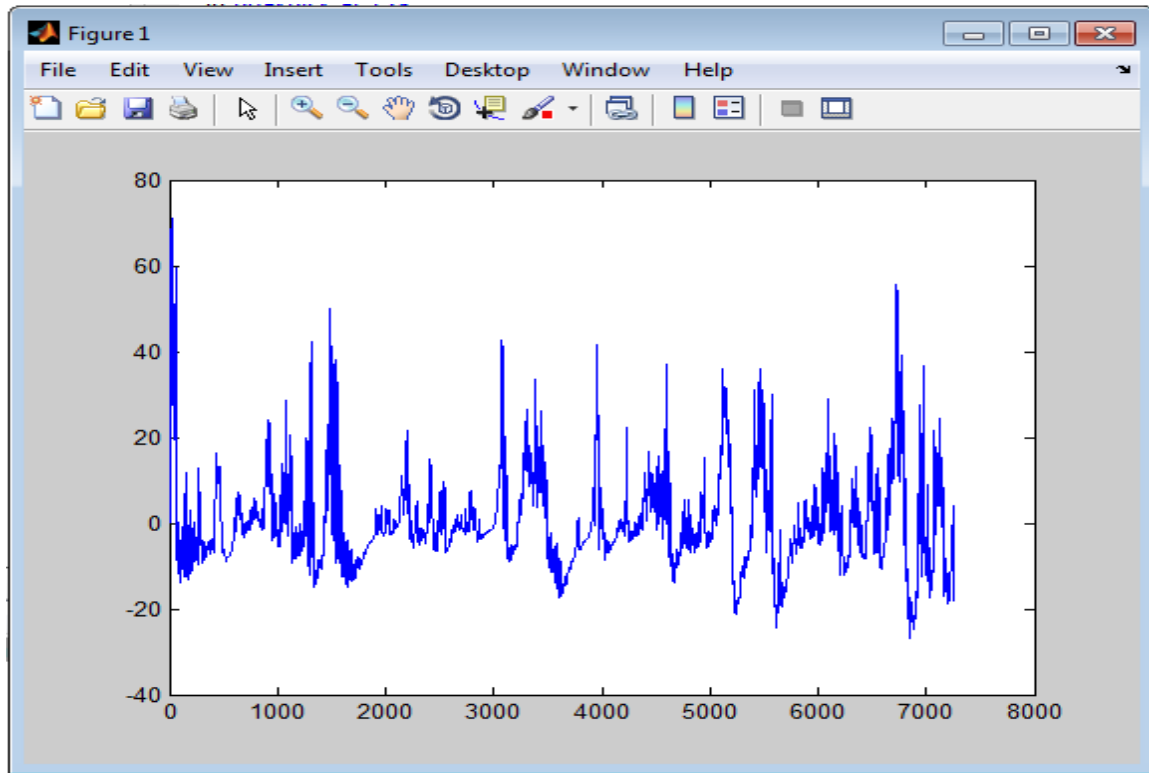
Εικόνα2.8: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(PD)-3ο κομμάτι



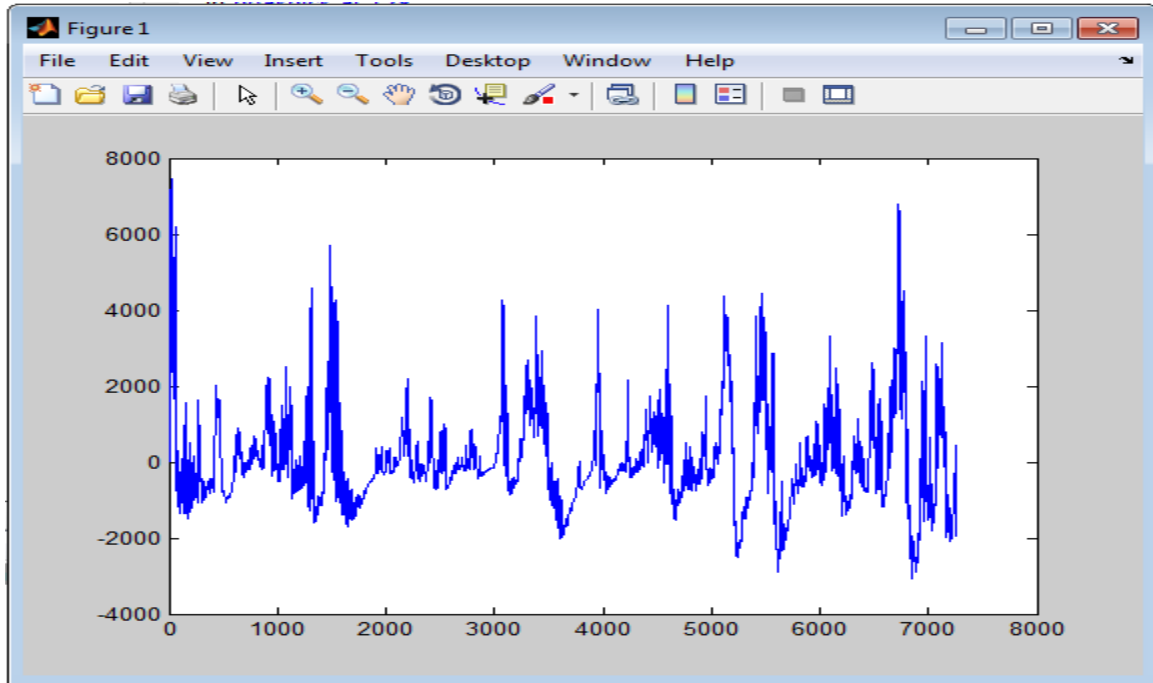
Εικόνα 2.9: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(RCD)-3ο κομμάτι



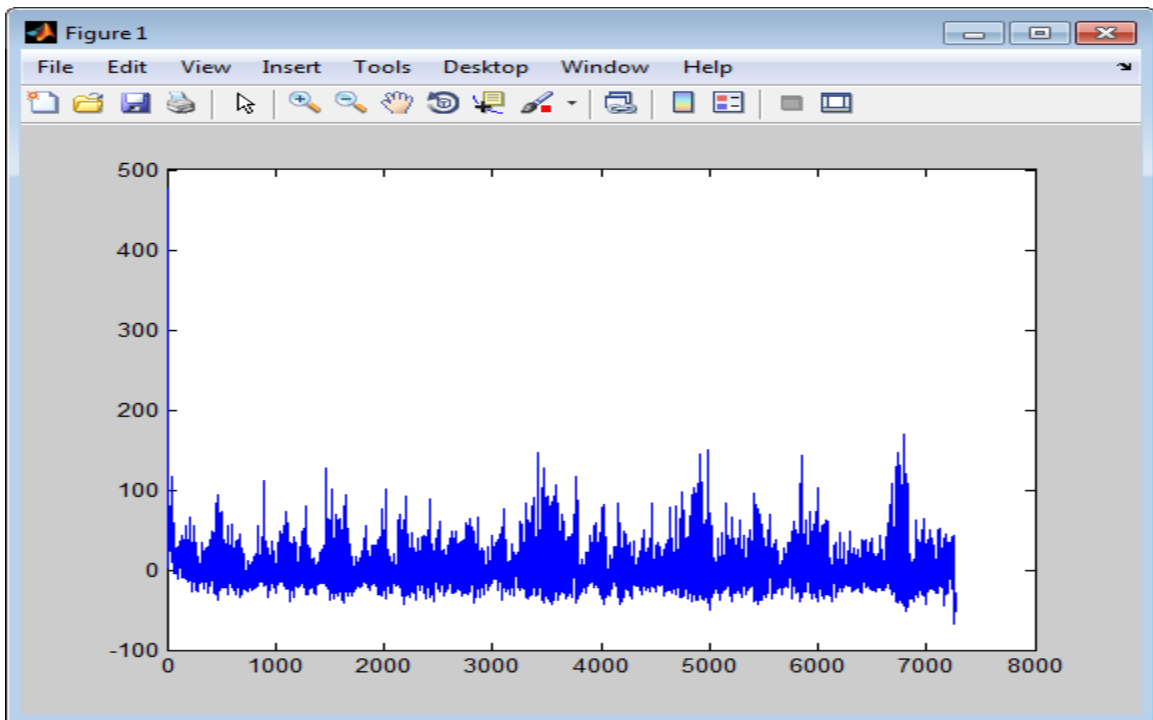
Εικόνα 2.10: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(**Duxbury**)-3ο κομμάτι



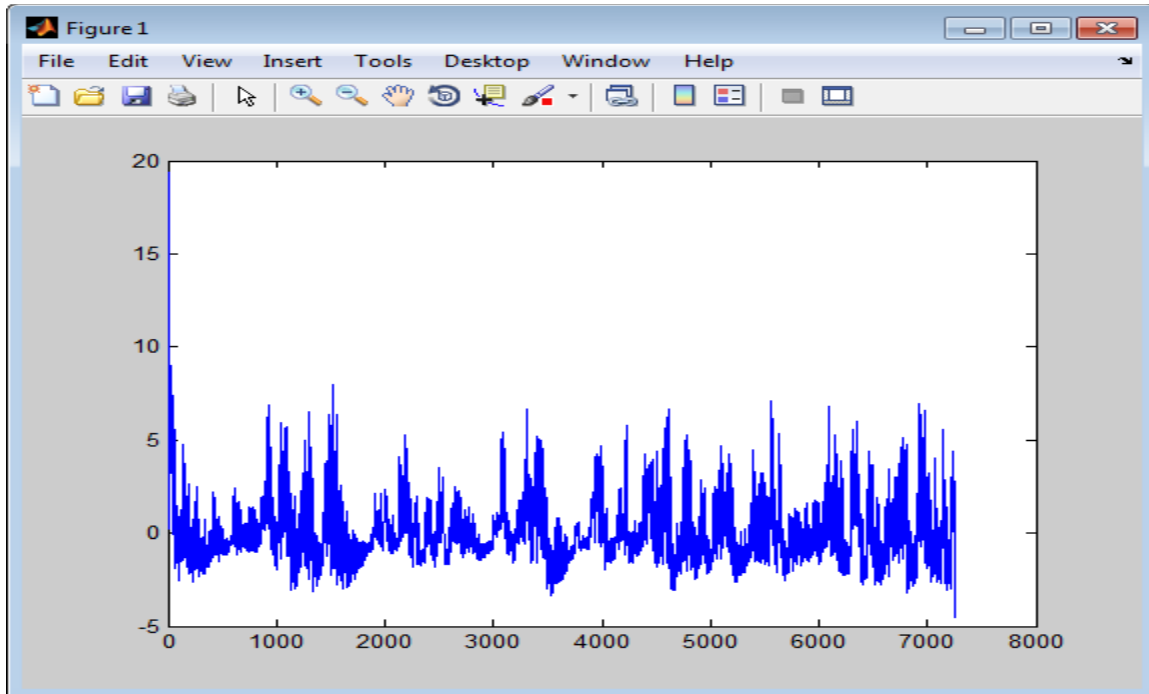
Εικόνα:2.11: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(**E**)-3ο κομμάτι



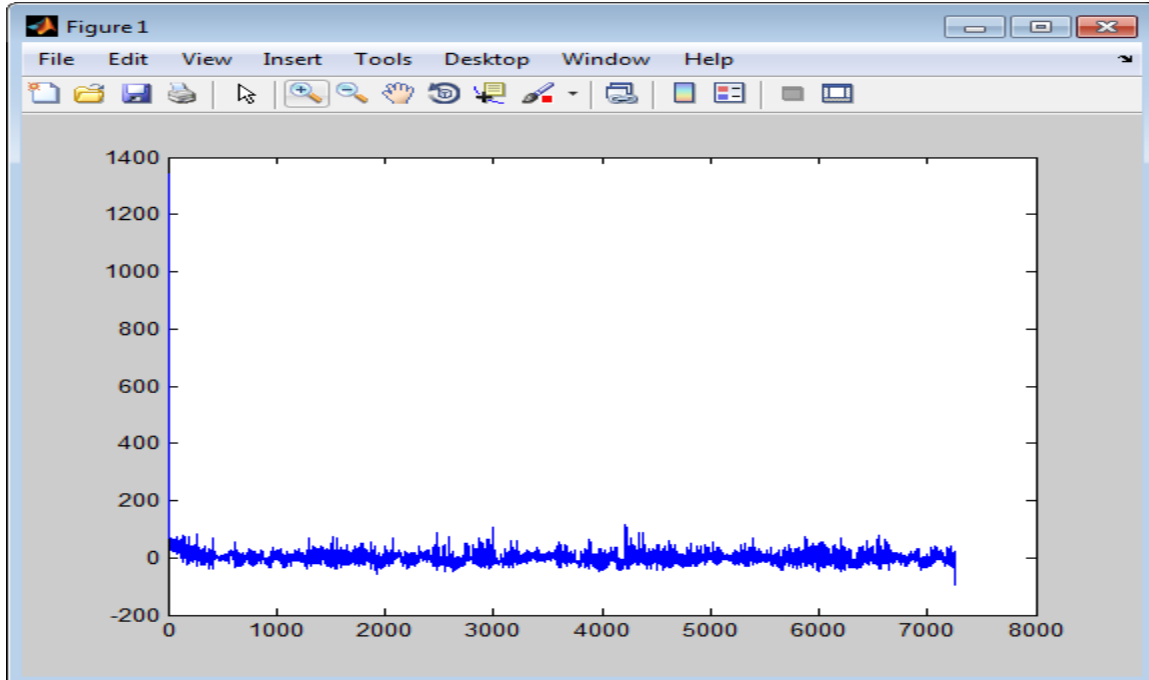
Εικόνα 2.12: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(**HFC**)-3ο κομμάτι



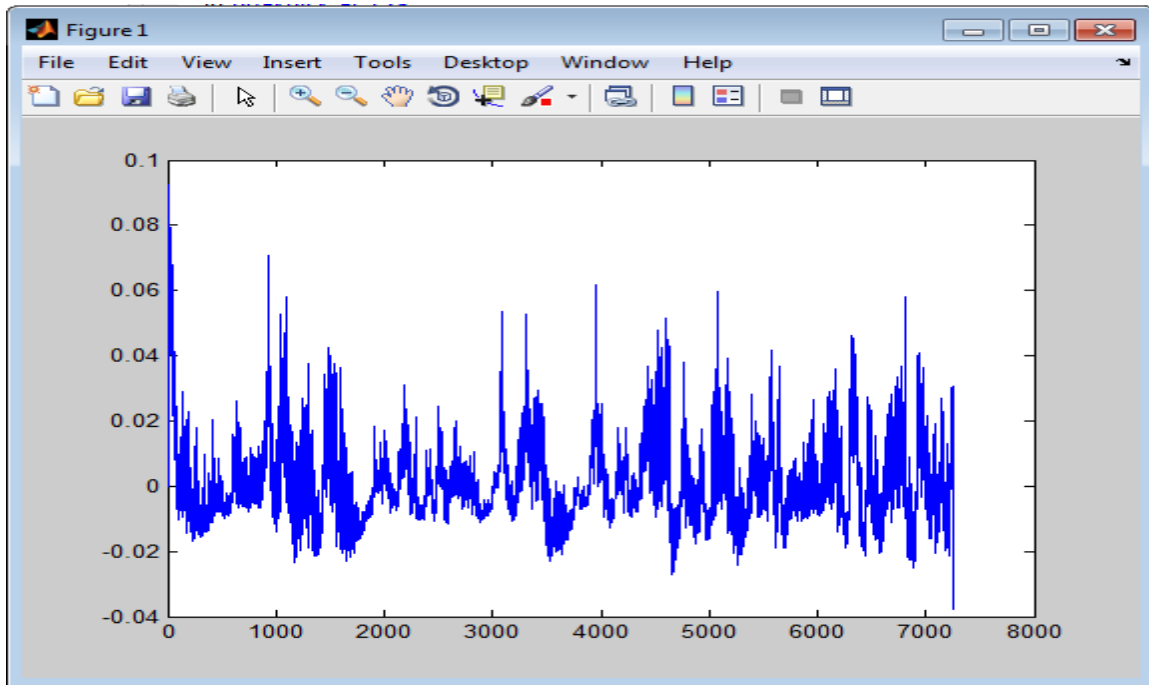
Εικόνα 2.13: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(**Kullback**)-3ο κομμάτι



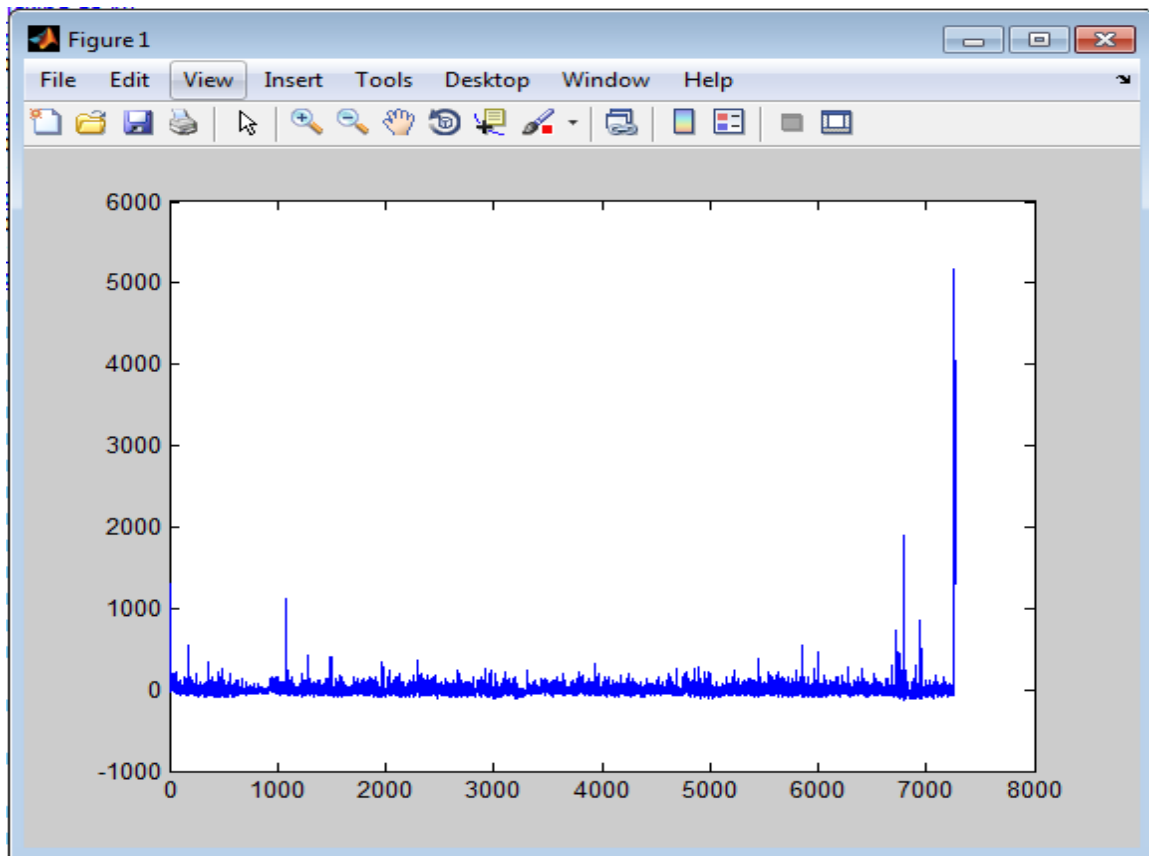
Εικόνα 2.14: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(**Masri**)-3ο κομμάτι



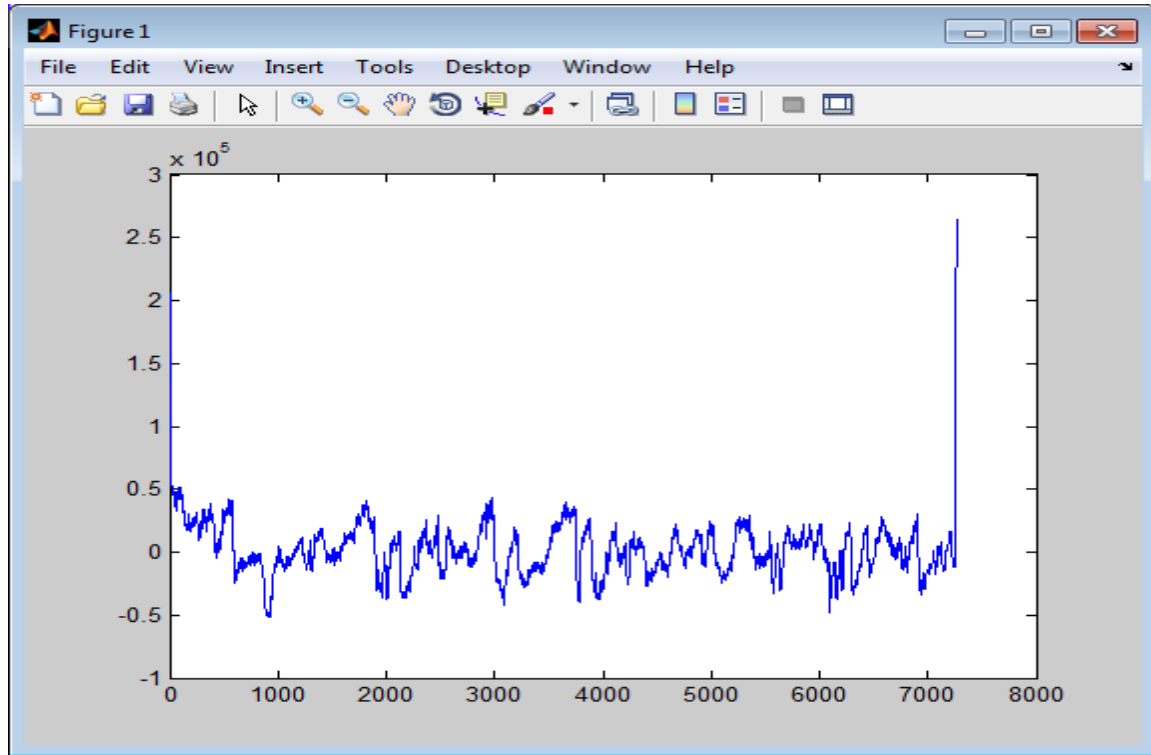
Εικόνα 2.15: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(**MoT**)-3ο κομμάτι



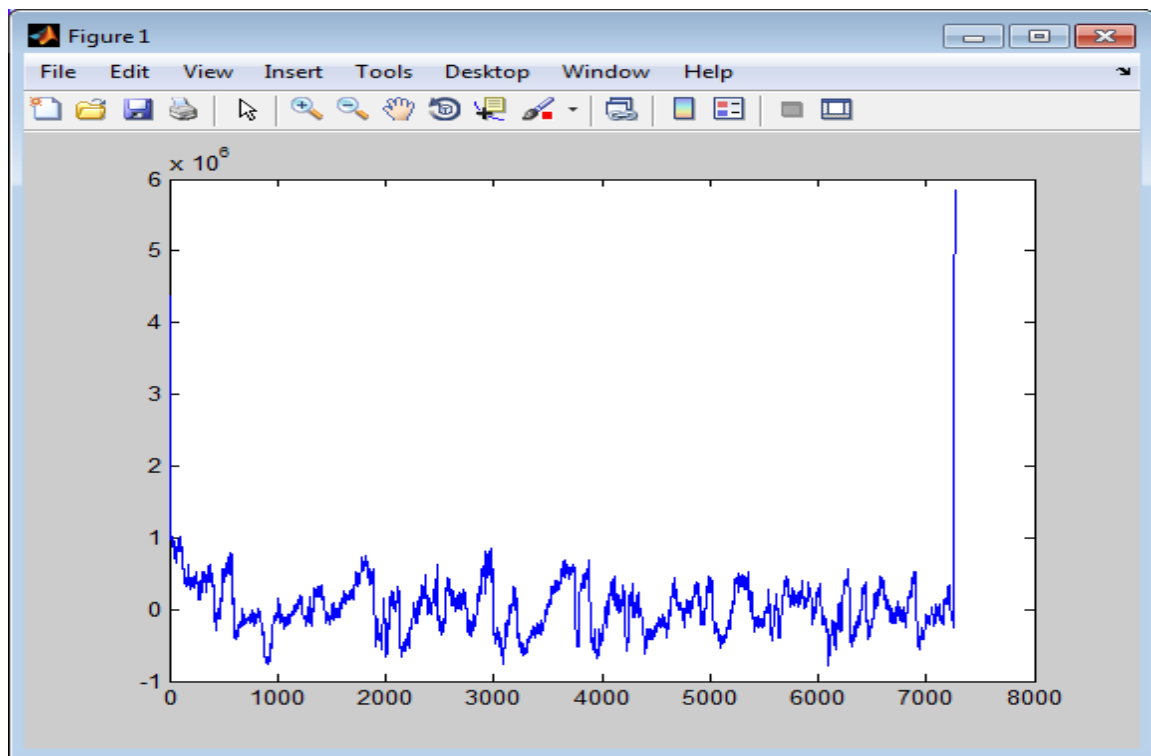
Εικόνα 2.16: OnsetCorrscollection-0-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(WPD)-3ο κομμάτι



Εικόνα 2.17: OnsetCorrscollection-1-1-40-0-0-xcorr-Ellis-Ellis-Ellis-0(Duxbury)-3ο κομμάτι



Εικόνα 2.18: OnsetCorrscollection-1-1-40-0-0-xcorr-Ellis-Ellis-Ellis-0(**E**)-3ο κομμάτι



Εικόνα 2.19: OnsetCorrscollection-1-1-40-0-0-xcorr-Ellis-Ellis-Ellis-0(**HFC**)-3ο κομμάτι

2.2.3.1 Ο αλγόριθμος που εφαρμόζεται

Συνάρτηση ανίχνευσης

Στη συνέχεια, σε κάθε ζώνη συχνοτήτων υπολογίζεται χωριστά μια συνάρτηση ανίχνευσης. Η συνάρτηση ανίχνευσης είναι στο μιγαδικό πεδίο και συνδυάζει πληροφορία και για την ενέργεια και για τη φάση του σήματος. [17]

Έστω ο μετασχηματισμός βραχέως χρόνου του σήματος, με τη χρήση ενός παραθύρου Hamming $w(m)$. Με τη βοήθεια του μετασχηματισμού υπολογίζεται μία αναπαράσταση χρόνου-συχνότητας. Εάν $X(n, k)$ αντιστοιχεί στην k -οστή συχνοτική συνιστώσα (frequency bin) του n -οστού πλαισίου, τότε:

$$X(n, k) = \sum_{m=-N/2}^{N/2-1} x(hn + m)w(m)e^{-\frac{2j\pi mk}{N}} \quad (2.1)$$

όπου N το μέγεθος του παραθύρου και h το hop size (αριθμός δειγμάτων μεταξύ δύο διαδοχικών παραθύρων fft).

Η φασματική ροή (spectral flux) είναι ένας τρόπος υπολογισμού της τοπικής φασματικής μεταβολής μεταξύ διαδοχικών πλαισίων. Δεδομένου ενός φασματογραφήματος η φασματική ροή είναι η απόσταση μεταξύ του φάσματος των διαδοχικών πλαισίων. Οι κορυφές στην καμπύλη της φασματικής ροής υποδεικνύουν τις χρονικές θέσεις σημαντικών αντιθέσεων στο φασματογράφημα. Με τη βοήθεια της φασματικής ροής ανιχνεύονται απότομες θετικές αλλαγές στην ενέργεια του σήματος που υποδεικνύουν τα attack μέρη νέων νοτών. Μετράει τη διαφορά στο πλάτος σε κάθε συχνοτική συνιστώσα, και αν περιοριστεί μόνο σε θετικές αλλαγές και αθροιστεί κατά μήκος όλων των συχνοτήτων, δίνει :

$$SF(n) = \sum_{k=-N/2}^{N/2-1} H(|X(n, k)| - |X(n-1, k)|) \quad (2.2)$$

Η στιγμιαία συχνότητα μπορεί να υπολογιστεί από την πρώτη διαφορά στη φάση του $X(n, k)$. Έστω $\psi(n, k)$ η φάση του $X(n, k)$, δηλαδή $X(n, k) = |X(n, k)|e^{j\psi(n, k)}$ όπου $-\pi <$

$\psi(n, k) \leq \pi$.

Τότε η στιγμιαία συχνότητα δίνεται από την :

$$\psi'(n, k) = \psi(n, k) - \psi(n-1, k) \quad (2.3)$$

αντιστοιχισμένη στο διάστημα $(-\pi, \pi]$. Η αλλαγή στη στιγμιαία συχνότητα, που είναι μία ένδειξη ενός πιθανού onset, δίνεται από τη δεύτερη διαφορά της φάσης:

$$\psi''(n, k) = \psi'(n, k) - \psi'(n-1, k) \quad (2.4)$$

αντιστοιχισμένη επίσης στο διάστημα $(-\pi, \pi]$.

Το πλάτος και η φάση μπορούν να θεωρηθούν από κοινού στην αναζήτηση αναχωρήσεων από τη συμπεριφορά σταθερής κατάστασης υπολογίζοντας το αναμενόμενο πλάτος και φάση της τρέχουσας συχνοτικής συνιστώσας $X(n, k)$ βασιζόμενοι στις δύο προηγούμενες συχνοτικές συνιστώσες $X(n-1, k)$ και $X(n-2, k)$. Η αναμενόμενη τιμή $XT(n, k)$ της τρέχουσας συνιστώσας εκτιμάται θεωρώντας σταθερό πλάτος και ρυθμό αλλαγής φάσης:

$$XT(n, k) = |X(n-1, k)|e^{j\psi(n-1, k) + \psi(n-1, k)} \quad (2.5)$$

και στη συνέχεια παίρνουμε στο μιγαδικό πεδίο τη συνάρτηση του αθροίσματος των απολύτων αποκλίσεων από τις αναμενόμενες τιμές :

Αυτή η ποσότητα υπολογίζεται σε κάθε ζώνη συχνοτήτων. Χρησιμοποιείται η νόρμα L1.

Onset Detection Functions

(Constuctions)

2.2.3.1.1 Masri

Μια πιο γενική προσέγγιση με βάση τις αλλαγές στο φάσμα είναι η διαμόρφωση της λειτουργία ανίχνευσης ως «απόσταση» μεταξύ 2 διαδοχικών βραχυπρόθεσμων φασμάτων Fourier, που αντιμετωπίζονται ως σημεία σε έναν τρισδιάστατο χώρο. Ανάλογα με το μέτρο που έχει επιλεγεί, υπολογίζει αυτή την απόσταση, διαφορετική φασματική διαφορά, ή φασματική ροή, οι λειτουργίες ανίχνευσης μπορεί να κατασκευαστούν με τις παρακάτω συναρτήσεις:

Masri χρησιμοποιεί την L1-νόρμα της διαφοράς μεταξύ των φασμάτων μεγέθους, ενώ Duxbury χρησιμοποιεί τη L2-νόρμα.

$$\text{Masri}(n) = \sum_{k=N/2}^{\frac{N}{2}-1} \{H(|X_n(n)| - |X_k(n-1)|)\} \quad (2.6)$$

2.2.3.1.2 Duxbury

$$D(n) = \sum_{k=N/2}^{\frac{N}{2}-1} \{H(|X_n(n)| - |X_k(n-1)|)\}^2 \quad (2.7)$$

Όπου $(x+|x|)/2$, μηδέν για αρνητικά arguments. Η διόρθωση έχει ως αποτέλεσμα την καταμέτρηση μόνο αυτών των συχνοτήτων, όπου υπάρχει μια αύξηση στον τομέα της ενέργειας, και έχει ως στόχο να τονίσει ενάρξεις όχι αντισταθμιστικά οφέλη.

2.2.3.1.3 Kullback

Μια πολύ γνωστή προσέγγιση βασίζεται στην διαδοχική δοκιμή λόγου πιθανότητας . Προϋποθέτει τα δείγματα σήματος $x(n)$ που παράγονται από ένα από τα δύο στατιστικά μοντέλα ή και από τα 2. Το log- likelihood αναλογία ορίζεται ως:

$$s = \log \frac{p_B(x)}{p_A(x)} \quad (2.8)$$

όπου $p_A(x)$, $p_B(x)$ είναι οι συναρτήσεις πυκνότητας πιθανότητας που συνδέεται με τα δύο μοντέλα. Η προσδοκία της παρατηρούμενης αναλογίας log- likelihood συγκύσεως εξαρτάται από το ποιο μοντέλο το σήμα είναι στην πραγματικότητα παρακάτω. Σύμφωνα με το μοντέλο, η προσδοκία είναι:

$$E_A(s) = - \int p_A(x) \log \frac{p_A(x)}{p_B(x)} dx = -D(p_A \parallel p_B) < 0 \quad (2.9)$$

όπου D : δηλώνει η απόκλιση Kullback-Leibler μεταξύ της μοντέλο και οι παρατηρούμενες κατανομές. Σύμφωνα με το μοντέλο, η προσδοκία είναι:

$$E_B(s) = \int p_B(x) \log \frac{p_B(x)}{p_A(x)} dx = D(p_B \parallel p_A) > 0. \quad (2.10)$$

2.2.3.1.4 E (Energy Function)

Η συνάρτηση της ενέργειας υπολογίζεται από το άθροισμα του πλάτους στο τετράγωνο για το συγκεκριμένο εύρος συχνοτήτων:

$$E = \sum_k \{|X(k)|\}^2 \quad (2.11)$$

Όπου: E= συνάρτηση ενέργειας για το τωρινό παράθυρο.
k= δείκτης συχνότητας (για τον FFT πίνακα)
X(k) είναι ο k-νιστός δείκτης του FFT.

2.2.3.1.5 HFC (High Frequency Content Function)

Η λειτουργία για τη μέτρηση υψηλών συχνοτήτων αυθαίρετα οριστεί σε μια σταθμισμένη συνάρτηση ενέργειας, γραμμικά προκατειλημμένη προς τις υψηλότερες συχνότητες:

$$HFC = \sum_{k=2}^{N/2+1} \left\{ |X(k)|^2 \cdot k \right\} \quad (2.12)$$

Όπου HFC είναι η συνάρτηση περιεχομένου υψηλών συχνοτήτων για το τρέχον πλαίσιο, τα άλλα σύμβολα όπως ορίζονται παραπάνω.

2.2.3.1.6 MoT (Measurement of Transience)

Η προϋπόθεση για την ανίχνευση συνδυάζει τα αποτελέσματα από κάθε ζεύγος διαδοχικών πλαισίων έτσι:

$$MoT = \frac{HFC_r}{HFC_{r-1}} * \frac{HFC_r}{E_r} \quad (2.13)$$

όπου ο δείκτης r δηλώνει την τρέχουσα πλαίσιο (ίσων τελευταία δύο σε λειτουργία ανίχνευσης), ο δείκτης r-1 δηλώνει το προηγούμενο πλαίσιο.

2.2.3.1.7 CD (Complex Domain)

Πλάτος και τη φάση μπορεί να θεωρηθεί από κοινού αναζήτηση για αναχωρήσεις από σταθερή κατάσταση συμπεριφοράς από τον υπολογισμό του αναμενόμενου πλάτους και φάσης του ρεύματος bin X (n, k), με βάση τα προηγούμενα δύο κάδους X (n - 1, k) και X (n - 2, k). Το XT τιμή στόχο (n, k) υπολογίζεται υποθέτοντας σταθερό πλάτος και το ρυθμό της αλλαγής φάσης:

$$X_T(n, k) = |X(n-1, k)| e^{\psi(n-1, k) + \psi'(n-1, k)}$$

και ως εκ τούτου, μια σύνθετη CD λειτουργία ανίχνευσης έναρξης τομέα μπορεί να οριστεί ως το άθροισμα των απόλυτων αποκλίσεων από τις τιμές-στόχους:

$$CD(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |X(n, k) - X_T(n, k)|$$

2.2.3.1.8 RCD (Rectified Complex Domain)

Ένα πρόβλημα με τη μέθοδο της CD είναι ότι δεν κάνει διάκριση μεταξύ στις αυξήσεις και μειώσεις του πλάτους του σήματος. καθόσον είναι σημαντικό να γίνει διάκριση στις ενάρξεις από τα αντισταθμιστικά οφέλη, προτείνουμε τη χρήση ενός παρόμοια ιδέα με αυτή που χρησιμοποιείται στη συνάρτηση SF, όπου ανόρθωση μισού κύματος χρησιμοποιείται για να διατηρήσει μόνο τις αυξήσεις στη φασματική ενέργεια κάδους. Η ιδέα αυτή μπορεί εύκολα να ενσωματωθεί με τη μέθοδο CD, δίνοντας μια (ημι-κύματος) διορθωμένο σύνθετο πεδίο (RCD) λειτουργία ανίχνευσης εμφάνιση ως εξής:

$$RCD(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} RCD(n, k)$$

where

$$RCD(n, k) = \begin{cases} |X(n, k) - X_T(n, k)|, & \text{if } |X(n, k)| \geq \\ & |X(n-1, k)| \\ 0, & \text{otherwise} \end{cases}$$

2.2.3.1.9 PD(Phase Deviation)

$$X(n, k) = |X(n, k)| e^{j\psi(n, k)}$$

where $-\pi < \psi(n, k) \leq \pi$. Then the instantaneous frequency is given by the first difference $\psi'(n, k)$:

$$\psi'(n, k) = \psi(n, k) - \psi(n - 1, k)$$

$$PD(n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |\psi''(n, k)|$$

2.2.3.1.10 WPD (Weighted Phase Deviation)

Στο υπόλοιπο αυτής της ενότητας θα προτείνει βελτιώσεις στην έναρξη λειτουργιών ανίχνευσης που περιγράφονται στην βιβλιογραφία. Η πρώτη ιδέα αντιμετωπίζει το πρόβλημα ότι η λειτουργία PD "είναι ευαίσθητη σε ... ο θόρυβος που εισάγεται από τα συστατικά χωρίς σημαντική ενέργεια ". Δηλαδή, η συνάρτηση θεωρεί ότι όλοι οι bins-συχνοτήτων k είναι ίσοι, αν η ενέργεια του σήματος συμπυκνώνεται εντός των κάδων που περιέχει τα partials των σήμερα ηχεί τόνους. προτείνουμε στάθμιση των bins συχνότητα από το μέγεθός τους, δίνοντας μια νέα λειτουργία ανίχνευσης εκδήλωση που καλούμε ως σταθμισμένη φάση απόκλισης (WPD):

$$WPD(n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |X(n, k) \psi''(n, k)|$$

2.2.3.1.11 NWPD (Normalised weighted phase deviation)

Παρόμοια με τη λειτουργία CD, το ότι το μέγεθος και η φάση θεωρούνται από κοινού, αλλά με διαφορετικό τρόπο συνδυασμού. Μια άλλη επιλογή είναι να ορίζουν μια κανονικοποιημένη απόκλιση σταθμισμένη φάση (NWPD) λειτουργία, όταν το άθροισμα των βαρών υπολογιστεί από να δώσει μια σταθμισμένη μέση απόκλιση φάσης:

$$NWPD(n) = \frac{\sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |X(n, k) \psi''(n, k)|}{\sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |X(n, k)|}$$

2.2.4 Περιγραφή Correlation

- Autocorrelation

Η **αυτοσυσχέτιση (autocorrelation)** ενός σήματος είναι η συσχέτιση ενός σήματος με τον εαυτό του και δείχνει την ομοιότητα μεταξύ παρατηρήσεων ως μια συνάρτηση των χρονικών διαφορών μεταξύ τους. Χρησιμοποιείται στα μαθηματικά και την επεξεργασία σήματος για εύρεση επαναλαμβανόμενων προτύπων, όπως την εύρεση κάποιου περιοδικού σήματος που περιέχει θόρυβο και την εύρεση της θεμελιώδους συχνότητας ενός σήματος [18].

Δεδομένου ενός σήματος $x(t)$ η συνεχής αυτοσυσχέτιση του $R_{ff}(\tau)$, ορίζεται ως το συνεχές ολοκλήρωμα ετεροσυσχέτισης του $x(t)$ με τον εαυτό του, στη χρονική στιγμή τ :

$$R_{ff}(\tau) = (x(t) * \bar{x}(-\tau))(\tau) = \int_{-\infty}^{+\infty} x(t + \tau) \bar{x}(t) dt = \int_{-\infty}^{+\infty} x(t) \bar{x}(t - \tau) dt$$

όπου \bar{x} το συζυγές σήμα του x και ο τελεστής $*$ αναπαριστά συνέλιξη.

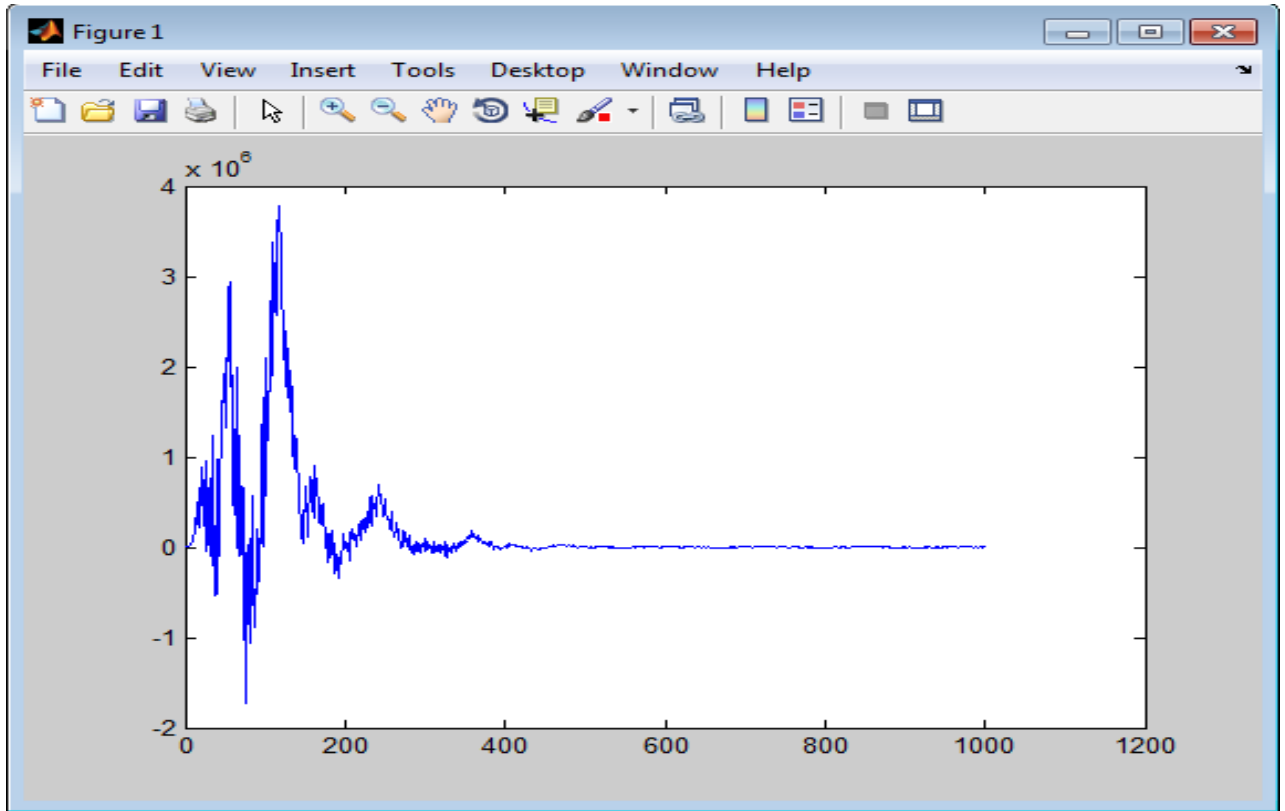
Για διακριτά σήματα ορίζεται:

$$R_{ff}(j) = \sum_n x_n \bar{x}_{n-j}$$

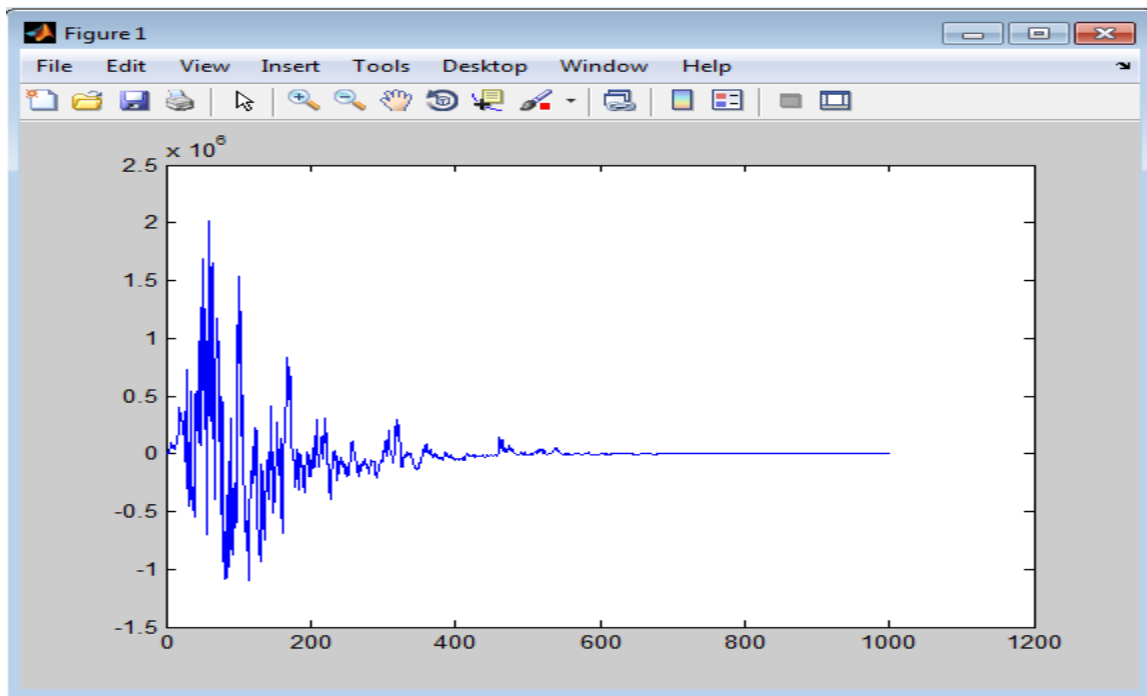
- **Cross-Correlation**

Η cross-correlation και οι περισσότερες άλλες διαχρονικές αναλύσεις, υποθέτουν ότι η συσχέτιση μεταξύ δύο μεταβλητών είναι σταθερή. Συνεπώς, ένα δείγμα από κάποια περιστασιακή μέτρηση, αναμένεται να αντιπροσωπευτεί από την σύνδεση μεταξύ των μεταβλητών, ανεξαρτήτως από την στιγμή εμφάνισης ή από τον αριθμό των περιπτώσεων στο δείγμα. Προτείνεται μία μέθοδος, για την ανάλυση της συσχέτισης μεταξύ δύο μεταβλητών όταν η υπόθεση της στασιμότητας δεν μπορεί να δικαιολογηθεί. Η μέθοδος οδηγεί σε εκτιμήσεις των δύο πιο δυνατών τιμών που εμφανίζεται κάποια κορυφή και της χρονικής καθυστέρησης, όταν πραγματοποιήθηκε η κορυφή σύνδεσης/διασταύρωσης για μια σειρά από αρχικές τιμές που παρήλθε ο χρόνος από την έναρξη του πειράματος.

Στις εικόνες 2.20 και 2.21 παρουσιάζεται η φάση του cross-correlation πάνω στο σήμα που περιγράφουμε από και στα προηγούμενα παραδείγματά μας.



Εικόνα 2.20 : CorrelationsCorrscollection-1-0-0-0-0-xcorr-Ellis-Ellis-Ellis-0(Masri)/Κομμάτι 3^ο



Εικόνα 2.21 : CorrelationsCorrscollection-1-1-40-0-0-0-xcorr-Ellis-Ellis-Ellis-0(Duxbury-)

2.2.5 Περιγραφή Peak-Picking

Αλγόριθμος Εύρεσης Κορυφών

Εφόσον έχει ολοκληρωθεί η μετεπεξεργασία της συνάρτησης Onset Detection Function για τον υπολογισμό του δυναμικού κατωφλίου, στη συνέχεια ξεκινά η διαδικασία Peak-Picking η οποία συνίσταται για την επιλογή των τοπικών μεγίστων που βρίσκονται πάνω από αυτό. Η ανίχνευση των τοπικών μεγίστων συνεπάγεται τη σύγκριση τουλάχιστον τριών διαδοχικών παρατηρήσεων, κάτι που απαιτεί γνώση τουλάχιστον μιας παρατήρησης μετά την κορυφή. Οι χρόνοι των onset τελικά ορίζονται ως κάθε τοπικό μέγιστο της συνάρτησης Peak-Picking:

$$\hat{D}[n] = D[n] = \tilde{\delta}[n]$$

Όπου $D[n]$ είναι οποιαδήποτε από τις συναρτήσεις onset det. που παρουσιάστηκαν παραπάνω.

3 Πειραματικό Μέρος-Υλοποίηση

3.1. Ανάπτυξη Λογισμικού-Υλοποίηση Αλγορίθμων

Για την Υλοποίηση των Αλγορίθμων του Πειραματικού Μέρους της παρούσας πτυχιακής Εργασίας, χρησιμοποιήσαμε το προγραμματιστικό περιβάλλον του Matlab. Συγκεκριμένα, οι εργασίες πραγματοποιήθηκαν στις εκδοχές του Matlab 2011a και 2011b. Αποτελεί ένα αριθμητικό υπολογιστικό περιβάλλον και μια 4ης γενιάς γλώσσα προγραμματισμού. Αναπτύχθηκε από τη Mathworks και η ονομασία του προέρχεται από τα αρχικά γράμματα των λέξεων MATrix LABoratory. Το MATLAB, είναι ένα εργαλείο που επιτρέπει τη γραφική σχεδίαση των λειτουργιών της και των δεδομένων της, την εφαρμογή αλγορίθμων, τη δημιουργία διεπαφών χρήστη και τη διασύνδεση με προγράμματα γραμμένα σε άλλες γλώσσες [Andrew Knight, 2000]. Ακόμη το MATLAB είναι εφοδιασμένο με πολλές επιλογές για γραφικά (δηλ. την κατασκευή γραφικών παραστάσεων) και προγράμματα γραμμένα στη δική του γλώσσα προγραμματισμού, για την επίλυση άλλων προβλημάτων. Το MATLAB προσφέρει στο χρήστη τη δυνατότητα να κατασκευάσει δικές του γραφικές διεπιφάνειες, (GUI). Η

χρησιμότητα της λειτουργίας αυτής είναι μεγάλη, επειδή τα προγράμματα - εφαρμογές που περιέχουν γραφική διεπιφάνεια γίνονται πιο φιλικά προς τον τελικό χρήστη [Βαρσάμης, 2010].

Το MATLAB είναι ένα διαδραστικό (interactive) πρόγραμμα για αριθμητικούς υπολογισμούς και οπτικοποίηση δεδομένων. Στις νεότερες εκδοχές, περιλαμβάνονται εργαλείοι όπως μπορεί να επιτραπούν συμβολικοί υπολογισμοί [Γ. Γεωργίου & Χ. Ξενοφώντος, 2007].

3.2. Βάση/Πειραματικά Δεδομένα

Επιλογή Dataset

Οι λόγοι για τους οποίους καταλήξαμε να εργαστούμε στο μουσικό dataset του Magnatagatune, είναι οι εξής:

1. Το dataset περιέχει μεγάλο αριθμό μουσικών κομματιών (5405), τα οποία είναι επιλεγμένα από 230 διαφορετικούς καλλιτέχνες και από πολλά μουσικά είδη.
2. Περιλαμβάνει σχολιασμούς κομματιών με την μορφή tags (188 μοναδικά tags) από χρήστες. Τα στοιχεία αυτά συλλέχτηκαν από το διαδικτυακό παιχνίδι του TagATune που δημιουργήθηκε από τον Edith Law.
3. Στο dataset του Magnatagatune, περιλαμβάνεται και η λεπτομερής ανάλυση της δομής ενός ηχητικού σήματος καθώς επίσης και η ανάλυση του μουσικού περιεχομένου των κομματιών, σε μορφή ενός xml αρχείου για κάθε ένα από τα μουσικά αποσπάσματα της συλλογής, η οποία προκύπτει από την υπηρεσία Echo Nest.
4. Τέλος, στο dataset υπάρχουν μουσικά αποσπάσματα των 29 δευτερολέπτων σε μορφή mp3 (mono, 16kHz, 32kbps). Κάθε ένα κομμάτι από αυτά προέρχεται από δημοσιευμένα μουσικά κομμάτια της υπηρεσίας Magnatune.com.

Στο dataset του Magnatagatune, περιέχονται συνολικά 25.863 αποσπάσματα των 29 sec. τα οποία προέρχονται από 5.405 διαφορετικά μουσικά κομμάτια. Για το καθένα μουσικό κομμάτι, παρέχεται και ο ανάλογος αριθμός αποσπασμάτων ανάλογα με τη διάρκειά του.

Το Magnatune, είναι μια διαδικτυακή δισκογραφική εταιρία η οποία ιδρύθηκε το 2003 από τον John Buckman (Καλιφόρνια) με κύριο στόχο του να ευνοήσει τους καλλιτέχνες να παρουσιάσουν το έργο τους στο κοινό παραβλέποντας τις νομικές και οικονομικές δυσκολίες που προκύπτουν κατά την έκδοση ενός μουσικού έργου από μια κανονική δισκογραφική εταιρία ή κατά τη διανομή του προσωπικού τους δίσκου από μουσικούς εμπόρους. Διαπιστώνοντας ότι συχνά οι καλλιτέχνες κερδίζουν ελάχιστα χρήματα από την πώληση των δίσκων τους και στη συνέχεια εγκλωβίζονται νομικά από τις δισκογραφικές τους εταιρίες, δημιούργησε το Magnatune που λειτουργεί με άδεια

Creative Commons (έναντι του συστήματος 'all rights reserved' που ακολουθούν όλες οι κλασσικές δισκογραφικές εταιρίες) και υποστηρίζει έτσι ότι χρεώνει στους καλλιτέχνες το 50% της τελικής τιμής των πωλήσεων των έργων τους όπως επίσης και το 50% του ποσού που επακολουθεί από τις συνδρομές των χρηστών.
(Αναφέρεται ότι το 25% των χρημάτων κάθε συνδρομής των on-line ακροατών μεταφέρεται στους καλλιτέχνες καθώς και 25% μοιράζεται στους καλλιτέχνες των οποίων ο χρήστης κατεβάζει τα μουσικά του κομμάτια.)

Στο Dataset του MagnaTagATune υπάρχει από ένα xml/μουσικό κομμάτι.

Το xml αυτό περιλαμβάνει περιγραφές μουσικού περιεχομένου, οι οποίες έχουν προκύψει από την ανάλυση του Echonest.

Το Echonest Analyzer παρέχει για το κάθε κομμάτι ένα μορφοποιημένο αρχείο με βάση το πρότυπο JSON. (JavaScript Object Notation).

Σε κάθε τέτοιο αρχείο περιέχονται οι παρακάτω πληροφορίες:

- ❖ **Μεταδεδομένα (metadata)** : πληροφορίες σχετικές με τη διαδικασία της ανάλυσης και πληροφορίες για το κομμάτι
- ❖ **Δεδομένα κομματιού (track data)**
- ❖ **ένδειξη ρυθμού (time signature)**: μια συνολικά εκτιμώμενη ένδειξη ρυθμού για το κομμάτι. Αποτελεί μια παγκόσμια σύμβαση για τον καθορισμό του αριθμού των beat που περιέχονται σε κάθε μέτρο (bar)
- ❖ **Κλίμακα (key)**: η συνολικά εκτιμώμενη κλίμακα του κομματιού.
- ❖ **mode** : Προσδιορίζει το modality (μείζον ή ελάσσων) του κομματιού, το είδος της κλίμακας από την οποία προήλθε το μελωδικό περιεχόμενο
- ❖ **Ρυθμός (tempo)** : μια συνολική εκτίμηση του ρυθμού του κομματιού σε beats ανά λεπτό (bpm) Αποτελεί τη ταχύτητα ενός κομματιού και προκύπτει κατευθείαν από τη μέση διάρκεια του beat
- ❖ **Ένταση (loudness)** : την συνολική ένταση του κομματιού σε dB
- ❖ **Διάρκεια (duration)** : η ακριβής διάρκεια του κομματιού όπως καταγράφεται από τον αποκωδικοποιητή ήχου
- ❖ **Τέλος του fade in (end of fade in)** : σε δευτερόλεπτα
- ❖ **Αρχή του fade out (start of fade out)** : σε δευτερόλεπτα
- ❖ **Διαδοχικά Δεδομένα (sequenced data)**: ο αναλυτής σπάει το ακουστικό κομμάτι σε μουσικά στοιχεία που εμφανίζονται σε αλληλουχίες στον χρόνο. Από τη μικρότερη στη μεγαλύτερη αυτές είναι:
 - ➔ **segments** : Ένα σύνολο από μουσικές οντότητες (συνήθως κάτω από ένα δευτερόλεπτο) η κάθε μια σχετικά ομοιόμορφη σε ηχόχρωμα και αρμονία.
 - ➔ **loudness_start** : υποδεικνύει το επίπεδο έντασης κατά την έναρξη του Segment
 - ➔ **loudness_max_time** : το αρχικό σημείο μέσα στο segment όπου εμφανίζεται η μέγιστη ένταση
 - ➔ **loudness_max** : η τιμή της μέγιστης έντασης
 - ➔ **Συντελεστές Τονικότητας (pitch coefficients)** : διάνυσμα 12 διαστάσεων με κάθε διάσταση να αντιστοιχεί σε μια από τις 12 τονικές

κλάσεις με τιμές [0,1]

- **Συντελεστές Ηχοχρώματος (timbral coefficients)** : διάνυσμα 12 διαστάσεων. (Αναλύεται περισσότερο στη συνέχεια)
- **Αντιληπτική Εμφάνιση (perceptual onset)**
- **Διάρκεια (duration)** : σε δευτερόλεπτα

→ **tatums** : μια λίστα από δείκτες tatum σε δευτερόλεπτα. Τα tatums αναπαριστούν τη χαμηλότερη ακολουθία παλμών που ο ακροατής συμπεραίνει ενστικτωδώς από τον χρονισμό των αντιλαμβανόμενων μουσικών γεγονότων.

→ **beats** : μια λίστα από δείκτες beat σε δευτερόλεπτα. Ένα beat αποτελεί τη βασική μονάδα χρόνου σε ένα μουσικό κομμάτι. Τα beat είναι συνήθως πολλαπλάσια των tatums

→ **Μέτρα (bars)** : μια λίστα από δείκτες μέτρων σε δευτερόλεπτα. Ένα μέτρο αποτελεί ένα κομμάτι του χρόνου ορισμένα σαν ένα δεδομένο αριθμό beat

→ **Ενότητες (sections)** : ένα σύνολο από δείκτες ενοτήτων σε δευτερόλεπτα. Οι ενότητες ορίζονται από μεγάλες μεταβολές στο ρυθμό ή το ηχόχρωμα.

3.3. Αλγόριθμοι

Αρχικά , δημιουργήσαμε parsers για να μπορούμε να αντιστοιχούμε τις πληροφορίες στα μουσικά αρχεία (mp3) και (ή) στα συνοδευτικά αρχεία τους (xml) .

1. **pathsXML.m [xmlNew.mat]**

Δημιούργησα την pathsXML συνάρτηση που παίρνει ως όρισμα το όνομα του φακέλου που περιέχει όλα τα xml αρχεία της συλλογής Magnatagatune που χρησιμοποιήσαμε για το πρακτικό μέρος την πτυχιακής εργασίας. Εφόσον, έχουμε πλέον πρόσβαση στα αρχεία αυτά μπορούμε να πάρουμε όποιο πεδίο μας ενδιαφέρει από τα περιεχόμενα των xml.

(Τα περιεχόμενα των xml, περιγράφηκαν παραπάνω).

Στην παρακάτω εικόνα διακρίνουμε ένα μέρος από την δομή των xml αρχείων.

```

<?xml version="1.0" encoding="UTF-8"?>

<analysis decoder="Quicktime" version="0x7608000">

  <track duration="29.12331" endOfFadeIn="2.10721" startOfFadeOut="29.12331" loudness="-17.387"
tempo="43.523" tempoConfidence="0.421" timeSignature="4" timeSignatureConfidence="0.721" key="9"
keyConfidence="0.575" mode="1" modeConfidence="0.721">

  <meter>

    <bar conf="-1">

      <beat conf="-1">

        <tatum conf="0">0.06104</tatum>

        <tatum conf="0">0.35361</tatum>

        <tatum conf="0">0.65130</tatum>

      </beat>

```

Εικόνα 3.1: Μέρος από την δομή ενός τυχαίου από τη βάση xml αρχείου

Από το αρχείο αυτά μας ενδιαφέρουν οι τιμές που υπάρχουν στα πεδία tempo και tempoConfidence τα οποία τα παίρνουμε με τις αντίστοιχες εντολές:

- tempo=str2num(readXML.getElementsByTagName('analysis').item(0).getElementsByTagName('track').item(0).getAttribute('tempo'));
- tempoConfidence=str2num(readXML.getElementsByTagName('analysis').item(0).getElementsByTagName('track').item(0).getAttribute('tempoConfidence'));

Τα δύο αυτά πεδία τα καταχωρούμε σε έναν πίνακα με όνομα Tempo.

Επίσης, παίρνουμε την τιμή που υπάρχει στο πεδίο meter, που αντιστοιχεί στην τιμή των beats.

Με τα δεδομένα που συλλέξαμε, δημιουργούμε ένα .mat αρχείο με όνομα **xmlNew.mat** που έχει αποθηκευμένα για τα 25.837 μουσικά κομμάτια τα:

- paths // που είναι αποθηκευμένα τα xml αρχεία
- Beats // που είναι η τιμή που πήραμε από το xml αρχείο (meter)
- Tempo // που περιέχει τις 2 τιμές (tempo, tempoConfidence)

2. pathsMp3.m [Mp3New.mat]

Για να μπορούμε να έχουμε πρόσβαση και στα Mp3 κομμάτια, δημιουργήσαμε μια αντίστοιχη συνάρτηση όπως κάναμε για τα xml, που δέχεται ως Input το όνομα του φακέλου που περιέχει όλα τα mp3 κομμάτια που χρησιμοποιήσαμε.

Στην συνέχεια φορτώνουμε τον ήδη έτοιμο πίνακα xmlNew.mat επειδή μπορούμε να πάρουμε από εκεί τα paths και τα ονόματα των μουσικών αρχείων, αφαιρώντας από το τέλος των ονομάτων τους την κατάληξη .xml.

Αυτό το πετυχαίνουμε με την εντολή :

```
mp3path=strcat(nameOfFile,'\',K.paths(1,j).pathXML(1:length(K.paths(1,j).pathXML)-4));
```

Από την συνάρτηση αυτή δημιουργούμε ένα .mat αρχείο (πίνακας) με όνομα **Mp3New.mat** , στον οποίο βρίσκονται τα paths των mp3 μουσικών κομματιών.

3. mesosBeats.m [Mesos.mat]

Δημιουργήσαμε την συνάρτηση mesosBeats που παίρνει ως όρισμα το xmlNew.mat και φτιάχνω από τα περιεχόμενα του αρχείου αυτού, τον μέσο όρο των Beats (xml).

Στη συνέχεια αποθηκεύω τον μέσο όρο των beats και το tempo σε ένα νέο αρχείο xml, με όνομα: **Mesos.mat** , με σκοπό να τα έχω έτοιμα όταν θα χρειαστεί να τα χρησιμοποιήσω παρακάτω στον κώδικά μου.

4. spect.m [Spectrograms]

Η συνάρτηση spect, δημιουργήθηκε με σκοπό να πάρουμε τα Spectrograms των μουσικών κομματιών, πάνω στα οποία θα γίνουν όλες οι ενέργειες για να εξάγουμε το BeatPeriod από την κάθε Onset Detection Function και από τις διάφορες παραμέτρους τους.

Δέχεται δύο ορίσματα, τα οποία είναι το window (μέγεθος παραθύρου) και overlap(επικάλυψη).

Φορτώνουμε από το ήδη έτοιμο mat αρχείο (Mp3New.mat) τα mp3 κομμάτια και στην συνέχεια δημιουργούμε τα spectrograms όλων των μουσικών κομματιών με την έτοιμη εντολή του Matlab:

```
[sp,F,T]=spectrogram(Song>window,overlap>window,FS);
```

Matlab: help spectrogram

`[S,F,T] = spectrogram(x,window,noverlap,F)` uses a vector `F` of frequencies in Hz. `F` must be a vector with at least two elements. This case computes the spectrogram at the frequencies in `F` using the Goertzel algorithm. The specified frequencies are rounded to the nearest DFT bin commensurate with the signal's resolution. In all other syntax cases where `nfft` or a default for `nfft` is used, the short-time Fourier transform is used. The `F` vector returned is a vector of the rounded frequencies. `T` is a vector of times at which the spectrogram is computed. The length

of \mathbb{F} is equal to the number of rows of \mathbb{S} . The length of \mathbb{T} is equal to k , as defined above and each value corresponds to the center of each segment.

Η συνάρτηση `spect` δίνει σαν `output` σε ξεχωριστά `mat` αρχεία τα `spectrograms` των 25.837 μουσικών κομματιών και αποθηκεύονται αυτόματα σε ένα φάκελο με όνομα `Spectrograms`.

5. `tss_spectral_difference.m` (έτοιμος κώδικας που χρησιμοποιήσαμε)
Στην συνάρτηση αυτή, υλοποιούνται οι τύποι των `Onset Det. Function`, που δημιουργούνται από την διαφορά φάσματος:

- **Masri**

```
%SD_Masri
SD_Masri = spectral_difference;
SD_Masri=SD_Masri(1:(size(SD_Masri,1)-2),:);
SD_Masri = 2*sum(abs(SD_Masri),1); % Only positive frequencies considered in K
SD_Masri(isnan(SD_Masri)) = 0;
SD_Masri(isinf(SD_Masri)) = 0;
```

- **Duxbury**

```
% Spectral difference: Duxbury uses the L2-norm on the rectified difference.
SD_Duxbury=spectral_difference;
SD_Duxbury=SD_Duxbury(1:(size(SD_Duxbury,1)-2),:);
SD_Duxbury = 2*sum(SD_Duxbury.^2,1);
SD_Duxbury(isnan(SD_Duxbury)) = 0;
SD_Duxbury(isinf(SD_Duxbury)) = 0;
```

- **Kullback**

```
% Kullback-Leibler distance measure
Kullback_Leibler_k = log2(current_spectral_magnitude./previous_spectral_magnitude);
Kullback_Leibler_k(Kullback_Leibler_k < 0) = 0;
Kullback_Leibler_k=Kullback_Leibler_k(1:(size(Kullback_Leibler_k,1)-2),:);
SD_Kullback_Leibler = 2*sum(Kullback_Leibler_k,1); % rectification
SD_Kullback_Leibler(isnan(SD_Kullback_Leibler)) = 0;
SD_Kullback_Leibler(isinf(SD_Kullback_Leibler)) = 0;
```

Στην συνέχεια, προσθέσαμε και άλλες τις `Onset Det. Function` που υλοποιούνται με την διαφορά φάσματος και είναι οι εξής:

- **E**

```
%SD_E
SD_E=spectral_magnitude.^2;
SD_E=SD_E(1:(size(SD_E,1)-2),:);
SD_E=sum(SD_E);
SD_E(isnan(SD_E)) = 0;
```

```
SD_E(isinf(SD_E)) = 0;
```

- **HFC**

```
%SD_HFC  
k=[size(spectral_magnitude,1):-1:1];  
k=k';  
k= repmat(k,1,size(spectral_magnitude,2));  
SD_HFC=(spectral_magnitude.^2).*k;  
SD_HFC=SD_HFC(1:(size(SD_HFC,1)-2),:);  
SD_HFC=sum(SD_HFC);  
SD_HFC(isnan(SD_HFC)) = 0;  
SD_HFC(isinf(SD_HFC)) = 0;
```

- **MoT**

```
%SD_MoT  
SD_HFC_current=SD_HFC(:,2:end);  
SD_HFC_previous=SD_HFC(:,1:end-1);  
SD_E_current=SD_E(:,2:end);  
SD_MoT=(SD_HFC_current./SD_HFC_previous).*(SD_HFC_current./SD_E_current);  
SD_MoT(isnan(SD_MoT)) = 0;  
SD_MoT(isinf(SD_MoT)) = 0;
```

6. **tss_phase_deviation.m** (έτοιμος κώδικας που χρησιμοποιήσαμε)

Στην συνάρτηση αυτή, υλοποιούνταν οι τύποι των Onset Det. Function, που δημιουργούνται από την διαφορά φάσης:

```
% Phase deviation calculation  
  
pd = 1/N_fft*sum(abs(instant_freq_change),1);  
weighted_pd = 1/N_fft*sum(abs(spectral_magnitude.*instant_freq_change),1);  
normalized_weighted_pd = sum(abs(spectral_magnitude.*instant_freq_change),1)...  
./sum(abs(spectral_magnitude),1);  
normalized_weighted_pd(isnan(normalized_weighted_pd)) = 0;  
normalized_weighted_pd(isinf(normalized_weighted_pd)) = 0;
```

7. **tss_complex_domain.m** (έτοιμος κώδικας που χρησιμοποιήσαμε)

Στην συνάρτηση αυτή, υλοποιούνται οι τύποι των Onset Det. Function, που δημιουργούνται από σύνθετα πεδία:

```
% Complex Domain detection function
cd = sum(abs(X-Xt),1);
```

```
% Rectified complex domain: only increases in the amplitude of the signal
% are considered
rcd_nk = abs(X-Xt);
rcd_nk(spectral_magnitude_0 < spectral_magnitude_1) = 0;
% decreases in amplitude set to 0
rectified_cd = sum(rcd_nk,1);
```

8. spectral_based_df.m (έτοιμος κώδικας που χρησιμοποιήσαμε)

Υπολογίζει όλες τις λειτουργίες ανίχνευσης φάσματος με βάση τις παραμέτρους τους παρακάτω:

- σύνθετο πεδίο (cd), το διορθωμένο σύνθετο πεδίο (RCD),
- φασματικές διαφορές (Masri, duxburi, Kullback,E,MoT,HFC),
- η απόκλιση φάσης (pd), απόκλιση σταθμισμένη φάση (wprd), ομαλοποιημένη και σταθμισμένη απόκλιση φάσης (nwprd)

9. BatchAnalyzeSpect.m (Script)

Δημιουργούμε συνδυασμούς , με τις παραμέτρους που υλοποιούνται στα Spectrograms των μουσικών κομματιών.

```
MEL
if isfield(Options,'MEL') && strcmp(Options.MEL,'1')

    m1mx = fft2melmx(swin,SR,Options.NMEL);
    D=m1mx(:,1:(swin/2+1))*abs(D);
end
if isfield(Options,'MEL') && strcmp(Options.MEL,'2') %Constant-Q
    TransfQ=logfmapQTransform(size(D,1),2,Options.NMEL);
    D=TransfQ*D; % D=SPQ
End
```

A popular formula to convert f hertz into m mel is:^[2]

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right)$$

dB

```
if isfield(Options,'dB') && strcmp(num2str(Options.dB),'1')
    D = 20*log10(max(1e-10,abs(D)));

    % Only look at the top 80 Options.dB
    D = max(D, max(max(D))-80);
end

if isfield(Options,'dB') && strcmp(num2str(Options.dB),'2')
    D=log(1 + 1000* abs(D));
end
```

4 Αποτελέσματα

Masri(0-0-0)

	/2	*1	*2	*4	mean	var
0-0-xcorr-Ellis-Ellis-Ellis-1	3205	16392	6236	4	0,168959	0,010458
0-0-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
0-0-xcorr-Ellis-SimpleMax-Ellis-1	421	2571	16174	6671	0,165855	0,009839
0-0-xcorr-cutOff-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-0-xcorr-cutOff-Ellis-mean-1	2609	15445	7772	11	0,167825	0,010357
0-1-xcorr-Ellis-Ellis-Ellis-1	353	1885	14445	9154	0,165456	0,009664
0-1-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
0-1-xcorr-cutOff-Ellis-Ellis-1	472	3153	16876	5336	0,168959	0,010458

Masri(1-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2797	15936	7097	7	0,168045	0,010453
0-0-xcorr-Ellis-Ellis-mean-1	4461	17457	3919	0	0,172207	0,010202
0-0-xcorr-Ellis-SimpleMax-Ellis-1	360	1948	14733	8796	0,165401	0,009674
0-0-xcorr-cutOff-Ellis-Ellis-1	485	3270	16924	5158	0,16803	0,009348
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	7591	16732	1514	0	0,17871	0,011986
0-1-xcorr-Ellis-Ellis-mean-1	10934	14440	463	0	0,183256	0,012983
0-1-xcorr-cutOff-Ellis-Ellis-1	13604	12148	85	0	0,184226	0,014037

Masri(1-1-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-0-xcorr-Ellis-Ellis-mean-1	3417	16698	5718	4	0,169804	0,01033
0-0-xcorr-Ellis-SimpleMax-Ellis-1	353	1885	14445	9154	0,165456	0,009664
0-0-xcorr-cutOff-Ellis-Ellis-1	364	2000	14895	8578	0,165312	0,009694
0-0-xcorr-cutOff-Ellis-mean-1	10934	14440	463	0	0,183256	0,012983
0-1-xcorr-Ellis-Ellis-Ellis-1	4768	17516	3553	0	0,172583	0,010364
0-1-xcorr-Ellis-Ellis-mean-1	11976	13671	190	0	0,18359	0,013121
0-1-xcorr-cutOff-Ellis-Ellis-1	10934	14440	463	0	0,183256	0,012983

Masri(1-2-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-SimpleMax-Ellis-1	6799	16914	2124	0	0,177905	0,01177
0-0-xcorr-cutOff-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-cutOff-Ellis-mean-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

Masri(2-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2364	14895	8565	13	0,167134	0,01018
0-0-xcorr-Ellis-Ellis-mean-1	2992	16174	6665	6	0,168242	0,010533
0-0-xcorr-Ellis-SimpleMax-Ellis-1	347	1779	14178	9533	0,165458	0,009652
0-0-xcorr-cutOff-Ellis-Ellis-1	333	1638	13729	10137	0,165655	0,009707
0-0-xcorr-cutOff-Ellis-mean-1	2364	14895	8565	13	0,167134	0,01018
0-1-xcorr-Ellis-Ellis-Ellis-1	5852	17371	2614	0	0,175466	0,010958
0-1-xcorr-Ellis-Ellis-mean-1	10934	14440	463	0	0,183256	0,012983
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

Masri(2-1-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-0-xcorr-Ellis-Ellis-mean-1	2609	15445	7772	11	0,167825	0,010357
0-0-xcorr-Ellis-SimpleMax-Ellis-1	353	1885	14445	9154	0,165456	0,009664
0-0-xcorr-cutOff-Ellis-Ellis-1	347	1779	14178	9533	0,165458	0,009652
0-0-xcorr-cutOff-Ellis-mean-1	2364	14895	8565	13	0,167134	0,01018
0-1-xcorr-Ellis-Ellis-Ellis-1	353	1885	14445	9154	0,165456	0,009664
0-1-xcorr-Ellis-Ellis-mean-1	3417	16698	5718	4	0,169804	0,01033
0-1-xcorr-cutOff-Ellis-Ellis-1	347	1779	14178	9533	0,165458	0,009652

Masri(2-2-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	21918	3919	0	0	0,218536	0,023018
0-0-xcorr-Ellis-Ellis-mean-1	16304	9509	24	0	0,188248	0,016534
0-0-xcorr-Ellis-SimpleMax-Ellis-1	6799	16914	2124	0	0,177905	0,01177
0-0-xcorr-cutOff-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-cutOff-Ellis-mean-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	16304	9509	24	0	0,188248	0,016534
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

Duxbury(0-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	3625	16876	5335	1	0,170651	0,010268
0-0-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
0-0-xcorr-Ellis-SimpleMax-Ellis-1	3417	16698	5718	4	0,169804	0,01033
0-0-xcorr-cutOff-Ellis-Ellis-1	527	3652	17412	4246	0,168441	0,009124
0-0-xcorr-cutOff-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-1-xcorr-Ellis-Ellis-Ellis-1	3625	16876	5335	1	0,170651	0,010268
0-1-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
0-1-xcorr-cutOff-Ellis-Ellis-1	527	3652	17412	4246	0,168441	0,009124

Duxbury (1-0-0)						
	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	8745	16036	1056	0	0,180185	0,012331
0-0-xcorr-Ellis-Ellis-mean-1	5460	17484	2893	0	0,174457	0,010663
0-0-xcorr-Ellis-SimpleMax-Ellis-1	403	2394	15936	7104	0,165806	0,00982
0-0-xcorr-cutOff-Ellis-Ellis-1	308	1234	11893	12402	0,168744	0,00935
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	7591	16732	1514	0	0,17871	0,011986
0-1-xcorr-Ellis-Ellis-mean-1	9466	15595	776	0	0,180959	0,012587
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

Duxbury (1-1-40)						
	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	4179	17412	4246	0	0,171825	0,010178
0-0-xcorr-Ellis-Ellis-mean-1	4179	17412	4246	0	0,171825	0,010178
0-0-xcorr-Ellis-SimpleMax-Ellis-1	5090	17509	3238	0	0,173497	0,01046
0-0-xcorr-cutOff-Ellis-Ellis-1	11976	13671	190	0	0,18359	0,013121
0-0-xcorr-cutOff-Ellis-mean-1	11976	13671	190	0	0,18359	0,013121
0-1-xcorr-Ellis-Ellis-Ellis-1	2364	14895	8565	13	0,167134	0,01018
0-1-xcorr-Ellis-Ellis-mean-1	10163	15076	598	0	0,181975	0,01278
0-1-xcorr-cutOff-Ellis-Ellis-1	10934	14440	463	0	0,183256	0,012983

Duxbury (1-2-40)						
	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	1587	12017	12148	85	0,168991	0,009425
0-0-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-SimpleMax-Ellis-1	1498	11592	12639	108	0,169571	0,009371
0-0-xcorr-cutOff-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-cutOff-Ellis-mean-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

Duxbury (2-0-0)						
	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2364	14895	8565	13	0,167134	0,01018
0-0-xcorr-Ellis-Ellis-mean-1	4461	17457	3919	0	0,172207	0,010202
0-0-xcorr-Ellis-SimpleMax-Ellis-1	342	1722	14004	9769	0,165527	0,009652
0-0-xcorr-cutOff-Ellis-Ellis-1	347	1779	14178	9533	0,165458	0,009652
0-0-xcorr-cutOff-Ellis-mean-1	2364	14895	8565	13	0,167134	0,01018
0-1-xcorr-Ellis-Ellis-Ellis-1	2797	15936	7097	7	0,168045	0,010453
0-1-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-1-xcorr-cutOff-Ellis-Ellis-1	1077	8733	15326	701	0,171583	0,009844

Duxbury (2-1-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2486	15129	8211	11	0,167388	0,010251
0-0-xcorr-Ellis-Ellis-mean-1	2609	15445	7772	11	0,167825	0,010357
0-0-xcorr-Ellis-SimpleMax-Ellis-1	376	2110	15129	8222	0,165429	0,00972
0-0-xcorr-cutOff-Ellis-Ellis-1	2486	15129	8211	11	0,167388	0,010251
0-0-xcorr-cutOff-Ellis-mean-1	2486	15129	8211	11	0,167388	0,010251
0-1-xcorr-Ellis-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-1-xcorr-Ellis-Ellis-mean-1	2609	15445	7772	11	0,167825	0,010357
0-1-xcorr-cutOff-Ellis-Ellis-1	2364	14895	8565	13	0,167134	0,01018

Duxbury (2-2-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	21918	3919	0	0	0,218536	0,023018
0-0-xcorr-Ellis-Ellis-mean-1	16304	9509	24	0	0,188248	0,016534
0-0-xcorr-Ellis-SimpleMax-Ellis-1	4179	17412	4246	0	0,171825	0,010178
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	16304	9509	24	0	0,188248	0,016534
0-1-xcorr-Ellis-Ellis-mean-1	10163	15076	598	0	0,181975	0,01278
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

Kullback (0-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2364	14895	8565	13	0,167134	0,01018
0-0-xcorr-Ellis-Ellis-mean-1	19166	6665	6	0	0,198722	0,019623
0-0-xcorr-Ellis-SimpleMax-Ellis-1	2797	15936	7097	7	0,168045	0,010453
0-0-xcorr-cutOff-Ellis-Ellis-1	2126	14178	9509	24	0,167021	0,010048
0-0-xcorr-cutOff-Ellis-mean-1	5852	17371	2614	0	0,175466	0,010958
0-1-xcorr-Ellis-Ellis-Ellis-1	13604	12148	85	0	0,184226	0,014037
0-1-xcorr-Ellis-Ellis-mean-1	19166	6665	6	0	0,198722	0,019623
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

Kullback (1-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-0-xcorr-Ellis-Ellis-mean-1	3417	16698	5718	4	0,169804	0,01033
0-0-xcorr-Ellis-SimpleMax-Ellis-1	392	2323	15793	7329	0,165818	0,009833
0-0-xcorr-cutOff-Ellis-Ellis-1	485	3270	16924	5158	0,16803	0,009348
0-0-xcorr-cutOff-Ellis-mean-1	2609	15445	7772	11	0,167825	0,010357
0-1-xcorr-Ellis-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-1-xcorr-Ellis-Ellis-mean-1	6799	16914	2124	0	0,177905	0,01177
0-1-xcorr-cutOff-Ellis-Ellis-1	2364	14895	8565	13	0,167134	0,01018

Kullback (1-1-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-0-xcorr-Ellis-Ellis-mean-1	3417	16698	5718	4	0,169804	0,01033
0-0-xcorr-Ellis-SimpleMax-Ellis-1	353	1885	14445	9154	0,165456	0,009664
0-0-xcorr-cutOff-Ellis-Ellis-1	2238	14445	9137	17	0,167147	0,010104
0-0-xcorr-cutOff-Ellis-mean-1	6799	16914	2124	0	0,177905	0,01177
0-1-xcorr-Ellis-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-1-xcorr-Ellis-Ellis-mean-1	2609	15445	7772	11	0,167825	0,010357
0-1-xcorr-cutOff-Ellis-Ellis-1	2238	14445	9137	17	0,167147	0,010104

Kullback (1-2-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-SimpleMax-Ellis-1	16304	9509	24	0	0,188248	0,016534
0-0-xcorr-cutOff-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-cutOff-Ellis-mean-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-Ellis-1	10163	15076	598	0	0,181975	0,01278
0-1-xcorr-Ellis-Ellis-mean-1	10163	15076	598	0	0,181975	0,01278
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

Kullback (2-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	19166	6665	6	0	0,198722	0,019623
0-0-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-SimpleMax-Ellis-1	5460	17484	2893	0	0,174457	0,010663
0-0-xcorr-cutOff-Ellis-Ellis-1	10934	14440	463	0	0,183256	0,012983
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	16304	9509	24	0	0,188248	0,016534
0-1-xcorr-Ellis-Ellis-mean-1	19166	6665	6	0	0,198722	0,019623
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

Kullback (2-1-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	347	1779	14178	9533	0,165458	0,009652
0-0-xcorr-Ellis-Ellis-mean-1	2126	14178	9509	24	0,167021	0,010048
0-0-xcorr-Ellis-SimpleMax-Ellis-1	342	1722	14004	9769	0,165527	0,009652
0-0-xcorr-cutOff-Ellis-Ellis-1	2019	13842	9945	31	0,167012	0,010019
0-0-xcorr-cutOff-Ellis-mean-1	2019	13842	9945	31	0,167012	0,010019
0-1-xcorr-Ellis-Ellis-Ellis-1	2364	14895	8565	13	0,167134	0,01018
0-1-xcorr-Ellis-Ellis-mean-1	19166	6665	6	0	0,198722	0,019623
0-1-xcorr-cutOff-Ellis-Ellis-1	6799	16914	2124	0	0,177905	0,01177

Kullback (2-2-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	18054	7772	11	0	0,194701	0,018472
0-0-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-SimpleMax-Ellis-1	2486	15129	8211	11	0,167388	0,010251
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	16304	9509	24	0	0,188248	0,016534
0-1-xcorr-Ellis-Ellis-mean-1	18054	7772	11	0	0,194701	0,018472
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

E (0-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	24598	1239	0	0	0,266654	0,031436
0-0-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-SimpleMax-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	360	1948	14733	8796	0,165401	0,009674
0-1-xcorr-Ellis-Ellis-mean-1	4179	17412	4246	0	0,171825	0,010178
0-1-xcorr-cutOff-Ellis-Ellis-1	308	1279	12017	12233	0,168328	0,009343

E (1-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	17259	8565	13	0	0,190978	0,01745
0-0-xcorr-Ellis-Ellis-mean-1	5460	17484	2893	0	0,174457	0,010663
0-0-xcorr-Ellis-SimpleMax-Ellis-1	4179	17412	4246	0	0,171825	0,010178
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

E (1-1-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-Ellis-Ellis-mean-1	6303	17225	2309	0	0,176735	0,011301
0-0-xcorr-Ellis-SimpleMax-Ellis-1	4179	17412	4246	0	0,171825	0,010178
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	5090	17509	3238	0	0,173497	0,01046
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

E (1-2-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-SimpleMax-Ellis-1	4461	17457	3919	0	0,172207	0,010202
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	6303	17225	2309	0	0,176735	0,011301
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

E (2-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-0-xcorr-Ellis-Ellis-mean-1	3899	17194	4744	0	0,171432	0,010304
0-0-xcorr-Ellis-SimpleMax-Ellis-1	376	2110	15129	8222	0,165429	0,00972
0-0-xcorr-cutOff-Ellis-Ellis-1	464	3052	16773	5548	0,167399	0,00943
0-0-xcorr-cutOff-Ellis-mean-1	2364	14895	8565	13	0,167134	0,01018
0-1-xcorr-Ellis-Ellis-Ellis-1	4461	17457	3919	0	0,172207	0,010202
0-1-xcorr-Ellis-Ellis-mean-1	8745	16036	1056	0	0,180185	0,012331
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

E (2-1-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	381	2166	15305	7985	0,165574	0,009736
0-0-xcorr-Ellis-Ellis-mean-1	3899	17194	4744	0	0,171432	0,010304
0-0-xcorr-Ellis-SimpleMax-Ellis-1	381	2166	15305	7985	0,165574	0,009736
0-0-xcorr-cutOff-Ellis-Ellis-1	472	3153	16876	5336	0,167732	0,009376
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	9466	15595	776	0	0,180959	0,012587
0-1-xcorr-Ellis-Ellis-mean-1	6303	17225	2309	0	0,176735	0,011301
0-1-xcorr-cutOff-Ellis-Ellis-1	307	1191	11592	12747	0,169109	0,00936

E (2-2-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	353	1885	14445	9154	0,165456	0,009664
0-0-xcorr-Ellis-Ellis-mean-1	6303	17225	2309	0	0,176735	0,011301
0-0-xcorr-Ellis-SimpleMax-Ellis-1	350	1824	14332	9331	0,165447	0,00965
0-0-xcorr-cutOff-Ellis-Ellis-1	308	1279	12017	12233	0,168328	0,009343
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	8745	16036	1056	0	0,180185	0,012331
0-1-xcorr-Ellis-Ellis-mean-1	10934	14440	463	0	0,183256	0,012983
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

HFC (0-0-0)	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	24598	1239	0	0	0,266654	0,031436
0-0-xcorr-Ellis-Ellis-mean-1	9466	15595	776	0	0,180959	0,012587
0-0-xcorr-Ellis-SimpleMax-Ellis-1	15482	10318	37	0	0,186028	0,015858
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	353	1885	14445	9154	0,165456	0,009664
0-1-xcorr-Ellis-Ellis-mean-1	5852	17371	2614	0	0,175466	0,010958
0-1-xcorr-cutOff-Ellis-Ellis-1	308	1234	11893	12402	0,168744	0,00935

HFC (1-0-0)	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	15482	10318	37	0	0,186028	0,015858
0-0-xcorr-Ellis-Ellis-mean-1	6303	17225	2309	0	0,176735	0,011301
0-0-xcorr-Ellis-SimpleMax-Ellis-1	4461	17457	3919	0	0,172207	0,010202
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	3899	17194	4744	0	0,171432	0,010304
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

HFC (1-1-40)	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	13604	12148	85	0	0,184226	0,014037
0-0-xcorr-Ellis-Ellis-mean-1	6799	16914	2124	0	0,177905	0,01177
0-0-xcorr-Ellis-SimpleMax-Ellis-1	4461	17457	3919	0	0,172207	0,010202
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	5460	17484	2893	0	0,174457	0,010663
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

HFC (1-2-40)	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	16304	9509	24	0	0,188248	0,016534
0-0-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-SimpleMax-Ellis-1	5852	17371	2614	0	0,175466	0,010958
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	6303	17225	2309	0	0,176735	0,011301
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

HFC (2-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	371	2057	15005	8404	0,165299	0,009715
0-0-xcorr-Ellis-Ellis-mean-1	4179	17412	4246	0	0,171825	0,010178
0-0-xcorr-Ellis-SimpleMax-Ellis-1	371	2057	15005	8404	0,165299	0,009715
0-0-xcorr-cutOff-Ellis-Ellis-1	472	3153	16876	5336	0,167732	0,009376
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	10163	15076	598	0	0,181975	0,01278
0-1-xcorr-Ellis-Ellis-mean-1	11976	13671	190	0	0,18359	0,013121
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

HFC (2-1-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	371	2057	15005	8404	0,165299	0,009715
0-0-xcorr-Ellis-Ellis-mean-1	4179	17412	4246	0	0,171825	0,010178
0-0-xcorr-Ellis-SimpleMax-Ellis-1	376	2110	15129	8222	0,165429	0,00972
0-0-xcorr-cutOff-Ellis-Ellis-1	485	3270	16924	5158	0,16803	0,009348
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	2797	15936	7097	7	0,168045	0,010453
0-1-xcorr-Ellis-Ellis-mean-1	6799	16914	2124	0	0,177905	0,01177
0-1-xcorr-cutOff-Ellis-Ellis-1	308	1234	11893	12402	0,168744	0,00935

HFC (2-2-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	353	1885	14445	9154	0,165456	0,009664
0-0-xcorr-Ellis-Ellis-mean-1	6303	17225	2309	0	0,176735	0,011301
0-0-xcorr-Ellis-SimpleMax-Ellis-1	347	1779	14178	9533	0,165458	0,009652
0-0-xcorr-cutOff-Ellis-Ellis-1	308	1279	12017	12233	0,168328	0,009343
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	10163	15076	598	0	0,181975	0,01278
0-1-xcorr-Ellis-Ellis-mean-1	13604	12148	85	0	0,184226	0,014037
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

MoT (0-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	6303	17225	2309	0	0,176735	0,011301
0-0-xcorr-Ellis-Ellis-mean-1	23713	2124	0	0	0,251386	0,029238
0-0-xcorr-Ellis-SimpleMax-Ellis-1	6303	17225	2309	0	0,176735	0,011301
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	18054	7772	11	0	0,194701	0,018472
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

MoT (1-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-Ellis-Ellis-mean-1	8745	16036	1056	0	0,180185	0,012331
0-0-xcorr-Ellis-SimpleMax-Ellis-1	4179	17412	4246	0	0,171825	0,010178
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	8745	16036	1056	0	0,180185	0,012331
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

MoT (1-1-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-Ellis-Ellis-mean-1	5460	17484	2893	0	0,174457	0,010663
0-0-xcorr-Ellis-SimpleMax-Ellis-1	4179	17412	4246	0	0,171825	0,010178
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	5460	17484	2893	0	0,174457	0,010663
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

MoT (1-2-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-SimpleMax-Ellis-1	6799	16914	2124	0	0,177905	0,01177
0-0-xcorr-cutOff-Ellis-Ellis-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-cutOff-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

MoT (2-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-Ellis-Ellis-mean-1	22599	3238	0	0	0,227307	0,024795
0-0-xcorr-Ellis-SimpleMax-Ellis-1	1382	10978	13318	159	0,170665	0,009353
0-0-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	0	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	18054	7772	11	0	0,194701	0,018472
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

MoT (2-1-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-Ellis-Ellis-mean-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-Ellis-SimpleMax-Ellis-1	1382	10978	13318	159	0,170665	0,009353
0-0-xcorr-cutOff-Ellis-Ellis-1	13604	12148	85	0	0,184226	0,014037
0-0-xcorr-cutOff-Ellis-mean-1	13604	12148	85	0	0,184226	0,014037
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	18054	7772	11	0	0,194701	0,018472
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

MoT (2-2-40)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	976	7769	16036	1056	0,171865	0,009839
0-0-xcorr-Ellis-Ellis-mean-1	2486	15129	8211	11	0,167388	0,010251
0-0-xcorr-Ellis-SimpleMax-Ellis-1	944	7487	16272	1134	0,171799	0,009859
0-0-xcorr-cutOff-Ellis-Ellis-1	4768	17516	3553	0	0,172583	0,010364
0-0-xcorr-cutOff-Ellis-mean-1	436	2769	16392	6240	0,166411	0,009701
0-1-xcorr-Ellis-Ellis-Ellis-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-Ellis-Ellis-mean-1	18054	7772	11	0	0,194701	0,018472
0-1-xcorr-cutOff-Ellis-Ellis-1	14356	11428	53	0	0,184818	0,014897

CD(0-0-0)

	/2	*1	*2	*4	mean	var
0-0-xcorr-Ellis-Ellis-Ellis-1	2992	16174	6665	6	0,168242	0,010533
0-0-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
0-0-xcorr-Ellis-SimpleMax-Ellis-1	412	2471	16034	6920	0,165806	0,009833
0-0-xcorr-cutOff-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-0-xcorr-cutOff-Ellis-mean-1	2609	15445	7772	11	0,167825	0,010357
0-1-xcorr-Ellis-Ellis-Ellis-1	360	1948	14733	8796	0,165401	0,009674
0-1-xcorr-Ellis-Ellis-mean-1	4179	17412	4246	0	0,171825	0,010178
0-1-xcorr-cutOff-Ellis-Ellis-1	472	3153	16876	5336	0,167732	0,009376

RCD (0-0-0)

	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	3205	16392	6236	4	0,168959	0,010458
0-0-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
0-0-xcorr-Ellis-SimpleMax-Ellis-1	421	2571	16174	6671	0,165855	0,009839
0-0-xcorr-cutOff-Ellis-Ellis-1	2609	15445	7772	11	0,167825	0,010357
0-0-xcorr-cutOff-Ellis-mean-1	2609	15445	7772	11	0,167825	0,010357
0-1-xcorr-Ellis-Ellis-Ellis-1	353	1885	14445	9154	0,165456	0,009664
0-1-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
0-1-xcorr-cutOff-Ellis-Ellis-1	485	3270	16924	5158	0,16803	0,009348

PD (0-0-0)	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	15482	10318	37	37	0,186028	0,015858
0-0-xcorr-Ellis-Ellis-mean-1	15482	10318	37	37	0,186028	0,015858
0-0-xcorr-Ellis-SimpleMax-Ellis-1	4768	17516	3553	3553	0,172583	0,010364
0-0-xcorr-cutOff-Ellis-Ellis-1	8745	16036	1056	1056	0,180185	0,012331
0-0-xcorr-cutOff-Ellis-mean-1	14356	11428	53	53	0,184818	0,014897
0-1-xcorr-Ellis-Ellis-Ellis-1	15482	10318	37	37	0,186028	0,015858
0-1-xcorr-Ellis-Ellis-mean-1	15482	10318	37	37	0,186028	0,015858
0-1-xcorr-cutOff-Ellis-Ellis-1	8745	16036	1056	1056	0,180185	0,012331

WPD (0-0-0)	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	2992	16174	6665	6	0,168242	0,010533
0-0-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
0-0-xcorr-Ellis-SimpleMax-Ellis-1	421	2571	16174	6671	0,165855	0,009839
0-0-xcorr-cutOff-Ellis-Ellis-1	464	3052	16773	5548	0,167399	0,00943
0-0-xcorr-cutOff-Ellis-mean-1	2609	15445	7772	11	0,167825	0,010357
0-1-xcorr-Ellis-Ellis-Ellis-1	2992	16174	6665	6	0,168242	0,010533
0-1-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
0-1-xcorr-cutOff-Ellis-Ellis-1	464	3052	16773	5548	0,167399	0,00943

NWPD (0-0-0)	/2	*1	*2	*4	Mean	Var
0-0-xcorr-Ellis-Ellis-Ellis-1	7591	16732	1514	0	0,17871	0,011986
0-0-xcorr-Ellis-Ellis-mean-1	12722	12993	122	0	0,183804	0,013453
0-0-xcorr-Ellis-SimpleMax-Ellis-1	6799	16914	2124	0	0,177905	0,01177
0-0-xcorr-cutOff-Ellis-Ellis-1	10934	14440	463	0	0,183256	0,012983
0-0-xcorr-cutOff-Ellis-mean-1	13604	12148	85	0	0,184226	0,014037
0-1-xcorr-Ellis-Ellis-Ellis-1	7591	16732	1514	0	0,17871	0,011986
0-1-xcorr-Ellis-Ellis-mean-1	12722	12993	122	0	0,183804	0,013453
0-1-xcorr-cutOff-Ellis-Ellis-1	10934	14440	463	0	0,183256	0,012983

5 Συμπεράσματα-Κατακλείδα

Από τα αποτελέσματα που συλλέξαμε στην ενότητα 4, μπορούμε να αποκόψουμε για την καθεμία Onset Detection Function και από καθένα συνδυασμό των παραμετροποιήσεων που οδηγούν στο Peak-Picking, μόνο τις γραμμές εκείνες στις οποίες βρήκαμε την βέλτιστη (μικρότερη) τιμή του πεδίου Mean. Από τους νέους αυτούς πίνακες που θα δημιουργήσουμε παρακάτω, μπορούμε να βγάλουμε κάποια συμπεράσματα για την καθεμία Onset Detection Function ξεχωριστά καθώς επίσης και να διακρίνουμε την βέλτιστη προσέγγιση του πεδίου Mean, από τα αποτελέσματα αυτά. Η νέα αυτή βέλτιστη προσέγγιση του mean που θα σημειώσουμε, δηλώνει ποιος συνδυασμός παραμετροποιήσεων είναι αυτός που δίνει την καλύτερη προσέγγιση του BeatPeriod και κατά συνέπεια του ρυθμού/tempo ενός μουσικού κομματιού. Τέλος, θα δημιουργήσουμε έναν τελικό πίνακα, στον οποίο θα έχουμε την βέλτιστη τιμή του mean για κάθε Onset Detection Function από όλους τους δυνατούς συνδυασμούς.

Masri

		/2	*1	*2	*4	Mean	Var
(0-0-0)-0-1-xcorr-Ellis-Ellis-Ellis-1	*	353	1885	14445	9154	0,165456	0,009664
(1-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1		360	1948	14733	8796	0,165401	0,009674
(1-1-40)-0-0-xcorr-cutOff-Ellis-Ellis-1		364	2000	14895	8578	0,165312	0,009694
(1-2-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	**	6799	16914	2124	0	0,177905	0,01177
(2-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1		347	1779	14178	9533	0,165458	0,009652
(2-1-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	*	353	1885	14445	9154	0,165456	0,009664
(2-1-40)-0-1-xcorr-Ellis-Ellis-Ellis-1	*	353	1885	14445	9154	0,165456	0,009664
(2-2-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-11	**	6799	16914	2124	0	0,177905	0,01177

*Παρατήρηση₁ → Για ίδιο συνδυασμό (dB,Mel,nMel) οι συνδυασμοί: 0-0-xcorr-Ellis-SimpleMax-Ellis-1 και 0-0-xcorr-Ellis-Ellis-Ellis-1 δίνουν το ίδιο αποτέλεσμα καθώς και για (0-0-0)-0-1-xcorr-Ellis-Ellis-Ellis-1.

**Παρατήρηση₂ →

Duxbury

		/2	*1	*2	*4	Mean	Var
(0-0-0)-0-1-xcorr-cutOff-Ellis-Ellis-1		527	3652	17412	4246	0,168441	0,009124
(0-0-0)-0-0-xcorr-cutOff-Ellis-Ellis-1		527	3652	17412	4246	0,168441	0,009124
(1-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1		403	2394	15936	7104	0,165806	0,00982
(1-1-40)-0-1-xcorr-Ellis-Ellis-Ellis-1		2364	14895	8565	13	0,167134	0,01018
(1-2-40)-0-0-xcorr-Ellis-Ellis-Ellis-1		1587	12017	12148	85	0,168991	0,009425
(2-0-0)-0-0-xcorr-cutOff-Ellis-Ellis-1		347	1779	14178	9533	0,165458	0,009652
(2-1-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1		376	2110	15129	8222	0,165429	0,00972
(2-2-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1		4179	17412	4246	0	0,171825	0,010178

Kullback

	/2	*1	*2	*4	Mean	Var
(0-0-0)-0-0-xcorr-cutOff-Ellis-Ellis-1	2126	14178	9509	24	0,167021	0,010048
(1-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	392	2323	15793	7329	0,165818	0,009833
(1-1-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	353	1885	14445	9154	0,165456	0,009664
(1-2-40)-0-0-xcorr-Ellis-Ellis-mean-1	7591	16732	1514	0	0,17871	0,011986
(2-2-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	5460	17484	2893	0	0,174457	0,010663
(2-1-40)-0-0-xcorr-Ellis-Ellis-Ellis-1	347	1779	14178	9533	0,165458	0,009652
(2-2-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	2486	15129	8211	11	0,167388	0,010251

E

	/2	*1	*2	*4	Mean	Var
(0-0-0)-0-1-xcorr-Ellis-Ellis-Ellis-1	360	1948	14733	8796	0,165401	0,009674
(1-0-0)-0-1-xcorr-Ellis-Ellis-mean-1	3625	16876	5335	1	0,170651	0,010268
(1-1-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	4179	17412	4246	0	0,171825	0,010178
(1-2-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	4461	17457	3919	0	0,172207	0,010202
(2-2-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	376	2110	15129	8222	0,165429	0,00972
(2-1-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	381	2166	15305	7985	0,165574	0,009736
(2-1-40)-0-0-xcorr-Ellis-Ellis-Ellis-1	381	2166	15305	7985	0,165574	0,009736
(2-2-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	350	1824	14332	9331	0,165447	0,00965
(2-2-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	350	1824	14332	9331	0,165447	0,00965

HFC

	/2	*1	*2	*4	Mean	Var
(0-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	347	1779	14178	9533	0,165458	0,009652
(1-0-0)-0-0-xcorr-Ellis-Ellis-Ellis-1	353	1885	14445	9154	0,165456	0,009664
(1-1-40)-0-0-xcorr-Ellis-Ellis-Ellis-1	371	2057	15005	8404	0,165299	0,009715
(1-2-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	371	2057	15005	8404	0,165299	0,009715
(2-2-0)-0-0-xcorr-Ellis-Ellis-Ellis-1	371	2057	15005	8404	0,165299	0,009715
(2-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	5852	17371	2614	0	0,175466	0,010958
(2-1-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	4461	17457	3919	0	0,172207	0,010202
(2-2-40)-0-1-xcorr-Ellis-Ellis-mean-1	3899	17194	4744	0	0,171432	0,010304
(2-2-40)-0-1-xcorr-Ellis-Ellis-Ellis-1	353	1885	14445	9154	0,165456	0,009664

MoT

	/2	*1	*2	*4	Mean	Var
(0-0-0)-0-0-xcorr-Ellis-Ellis-Ellis-1	6303	17225	2309	0	0,176735	0,011301
(0-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	6303	17225	2309	0	0,176735	0,011301
(1-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	4179	17412	4246	0	0,171825	0,010178
(1-1-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	4179	17412	4246	0	0,171825	0,010178
(1-2-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	6799	16914	2124	0	0,177905	0,01177
(2-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	1382	10978	13318	159	0,170665	0,009353
(2-1-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	1382	10978	13318	159	0,170665	0,009353
(2-2-40)-0-0-xcorr-cutOff-Ellis-mean-1	436	2769	16392	6240	0,166411	0,009701

CD

	Mean	Var

	/2	*1	*2	*4		
(0-0-0)-0-1-xcorr-Ellis-Ellis-Ellis-1	360	1948	14733	8796	0,165401	0,009674

RCD

	/2	*1	*2	*4	Mean	Var
(0-0-0)-0-1-xcorr-Ellis-Ellis-Ellis-1	353	1885	14445	9154	0,165456	0,009664

PD

	/2	*1	*2	*4	Mean	Var
(0-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	4768	17516	3553	3553	0,172583	0,010364

WPD

	/2	*1	*2	*4	Mean	Var
(0-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	421	2571	16174	6671	0,165855	0,009839

NWPD

	/2	*1	*2	*4	Mean	Var
(0-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	6799	16914	2124	0	0,177905	0,01177

ΤΕΛΙΚΟΣ ΠΙΝΑΚΑΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

<u>Onset</u>							
<u>Detection</u>	Συνδυασμός	/2	*1	*2	*4	Mean	Var
<u>Function</u>							
Masri	(1-1-40)-0-0-xcorr-cutOff-Ellis-Ellis-1	364	2000	14895	8578	0,165312	0,009694
Duxbury	(2-1-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	376	2110	15129	8222	0,165429	0,00972
Kullback	(1-1-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	353	1885	14445	9154	0,165456	0,009664
E	(0-0-0)-0-1-xcorr-Ellis-Ellis-Ellis-1	360	1948	14733	8796	0,165401	0,009674
HFC	(1-1-40)-0-0-xcorr-Ellis-Ellis-Ellis-1	371	2057	15005	8404	0,165299	0,009715
HFC	(1-2-40)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	371	2057	15005	8404	0,165299	0,009715
HFC	(2-2-0)-0-0-xcorr-Ellis-Ellis-Ellis-1	371	2057	15005	8404	0,165299	0,009715
MoT	(2-2-40)-0-0-xcorr-cutOff-Ellis-mean-1	436	2769	16392	6240	0,166411	0,009701
CD	(0-0-0)-0-1-xcorr-Ellis-Ellis-Ellis-1	360	1948	14733	8796	0,165401	0,009674
RCD	(0-0-0)-0-1-xcorr-Ellis-Ellis-Ellis-1	353	1885	14445	9154	0,165456	0,009664
PD	(0-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	4768	17516	3553	3553	0,172583	0,010364
WPD	(0-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	421	2571	16174	6671	0,165855	0,009839
NWPD	(0-0-0)-0-0-xcorr-Ellis-SimpleMax-Ellis-1	6799	16914	2124	0	0,177905	0,01177

Στη συγκεκριμένη πτυχιακή εργασία, θέσαμε το ερευνητικό υπόβαθρο για την αξιολόγηση ενός εύρους μεθόδων και των παραμέτρων τους όσον αφορά την Onset Det. για την εκτίμηση ρυθμού (Beat Tracking).

5.2 Δεν υλοποιήθηκαν

Δεν υλοποιήθηκαν οι παράμετροι:

➤ **MEL=2 (Constant-Q)**

```
if isfield(Options,'MEL') && strcmp(Options.MEL,'2') %Constant-Q
    TransfQ=logfmapQTransform(size(D,1),2,Options.NMEL);
    D=TransfQ*D; % D=SPQ
end
```

➤ **THRES=1**

```
if isfield(Options,'THRES') && Options.THRES>0
```

```
    CurrentFunction(CurrentFunction>Options.THRES)=Options.THRES+log(CurrentFunction(Curre
ntFunction>Options.THRES)-Options.THRES+1);
```

```
end
```

➤ **Correlation ='fourier'**

```
if strcmp(Options.CORRELATION,'fourier')
```

```
    xcr=(abs(fft(CurrentFunction)));
    xcr=xcr(1:ceil(length(xcr)/2));
    xcr=single(xcr);
    CurrentFunction=xcr;
```

```
end
```

5.3 Μελλοντικές Επεκτάσεις

Στη συγκεκριμένη πτυχιακή εργασία, θέσαμε το ερευνητικό υπόβαθρο για την αξιολόγηση ενός εύρους μεθόδων και των παραμέτρων τους όσον αφορά την Onset Det. για την εκτίμηση ρυθμού μουσικών αρχείων Mp3 .(Beat Tracking).

Ως μελλοντική εργασία, θεωρούμε πάντα την βελτίωση των ήδη υπάρχουσών εργασιών καθώς και την υλοποίηση όλων των κομματιών που δεν υλοποιήθηκαν στην συγκεκριμένη πτυχιακή εργασία.

6 Βιβλιογραφία

- [1] Content-Based Music Information Retrieval: Current Directions and Future Challenges By Michael A. Casey, Member IEEE, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney, Senior Member IEEE
- [2] METHODOLOGY AND RESOURCES FOR THE STRUCTURAL SEGMENTATION OF MUSIC PIECES INTO AUTONOMOUS AND COMPARABLE BLOCKS, *Frédéric BIMBOT, Emmanuel DERUTY, Gabriel SARGENT, Emmanuel VINCENT*, 12th International Society for Music Information Retrieval Conference (ISMIR 2011)
- [3] Melody Transcription From Music Audio: Approaches and Evaluation, Graham E. Poliner, *Student Member, IEEE*, Daniel P. W. Ellis, *Senior Member, IEEE*, Andreas F. Ehmann, Emilia Gómez, Sebastian Streich, and Beesuan Ong, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 15, NO. 4, MAY 2007
- [4] The Timbre Toolbox: Extracting audio descriptors from musical signals.
- [5] “Copy and Scale” Method for Doing Time-Localized M.I.R. Estimation: Application to Beat-tracking, Geoffroy Peeters IRCAM – CNRS 1, pl. Igor Stravinsky - 75004 Paris – France peeters@ircam.fr
- [6] Audio Content-Based Music Retrieval, Peter Grosche¹, Meinard Müller*¹, and Joan Serra†² ¹ Saarland University and MPI Informatik Campus E1-4, 66123 Saarbrücken, Germany pgrosche@mpi-inf.mpg.de, meinard@mpi-inf.mpg.de ² Artificial Intelligence Research Institute (IIIA-CSIC) Campus UAB s/n, 08193 Bellaterra, Barcelona, Spain jserra@iiia.csic.es
- [7] Music Retrieval: A Tutorial and Review, Nicola Orio *Department of Information Engineering, University of Padova Via Gradenigo, 6/b, Padova 35131, Italy, orio@dei.unipd.it*, Foundations and Trends^R in Information Retrieval Vol. 1, No 1 (November 2006) 1–90_c November 2006 Nicola Orio
- [8] Beat Tracking by Dynamic Programming, Daniel P.W. Ellis LabROSA, Columbia University, New York July 16, 2007

- [9] Reliability-Informed Beat Tracking of Musical Signals Norberto Degara, *Student Member, IEEE*, Enrique Argones Rúa, *Member, IEEE*, Antonio Pena, *Member, IEEE*, Soledad Torres-Guijarro, Matthew E. P. Davies, *Member, IEEE*, and Mark D. Plumbley, *Member, IEEE*
- [10] ONSET DETECTION REVISITED Simon Dixon Austrian Research Institute for Artificial Intelligence Freyung 6/6, Vienna 1010, Austria simon.dixon@ofai.at
- [11] [MIREX, 2005] http://www.music-ir.org/mirex/wiki/2005:Main_Page
- [12] [MIREX, 2009] http://www.music-ir.org/mirex/wiki/2009:Main_Page
- [13] [MIREX, 2014] http://www.music-ir.org/mirex/wiki/2014:Main_Page
- [14] A Tutorial on Onset Detection in Music Signals Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler, *Senior Member, IEEE*
- [15] RENI:Real Time Beat Tracking andMetrical AnalysisDonal Mulvihill, Master of Science Artificial Intelligence School of Informatics University of Edinburgh 2008
- [16] Evaluation of Audio Beat Tracking and Music Tempo Extraction Algorithms, M. F. McKinney¹, D. Moelants², M. E. P. Davies³ and A. Klapuri, ¹Philips Research Laboratories, Eindhoven, The Netherlands; ²Ghent University, Belgium; ³Queen Mary University of, London, UK; ⁴Tampere University of Technology, Finland
- [17] Αυτόματη Καταγραφή Μουσικής,, Ελένης Τζιρίτα Ζαχαράτου, 2013
- [18] A. V. Oppenheim, A. S. Willsky and H. S. Nawab, Signals and Systems, Prentice Hall, 1997.
- [19] Improved Modelling of Attack Transients in Music Analysis-Resynthesis

Παράρτημα Α: Πηγαίος Κώδικας

BatchAnalyzeSpects.m (Script)

```
Options(1).dB='0';
Options(1).MEL='1';
Options(1).NMEL=70;

Options(2).dB='0';
Options(2).MEL='0';
Options(2).NMEL=70;

Options(3).dB='1';
Options(3).MEL='0';
Options(3).NMEL=70;

Options(4).dB='2';
Options(4).MEL='0';
Options(4).NMEL=70;

Options(5).dB='2';
Options(5).MEL='1';
Options(5).NMEL=70;

Options(6).dB='1';
Options(6).MEL='1';
Options(6).NMEL=70;

Options(7).dB='0';
Options(7).MEL='1';
Options(7).NMEL=13;

Options(8).dB='0';
Options(8).MEL='0';
Options(8).NMEL=13;

Options(9).dB='1';
Options(9).MEL='0';
Options(9).NMEL=13;

Options(10).dB='2';
Options(10).MEL='0';
Options(10).NMEL=13;

Options(11).dB='2';
Options(11).MEL='1';
Options(11).NMEL=13;

for ii=1:11
    spectAnalysis('collection',Options(ii));
end
```

spectral_based_df.m (function)

```
function [df] = spectral_based_df(Spect,Options)
% Calculates all the spectral based detection functions based on the
% parameters above: complex domain (cd), rectified complex domain (rcd),
% spectral differences (masri, duxburi, kullback), phase deviation (pd),
% weighted phase deviation (wpd), normalized and weighted phase deviation
% (nwpd)

% % Divide original waveform in frames
% n_start = 1;
% n_end = length(signal);
% [frames,t_frames] = segment_signal(signal,fs,n_start,n_end,frame_length,hop_length);
%
% % STFT calculation
% [spectral_magnitude,spectral_phase] = stft(frames>window_handle,N_fft);

%Αλλαγή :
spectral_magnitude = abs(Spect);

spectral_magnitudeChange=Spect;
% % Time axis for the detection functions
% t_df = t_frames;

df = [];

if strcmp(Options.dB,'0') && strcmp(Options.MEL,'0')
    % Complex domain detection function
    spectral_magnitude_0 = spectral_magnitude(:,3:end);
    spectral_magnitude_1 = spectral_magnitude(:,2:end-1);

    spectral_phase = angle(Spect);
    spectral_phase_0 = spectral_phase(:,3:end);
    spectral_phase_1 = spectral_phase(:,2:end-1);
    spectral_phase_2 = spectral_phase(:,1:end-2);

    method = 'all';% complex domain and rectified complex domain
    df_tmp = tss_complex_domain(method,spectral_phase_0,spectral_phase_1,...
        spectral_phase_2,spectral_magnitude_0,spectral_magnitude_1,[],[],[]);
    df_cd = df_tmp.cd;
    df_rcd = df_tmp.rcd;

    df_cd = [0 0 df_cd]; % pad with 0 first 2 values
    df_rcd = [0 0 df_rcd];
    clear spectral_magnitude_0 spectral_magnitude_1 spectral_phase_0 spectral_phase_1
    spectral_phase_2

    df.cd = df_cd;
    df.rcd = df_rcd;

    % Phase deviation
```

```

spectral_magnitude_0 = spectral_magnitude(:,3:end);
spectral_phase_0 = spectral_phase(:,3:end);
spectral_phase_1 = spectral_phase(:,2:end-1);
spectral_phase_2 = spectral_phase(:,1:end-2);

method = 'all';
df_tmp = tss_phase_deviation(method,...
    spectral_phase_0,spectral_phase_1,spectral_phase_2,spectral_magnitude_0,...
    [],[],[]);
df_pd = df_tmp.pd;
df_wpd = df_tmp.wpd;
df_nwpd = df_tmp.nwpd;

df_pd = [0 0 df_pd];
df_wpd = [0 0 df_wpd];
df_nwpd = [0 0 df_nwpd];
clear spectral_magnitude_0 spectral_phase_0 spectral_phase_1 spectral_phase_2
df.pd = df_pd;
df.wpd = df_wpd;
df.nwpd = df_nwpd;
end

% spectral differences
spectral_magnitude_0 = spectral_magnitude(:,2:end);
spectral_magnitude_1 = spectral_magnitude(:,1:end-1);
spectral_magnitudeCh_0=spectral_magnitudeChange(:,2:end);
spectral_magnitudeCh_1=spectral_magnitudeChange(:,1:end-1);

spectral_magnitude_cut=spectral_magnitude(:,11:end-20);

method = 'all';
df_tmp = tss_spectral_difference(method,...

spectral_magnitude_0,spectral_magnitude_1,spectral_magnitude,spectral_magnitude_cut,spectral_
magnitudeCh_0,spectral_magnitudeCh_1,[],[],[]);

%df_sd_masriChange = df_tmp.sd_masriChange;
df_sd_masri = df_tmp.sd_masri;
df_sd_duxbury = df_tmp.sd_duxbury;
df_sd_kullback = df_tmp.sd_kullback;
df_sd_E = df_tmp.sd_E;
df_sd_HFC = df_tmp.sd_HFC;
df_sd_MoT = df_tmp.sd_MoT;
%df_sd_d_rel = df_tmp.sd_d_rel;

%df_sd_masriChange=[0 df_sd_masriChange];
df_sd_masri = [0 df_sd_masri];
df_sd_duxbury = [0 df_sd_duxbury];
df_sd_kullback = [0 df_sd_kullback];
df_sd_MoT = [0 df_sd_MoT];
%df_sd_E = [0 df_sd_E];
%df_sd_d_rel = [0 df_sd_d_rel];
clear spectral_magnitude_0 spectral_magnitude_1

```

```

df.sd_masri = df_sd_masri;
%df.sd_masriChange = df_sd_masriChange;
df.sd_duxbury = df_sd_duxbury;
df.sd_kullback = df_sd_kullback;
df.sd_E = df_sd_E;
df.sd_HFC = df_sd_HFC;
df.sd_MoT = df_sd_MoT;
%df.df_sd_d_rel = df_sd_d_rel;

% Detection function struct

```

spectAnalysis.m (function)

```

% Συνάρτηση η οποία παίρνει ως ορίσματα :
% 1) όνομα μιας Onset Function (πχ: SF...)
% 2) Αν θα είναι σε μορφή Options.dB ή όχι.
% 3) Αν θα είναι το spectrogram MEL ή όχι
% 4) Αν το Correlation θα είναι: xcorr or fourier
% Θα επιστρέφει το xcr

%function [mm]=spectAnalysis(typeOfOnset,Options)
function [xcr,mm]=spectAnalysis(typeOfOnset,Options)

% Options.MEL: '1' for mel, '2' for ConstantQ, anything else for plain
% spectrogram
% Options.nmel: if mel, number of channels. If ConstantQ, the upper
% frequency taken into account (the lower is always 2).
% Options.dB: '1' for dB conversion, '2' for generic log conversion.
% Options.movavg: '1' for applying a moving average on the signal prior to
% correlation/fourier

%D = dir('spectroFirst3'); %Δοκιμή στα 10 πρώτα αρχεία .mat
D = dir('I:\Spectrograms'); %Για τα 25.837 αρχεία
SR=8000;%SR=SampleRate
swin=256;

%SavePath=strcat('I:\Corrs','./Corrs',typeOfOnset,'-',Options.dB,'-',Options.MEL,'-
',num2str(Options.NMEL));
SavePath=strcat('\Corrs','./Corrs',typeOfOnset,'-',Options.dB,'-',Options.MEL,'-
',num2str(Options.NMEL));

mkdir(SavePath);

%for ii=1:50% :size(D,1)-2
for ii=1:size(D,1)-2
%for ii=15039:size(D,1)-2

```

```

if mod(ii,500)==0
    disp(ii)
end

%g=strcat('spectroFirst3\',num2str(ii),'.mat');
g=strcat('I:\Spectrograms\',num2str(ii),'.mat'); %Για τα 25.837 αρχεία
sp=load(g);

D =sp.spectro.SP;

if isfield(Options,'MEL') && strcmp(Options.MEL,'1')
    mlmx = fft2melmx(swin,SR,40);
    %mlmx = fft2melmx(swin,SR,Options.NMEL);
    % disp(mlmx)
    D=mlmx(:,1:(swin/2+1))*abs(D);
end
if isfield(Options,'MEL') && strcmp(Options.MEL,'2') %Constant-Q
    TransfQ=logfmapQTransform(size(D,1),2,Options.NMEL);
    D=TransfQ*D; % D=SPQ
end
%disp(mlmx)

if isfield(Options,'dB') && strcmp(num2str(Options.dB),'1')
    D = 20*log10(max(1e-10,abs(D)));

    % Only look at the top 80 Options.dB
    D = max(D, max(max(D))-80);
%    figure;
%    surf(imresize(double(D),[129 100]));
%    size(D)

end

if isfield(Options,'dB') && strcmp(num2str(Options.dB),'2')
    D=log(1 + 1000* abs(D));
end

% The raw onset decision waveform

if strcmp(typeOfOnset,'collection')
    Corrs= spectral_based_df(D,Options);
end

save(strcat(SavePath,'\ ',num2str(ii),'.mat'),'Corrs');

% save(strcat(SavePath,'\ ',Options,'.mat'),'Corrs');
end
end

```

spect.m (function)

% Συνάρτηση η οποία παίρνει ως παραμέτρους: το μέγεθος παραθύρου και την επικάλυψη και επιστρέφει τα spectrogram.

```
function spect(window,overlap)
count=0;
load('Mp3New.mat');
sro=8000;

for ii=1:length(paths)
    count=count+1
    t=strcat('MP3','\ ',paths(1,ii).pathMP3);
    [Song,FS]=mp3read(t);
    Song=mean(Song,2);

    % resample to 8 kHz
    if (FS ~= sro)
        gg = gcd(sro,FS);
        Song = resample(Song,sro/gg,FS/gg);
        FS = sro;
    end

    [sp,F,T]=spectrogram(Song>window,overlap>window,FS);
    SPs=single(sp);
    % InformationLoss=spectrograms.sp-SPs;
    % mn=mean(mean(InformationLoss))/mean(mean(sp));
    spectro.SP=SPs;
    spectro.F=F;
    spectro.T=T;
    name=strcat('C:\Spectrograms\',num2str(ii),'.mat');
    save(name,'spectro');
end
end
```

tss_spectral_difference.m

```
function df = tss_spectral_difference(method,...
```

```
current_spectral_magnitude,previous_spectral_magnitude,spectral_magnitude,spectral_magnitude_
cut,spectral_magnitudeCh_0,spectral_magnitudeCh_1,f_min,f_max,f_weight)
```

```
% TSS_SPECTRAL_DIFFERENCE detection function as a distance between successive
% short-term Fourier spectra.
%
% *Syntax*
%
```



```

% df = tss_spectral_difference(method,...
%   current_spectral_magnitude,previous_spectral_magnitude,varargin)
%
% *Description*
%
% TSS_SPECTRAL_DIFFERENCE detection function as a distance between successive
% short-term Fourier spectra, treating them as points in an N-dimensional space [1].
% Masri uses the L1-norm of the difference between magnitude spectra. Duxbury
% uses the L2-norm on the rectified difference. Rectification has the effect of
% counting only those frequencies where there is an increase in energy and is
% intended to emphasize onsets rather than offsets.
%
% *Inputs*
%
% current_spectral_magnitude - spectral magnitude of current frame
% previous_spectral_magnitude - spectral magnitude of the previous frame
% spectral_range           - spectral range over which the distance is
%                           evaluated (k indexes, default 1:N_fft/2)
%
% *Outputs*
%
% SD_Duxbury   - Duxbury spectral difference
% SD_Masri     - Masri spectral difference
%
% *References*
%
% [1] Bello, J.P.; Daudet, L.; Abdallah, S.; Duxbury, C.; Davies, M.; Sandler,
% M.B., "A Tutorial on Onset Detection in Music Signals," Speech and Audio
% Processing, IEEE Transactions on , vol.13, no.5, pp. 1035-1047, Sept. 2005

SD_Duxbury = 0;
SD_Masri = 0;
SD_Kullback_Leibler = 0;
%SD_D_rel = 0;
%SD_MasriChange=0;
SD_E=0; %E=Energy Function
SD_HFC=0;
SD_MoT=0;
% Check for empty input arguments
if isempty(current_spectral_magnitude) | isempty(previous_spectral_magnitude)
    return
end

% Spectral range
% N_fft = size(current_spectral_magnitude,1);
% K = 1:N_fft/2;
% % if nargin > 3
% %   spectral_range = varargin{1};
% %   K = spectral_range;
% % end

% spectral difference
spectral_difference = abs(current_spectral_magnitude)-...

```

```

abs(previous_spectral_magnitude);

%KK = repmat((1:N_fft/2)',1,size(spectral_difference,2)); % linearly emphasises the contribution of
each frequency bin
%spectral_differenceChange = spectral_magnitudeCh_0-spectral_magnitudeCh_1;
% figure;
% plot(spectral_differenceChange)

% Rectification has the effect of counting only those frequencies where there is
% an increase in energy and is intended to emphasize onsets rather than offsets.
spectral_difference(spectral_difference < 0) = 0;
%spectral_differenceChange(spectral_differenceChange < 0) = 0;

% Spectral difference: Masri uses the L1-norm of the difference between
% magnitude spectra.

%SD_E
SD_E=spectral_magnitude.^2;
SD_E=SD_E(1:(size(SD_E,1)-2),:);
SD_E=sum(SD_E);
SD_E(isnan(SD_E)) = 0;
SD_E(isinf(SD_E)) = 0;

%SD_HFC
k=[size(spectral_magnitude,1):-1:1];
k=k';
k=repmat(k,1,size(spectral_magnitude,2));
SD_HFC=(spectral_magnitude.^2).*k;
SD_HFC=SD_HFC(1:(size(SD_HFC,1)-2),:);
SD_HFC=sum(SD_HFC);
SD_HFC(isnan(SD_HFC)) = 0;
SD_HFC(isinf(SD_HFC)) = 0;

%SD_MoT
SD_HFC_current=SD_HFC(:,2:end);
SD_HFC_previous=SD_HFC(:,1:end-1);
SD_E_current=SD_E(:,2:end);
SD_MoT=(SD_HFC_current./SD_HFC_previous).*(SD_HFC_current./SD_E_current);
SD_MoT(isnan(SD_MoT)) = 0;
SD_MoT(isinf(SD_MoT)) = 0;

% % %SD_MasriChange
% % SD_MasriChange = spectral_differenceChange;
% % SD_MasriChange=SD_MasriChange(1:(size(SD_MasriChange,1)-2),:);
% % SD_MasriChange = mean(abs(SD_MasriChange),1);
% % SD_MasriChange(isnan(SD_MasriChange)) = 0;
% % SD_MasriChange(isinf(SD_MasriChange)) = 0;
% % % figure;
% % % plot(SD_MasriChange)

%SD_Masri
SD_Masri = spectral_difference;

```

```

SD_Masri=SD_Masri(1:(size(SD_Masri,1)-2),:);
SD_Masri = 2*sum(abs(SD_Masri),1); % Only positive frequencies considered in K
SD_Masri(isnan(SD_Masri)) = 0;
SD_Masri(isinf(SD_Masri)) = 0;

% Spectral difference: Duxbury uses the L2-norm on the rectified difference.
SD_Duxbury=spectral_difference;
SD_Duxbury=SD_Duxbury(1:(size(SD_Duxbury,1)-2),:);
SD_Duxbury = 2*sum(SD_Duxbury.^2,1);
SD_Duxbury(isnan(SD_Duxbury)) = 0;
SD_Duxbury(isinf(SD_Duxbury)) = 0;

% Kullback-Leibler distance measure
Kullback_Leibler_k = log2(current_spectral_magnitude./previous_spectral_magnitude);
Kullback_Leibler_k(Kullback_Leibler_k < 0) = 0;
Kullback_Leibler_k=Kullback_Leibler_k(1:(size(Kullback_Leibler_k,1)-2),:);
SD_Kullback_Leibler = 2*sum(Kullback_Leibler_k,1); % rectification
SD_Kullback_Leibler(isnan(SD_Kullback_Leibler)) = 0;
SD_Kullback_Leibler(isinf(SD_Kullback_Leibler)) = 0;

% Onset Detection d_rel(i)=(o_i-0_il)*o_ir
%SD_D_rel=(spectral_magnitude_cut-meanprevSpec_mag).*meanNextSpec_mag;

% for ii=1:length(spectral_magnitude_cut)
%   % disp(ii)
%   prev_spect_magnitude=spectral_magnitude(:,ii:9+ii); %10 προηγούμενα δείγματα
%   current_spect_magnitude=spectral_magnitude_cut(:,ii); %Το μέτρο κάθε στιγμή
%   next_spect_magnitude=spectral_magnitude(:,11+ii:30+ii); %20 επόμενα δείγματα
%
%   prev_spect_magnitude=mean(prev_spect_magnitude,2);% μέσος των 10 προηγούμενων
δειγμάτων
%   next_spect_magnitude=mean(next_spect_magnitude,2);% μέσος των 20 επόμενων δειγμάτων
%   SD_D_rel(:,ii)=(current_spect_magnitude-prev_spect_magnitude).*next_spect_magnitude;
%   %SD_D_rel(isnan(SD_D_rel)) = 0;
%   %SD_D_rel(isinf(SD_D_rel)) = 0;
% end
% Selection of the detection function
switch lower(method)
case 'masri'
    df = SD_Masri;
case 'duxbury'
    df = SD_Duxbury;
case 'kullback_leibler'
    df = SD_Kullback_Leibler;
case 'E'
    df = SD_E;
case 'HFC'
    df = SD_HFC;
case 'MoT'
    df = SD_MoT;
    %case 'd_rel'
    %df = SD_D_rel;

    %case 'masriCh'

```

```

%df = SD_MasriChange;

case 'all'
    df.sd_masri = SD_Masri; %L1
    df.sd_duxbury = SD_Duxbury; %L2
    df.sd_kullback = SD_Kullback_Leibler;
    df.sd_E = SD_E;
    df.sd_HFC = SD_HFC;
    df.sd_MoT = SD_MoT;
    %df.sd_d_rel = SD_D_rel;
    %df.sd_masriChange = SD_MasriChange;
otherwise
    df = [];
end
% % % % Nor: weighting spectral difference as in HFC
% % % SD_Nor = 2*sum(KK.*abs(spectral_difference),1); % Only positive frequencies considered
in K

```

tss_phase_deviation.m

```

function df = tss_phase_deviation(method,...
    spectral_phase_0,spectral_phase_1,spectral_phase_2,spectral_magnitude,...
    f_min,f_max,f_weight)
% TSS_PHASE_DEVIATION detection function that measures the change of
% instantaneous frequency.
%
% *Syntax*
%
% [pd,weighted_pd,normalized_weighted_pd] = tss_phase_deviation(...
% spectral_phase_0,spectral_phase_1,spectral_phase_2)
%
% *Description*
%
% The rate of change of phase in an STFT frequency bin is
% an estimate of the instantaneous frequency of that component and the
% change of instantaneous frequency is an indicator of possible onset [1].
%
% *Inputs*
%
% spectral_phase_0 - spectral phase at instant n
% spectral_phase_1 - spectral phase at instant n-1
% spectral_phase_2 - spectral phase at instant n-2
%
% *Outputs*
%
% pd - phase deviation
% weighted_pd - weighted phase deviation
% normalized_weighted_pd - normalized and weighted phase deviation
%
% *References*
%

```

```

% [1] Juan Pablo Bello Correa. "Towards the Automated Analysis of Simple
% Polyphonic Music: A Knowledge-based Approach", Department of Electronic
% Engineering, Queen Mary, University of London. PhD Dissertation

```

```

pd = 0;
weighted_pd = 0;
normalized_weighted_pd = 0;

```

```

% FFT length
N_fft = size(spectral_magnitude,1);

```

```

% Instantaneous frequency
inst_freq_0 = princarg(spectral_phase_0-spectral_phase_1);
inst_freq_1 = princarg(spectral_phase_1-spectral_phase_2);

```

```

% Change in instantaneous frequency
instant_freq_change = princarg(inst_freq_0-inst_freq_1);

```

```

% Phase deviation calculation
pd = 1/N_fft*sum(abs(instant_freq_change),1);
weighted_pd = 1/N_fft*sum(abs(spectral_magnitude.*instant_freq_change),1);
normalized_weighted_pd = sum(abs(spectral_magnitude.*instant_freq_change),1)...
    ./sum(abs(spectral_magnitude),1);
normalized_weighted_pd(isnan(normalized_weighted_pd)) = 0;
normalized_weighted_pd(isinf(normalized_weighted_pd)) = 0;

```

```

% Selection of the detection function
switch lower(method)
case 'pd'
    df = pd;
case 'wpd'
    df = weighted_pd;
case 'nwpd'
    df = normalized_weighted_pd;
case 'all'
    df.pd = pd;
    df.wpd = weighted_pd;
    df.nwpd = normalized_weighted_pd;
otherwise
    df = [];
end

```

```

% Helper function

```

```

%-----
function phase_out = princarg(phase_in)
% princarg, function to compute the principal phase argument in the
% range [-pi, pi].
%
```

```

phase_out = mod(phase_in+pi,-2*pi)+pi;

```

tss_complex_domain.m

```
function df = tss_complex_domain(method,spectral_phase_0,spectral_phase_1,...
    spectral_phase_2,spectral_magnitude_0,spectral_magnitude_1,f_min,f_max,f_weight)
% TSS_COMPLEX_DOMAIN this onset detection approach works with Fourier
% coefficients in the complex domain.
%
% *Syntax*
%
% df = tss_complex_domain(method,spectral_phase_0,spectral_phase_1,...
%     spectral_phase_2,spectral_magnitude_0,spectral_magnitude_1)
%
% *Description*
%
% TSS_COMPLEX_DOMAIN quantifies the stationarity of the kth spectral bin by
% calculating the euclidean distance between the observed  $X_k(n)$  and the
% predicted by the previous frames, eq. 10 [1].
%
% *Inputs*
%
%     method - 'cd' for complex domain and 'rcd' for rectified complex
%     domain
%     spectral_phase_i - matrix of spectral phases delayed by i
%     spectral_magnitude_i - matrix of spectral magnitudes delayed by i
%
% *Outputs*
%
%
% *References*
%
% [1] Bello, J.P.; Daudet, L.; Abdallah, S.; Duxbury, C.; Davies, M.; Sandler,
% M.B., "A Tutorial on Onset Detection in Music Signals," Speech and Audio
% Processing, IEEE Transactions on , vol.13, no.5, pp. 1035-1047, Sept. 2005

df = []; % detection function

X = spectral_magnitude_0.*exp(j*spectral_phase_0);

% The predicted magnitude for the current frame is approximated by the magnitude
% of the previous frame
inst_freq_1 = princarg(spectral_phase_1-spectral_phase_2); % instant frequency
Xt = spectral_magnitude_1.*exp(j*(spectral_phase_1+inst_freq_1));

% Complex Domain detection function
cd = sum(abs(X-Xt),1);

% Rectified complex domain: only increases in the amplitude of the signal
% are considered
rcd_nk = abs(X-Xt);
rcd_nk(spectral_magnitude_0 < spectral_magnitude_1) = 0; % decreases in amplitude set to 0
rectified_cd = sum(rcd_nk,1);

% Selection of the detection function
```

```

switch lower(method)
  case 'cd'
    df = cd;
  case 'rcd'
    df = rectified_cd;
  case 'all'
    df.cd = cd;
    df.rcd = rectified_cd;
  otherwise
    df = [];
end

% Helper function
%-----
function phase_out = princarg(phase_in)
% princarg, function to compute the principal phase argument in the
% range [-pi, pi].
%
phase_out = mod(phase_in+pi,-2*pi)+pi;

```

peakpick.m

```

corrs = dir('I:\Corrs\old_full')
for ii=3:size(corrs,1)
  disp(ii)
  corrs = dir('I:\Corrs\old_full')
  k=corrs(ii).name;
  disp(k)
  % k='Corrscollection-1-0-40'

Options(1).MOVAVG=0;
Options(1).THRES=0;
Options(1).CORRELATION='xcorr';
Options(1).windowMethod='cutOff';
Options(1).PICKMETHOD='Ellis';
Options(1).normalization='Ellis';
Options(1).normalization2='1';

Options(2).MOVAVG=0;
Options(2).THRES=1;
Options(2).CORRELATION='xcorr';
Options(2).windowMethod='Ellis';
Options(2).PICKMETHOD='Ellis';

```

```
Options(2).normalization='mean';  
Options(2).normalization2='1';
```

```
Options(3).MOVAVG=0;  
Options(3).THRES=1;  
Options(3).CORRELATION='xcorr';  
Options(3).windowMethod='Ellis';  
Options(3).PICKMETHOD='Ellis';  
Options(3).normalization='Ellis';  
Options(3).normalization2='1';
```

```
Options(4).MOVAVG=0;  
Options(4).THRES=1;  
Options(4).CORRELATION='xcorr';  
Options(4).windowMethod='cutOff';  
Options(4).PICKMETHOD='Ellis';  
Options(4).normalization='Ellis';  
Options(4).normalization2='1';
```

```
Options(5).MOVAVG=0;  
Options(5).THRES=0;  
Options(5).CORRELATION='xcorr';  
Options(5).windowMethod='Ellis';  
Options(5).PICKMETHOD='SimpleMax';  
Options(5).normalization='Ellis';  
Options(5).normalization2='1';
```

```
Options(6).MOVAVG=0;  
Options(6).THRES=0;  
Options(6).CORRELATION='xcorr';  
Options(6).windowMethod='Ellis';  
Options(6).PICKMETHOD='Ellis';  
Options(6).normalization='Ellis';  
Options(6).normalization2='1';
```

```
Options(7).MOVAVG=0;  
Options(7).THRES=0;  
Options(7).CORRELATION='xcorr';  
Options(7).windowMethod='Ellis';  
Options(7).PICKMETHOD='Ellis';  
Options(7).normalization='mean';  
Options(7).normalization2='1';
```



```

Options(8).MOVAVG=0;
Options(8).THRES=0;
Options(8).CORRELATION='xcorr';
Options(8).windowMethod='cutOff';
Options(8).PICKMETHOD='Ellis';
Options(8).normalization='mean';
Options(8).normalization2='1';

for jj=1:8
    disp(jj)
    [Differences] = PeakPicking(k,Options(jj));
end
end

```

PeakPicking.m

```

function [Output] = PeakPicking( folderName, Options )
%PEAKPICKING Extract peaks from stored correlation/fourier files
% Detailed explanation goes here
L=load('Mesos.mat');
SRo=8000;

SRSpect=SRo/32;
acmax=round(4*SRSpect);
disp('test');

MinBoundary=0.125; %Highest beat is 8bps
disp('test');
MaxBoundary=1; %Highest beat is 1bps
WindowMean=240;
cutOffLow=0;
cutOffHigh=0;
maxd = 60;
maxt = 120;

fileList = dir(folderName);
SavePath=strcat('\Differences', './Differences', folderName, '-', num2str(Options.MOVAVG), '-
', num2str(Options.THRES), '-', Options.CORRELATION, '-', Options.windowMethod, '-
', Options.PICKMETHOD, '-', Options.normalization, '-', num2str(Options.normalization2));
mkdir(SavePath);
% SavePathh=strcat('\Correlations', './Correlations', folderName, '-', num2str(Options.MOVAVG), '-
', num2str(Options.THRES), '-', Options.CORRELATION, '-', Options.windowMethod, '-
', Options.PICKMETHOD, '-', Options.normalization, '-', num2str(Options.normalization2));
% mkdir(SavePathh);
% SavePathhh=strcat('\Onset', './Onset', folderName, '-', num2str(Options.MOVAVG), '-
', num2str(Options.THRES), '-', Options.CORRELATION, '-', Options.windowMethod, '-
', Options.PICKMETHOD, '-', Options.normalization, '-', num2str(Options.normalization2));
% mkdir(SavePathhh);

```

```

size(fileList,1)
for ii=1:3%:size(fileList,1)-2 %25.837
    if mod(ii,500)==0
        disp(ii)
    end
%for ii=1:size(fileList,1)-2
    Input=load(strcat(folderName,'\ ',num2str(ii),'.mat'));
    fieldList=fieldnames(Input);
    Input=getfield(Input,fieldList{1});
    fieldList=fieldnames(Input);

    %fieldList=[];
    %fieldList{1}='sd_masri';
    %fieldList{2}='sd_masriChange';

    TempBeatPeriod=Input;

    for jj=1:length(fieldList)

%        SubFolderPathh=strcat(SavePathh,'./',fieldList{jj});%Correlations/onset_Detections
%        mkdir(SubFolderPathh);
%
%        SubFolderPathhh=strcat(SavePathhh,'./',fieldList{jj});%Onset/onset_Detections
%        mkdir(SubFolderPathhh);

        CurrentFunction=getfield(Input,fieldList{jj});
        CurrentFunction=CurrentFunction(:,2:end);
%        figureTitle=strcat(num2str(ii),'.mat\ ',fieldList{jj,1});
%        figure;
%        plot(CurrentFunction)
%        title(figureTitle,'interpreter','none')
%        %surf(imresize(double(CurrentFunction),[129 100]));

        if isfield(Options,'MOVAVG') && Options.MOVAVG>0
            Smoothed=tsmovavg(CurrentFunction,'s',Options.MOVAVG);
            %Smoothed=tsmovavg(CurrentFunction,'s',WindowMean);
            Smoothed=CurrentFunction-Smoothed;
            CurrentFunction=Smoothed;
            %CurrentFunction(CurrentFunction < 0) = 0;
        end
%        figureTitlee=strcat(' with MOVAVG_',num2str(Options.MOVAVG),' ',num2str(ii),'.mat\
',fieldList{jj,1});
%        figure;
%        plot(CurrentFunction)
%        title(figureTitlee,'interpreter','none')
%        % surf(imresize(double(CurrentFunction),[129 100]));

        if isfield(Options,'THRES') && Options.THRES>0

        CurrentFunction(CurrentFunction>Options.THRES)=Options.THRES+log(CurrentFunction(Curre
ntFunction>Options.THRES)-Options.THRES+1);
        end

```



```

if strcmp(Options.CORRELATION,'fourier')

    xcr=(abs(fft(CurrentFunction)));
    xcr=xcr(1:ceil(length(xcr)/2));
    xcr=single(xcr);
    CurrentFunction=xcr;
end
% disp(CurrentFunction)
% figureTitleee=strcat('with Correlation ',Options.CORRELATION,' ',num2str(ii),'.mat\
',fieldList{jj,1});
% figure;
% plot(CurrentFunction)
% title(figureTitleee,'interpreter','none')

if strcmp(Options.windowMethod, 'cutOff')
    cutOffLow=round(SRSpect*MinBoundary);
    cutOffHigh=round(SRSpect*MaxBoundary);
    CurrentFunction=CurrentFunction(cutOffLow+1:cutOffHigh);
end
%figureTitle=strcat(num2str(ii),'\',fieldList{jj,1});
%figure;
%plot(CurrentFunction)
%title(figureTitle,'interpreter','none')

if strcmp(Options.windowMethod, 'Ellis')
    bpms = 60*SRSpect./([0:acmax]+0.1);
    xcrwin = exp(-.5*((log(bpms/WindowMean)/log(2)/1).^2));
    CurrentFunction = CurrentFunction.*xcrwin;
end
% if jj==1
% %%%%%%%%%%%%%%%
save(strcat(SubFolderPathh,'',num2str(ii),'.mat'),'CurrentFunction'); %save .mat files into the file
of Correlations
%end

if strcmp(Options.PICKMETHOD, 'SimpleMax')
    [Value, Coord]=max(CurrentFunction);
end

if strcmp(Options.PICKMETHOD, 'Ellis')
    lxcr = length(CurrentFunction);
    xcr00 = [0, CurrentFunction, 0];
    xcr2 =
CurrentFunction(1:ceil(lxcr/2))+.5*(.5*xcr00(1:2:lxcr)+xcr00(2:2:lxcr+1)+.5*xcr00(3:2:lxcr+2));
    xcr3 =
CurrentFunction(1:ceil(lxcr/3))+.33*(xcr00(1:3:lxcr)+xcr00(2:3:lxcr+1)+xcr00(3:3:lxcr+2));

    if max(xcr2) > max(xcr3)
        [Value, Coord] = max(xcr2);
    end
end

```

```

else
    [Value, Coord] = max(xcr3);
end

Coord=Coord-1;

end
%figure;
%plot(CurrentFunction)

Estimate=(Coord+cutOffLow)/SRSpect;
TempBeatPeriod=setfield(TempBeatPeriod,fieldList{jj},Estimate);
%BeatPeriodString=strcat('BeatPeriod(',num2str(ii),'.', fieldList{jj});
%eval(sprintf('%s=%s',BeatPeriodString, 'Estimate'));
%figureTitle=strcat(num2str(ii),'.mat\ ',fieldList{jj,1});
%figure;
%plot(CurrentFunction)
%title(figureTitle,'interpreter','none')
end

BeatPeriod(ii)=TempBeatPeriod;

end

for m=1:length(BeatPeriod)
    disp(BeatPeriod)
    disp(length(BeatPeriod))

    fieldListt=fieldnames(BeatPeriod);
    TempDiffers=BeatPeriod(m);

    for kk=1:length(fieldListt)
        disp(fieldListt)
        disp(length(fieldListt))
        %nameOfDiff=strcat('Difference(',fieldListt(kk,1),')');
        fieldIn=getfield(BeatPeriod,fieldListt{kk});
        Benchmark=L.mesos(1,m).mesos;
        %[fieldIn Benchmark]
        differ=abs([0 Benchmark/2 Benchmark Benchmark*2 Benchmark*4]-fieldIn)./[1 Benchmark/2
Benchmark Benchmark*2 Benchmark*4];
        [minDiffer,position]=min(differ);
        %    if position==1
        %        pos='0';
        %    end
        %    if position==2
        %        pos='Benchmark/2';
        %    end
        %    if position==3
        %        pos='Benchmark';
        %    end
        %    if position==4
        %        pos='Benchmark*2';

```

```
% end
% if position==5
%     pos='Benchmark*4';
%     end
differ=[differ minDiffer position];

TempDiffers=setfield(TempDiffers,fieldListt{kk},differ);

end

Differs(m)=TempDiffers;

end
Output=Differs;
save(strcat(SavePath,'/',num2str(ii),'.mat'),'Output');

%save(strcat(SavePath,'/',Options,'.mat'),'Output');
end
```