ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ
ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Ηχοποίηση Βίντεο: Η Παράμετρος του Ρυθμού
Θεωρητική Προσέγγιση και
Υλοποίηση Ρυθμικής Συσχέτισης Βίντεο και Μουσικής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Φοίβος-Δημήτριος Γκούβας

Επιβλέπων: Γεώργιος Καμπουράκης
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάιος 2015

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ
ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Ηχοποίηση Βίντεο: Η Παράμετρος του Ρυθμού
Θεωρητική Προσέγγιση και
Υλοποίηση Ρυθμικής Συσχέτισης Βίντεο και Μουσικής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Φοίβος-Δημήτριος Γκούβας

Επιβλέπων: Γεώργιος Καμπουράκης
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή επιτροπή την 06/05/2015

.....	.....	.....
Καμπουράκης Γεώργιος	Λούμος Βασίλειος	Κουκούτσης Ηλίας
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα,Μάιος 2015

.....

Φοίβος-Δημήτριος Γκούβας

Διπλωματούχος Ηλεκτρολόγος Μηχανικός & Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Παρ' όλη την ευρέως διαδεδομένη αντίληψη πως το σύνολο της μουσικής δημιουργίας είναι προϊόν αισθητικής έμπνευσης και καλλιτεχνικών διαδικασιών, η μουσική σύνθεση πολύ συχνά στηρίζεται σε τυποποιημένους μηχανισμούς. Η αλγοριθμική σύνθεση μάλιστα στοχεύει ουσιαστικά στην εξ' ολοκλήρου δημιουργία μουσικής με χρήση τέτοιων αυτοματοποιημένων μεθόδων.

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η υλοποίηση μιας διαδικασίας αλγοριθμικής σύνθεσης ρυθμικών δομών. Αρχικά, γίνεται μια προσπάθεια να οριστεί με βάση προϋπάρχουσα έρευνα, ο ρυθμός ενός βίντεο ώστε να επιτευχθεί μηχανικά η σύνδεση του με το μουσικό ρυθμό. Ο αλγόριθμος που θα υλοποιεί τη σύνθεση στηρίζεται στη μετατροπή ενός video σε ήχο, μέσω μιας διαδικασίας που να μπορεί να εντάσσεται στα πλαίσια του video sonification. Έτσι, το τελικό αποτέλεσμα θα μπορεί να συνοδεύσει το οπτικό περιεχόμενο ή και να εξετάζεται ως αυτοτελής μουσική οντότητα. Το αποτέλεσμα αυτό προκύπτει ως εξής: Μέσω συγκεκριμένης ανάλυσης του βίντεο εισόδου σε MATLAB παράγεται ως ενδιάμεση έξοδος ένα αρχείο MIDI το οποίο κατόπιν τροφοδοτείται στο interface υλοποίησης, Ableton Live, όπου και παράγεται το τελικό αποτέλεσμα, "ερμηνευμένο" στα τύμπανα.

Λέξεις-κλειδιά

Ηχοποίηση, βίντεο, αλγοριθμική σύνθεση, ηχητική αντίληψη, ρυθμός, τύμπανα, MIDI

Abstract

Despite the popular belief that the whole of music creation is a result of aesthetic inspiration and artistic procedures, music composition is often based on standarized mechanisms. It is notable that the fundamental aim of algorithmic composition is the creation of music using exclusively such automated methods.

The objective of the present diploma thesis is to implement a procedure of algorithmic composition of rhythmic structures. Firstly, an effort is made to define the rhythm of a video, based on past research, so that we can automatically connect it with musical rhythm. The algorithm used to realize the composition is based on the transformation of a video into sound, through a procedure that can be characterized as video sonification. Thus, the result can accompany the visual content or be examined as an independent music form and it is produced as follows: An initial output in the form of a MIDI file is produced through a specific analysis of an input video in MATLAB, and it is consequently fed into the implementation interface, Ableton Live, where the final result is produced, "interpreted" on drums.

Keywords

Sonification, video, algorithmic composition, auditory perception, rhythm, pace, drums, MIDI

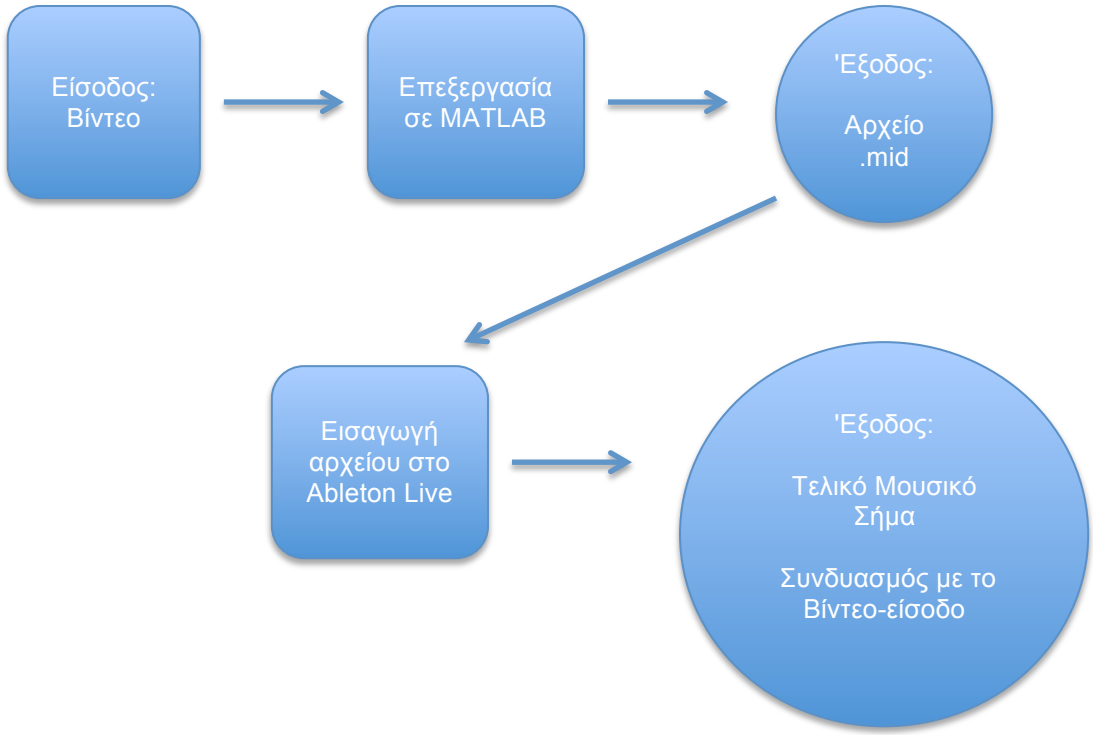
Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή
κ. **Γεώργιο Καμπουράκη** καθώς και τον υποψήφιο διδάκτορα
Κωνσταντίνο Μπακογιάννη για την υπερπολύτιμη βοήθειά τους,
την καθοδήγηση και τη στήριξή τους αλλά και τις κατευθυντήριες
προτάσεις τους καθ'όλη τη διαδικασία σύλληψης και συγγραφής της
παρούσας διπλωματικής εργασίας.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΚΕΦΑΛΑΙΟ 1 :
ΕΙΣΑΓΩΓΗ

1.1. ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ

Η συγκεκριμένη διπλωματική εργασία πραγματεύεται την αυτοματοποιημένη δημιουργία μουσικής μέσω της επεξεργασίας συγκεκριμένων παραμέτρων ενός βίντεο, το οποίο θεωρούμε ως είσοδο του συστήματος. Για το μουσικό αποτέλεσμα επιδιώκουμε κυρίως να έχει ρυθμικό περιεχόμενο και δευτερευόντως μελωδικό. Για την επίτευξη του σκοπού αυτού, θα γίνει χρήση του περιβάλλοντος MATLAB για την επεξεργασία της εισόδου και την παραγωγή ενός μουσικού αρχείου το οποίο θα είναι καθολικά "αναγνώσιμο" από μουσικά λογισμικά. Η έξοδος λοιπόν θα είναι ένα αρχείο .mid του πρωτοκόλλου MIDI, το οποίο θα μπορεί να δοθεί ως είσοδος στο γνωστό μουσικό λογισμικό Ableton Live. Μετά και από αυτό το στάδιο, το τελικό αποτέλεσμα θα είναι ένα ηχητικό σήμα προερχόμενο από ένα αυτόματο "παίξιμο" του ρυθμικού οργάνου της επιλογής μας (τύμπανα, μπάσο κ.ά.).



Ως δευτερεύουσα εφαρμογή, θα υλοποιηθούν επίσης παλαιότεροι ορισμοί για το μέγεθος του Ρυθμού ενός βίντεο. Αφού προσαρμοστούν στους περιορισμούς αλλά και τους στόχους της εν λόγω διπλωματικής εργασίας, με τη βοήθεια και πάλι του MATLAB οι ορισμοί αυτοί θα εφαρμοστούν σε τέσσερα διαφορετικά αποσπάσματα βίντεο με στόχο την εξαγωγή συμπερασμάτων σχετικά με τη λογική ορθότητα των ορισμών και το βαθμό σύνδεσης τους με τον αντίστοιχο μουσικό ρυθμό που θα μπορούσε να συνοδεύει τα συγκεκριμένα αποσπάσματα.

2. ΚΕΦΑΛΑΙΟ 2 :
ΑΝΑΣΚΟΠΗΣΗ SONIFICATION - ΑΛΓΟΡΙΘΜΙΚΗΣ ΜΟΥΣΙΚΗΣ

2.1. SONIFICATION

Η ελληνική απόδοση του όρου sonification θα ήταν ηχοποίηση αλλά λόγω της περιορισμένης χρήσης του, παρακάτω θα χρησιμοποιείται η αγγλική λέξη. Sonification ονομάζουμε γενικά τη μετατροπή δεδομένων σε ήχο. Παρ'όλο λοιπόν που σαν όρος δεν χρησιμοποιείται για μεγάλο χρονικό διάστημα, ως παραδείγματα "εφαρμογής" του μπορούν να θεωρηθούν διάφορα τετριμμένα φαινόμενα: από το κούρδισμα μιας κιθάρας (και την μετατροπή της μεταβολής της τάσης μιας χορδής της σε διαφορά στον τόνο) έως την ενεργοποίηση ενός συναγερμού σε περίπτωση παραβίασης μιας κλειδαριάς. Ο πιο ευρέως αποδεκτός ορισμός του sonification είναι αυτός του T.Hermann [1] :

Μια τεχνική που χρησιμοποιεί δεδομένα ως είσοδο και τα μετατράπει σε ηχητικά σήματα μπορεί να ονομάζεται sonification, εάν και μόνο εάν πληρούνται τα παρακάτω:

- *Ο ήχος αντανακλά αντικειμενικές ιδιότητες στα δεδομένα εισόδου*
- *Η μετατροπή γίνεται με συστηματικό τρόπο, δηλαδή υπάρχει ακριβής ορισμός του τρόπου που τα δεδομένα οδηγούν στην μεταβολή του ήχου*
- *Το sonification είναι ανεξάρτητο της διαδικασίας (reproducible), δηλαδή με τα ίδια δεδομένα εισόδου και τις ίδιες διαδράσεις ο εξαγόμενος ήχος έχει ακριβώς την ίδια δομή κάθε φορά που εφαρμόζεται η διαδικασία.*
- *Το σύστημα μπορεί να χρησιμοποιηθεί κατά τη βούληση του χρήστη και με διαφορετικά δεδομένα ή κατά εξακολούθηση με τα ίδια δεδομένα.*

Το sonification ως πρακτική είναι ακριβώς αντίστοιχη με την οπτικοποίηση (visualization) δεδομένων. Εκεί, τα στοιχεία της εισόδου αντιστοιχίζονται με παραμέτρους που μπορεί να αντιληφθεί το ανθρώπινο μάτι όπως το σχήμα και το χρώμα. Ένα πολύ απλό παράδειγμα είναι η δημιουργία γραφημάτων για την αντίληψη της μεταβολής ενός φυσικού μεγέθους συναρτήσει του χρόνου.Καθώς μάλιστα, το αυτί ως όργανο αντίληψης έχει μεγαλύτερο εύρος ζώνης από το μάτι, το sonification μπορεί να βοηθήσει να αντιληφθούμε φαινόμενα τα οποία δεν είναι εύκολο να συλλάβουμε οπτικά. Επίσης, έχει παρατηρηθεί ότι ο εγκέφαλος μπορεί να παρακουθήσει ευκολότερα ταυτόχρονα δεδομένα έαν γίνονται αντιληπτά ηχητικά παρά οπτικά. Το λεγόμενο *background listening (ακούγοντας στο παρασκήνιο)* βοηθά να αντιληφθούμε πολλαπλά "επίπεδα" ήχου. Γίνεται λοιπόν εμφανές ότι η τεχνική του sonification έχει πολύ μεγάλη ισχύ και για το λόγο αυτό χρησιμοποιείται σε πάμπολλες εφαρμογές: σε συναγερμούς και ειδοποιήσεις, σε συστήματα υποστήριξης για άτομα με προβλήματα όρασης καθώς και για διδακτικούς και καλλιτεχνικούς σκοπούς. Τέλος, ενδιαφέρον παρουσιάζει η εφαρμογή του ερευνητή Robert Alexander του πανεπιστημίου του Michigan, την οποία στήριξε σε δεδομένα της NASA από παρατηρήσεις σχετικά με τον ήλιο [2]. Τα δεδομένα αυτά, που λήφθηκαν σε διάστημα 43 χρόνων, περιείχαν και πληροφορίες από μη ορατά κομμάτια του φάσματος τα οποία είχαν αποτυπωθεί ελλιπώς με ψευδοχρώματα. Μέσω του sonification λοιπόν, η επιστημονική κοινότητα ήταν σε θέση να μελετήσει εκτενέστερα τα συγκεκριμένα δεδομένα ενώ ο Alexander ήταν σε θέση να παράγει και καλλιτεχνικό έργο βασιζόμενος στα ίδια δεδομένα.

2.2. VISUAL MUSIC

Θα χρησιμοποιούμε τον αγγλικό όρο καθώς οποιαδήποτε ελληνική απόδοση του δεν θα ήταν δόκιμη. Visual music ονομάζουμε τη χρήση μουσικών δομών για τη δημιουργία οπτικών απεικονίσεων, που μπορούν να περιλαμβάνουν βουβά φιλμ ή εικόνες που προκύπτουν από τη έντεχνη χρήση του φωτός (γνωστή ως πρακτική ως Lumia) [3]. Ο όρος χρησιμοποιήθηκε πρώτη φορά από τον καλλιτέχνη και κριτικό τέχνης Roger Fry για τη "μετάφραση" της μουσικής σε ζωγραφική, σε μια προσπάθεια να περιγραφεί το έργο του Wassily Kandinsky. Επίσης, χρησιμοποιείται για την αναφορά σε συστήματα που μετατρέπουν τη μουσική σε οπτικές μορφές (φιλμ,βίντεο ή γραφικά υπολογιστή) είτε με μηχανικό τρόπο είτε μέσω της ανθρώπινης ερμηνείας από έναν καλλιτέχνη. Πολλοί κινηματογραφιστές έχουν εργαστεί σε αυτήν την κατεύθυνση ή στην αντίστροφη, τη μετατροπή δηλαδή εικόνων σε ήχο,κάτι που με σημερινούς όρους θα μπορούσαμε να αποκαλέσουμε image sonification. Η πρακτική συχνά ονομάζεται και color music καθώς ιστορικά συνδέεται και με τη δημιουργία και τον πειραματισμό με τα "όργανα χρωμάτων" (color organs),τα οποία μπορούσαν να "παράγουν" φως και να το διαμορφώσουν με τρόπο και ροή αντίστοιχη με αυτή ενός μουσικού έργου. Ορισμένοι σημαντικοί καλλιτέχνες που έχουν ασχοληθεί με το αντικείμενο είναι ο Walter Ruttmann με τα *Lichtspiel: Opus I* και *Opus II* (στα ελληνικά "Παιχνίδι με το φως"), ο Viking Eggeling που έχει συνδεθεί με το κίνημα Dada καθώς και ο Oskar Fischinger με τα κλασικά έργα *Motion Painting No.1* και An *Optical Poem* μεταξύ άλλων. Τα δύο τελευταία ήταν ουσιαστικά προσπάθειες του Fischinger να αποδόσει οπτικά και να συνδέσει με εικόνα τα μουσικά έργα *Brandenburg Concerto no. 3, BWV 1048* του Johann Sebastian Bach και *Second Hungarian Rhapsody* του Franz Liszt αντίστοιχα. Στην εισαγωγή του *An Optical Poem* μάλιστα ο Fischinger γράφει:

Στους περισσότερους από εμάς η μουσική υπονοεί συγκεκριμένες νοητικές εικόνες μορφών και χρωμάτων. Το φιλμ που θα δείτε αποτελεί ένα πρωτότυπο επιστημονικό πείραμα - το αντικείμενο του είναι να μεταφέρει αυτές τις νοητικές εικόνες σε οπτική μορφή.

Εικόνα 2.1 : Στιγμιότυπο από το An Optical Poem

Πλέον είναι πολύ διαδεδομένα διάφορα λογισμικά που έχουν τη δυνατότητα να μετατρέπουν ένα ηχητικό σήμα σε κινούμενη εικόνα ή να δουλεύουν αντίστροφα μετατρέποντας εικόνες σε μουσικά αποσπάσματα. Η διάδοση της visual music μάλιστα είναι τέτοια που ένας καινούριος "τομέας" καλλιτεχνικής δημιουργίας ασχολείται με τη ζωντανή εκτέλση αυτής. Ο τομέας αυτός είναι το VJing.

Ο όρος αυτός προκύπτει ακριβώς αντίστοιχα με την πρακτική του DJing από το αγγλικό Video Jockey (όπως ο DJ προκύπτει από το Disc Jockey) κι η αρχική του ερμηνεία αφορά την επιλογή και την αναπαραγωγή οπτικών εφέ και βίντεο. Το VJing ωστόσο με την εξέλιξη και τη δημοφιλία του έχει καταλήξει ξεχωριστή πρακτική που αφορά τη δημιουργία και διαμόρφωση οπτικών εφέ σε συγχρονισμό με μουσική σε ζωντανό χρόνο ενώ βρίσκει εφαρμογές σε συναυλίες, DJ sets και κάθες είδους παραστάσεις. "Πρόγονος" του VJing και της τεχνολογίας που το συνοδεύει θεωρούνται τα color organs που προαναφέρθηκαν.

Στην πιο δίπλα φωτογραφία φαίνεται η πιανίστρια Mary Hallock-Greenewalt με την εφεύρεση της Serabat, για τη δημιουργίας της visual music που η ίδια αποκαλούσε *Nourathar* ("υπόσταση του φωτός" στα αραβικά). Το Sarabet εφευρέθηκε μεταξύ 1919 και 1927 και μπορούσε να παράγει μια κλίμακα εντάσεων φωτός και χρωμάτων και ο χειρισμός του έμοιαζε με τον μοντέρνο επαγγελματικό εξοπλισμό μουσικής μίξης: Στηριζόταν σε ολισθητήρες (sliders) για την ρύθμιση των εντάσεων καθώς και σε πετάλια για τα πόδια όμοια με του πιάνου και διακόπτες δύο θέσεων.

Εικόνα 2.2: Η Mary-Hallock Greenewalt και το Serabat

Επίσης, τα οπτικά σόου που συνόδευαν τις συναυλίες από τη δεκαετία του 1960 ήδη μπορούν να θεωρηθούν ως ένα προστάδιο του VJing: Οι καλλιτεχνικές ομάδες *The Joshua Light Show* και *The Brotherhood of Light* είχαν αναλάβει να συνοδεύουν τις ζωντανές εμφανίσεις των Grateful Dead ενώ ο Andy Warhol την ίδια περίοδο διοργάνωνε τα Exploding Plastic Inevitable, εκδηλώσεις δηλαδή για τη σύνδεση μουσικής, χορού και βίντεο και οπτικών εφέ. Έκτοτε, η όλο και μεγαλύτερη διάδοση του βίντεο, η ανάδειξη της κουλτούρας του βίντεο κλιπ και η τεράστια επιτυχία των μουσικών καναλιών MTV και VH1 οδήγησαν στην άρρηκτη σύνδεση της μουσικής με το οπτικό περιεχόμενο. Σε αυτό προφανώς συντέλεσε και η ανάπτυξη πρωτοποριακής τεχνολογίας όπως τα video mixers (μείκτες βίντεο) WJ-MX50 της Panasonic και το Videonics MX-1 αλλά και τα πάμπολλα λογισμικά για VJing με αρχαιότερο το Vujak του 1992. Έτσι, το VJing έχει φτάσει να είναι αναπόσπαστο κομμάτι πολλών συναυλιών μεγάλων ονομάτων της μουσικής βιομηχανίας παράγοντας συχνά έργα πρωτοποριακής αισθητικής (όπως το σόου ISAM του γνωστού παραγωγού ηλεκτρονικής μουσικής Amon Tobin) ενώ ταυτόχρονα ανθίζει και ως "αυτόνομη μορφή" τέχνης.

2.3. ΜΟΥΣΙΚΗ

Γενικά, η μουσική είναι η μορφή τέχνης εκείνη που έχει ως μέσο τον ήχο και στηρίζεται στην οργάνωση του με σκοπό την σύνθεση,εκτέλεση και ακρόαση ενός έργου. Το όνομα προέρχεται από την αρχαία Ελλάδα και τις εννά Μούσες. Θεωρούνταν ως η *Απολλώνια Τέχνη* και περίκλειε την *Ποίηση*, το *Μέλος* και το *Χορό* ως αναπόσπαστα κομμάτια της. Ένας θεμελιώδης λοιπόν ορισμός της είναι αυτός που δίνεται από το Oxford Universal Dictionary [4]:

Μουσική είναι εκείνη εκ των καλών τεχνών που αφορά το συνδυασμό των ήχων με στόχο την ομορφιά της μορφής και την έκφραση σκέψεων ή συναισθημάτων.

Ωστόσο, ο παραπάνω ορισμός είναι κάτι παραπάνω από αμφιλεγόμενος εάν αναλογιστούμε πως μεγάλο μέρος της μοντέρνας μουσικής δημιουργίας περικλείει το υποείδος που είναι γνωστό ως noise music ("μουσική θορύβου"), το οποίο αμφισβητεί την κυρίαρχη ιδέα περί των αισθητικών στοιχείων της μουσικής. Διάσημο παράδειγμα του διλήμματος που προκύπτει είναι το έργο του John Cage με τίτλο *4'33"*. Το έργο αυτό έχει τρία μέρη και κατευθύνει τους εκτελεστές του να εμφανιστούν επί σκηνής,
να σηματοδοτήσουν με κάποια χειρονομία την έναρξη του έργου και έπειτα να μην παράγουν κανέναν ήχο μέχρι το τέλος του έργου που σηματοδοτείται πάλι με κάποια χειρονομία. Πολλοί θεωρούν πως το έργο αυτό δεν εμπίπτει στα όρια της μουσικής καθώς δεν περιέχει κανέναν ήχο εν γένει "μουσικό" και ο εκτελεστής του δεν έχει κάποιον έλεγχο επί των όποιων ήχων ακούγονται (από κινήσεις του ακροατηρίου ή του εκτελεστή ή από αντιδράσεις των θεατών). Άλλοι πάλι επιχειρηματολογούν πως τα όρια της συμβατικής μουσικής είναι περιορισμένα με λανθασμένα και αυθαίρετα κριτήρια και για το λόγο αυτό δεν υπάρχει λόγος να αποκλείσουμε το 4'33" από τον χαρακτηρισμό "μουσική". 'Ετσι λοιπόν έχουν υπάρξει πολλοί νεότεροι ορισμοί περί της φύσης της μουσικής: Ο μοντερνιστής συνθέτης Edgar Varèse λόγω της ίδιας του της αισθητικής και του οράματος του για τον ήχο ως μια "ζώσα ύλη", αναφέρεται στη μουσική απλά ως *"οργανωμένο ήχο"* . Ακόμη, η διάσημη Encyclopædia Britannica αναφέρει στην 15η έκδοσή της πως

"ενώ δεν υπάρχουν ήχοι οι οποίοι δεν μπορούν να περιγραφούν ως εγγενώς μουσικοί, οι μουσικοί καλλιτέχνες στις διάφορες κουλτούρες τείνουν να περιορίζουν το εύρος των ήχων που αποδέχονται ως τέτοιους" .

Άλλοι πάλι έχουν επιχειρήσει να ερμηνεύσουν τη φύση της μουσικής με πιο κοινωνικά κριτήρια. Χαρακτηριστικά παραδείγματα ο μουσικολόγος Jean-Jacques Nattiez ο οποίος αναφέρει μεταξύ άλλων:

"όπως ακριβώς μουσική είναι ότι οι άνθρωποι επιλέγουν να αναγνωρίσουν ως τέτοια, έτσι ακριβώς και θόρυβος είναι οτιδήποτε αναγνωρίζεται ως ενοχλητικό, δυσάρεστο ή και τα δύο" , καθώς επίσης κι ότι

"τα όρια μεταξύ μουσικής και θορύβου ορίζονται πάντα πολιτισμικά, που σημαίνει ότι ακόμη και εντός της ίδιας κοινωνίας τα όρια αυτά δεν ισχύουν όμοια για όλους, εν συντομία δηλαδή σπάνια υπάρχει ομοφωνία. Σίγουρα πάντως δεν υπάρχει μοναδική ή διαπολιτισμική οικουμενική αντίληψη που να ορίζει τι είναι μουσική" [5]

αλλά και ο Jean Molino:

"η μουσική, που συχνά είναι τέχνη/ψυχαγωγία, είναι ένα απόλυτα κοινωνικό γεγονός ο ορισμός του οποίου μεταβάλλεται ανάλογα με την εποχή και την κουλτούρα" [6] .

Τέλος αξίζει να παραθέσουμε και τον (φιλοσοφικό) ορισμό του Ιάννη Ξενάκη από το έβδομο κεφάλαιο του *Formalized Music*, με τίτλο *Towards a Metamusic* ("*Προς μια Μεταμουσική*") [7]:
Η μουσική:
1. Είναι ένα είδος συμπεριφοράς απαραίτητο για όποιο την σκέφτεται και την δημιουργεί
2. Είναι ένα ξεχωριστό πλήρωμα (στα αγγλικά pleroma), μία πραγμάτωση
3.Είναι ένας καθορισμός σε ήχο νοητικών εικονικοτήτων(κοσμολογικών, φιλοσοφικών)
4. Είναι κανονιστική, δηλαδή υποσυνείδητα είναι ένα μοντέλο για το τι είναι ή πράττει κάποιος με συμπονετική διάθεση.
5. Είναι καταλυτική:η ίδια της η ύπαρξη επιτρέπει εσωτερικές ψυχικές ή νοητικές μεταμορφώσεις με τον ίδιο τρόπο που επιτρέπει η κρυστάλλινη μπάλα του υπνωτιστή.
6. Είναι το άσκοπο παιχνίδι παιχνίδι ενός παιδιού.
7. Είναι μια μυστικιστική (αλλά αθεϊστική) ασκητική. Γι'αυτό εκφράσεις λύπης,χαράς,αγάπης καθώς και δραματικές καταστάσεις δεν είναι παρά κάποια πολύ περιορισμένα στιγμιότυπα.

2.4. ΑΛΓΟΡΙΘΜΙΚΗ ΜΟΥΣΙΚΗ

2.4.1 ΟΡΙΣΜΟΣ

Παρ'όλο που συχνά χρησιμοποείται ο όρος *αλγοριθμική μουσική*, πιο δόκιμο θα ήταν το *υπολογιστική μουσική* ή *αλγοριθμική σύνθεση*. Όπως είναι γνωστό *αλγόριθμος* ονομάζεται μια σειρά εντολών αυστηρά δομημένων, σαφών και εκτελέσιμων που σκοπό έχουν την επίλυση ενός προβλήματος. Έτσι λοιπόν, "σκοπός" της αλγοριθμικής σύνθεσης είναι η επίλυση του "προβλήματος " της σύνθεσης ενός μουσικού έργου τελείως αυτοματοποιημένα με τη βοήθεια ενός αλγορίθμου όπως περιγράφηκε παραπάνω. Ο όρος σήμερα συνδέεται κυρίως με την παραγωγή μουσικής χωρίς την ανθρώπινη παρέμβαση, με χρήση ηλεκτρονικών υπολογιστών ή με διαδικασίες που εμπλέκουν την τυχαιότητα. Τυπικά ωστόσο, προϋπήρχε ως διαδικασία των ηλεκτρονικών υπολογιστών: Στη δυτική αντίστιξη για παράδειγμα, ο καθορισμός των μελοδικών κινήσεων των διαφορετικών φωνών, η λεγόμενη φωνοδήγηση (voice-leading), μπορούμε να θεωρήσουμε ότι προκύπτει μέσα από την εκτέλεση ενός αλγορίθμου.

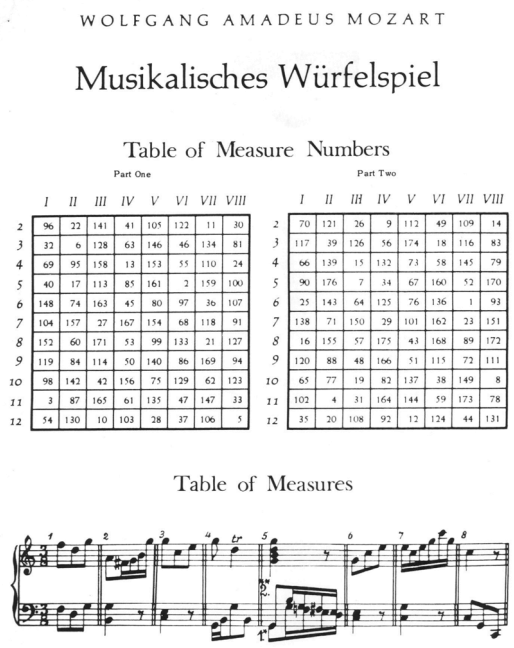
2.4.2 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ

Ως προάγγελος της ιδέας για σύνδεση της μουσικής με μία επιστημονική, μηχανοποιημένη διαδικασία μπορεί να θεωρηθεί η πυθαγόρεια φιλοσοφία. Ο Πυθαγόρας θεωρούσε πως η μελέτη της μουσικής θα έπρεπε να είναι αδιάσπαστη από τη μελέτη των μαθηματικών στο πλαίσιο μιας ευρύτερης προσπάθειας κατανόησης και ερμηνείας της φύσης. Το γεγονός όμως ότι όλη η αρχαία ελληνική μουσική παραγωγή προέκυπτε κατά βάση αυτοσχεδιαστικά,

χωρίς κάποια συγκεκριμένη μεθοδολογία σύνθεσης, αποτρέπει την απευθείας σύνδεση της αλγοριθμικής μουσικής με την αρχαία Ελλάδα.

Ως πρώτο παράδειγμα συστηματοποιημένης σύνθεσης ωστόσο, μπορούμε να θεωρήσουμε τη σύνθεση με τη χρήση ενός *κανόνα* προς τα τέλη του 15ου αιώνα: Ο συνθέτης αφότου συνέθετε τη μελωδία της πρώτης φωνής, επέλεγε έναν κανόνα για το πως θα προέκυπταν οι υπόλοιπες φωνές. Για παράδειγμα, μπορούσε να επιλέξει η δεύτερη φωνή να μιμηθεί την πρώτη ξεκινώντας κάποια μέτρα μετά ή να αποτελεί αναστροφή της πρώτης.

Στη δυτική μουσική παράδοση επίσης μπορούμε να εντοπίσουμε και αυτοματοποιημένες διαδικασίες σύνθεσης που περιλάμβαναν το στοιχείο της τυχαιότητας, ώστε ίδιες εκτελέσεις του "αλγορίθμου" να παράγουν διαφορετικά αποτελέσματα κάθε φορά. Ο Wolfgang Amadeus Mozart λέγεται πως ήταν εμπνευστής ενός τέτοιου μη ντετερμινιστικού συστήματος (αν και ο ισχυρισμός δεν έχει εξακριβωθεί), το οποίο ονομαζόταν Musikalisches Wurfelspiel, δηλαδή μουσικό παιχνίδι ζαριών. Έχοντας στη διάθεση του διάφορα μουσικά αποσπάσματα, ο παίχτης ρίχνοντας τα ζάρια επέλεγε τυχαία ποια μέρη να συνδυάσει και με ποιο τρόπο ώστε να προκύψει η τελική σύνθεση. Το "παιχνίδι" είχε σχεδιαστεί για τη σύνθεση βαλς, έχοντας τη δυνατότητα να παράγει $11^{16} = 45949729863572161$ διαφορετικά αλλά όμοια αποτελέσματα. Ο ίδιος ο Mozart μάλιστα υποτίθεται ότι είχε ονομάσει κάποιες από τις συνθέσεις του *"Anleitung zum Componieren von Walzern so viele man will vermittelst zweier Würfel, ohne etwas von der Musik oder Composition zu verstehen"* δηλαδή *"οδηγίες για τη σύνθεση όσων βαλς επιθυμείτε με δύο ζάρια, χωρίς να καταλαβαίνετε το παραμικρό για μουσική και σύνθεση"* .



WOLFGANG AMADEUS MOZART

Musikalisches Würfelspiel

Table of Measure Numbers

Part One

	I	II	III	IV	V	VI	VII	VIII
2	96	22	141	41	105	122	11	30
3	32	6	128	63	146	46	134	81
4	69	95	158	13	153	55	110	24
5	40	17	113	85	161	2	159	100
6	148	74	163	45	80	97	36	107
7	104	157	27	167	154	68	118	91
8	152	60	171	53	99	133	21	127
9	119	84	114	50	140	86	169	94
10	98	142	42	156	75	129	62	123
11	3	87	165	61	135	47	147	33
12	54	130	10	103	28	37	106	5

Part Two

	I	II	III	IV	V	VI	VII	VIII
2	70	121	26	9	112	49	109	14
3	117	39	126	56	174	18	116	83
4	66	139	15	132	73	58	145	79
5	90	176	7	34	67	160	52	170
6	25	143	64	125	76	136	1	93
7	138	71	150	29	101	162	23	151
8	16	155	57	175	43	168	89	172
9	120	88	48	166	51	115	72	111
10	65	77	19	82	137	38	149	8
11	102	4	31	164	144	59	173	78
12	35	20	108	92	12	124	44	131

Table of Measures

Εικόνα 2.3: Musikalisches Wurfelspiel

Έκτοτε, η αλγοριθμική συνθεση γνώρισε μεγάλη άνοδο ιδιαίτερα την περίοδο μετά τον Β' Παγκόσμιο Πόλεμο. Ανάμεσα στις εξελίξεις που σημειώθηκαν ξεχωρίζουν οι μέθοδοι του δωδεκαφθογγισμού (twelve-tone technique) και γενικότερα του σειριαλισμού (serialism) που προέκυψαν από το έργο του αυστριακού Arnold Schoenberg. Οι τεχνικές αφορούν στον καθορισμό πολλών παραμέτρων ενός κομματιού, νότες, ρυθμό ή και δυναμικές, ώστε να εξασφαλίζεται ότι όλη η χρωματική κλίμακα χρησιμοποιείται εξίσου χωρίς να δίνεται ιδιαίτερη έμφαση σε κάποια συγκεκριμένη νότα. Ομολογουμένως όμως, η μεγαλύτερη πρόοδος της αλγοριθμικής σύνθεσης προήλθε με την εξέλιξη και τη χρησιμοποίηση των ηλεκτρονικών υπολογιστών.

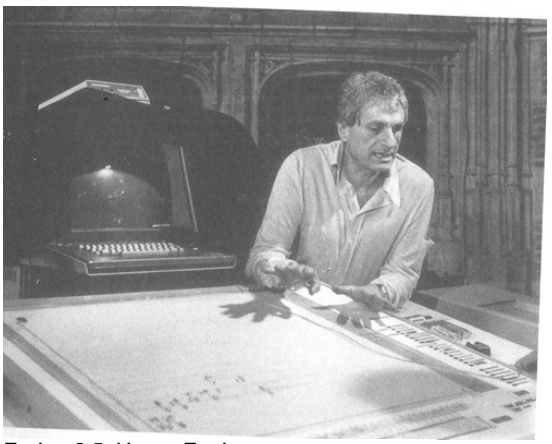
Η σημασία που θα είχε ο ηλεκτρονικός υπολογιστής για μια πιθανή αυτόματη σύνθεση μουσικής είχε γίνει αντιληπτή ακόμα και πριν την εφεύρεση του:
Η Ada Lovelace, εφευρέτρια της υπολογιστικής μηχανής-προδρόμου του Η/Υ, είχε προβλέψει ότι ένας μελλοντικός υπολογιστής θα μπορεί πιθανώς να συνθέτει λεπτομερή και επιστημονικά μουσικά κομμάτια οποιασδήποτε πολυπλοκότητας και μεγέθους.

Ο πρώτος που παρουσίασε μουσικό έργο έχοντας κάνει χρήση ηλεκτρονικού υπολογιστή ήταν ο Lejaren Hiller μαζί με τον Leonard Isaacson στο πανεπιστήμιο του Illinois το 1955. Χρησιμοποιώντας τον ψηφιακό υπολογιστή Illiac, κατάφεραν να συνθέσουν το *Illiac Suite* για κουαρτέτο εγχόρδων. Ουσιαστικά, ο αλγόριθμος τους παρήγαγε τυχαίους αριθμούς που αντιστοιχούσαν σε διαφορετικές νότες με έναν προκαθορισμένο τρόπο και έπειτα με ρυθμικά μοτίβα ή δυναμικές. Έπειτα, βάση της μουσικής θεωρίας ορίζονταν συγκεκριμένοι έλεγχοι ώστε τα αυθαίρετα αρχικά αποτελέσματα να φιλτράρονται για να παραχθεί το τελικό αποτέλεσμα.
Το έργο αποτελούνταν από τέσσερις κινήσεις, αντίστοιχων τεσσάρων ανεξάρτητων πειραμάτων: Το πρώτο αφορούσε στον καθορισμό κάποιων σταθερών μελωδιών (cantus firmi), το δεύτερο στην παραγωγή αποσπασμάτων με τέσσερις φωνές με διάφορους κανόνες, το τρίτο στα ρυθμικά μοτίβα, τις δυναμικές και τις οδηγίες εκτέλεσης και το τέταρτο σε μοντέλα για γενετικές γραμματικές και αλυσίδες Markov.
Ο Hiller συνεργάστηκε επίσης με τον Robert Baker και με τη χρήση ενός άλλου υπολογιστικού συστήματος, του MUSICOMP,δημιούργησαν το έργο Computer Cantata.

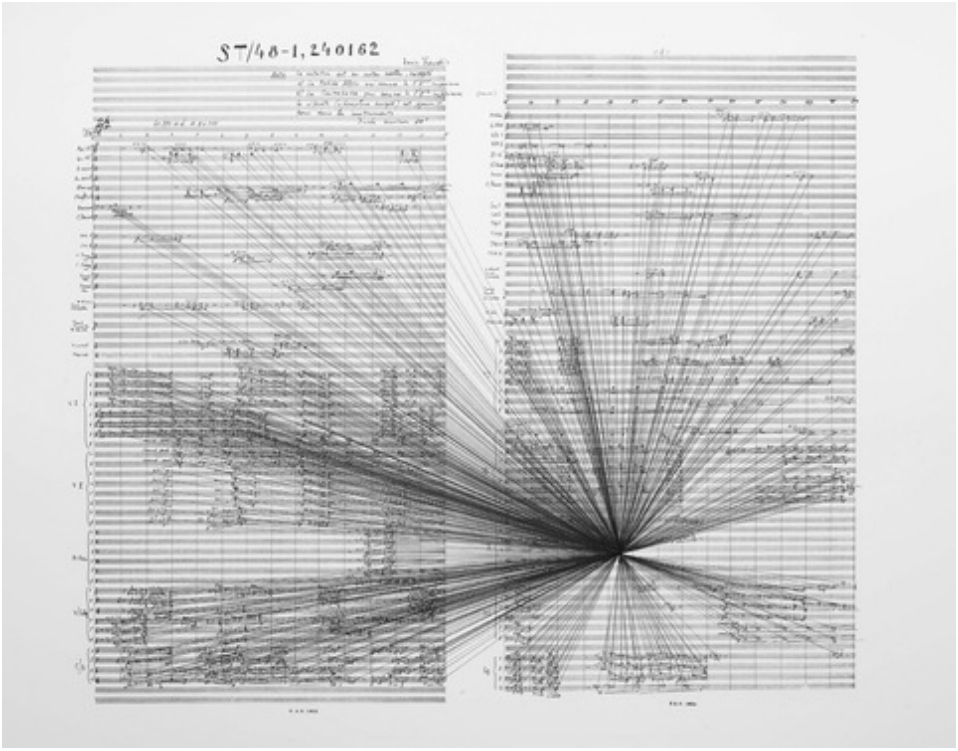
Εικόνα 2.4:Ο Lejaren Hiller και ο Illiac

Για την ολοκλήρωσή του στηρίχτηκαν σε καθαρά προγραμματιστικές τεχνικές, χρησιμοποιώντας για παράδειγμα υπορουτίνες ανεξάρτητες η μία από την άλλη, για τη σύνθεση μικρών αποσπασμάτων, τα οποία στο τέλος συνενώνονταν για να δώσουν το τελικό αποτέλεσμα.

Λίγο αργότερα, τη δεκαετία του 1960 ξεχωρίζουμε το έργο του σπουδαίου Ιάννη Ξενάκη. Το 1962 συγκεκριμένα, ο Ξενάκης παρουσίασε τα *ST/4, ST/10, Atrées* και *Morsima-Amorsima* στηριζόμενος σε στοχαστικές διαδικασίες: Χρησιμοποιώντας στατιστικά μοντέλα, έδινε ως είσοδο μια λίστα από νότες και ορισμένα πιθανοτικά βάρη και ο υπολογιστής παρήγαγε ως έξοδο την παρτιτούρα ενός νέου κομματιού. Δευτερεύουσες αποφάσεις σχετικά με τη δομή του έργου προέρχονταν πάλι από τον υπολογιστή με βάση τυχαίους αριθμούς. Το τελικό έργο ερμηνευόταν από ζωντανούς μουσικούς σε αντίθεση με το *Illiac suite* που εκτελούνταν από τον ίδιο τον Illiac στο πλαίσιο της προσπάθειας του Hiller για πλήρη "δημιουργική" παραγωγή του υπολογιστή.



Εικόνα 2.5: Ιάννης Ξενάκης



Εικόνα 2.6: Η "παρτιτούρα" για το *ST/48* του Ξενάκη

Σημαντική θεωρείται και η συνεισφορά του John Cage που αναφέρθηκε παραπάνω για το έργο *4'33"* και του Karlheinz Stockhausen. Και οι δύο σημαντικοί συνθέτες ασχολήθηκαν εκτεταμένα με την τυχαιότητα και την ένταξη της στην καλλιτεχνική παραγωγή: Ο Cage συγκεκριμένα συνεργάστηκε με τον Hiller για τη σύνθεση του *HPSCHD* που στην ουσία είναι ένα κολάζ διάφορων αποσπασμάτων έργων του Mozart παιγμένων σε σειρά που καθόριζε ένας υπολογιστής μέσω γεννητριών τυχαίων αριθμών, ενώ ο Stockhausen στο *Klaveirstucke XI* ορίζει πάλι πως διάφορα αποσπάσματα πρέπει να εκτελεστούν από τους μουσικούς σε τυχαία σειρά.

Αξίζει τέλος να αναφερθούμε σε ορισμένες πιο σύγχρονες μεθόδους που χρησιμοποιούνται στην αλγοριθμική σύνθεση. Μία από αυτές είναι η χρήση της Τεχνητής Νοημοσύνης. Γενικά, ένα σύστημα τεχνητής νοημοσύνης χαρακτηρίζεται από ορισμένους προκαθορισμένους κανόνες και μια δική του γραμματική έχοντας έτσι τη δυνατότητα να παράγει με αυτόματες μεθόδους συλλογιστικής νέα "γνώση" εντός του αρχικού συστήματος. Το σύστημα που ανέπτυξε ο David Cope, *Experiments in Musical Intelligence (EMI)* αποτελεί ένα παράδειγμα μουσικής εφαρμογής όλης αυτής της θεωρίας [8]. Το EMI περιλαμβάνει μια μεγάλη βάση δεδομένων για την περιγραφή μουσικών ειδών και κανόνες για διάφορες "στρατηγικές" σύνθεσης. Επίσης, δοθέντων έργων ενός συνθέτη μπορεί να επεκταθεί και να παράγει μια δικιά του γραμματική και βάση κανόνων, για την παραγωγή νέων μουσικών έργων. Συγκεκριμένα, το EMI έχει παράγει έργα που προσέγγιζαν επιτυχώς τα ιδιαίτερο στυλ των Bach, Mozart, Brahms κ.ά.Άλλο παράδειγμα της Τεχνητής Νοημοσύνης στη μουσική σύνθεση αποτελεί ο *γενετικός προγραμματισμός (genetic programming).* Σ'αυτόν τον τύπο προγραμματισμού, ο χρήστης εισάγει εργαλεία στο σύστημα, όπως βιβλιοθήκες συναρτήσεων, και διαμορφώνει μια συνάρτηση-"κριτή" για την αξιολόγηση των αποτελεσμάτων. Με βάση λοιπόν τις εισόδους και τις προδιαγραφές του "κριτή" το πρόγραμμα προσπαθεί να παράγει επιθυμητά αποτελέσματα και αναλαμβάνει το ίδιο να τα "κρίνει": Παράγονται διάφορες συνθέσεις,αξιολογούνται κι η όλη διαδικασία επαναλαμβάνεται έως ότου προκύψει το βέλτιστο αποτέλεσμα. Ο γενετικός προγραμματισμός λοιπόν είναι μια ακραία μορφή αλγοριθμικής σύνθεσης αφού υπάρχει ένα νέο επίπεδο αφαίρεσης: Τόσο το τελικό αποτέλεσμα, όσο και η μηχανοποιημένη διαδικασία που το παράγει δημιουργείται αυτόματα.

Συνοψίζοντας, θα πρέπει σε κάθε περίπτωση και σε κάθε μορφή αλγοριθμικής σύνθεσης να θυμόμαστε πως ο τελικός ακροατής δεν είναι σε θέση να γνωρίζει τη σκέψη ή τη δημιουργική διαδικασία του καλλιτέχνη παρά μόνο το τελικό αποτέλεσμα. Για το λόγο αυτό, κατά καιρούς έχουν διατυπωθεί διάφορα κριτήρια τα οποία θα πρέπει να πληρεί ένας αλγόριθμος σύνθεσης και τα αποτελέσματα του, με χαρακτηριστικό παράδειγμα αυτα του Donald Knuth [9]: Ο αλγόριθμος θα πρέπει να περιέχει απλότητα, οικονομία, κομψότητα και να χειρίζεται σχετικά εύκολα. Ακόμα και αυτά τα υποτυπώδη πάντως να μην πληρούνται, τις περισσότερες φορές θεωρείται απαραίτητο ο αλγόριθμος να παρουσιάζει μια πρωτοτυπία και το ηχητικό του εξαγόμενο να είναι, όσο δύσκολο κι αν είναι να οριστεί αντικειμενικά, ικανοποιητικό.

2.5. BEAT TRACKING

Άλλη μια αυτοματοποιημένη διαδικασίας εξαγωγής μουσικών αποτελεσμάτων, που θα μπορούσε να θεωρηθεί μέρος της αλγοριθμικής μουσικής είναι η διαδικασία του Beat Tracking, που συνδέεται άμεσα με το αντικείμενο της παρούσας εργασίας.

Η αναγνώριση των χρονικών θέσεων των πιο ισχυρών χτύπων ενός μουσικού αποσπάσματος (των άρσεων και θέσεων ή upbeats και downbeats στα αγγλικά) είναι μια θεμελιώδης μουσική ικανότητα η οποία στους περισσότερους ανθρώπους προκύπτει φυσικά: Όλοι μπορούμε ακόμα και σε σχετικά ρυθμικά περίπλοκα κομμάτια να "ακολουθήσουμε" τη μουσική με ένα απλό τρόπο, είτε χτυπώντας παλαμάκια είτε χτυπώντας το πόδι, κάνοντας δηλαδή "foot tapping".

Ένα σύστημα ικανό να διακρίνει και να παρακολουθεί το ρυθμό ενός μουσικού αποσπάσματος τελείως αυτοματοποιημένα, είναι πολύ σημαντικό για συστήματα αυτοσχεδιασμού μεταξύ ανθρώπου και υπολογιστή, για την καταγραφή μουσικών κομματιών, την επεξεργασία τους και το συγχρονισμό τους με άλλα καθώς και για μουσικολογικές μελέτες. Μάλιστα, είναι γνωστό πως για έναν μουσικό είναι δυσκολότερο να ακολουθήσει το ρυθμό του υπόλοιπου συνόλου παρά να επιβάλλει ένα δικό του τέμπο για να ακολουθήσουν οι υπόλοιποι. Έτσι, με ένα αυτόματο σύστημα beat tracking ο υπολογιστής και τα διάφορα μουσικά λογισμικά μπορούν να είναι πλέον αυτοί που ακολουθούν το ρυθμό και όχι αυτοί που τον ορίζουν, κάνοντας τη διάδραση με ανθρώπους μουσικούς σαφώς ευκολότερη.

Στην κατεύθυνση αυτή, έχει ήδη υπάρξει σημαντική έρευνα αλλά και πρακτικά συστήματα ως εφαρμογή αυτής. Μία από τις πρώτες προσπάθειες ήταν αυτή του Longuet-Higgins [10]. Η κεντρική ιδέα της προσέγγισης του, είναι πως οποιαδήποτε νότα που παίζεται κοντά σε ένα αναμενόμενο beat, θα πρέπει να είναι όντως beat. Ορίζοντας ως T_i έναν υποψήφιο "χτύπο" και ε την επιθυμητή απόκλιση που θα ορίζει την έννοια του "κοντά", οποιαδήποτε νότα παρατηρείται στο διάστημα $[T_i - \varepsilon\ ,\ T_i + \varepsilon]$ θα πρέπει να θεωρηθεί χτύπος. Τα υποψήφια beats προκύπτουν μέσα από το συνολικό έργο που ανέπτυξε ο Longuet-Hollins κι ήταν μέρος ενός υπολογιστικού μοντέλου για την αντίληψη της δυτικής κλασσικής μελωδίας.

Έπειτα από αυτή την πρώτη προσέγγιση, έχουν ακολουθηθεί διάφορες μέθοδοι: Το 1990 οι Allen και Dannenberg [11] επιχείρησαν να εντοπίσουν το beat ενός μουσικού αποσπάσματος σε ζωντανό χρόνο χρησιμοποιώντας μεθοδολογία της Τεχνητής Νοημοσύνης: Παραμετροποιώντας τις πιθανές χρονικές θέσεις των χτύπων μέσω ενός ιστορικού προηγούμενης ανάλυσης αλλά και βελτιώσεων του ανάλογα με τα νέα αποτελέσματα, εφαρμόζεται ο αλγόριθμος αναζήτησης beam search ώστε να βρεθούν ταυτόχρονα πολλαπλά "σενάρια" για το υπάρχον beat ώστε πάντα να τηρούνται ορισμένα μουσικά κριτήρια.

Ακόμη έχουν αναπτυχθεί πιθανοτικές μέθοδοι βασιζόμενοι είτε σε φίλτρα Kalman[12] είτε σε μοντέλα Bayes [13] . Ακόμα, οι Sethares et al [14] στηρίζονται καθαρά στην επεξεργασία του ήχου ως κυματομορφή και βασιζόμενοι στα θεμελιώδη χαρακτηριστικά του όπως η ενέργεια του και η ανάλυση του τόσο στο πεδίο του χρόνου όσο και της συχνότητας. Τέλος, το 2009 ο Daniel Ellis στηρίζεται στη χρήση δυναμικού προγραμματισμού για να κάνει beat tracking [15]: Θεωρώντας πως δίνεται εκ των προτέρων ένας σταθερός ρυθμός, στοχεύει να υπολογίσει μια ακολουθία θέσεων των χτύπων ενός ρυθμού που αντιστοιχούν στους χτύπους του ηχητικού αποσπάσματος-εισόδου ενώ αποτελούν από μόνοι μια ρυθμική οντότητα. Χρησιμοποιώντας λοιπόν μια κατάλληλα διαμορφωμένη ευριστική συνάρτηση εξάγει την ακολουθία αυτή, επιδιώκοντας τη μεγιστοποίηση της ευριστικής χρησιμοποιώντας δυναμικό προγραμματισμό.

3. ΚΕΦΑΛΑΙΟ 3 : ΡΥΘΜΟΣ

3.1. ΟΡΙΣΜΟΣ ΡΥΘΜΟΥ

Σύμφωνα με τον αρχαιοελληνικό ορισμό, ρυθμός είναι κάθε επαναλαμβανόμενη κίνηση ή και η συμμετρία. Το έγκυρο λεξικό του πανεπιστημίου της Οξφόρδης για την αγγλική γλώσσα ορίζει πως ως ρυθμό αντιλαμβανόμαστε γενικά *"την κίνηση που σηματοδοτείται από την οργανωμένη διαδοχή ισχυρών και αδύναμων στοιχείων, ή στοιχείων αντίθετων μεταξύ τους ή με διαφορετικές συνθήκες"* [4] . Ο ορισμός αυτός είναι προφανώς πολύ γενικός και μπορεί να εφαρμοστεί ακόμα και σε φυσικά φαινόμενα που παρουσιάζουν μια περιοδικότητα.

Στις τέχνες γενικά, απλοποιημένα,αφορά στο χρονισμό των γεγονότων σε μια ανθρώπινη πάντα κλίμακα: Στη μουσική των ήχων και των σιωπών, των βημάτων ενός χορού ή του μέτρου στην ποίηση. Ο ρυθμός αφορά συχνά και την οπτική αναπαράσταση και όπως θα δούμε και παρακάτω έχουν υπάρξει ορισμοί και για το ρυθμό ενός βίντεο.

Πιο συγκεκριμένα, όσον αφορά το μουσικό ρυθμό, διάφορες σύγχρονες θεωρίες (των Cooper [16] και Meyer, του Epstein, του Kramer [17], του Yeston [18]) τον ορίζουν ως την οργάνωση και δόμηση των μουσικών ήχων σε μέρη με περισσότερο ή λιγότερο "προεξέχοντα" στοιχεία τα οποία βρίσκονται σε διαρκή αλληλεπίδραση με μια ιεραρχία χτύπων (beats). Η δόμηση αυτών των μερών μπορεί να γίνεται με διάφορα κριτήρια :Είτε φυσικά, όπως για παράδειγμα ο δυναμικός τονισμός, οι αρμονικές αλλαγές, η αντιπαραβολή υψηλών με χαμηλές συχνότητες, μεγάλων νοτών με πιο κοφτές κτλ, είτε με κριτήρια αντίληψης εκ μέρους του ακροατή.Η δόμηση με τον τελευταίο τρόπο, σχετίζεται άμεσα με τη μορφή, την περίφημη Gestalt: Δύο βασικές αρχές της ψυχολογίας Gestalt είναι πως αντικείμενα που βρίσκονται κοντά το ένα στο άλλο ή ομοιάζουν μεταξύ τους, τείνουν να γίνονται αντιληπτά ως ενιαίο σύνολο.

3.2. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΜΟΥΣΙΚΟΥ ΡΥΘΜΟΥ

3.2.1 ΠΑΛΜΟΣ ΚΑΙ ΜΕΤΡΟ

Παλμό ονομάζουμε την επαναλαμβονόμενη ακολουθία των όμοιων ερεθισμάτων, που συνήθως είναι μικρής διάρκειας και ο ακροατής τα αντιλαμβάνεται ως σημεία στο χρόνο. Ο παλμός μπορεί να μην είναι απαραίτητα το πιο γρήγορο ή πιο αργό στοιχείο της ηχητικής ακολουθίας, είναι όμως εκείνο που καταλαβαίνουμε ως πιο "ισχυρό": Το στοιχείο εκείνο που κάνει τον ακροατή να "ακολουθήσει" το κομμάτι είτε με παλαμάκια είτε με χορευτικά βήματα και κινήσεις. Στο μεγαλύτερο εύρος της κλασσική δυτικής μουσικής ένας παλμός ισοδυναμεί με ένα τέταρτο. Η αντίληψη του παλμού βοηθά στη δόμηση του "μέτρου" το οποίο ο MacPherson περιγράφει ως *"χρόνο"* ή *"ρυθμικό σχήμα"* . Όπως αναφέρει και ο Lester το 1986, όταν εδραιωθεί μια τέτοια μετρική ιεραρχία, ο ακροατής τη διατηρεί και την

ακολουθεί για όσο του δίνονται ενδείξεις από το ηχητικό σήμα. Ο παλμός συχνά αναφέρεται και ως χτύπος (beat), αν και ο όρος συχνά είναι παραπλανητικός αφού στα αγγλικά χρησιμοποιείται και για την περιγραφή συγκεκριμένων ρυθμών παιγμένων σε κρουστά κυρίως όργανα ή για την αναφορά στην αίσθηση του *groove*, της "εύφορης" δηλαδή διάθεσης που πολύ συχνά προκαλεί ένα ρυθμικό μουσικό απόσπασμα. Επίσης στα αγγλικά συνηθίζεται και ο όρος *tactus.* Στη δυτική μουσική, χρησιμοποιείται ευρέως ως βασικό στοιχείο της μετρικής δομής η έννοια του *time signature,* που στα ελληνικά συνήθως αποδίδεται απλά ως ρυθμός. Ο όρος ορίζει κάθε φορά σε τι χρονική αξία αντιστοιχεί ο ένας παλμός και πόσοι τέτοιοι παλμοί θα υπάρχουν συνολικά στο ένα *μέτρο*, τη θεμελιώδη υποδιαίρεση του συνολικού έργου. Έτσι, για παράδειγμα το πολύ συνηθισμένο time signature $^4/_4$ (τέσσερα τέταρτα) ορίζει πως η χρονική αξία είναι το τέταρτο και σε κάθε μέτρο θα υπάρχουν τέσσαρα συνολικά τέτοια τέταρτα. Μαζί με τα $^3/_4$ ή τα $^2/_4$, θεωρούνται *απλά* time signatures. Υπάρχουν επίσης *σύνθετα,* όταν ο αριθμός των παλμών (ο αριθμητής δηλαδή του κλάσματος) είναι πολλαπλάσιο του 2 ή του 3. Συνήθως στον έναν παλμό δίνεται χρονική αξία ογδοού, π.χ. $^9/_8$, $^{12}/_8$. Ακόμη, συναντάμε *μη κανονικά σύνθετα* τα οποία μπορούν να χωριστούν σε υποομάδες άνισης διάρκειας, για παράδειγμα $^5/_4$ ή $^7/_4$, *μικτά* ή και *κλασματικά* time signatures, π.χ. $\frac{2^1/_2}{4}$.

3.2.2 ΜΟΝΑΔΑ (UNIT) ΚΑΙ ΚΙΝΗΣΗ (GESTURE)

Ένα μοτίβο που κατά κάποιον τρόπο συσχετίζεται χρονικά με τον παλμό ονομάζεται ρυθμική μονάδα (rhythmic unit). Στην αγγλική ορολογία διακρίνονται σε *metric* όταν είναι ίσα μεταξύ τους, *intrametric* όταν τονίζουν ένα παλμό όπως π.χ. ένα μοτίβο swing, *contrametric* όταν δεν τονίζουν ή είναι κατά κάποιο τρόπο σε μη προφανώς αναμενόμενη σχέση με τον παλμό (syncopation) και σε *extrametric* όταν είναι τελείως ασυνήθιστες, όπως οι άρρητοι ρυθμοί. Ένα μοτίβο αντίθετα το οποίο δεν έχει την ίδια χρονική έκταση με έναν παλμό ονομάζεται ρυθμική κίνηση (rhythmic gesture). Περιγράφεται ανάλογα με την αρχή και το τέλος της ή με τις ρυθμικές μονάδες που περιλαμβάνει. Χρησιμοποιώντας και πάλι την αγγλική ορολογία διακρίνονται σε: Ανάλογα με το αν αρχίζουν σε δυνατό παλμό, ασθενή παλμό ή μετά από παύση σε *thetic, weak* και *initial rest* και ανάλογα με το αν τελειώνουν σε ισχυρό ή παλμό σε *strong* και *weak*.

3.2.3 ΕΝΑΛΛΑΓΗ ΚΑΙ ΕΠΑΝΑΛΗΨΗ

Αναφέρθηκε πιο πάνω πως ο ρυθμός σηματοδοτείται από την οργανωμένη διαδοχή αντίθετων στοιχείων, για παράδειγμα δηλαδή τις δυναμικές μεταξύ του ισχυρού και ασθενούς παλμού ή της μεγάλης με τη σύντομη νότα. Εξίσου σημαντικό με το να αντιληφθούμε το ρυθμό είναι να μπορούμε να τον αναμένουμε ή και να προβλέπουμε την εξέλιξη του. Η δυνατότητα μας αυτή βασίζεται στην επανάληψη ενός μοτίβου το οποίο έχει κατάλληλο μέγεθος ώστε να είναι εύκολο να απομνημονευθεί. Η εναλλαγή και η επανάληψη λοιπόν είναι θεμελιώδεις για την περιγραφή ενός ρυθμού. Με μουσικούς όρους, η εύρεση της άρσης και της θέσης (ή upbeat και downbeat) και η επανάληψη της διαδοχής τους κρύβεται πίσω από τη συντριπτική πλειοψηφία της μουσικής δημιουργίας. Σύμφωνα με θεωρητικούς όπως ο MacPherson και ο Scholes όλα οι μετρικές δομές, ακόμα και οι πιο σύνθετες μπορούν να εκφραστούν με διπλούς ή τριπλούς παλμούς μέσω πρόσθεσης ή

διαίρεσης. Για τον Pierre Boulez οι ρυθμικές δομές με χτύπους πέραν των τεσσάρων είναι "απλά μη φυσικές". Η θέση αυτή που με μια πρώτη εξέταση συμφωνεί με το μεγαλύτερο μέρος της δυτικής μουσικής παράδοσης, έρχεται σε σύγκρουση με τη μουσική παράδοση πολλών λάων. Η μουσική Yakshagana για παράδειγμα της Ινδίας περιλαμβάνει μέχρι και κλασματικούς παλμούς.

3.2.4 ΤΕΜΠΟ ΚΑΙ ΔΙΑΡΚΕΙΑ

Το τέμπο (tempo) ενός κομματιού είναι απλά η ταχύτητα ή η συχνότητα του tactus,δηλαδή του παλμού και είναι ένα μέτρο του πόσο "ρέει" και εξελίσσεται ένας ρυθμός. Κατά κύριο λόγο μετράται σε *"χτύπους ανά λεπτό"* (*beats per minute, bpm*) : Ένα τέμπο 60bpm υποδηλώνει 1 χτύπο ανά δευτερόλεπτο δηλαδή συχνότητα ίση με 1Hz. Όσον αφορά τη διάρκεια ενός ρυθμικού μουσικού γεγονότος, ο Michael Moravcsik διακρίνει τις εξής κατηγορίες [19]

- *Υπερσύντομο (supershort) ,* της τάξης του 1/30-1/10000sec , που ακούγονται ως συνεχόμενοι τόνοι
- *Σύντομο (short) ,* της τάξης του 1sec που αντιστοιχεί στο χτύπο της ανθρώπινης καρδιάς (~60 bpm) και στη διάρκεια ενός βήματος .
- *Μέτριο (medium),* της τάξης των μερικών δευτερολέπτων που επιτρέπει την ανάπτυξη ενός ρυθμικού μοτίβου.
- *Μεγάλο (Long) ,* της τάξης σχεδόν του ενός λεπτού για τη δόμηση πιο σύνθετων μουσικών φράσεων και
- *Πολύ μεγάλο (Very Long),* της τάξης των ωρών. Σε αυτή την κατηγορία, δεν ανήκουν ρυθμικές δομές αφού είτε δεν υπάρχει επαναληπτικότητα, είτε υπάρχει μα η περίοδος της είναι πολύ μεγάλη ώστε ο ανθρώπινος εγκέφαλος να την αντιληφθεί.

3.2.5 ΜΕΤΡΙΚΗ ΔΟΜΗ

Η μετρική δομή περιλαμβάνει το μέτρο, το τέμπο και όλα τα ρυθμικά στοιχεία που παράγουν μια χρονική κανονικότητα πάνω στην οποία "πατάνε" και προβάλλονται τα διάφορα μοτίβα με τις ποικίλλες διάρκειες. Η χορευτική μουσική έχει συνήθως άμεσα αναγνωρίσιμη μετρική δομή: Το τανγκό για παράδειγμα έχει συνήθως χρόνο 2/4 και τέμπο περίπου 66bpm. Γενικά, μπορούμε να κάνουμε τους εξής διαχωρισμούς: Ανάλογα με τον αν οι χρονικές αξίες είναι πολλαπλάσια και υποδιαιρέσεις του βασικού παλμού ή όχι οι ρυθμοί μπορεί να είναι *μετρικοί* (metrical,divisive) ή *ελεύθεροι (free)* . Παράδειγμα ελεύθερου ρυθμού αποτελούν κάποιοι χριστιανικοί ψαλμοί οι οποίοι έχουν ένα θεμελιώδη παλμό-χτύπο αλλά πολύ λιγότερο αυστηρή οργάνωση. Τέλος, υπάρχουν μουσικά έργα στα οποία δεν διακρίνεται καν ένα θεμελιώδης παλμός, και χρησιμοποιείται μόνο μια χρονική οργάνωση για να μετρηθεί πόσο θα διαρκέσει μια συγκεκριμένη φράση. Για τα έργα αυτά χρησιμοποιείται συχνά ο ιταλικός όρος *senza misura*, δηλαδή "χωρίς μέτρο".

Όλα αυτά τα χαρακτηριστικά του ρυθμού, έχουν προφανώς άμεση σχέση με τον τρόπο που ο ανθρώπινος εγκέφαλος αντιλαμβάνεται το ρυθμό, αφού είναι τα στοιχεία εκείνα που τον βοηθούν να τον καταλαβαίνει και να τον περιγράφει. Για αυτό αξίζει να αναφερθούμε πιο αναλυτικά σε ορισμένες θεωρίες για τον τρόπο που ο άνθρωπος επεξεργάζεται ένα ρυθμικό ερέθισμα.

3.3. ΡΥΘΜΙΚΗ ΑΝΤΙΛΗΨΗ

Ο τρόπος με τον οποίο αντιλαμβανόμαστε το ρυθμό είναι εξαιρετικά δύσκολος να οριστεί αφενός διότι αφορά σε θεμελιώδεις ικανότητες του ανθρώπινου εγκεφάλου οι οποίες δεν έχουν χαρτογραφηθεί ικανοποιητκά και αφετέρου διότι όπως προείπαμε δεν υπάρχει ένας καθολικά αποδεκτός ορισμός για το τι είναι στην ουσία του ο ρυθμός. Είναι χαρακτηριστικό πως σε πολλές γλώσσες της Υποσαχάριας Αφρικής (η οποία είναι περιοχή με τεράστιο ρυθμικό και μουσικό πλούτο που συντέλεσε στην ανάπτυξη πάμπολλων ειδών μουσικής της Λατινικής Αμερικής και της Αφροαμερικάνικης μουσικής : *blues,jazz,reggae*) δεν υπάρχει λέξη για το ρυθμό. Και αυτό διότι αυτό που κατανοούμε ως ρυθμό για τους λαούς αυτούς είναι μια απεικόνιση της ζωής, της αλληλεξάρτησης μεταξύ των ανθρώπων και της διάδρασης των σχέσεων.

Η μελέτη μάλιστα της ανθρώπινης αντίληψης του ρυθμού αποτελεί ένα τα πρωταρχικά στάδια της μελέτης του εγκεφάλου διότι στηρίζεται στη μελέτη της αντίληψης του χρόνου και της αναπαράστασης του σε νοητά διαστήματα. Ένα από τα πρώτα ψυχοφυσιολογικά πειράματα σχετικά με το ρυθμό διεξήχθη το 1894 από τον Thaddeus Bolton [20]. Στα άτομα που αποτελούσαν το δείγμα δόθηκαν ακουστικά στα οποία τροφοδοτούνταν σε σταθερά χρονικά διαστήματα σύντομοι ήχοι (κλικ) της ίδιας ακριβώς έντασης. Όταν ρωτήθηκαν από τους πειραματιστές, τα υποκείμενα απεφάνθησαν πως ενώ στην αρχή άκουγαν ένα σταθερό ηχητικό σήμα, στη συνέχεια μπορούσαν να διακρίνουν μια οργάνωση των ήχων σε ομάδες των δύο, τριών ή τεσσάτων χτύπων. Όταν μάλιστα τους έγινε η ψευδής νύξη πως θα ακούσουν ομαδοποιημένους ήχους, πολλοί όντως δήλωσαν ότι αντιλαμβανόντουσαν τον ήχο όπως τους ειπώθηκε ενώ άλλοι ομαδοποιούσαν τα κλικ με προσωπικό τρόπο. Το αποτέλεσμα της έρευνας του Bolton δεν είναι προφανώς κάτι ανήκουστο, αφού συχνά και στην καθημερινή μας ζωή παρατηρούμε παραδείγματα που διαφορετικά άτομα αντιλαμβάνονται εντελώς διαφορετικά την μετρική οργάνωση ενός μουσικού έργου. Διάσημα παραδείγματα από την κλασσική αλλά και την πιο σύχγρονη μουσική δημιουργία αποτελούν η Συμφωνία Νο 92 σε Σολ Ματζόρε του Haydn και το κομμάτι Hang up your Hang Ups του Herbie Hancock. Στο πρώτο παράδειγμα για πολλά μέτρα είναι τελείως ασαφές που "βρίσκεται" η θέση και η άρση ενώ στο δεύτερο οι περισσότεροι ακροατές ορίζουν διαισθητικά μια μετρική δομή αντίθετη σε αυτήν που χρησιμοποιούν τα τύμπανα, το κατεξοχήν όργανο του ρυθμικού μέρους, όταν μπαίνουν στο κομμάτι.

Πολλοί ερεύνητες στηρίχτηκαν στο έργο του Bolton τα μετέπειτα χρόνια, ορίζοντας μάλιστα το *διάστημα αδιαφορίας* (*indifference interval*) εντός του διαστήματος μεταξύ δύο ερεθισμάτων (*interstimulus interval*) το οποίο επηρεάζει τον τρόπο που κατανοούμε τον ήχο [20]. Για παράδειγμα οι συλλαβές της αγγλικής γλώσσας *bit* και *ter* όταν προφερθούν διαδοχικά,σε διάστημα μικρότερο των 1.5-2 sec γίνονται αντιληπτές ως η λέξη *bitter* ενώ αν μεσολαβήσει μεγαλύτερο διάστημα, ο ανθρώπινος εγκέφαλος τις αντιλαμβάνεται ξεχωριστά. Επίσης, προέκυψε η έννοια του *"αυθόρμητου tempo" (spontaneous tempo):* Εάν ζητηθεί από ένα άτομο να χτυπήσει ρυθμικά το πόδι του στη συχνότητα που του φαίνεται πιο "φυσική", το αποτέλεσμα για κάθε άτομο είναι διαφορετικό και συνήθως σταθερό. Επίσης, μελετήθηκε ο ρόλος που έχει η διάρκεια ενός ήχου στην αντίληψη των τονισμών (accents) και παρ'όλες τις διάφορες διαφωνίες μεταξύ των ερευνητών έχει

καταρτιστεί η εξής λίστα των στοιχείων που επηρεάζουν και τροποποιούν τη χρονική και ρυθμική οργάνωση από άτομο σε άτομο:

- *Ο υποκειμενικός ρυθμός (subjective rhythm) :* Δοθέντος ενός ακουστικού ερεθίσματος αποτελούμενου από μια σειρά ισαπέχοντων παλμών, διαφορετικά άτομα το αντιλαμβάνεται οργανωμένο σε γκρουπ των 2,3,4,6 ή 8 (με τα γκρουπ των 2,3 και 4 να είναι πιο συχνά). [21]
- *Ο τονισμός της έντασης (intensity accentuation) :* Εάν ο κάθε ν-οστός παλμός μιας ισόχρονης παλμοσειράς είναι έστω και λίγο πιο έντονος από τους υπόλοιπους, τότε όλοι οι παλμοί ακούγονται ως ομάδες των ν στο σύνολο ήχων, με τον πιο έντονο κάθε φορά να είναι πρώτος.Επίσης, το "κενό" διάστημα αμέσως μετά τον ισχυρό παλμό ακούγεται ως κατάτι συντομότερο [22] .
- *Ο τονισμός της διάρκειας (duration accentuation) :* Εάν ο ν-οστός παλμός μιας ισόχρονης παλμοσειράς είναι έστω και λίγο πιο μεγάλος σε διάρκεια από τους υπόλοιπους, τότε όλοι οι παλμοί ακούγονται ως ομάδες των ν, με τον πιο "μεγάλο" κάθε φορά να είναι ο τελευταίος [23]
- *Οι διαφορές στα διαστήματα (interval differences) :* Εάν το διάστημα μεταξύ του ν-οστού παλμού μιας ισόχρονης παλμοσειράς και του προηγούμενου του είναι έστω και λίγο πιο μεγάλο από τα υπόλοιπα, τότε όλοι οι παλμοί ακούγονται ομαδοποιημένοι σε σύνολα των ν με το μεγάλο διάστημα μεταξύ τους [24] .
- *Οι διαφορές στη συχνότητα (frequency differences) :* Η συχνότητα επηρεάζει την αντιληπτική ομαδοποίηση κατά πολλούς, αλληλένδετους τρόπους. Όλα τα παρακάτω συνήθως ακούγονται ως οι αρχές μιας ομάδας ήχων: Οι παλμοί υψηλότερου *τονικού ύψους (pitch),* οι παλμοί με τα λιγότερα συχνά ύψη-pitch, οι παλμοί - "σημεία καμπής" σε ένα σύνολο που μεταβάλλεται από αυξάνοντα σε μειούμενα ύψη καθώς και το τονικό ύψος που ακολουθεί ένα άλμα (δηλαδή μια απότομη μετάβαση) ανάμεσα σε pitches [25] .

Τα παραπάνω προφανώς αποτελούν μέρος μιας γνωσιακής προσέγγισης για την ερμηνεία της ανθρώπινης κατανόησης του ρυθμού. Αξίζει λοιπόν να δούμε, έστω και σύντομα, ορισμένα συμπεράσματα που παρήγαγαν αντίστοιχες νευροφυσιολογικές μελέτες.

Στηριζόμενοι σε εγκεφαλογραφήματα και φυσιολογικές αποκρίσεις του ανθρώπου σε ρυθμικά ερεθίσματα, οι μελετητές Thaut, Trimarchi και Parsons στο άρθρο τους Human Brain Basis of Musical Rhythm Perception: Common and Distinct Neural Substrates for Meter,Tempo and Pattern [26], κατέληξαν εν ολίγοις στα εξής: Ο ανθρώπινος εγκέφαλος στηρίζεται σε διαφορετικές λειτουργίες για την αναγνώριση και αναπαράσταση του μέτρου, του τέμπο και των ρυθμικών μοτίβων. Το μέτρο γίνεται αντιληπτό με λειτουργίες βασιζόμενες πιο πολύ σε αφηρημένες αναπαραστάσεις: Η θεωρητική οργάνωση του ήχου σε χρονικές ομάδες γίνεται στις προμετωπιαίες και μετωπιαίες περιοχές του δεξιού ημισφαιρίου. Αντίθετα, για την αντίληψη του τέμπο και κάποιων επαναλαμβανόμενων μοτίβων ο εγκέφαλος λειτουργεί με τον τρόπο που χειρίζεται πρωταρχικά οποιαδήποτε ηχητική πληροφορία. Πιο συγκεκριμένα, για να επεξεργαστεί την έννοια του τέμπο, ο άνθρωπος πρέπει να επιστρατεύσει λειτουργίες που αφορούν τόσο τις σωματικές και αισθητικές του αποκρίσεις όσο και ορισμένες συναισθηματικές. Χρησιμοποιεί λοιπόν τις περιοχές του εγκεφάλου που ονομάζονται *έλικα (gyrus)* και *νησιωτικός*

φλοιός (ή Νήσος του Reil, insula). Τέλος, για την επεξεργασία μοτίβων ο εγκέφαλος αναπαριστά χρονικά διαστήματα που ποικίλλουν σε πολλά σημεία και απαιτούν μια οργάνωση υψηλότερου επιπέδου. Αυτές οι λειτουργίες επικεντρώνονται σε ορισμένες περιοχές του κροταφικού λοβού. Οι ερευνητές παρατήρησαν επίσης δραστηριότητα και στην παρεγκεφαλίδα, σε περιοχές που ήταν κοινές και για το μέτρο και για το τέμπο και τα ρυθμικά μοτίβα.Όπως γίνεται εμφανές για την κατανόηση της ανθρώπινης αντιμετώπισης των μουσικών και ρυθμικών εννοιών, σημαίνων ρόλο κατέχει το *ερέθισμα* και πως το αντιλαμβάνεται ο άνθρωπος.Σε αυτό το σημείο λοιπόν, αξίζει να αναφέρουμε ορισμένα σημεία της Θεωρίας της Πληροφορίας που σκοπεύουν στην ποσοτικοποίηση ενός ερεθίσματος, της πληροφορίας που μπορεί να εμπεριέχει και της απόκρισης του ανθρώπου σε αυτό.

3.4. ΘΕΩΡΙΑ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΜΟΥΣΙΚΗ

Εισαγωγικά, πρέπει να αναφέρουμε πως ο κλάδος της ντετερμινιστικής ψυχολογίας ορίζει πως

Ερέθισμα + Οργανισμός = Αντίδραση

Τα ερεθίσματα είναι μηνύματα που λαμβάνει ένα άτομο διαμέσω διάφορων διαύλων. *Μήνυμα* είναι μια πεπερασμένη και διατεταγμένη ομάδα αντιληπτικών στοιχείων που αντλούνται από ένα μητρώο και συνδυάζονται σε μια δομή [27]. Η μουσική για παράδειγμα, όπως και ο χρόνος είναι ένα καθαρά χρονικό μήνυμα. Ο άνθρωπος-δέκτης του μηνύματος, για να μπορέσει να αντιδράσει στο μήνυμα, θα πρέπει αυτός να βρίσκεται εντός ενός πλαισίου έντασης: Μεταξύ του ορίου της ευαισθησίας και του ορίου του κορεσμού. Σύμφωνα με το νόμο Βέμπερ-Φέχνερ η αίσθηση που προκαλεί στο δέκτη, ένα τυχαίο ερέθισμα μεταβάλλεται ανάλογα προς το λογάριθμο αυτού:

$$S = K \log E \ .$$

Ένα ηχητικό μήνυμα θεωρούμε ότι μπορεί να περιγραφεί με τις εξής διαστάσεις:
Φυσικές διαστάσεις του είναι το *πλάτος* (σε βαρίδες), η *συχνότητα* (σε Hertz) και το *μήκος* (σε seconds) ενώ *αντιληπτικές διαστάσεις* είναι η *ένταση* (σε decibel,dB), το *ύψος* (σε οκτάβες) και η *διάρκεια* (σε *log t*).

Ο πρωτοπόρος μηχανικός *Abraham Moles*, ο οποίος ήταν από τους πρώτους που επιχείρησε τη σύνδεση της Ακουστικής, της Αισθητικής και της Θεωρίας της Πληροφορίας, καταλήγει στο βιβλίο *"Θεωρία της Πληροφορίας και Αισθητική Αντίληψη"* [27] μεταξύ άλλων στα εξής, συνοψισμένα, συμπεράσματα:

- Η μελέτη του μουσικού μηνύματος με αισθητικούς όρους δεν μπορεί να θεμελιωθεί στη μουσική θεωρία, η οποία έχει συχνά κατηγορηθεί ως δογματική και ανεπαρκής.
- Η μουσική αντίληψη στηρίζεται στη *χρονική ηχητική ύλη,* που λαμβάνει υπόσταση και μετατρέπεται σε παρατηρήσιμο αντικείμενο, μέσω μιας ηχογράφησης.
- Μια ηχογράφηση αποτελεί προσαρμογή του χρόνου στο χώρο, δίνοντας στον ήχο τις ιδιότητες της αναπαραγωγισιμότητας, της μονιμότητας, της αναστρεψιμότητας και της διαιρετότητας

- Η χρονική ηχητική υπόσταση αναπαρίσταται με τρισδιάστατα διαγράμματα όπου οι εντάσεις και τα ύψη περιγράφουν τη χρονική εξέλιξη της χροιάς.
- Ο ήχος υποδιαιρείται σε *ηχητικά αντικείμενα* με αυτόνομο κέντρο ενδιαφέροντος γύρω από το οποίο οργανώνεται η αντίληψη της διάρκειας του μηνύματος.
- Τα ηχητικά αντικείμενα που καταγράφει μια ηχογράφηση πραγματώνονται ανεξάρτητα από την πηγή τους: Κάθε ήχος και κάθε θόρυβος μπορεί να έχει τη θέση του σε μια υποθετική, "πειραματική" ορχήστρα.
- Οι δομές ενός ηχητικού μηνύματος περιλαμβάνουν μια *στοιχειώδη δομή*, μια *μικροδομή*, μια *ενδιάμεση δομή* και μια *μακροδομή.* Ο ρυθμός μπορεί να θεωρηθεί παράδειγμα μικροδομής ως η συγκέντρωση κάποιων *συμβόλων* που συνεισφέρουν σε ένα *ηχητικό αντικείμενο*.

4. ΚΕΦΑΛΑΙΟ 4 : ΒΙΝΤΕΟ ΚΑΙ ΧΡΩΜΑ

4.1. ΟΡΙΣΜΟΣ ΒΙΝΤΕΟ

Με τον όρο βίντεο (video) εννοούμε ένα ηλεκτρονικό μέσο για την καταγραφή και αναπαραγωγή και μετάδοση κινούμενων εικόνων και ήχου. Το βίντεο σαν έννοια είναι άρρηκτα συνδεδεμένο με την ανάπτυξη της τηλεόρασης. Και αυτό διότι ο John Logie Baird από το 1926 κιόλας όταν παρουσίασε δημοσίως την τηλεόραση, είχε υποβάλλει αίτηση για να κατοχυρώσει μια ευρισιτεχνία που ο ίδιος ονόμαζε Phonivisor, ένα σύστημα που υπό σημερινούς όρους θα λέγαμε ότι στόχευε στην καταγραφή βίντεο. Ο Baird κατάφερε να καταγράψει εικόνες σε δίσκους αλλά ποτέ δεν τις αναπαρήγαγε δημόσια. Πρόσφατα ωστόσο μέσω κατάλληλης ψηφιακής επεξεργασίας, ερευνητές κατάφεραν να ανακτήσουν το περιεχόμενο των δίσκων. Η πρώτη μορφή τεχνολογίας βίντεο που μοιάζει με τις σημερινές αναπτύχθηκε για χρήση σε τηλεοράσεις καθοδικού σωλήνα (cathod ray tube-CRT television) και από τότε έχουν πολλές βελτιώσεις. Ο Charles Ginsburg ανέπτυξε ένα από τα πρώτα συστήματα καταγραφής βίντεο σε κασέτα (video tape recorder-VTR) ενώ το 1951 επετεύχθη η καταγραφή βίντεο από "ζωντανές" τηλεοπτικές κάμερες μέσω της μετατροπής των ηλεκτρικών σημάτων και της αποθήκευσης τους σε μαγνητικές βιντεοκασέτες. Το 1971 η Sony εισήγαγε στην αγορά το VCR (videocassette recorder) κονσόλες και κασέτες φέρνοντας την καταγραφή και αναπαραγωγή βίντεο σε πολύ ευρύτερο κοινό. Επίσης, η εισαγωγή του DVD το 1997 και του δίσκου Blu-Ray το 2006 καθώς και η υποστήριξη βίντεο τεχολογιών από ηλεκτρονικούς υπολογιστές έχουν οδηγήσει σε μια διαρκώς αυξανόμεση διάδοση και χρήση του μέσου αυτού [28].

4.2. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΒΙΝΤΕΟ

4.2.1 ΚΑΡΕ

Καρέ ονομάζουμε ένα οπτικό στιγμιότυπο ενός βίντεο. Προκύπτει εάν με οποιοδήποτε τρόπο "παγώσουμε" την εικόνα και το σύνολο τους είναι αυτά που στην ουσία συνθέτουν το οπτικό περιεχόμενο του βίντεο.

4.2.2 FRAME RATE

Με τον όρο frame rate εννοούμε τον αριθμό καρέ που προβάλλονται ανά δευτερόλεπτο. Οι παλαιότερες μηχανικές κάμερες υποστήριζαν frame rate των 6 ή 8 καρέ ανά δευτερόλεπτο (frames per second-fps) ενώ οι σημερινές φτάνουν και σε frame rate των 120fps. Ανάλογα με τα διαφορετικά πρότυπα, σε κάποιες χώρες έχουν καθιερωθεί τα 25fps (πρότυπα PAL στην Ευρώπη, την Ασία και την Αυστραλία και SECAM στη Ρωσία, τη Γαλλία και κάποιες χώρες της Αφρικής) ή τα 29.97fps (πρότυπα NTSC στις Η.Π.Α, τον Καναδά, την Ιαπωνία και άλλες χώρες). Οι κινηματογραφικές ταινίες σκηνοθετούνται στα 24fps που κάνει τη μετατροπή τους σε βίντεο αρκετά περίπλοκη ενώ το ελάχιστο frame rate για να δημιουργείται στο ανθρώπινο μάτι η ψευδαίσθηση της κίνησης είναι τα 16 καρέ ανά δευτερόλεπτο.

4.2.3 ASPECT RATIO

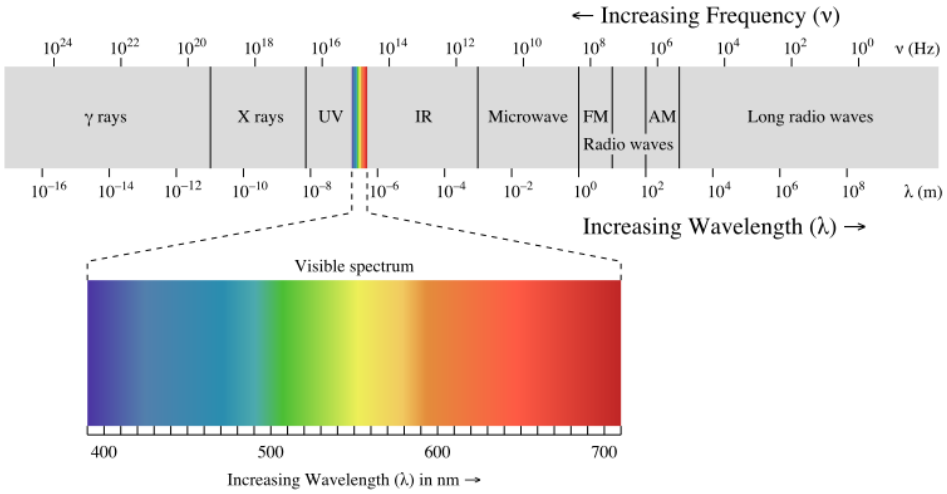
Ο όρος αυτός αφορά στις διαστάσεις της "οθόνης" που προβάλλεται καθώς και των στοιχείων μέσα σε αυτήν. Εφόσον μάλιστα όλα τα δημοφιλή είδη βίντεο είναι παραλληλόγραμμα, χρησιμοποιείται ο λόγος του πλάτους προς το ύψος της οθόνης. Οι "παραδοσιακές" τηλεοράσεις είχαν aspect ratio οθόνης $4:3$ ενώ οι πιο σύγχρονες, υψηλής ευκρίνειας έχουν πλέον $16:9$.

4.2.4 ΜΟΝΤΕΛΟ ΧΡΩΜΑΤΩΝ ΚΑΙ BITS ΑΝΑ PIXEL

Ο όρος μοντέλο χρωμάτων αφορά τον τρόπο αναπαράστασης των χρωμάτων στις διάφορες τεχνολογίες βίντεο. Πιο κάτω γίνεται εκτενής αναφορά στο τι είναι χρώμα και πως προσομοιώνεται σε έναν ηλεκτρονικό υπολογιστή. Τα bits ανά pixel καθορίζουν τον αριθμό των διαφορετικών χρωμάτων που μπορεί να αναπαριστά ένα pixel. Γι' αυτό και ένας απλός τρόπος να μειωθεί η ποσότητα πληροφορίας σε ένα ψηφιακό βίντεο είναι να αποδίδουμε χρωματικές τιμές σε μια ολόκληρη γειτονιά από pixel παρά σε κάθε ένα ξεχωριστά.

4.3. ΟΡΙΣΜΟΣ ΧΡΩΜΑΤΟΣ

Ως γνωστόν, το χρώμα δεν είναι τίποτα άλλο παρά ηλεκτρομαγνητική ακτινοβολία η οποία είναι ορατή από το ανθρώπινο μάτι και είναι υπεύθυνη για την αίσθηση της όρασης. Το ορατό φως δεν είναι παρά μια μικρή περιοχή του φάσματος της ηλεκτρομαγνητικής ακτινοβολίας με μήκος κύματος από 400nm έως και 700nm. Όλα αυτά τα μήκη κύματος μαζί δίνουν το λευκό φως, ενώ εάν περάσουμε τις ακτίνες του ήλιου μέσα από ένα πρίσμα θα τις δούμε να αναλύονται σε όλα τα χρώματα που μπορεί να ανιχνεύσει το ανθρώπινο μάτι.



Εικόνα 4.1: Φάσμα Η/Μ ακτινοβολίας

4.4. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΧΡΩΜΑΤΩΝ

4.4.1 HUE

Η ελληνική απόδοση είναι *χροιά* και αφορά στην ψυχοσωματική κατά βάση αντίδραση του ανθρώπου προς το κυρίαρχο μήκος κύματος ενός χρώματος. Είναι για παράδειγμα το στοιχείο εκείνο που κάνει τον καθέναν να αναγνωρίζει ένα χρώμα ως για παραδειγμά κόκκινο (ή κοκκινωπό) αν και είναι προφανές ότι δεν υπάρχει μοναδικό κόκκινο χρώμα, αλλά κοντά στις 100 αποχρώσεις του.

4.4.2 BRIGHTNESS

Αποδίδεται ως *λάμψη* και είναι χαρακτηριστικό των πηγών φωτός. Πρακτικά, είναι η "ένταση" του φωτός που αντιλαμβάνεται ο άνθρωπος και προφανώς σχετίζεται με το πόσο μαύρο έχει "αναμειχθεί" στην ακτινοβολία. Εξαρτάται λοιπόν από το μήκος κύματος και επομένως χρώματα με την ίδια *χροιά*, ενώ θα περιγράφονται από το ίδιο κυρίαρχο μήκος κύματος, μπορούν και να έχουν διαφορετική λάμψη.

4.4.3 LIGHTNESS

Η *φωτεινότητα* έχει άμεση σχέση με τη λάμψη και στα αγγλικά αναφέρεται και ως value (τιμή) ή tone (τόνος). Θεωρείται ως η αναπαράσταση της μεταβολής στην αντίληψη στη λάμψη: Μπορούμε να πούμε ότι είναι η λάμψη ενός τυχόντος αντικειμένου σε σχέση με τη λάμψη ενός άλλου αντικειμένου που υπό τις ίδιες ακριβώς συνθήκες γίνεται αντιληπτό ως λευκό.

4.4.4 COLORFULNESS

Μπορούμε να χρησιμοποιήσουμε τον όρο *πληρότητα χρώματος* και είναι ουσιαστικά το περιεχόμενο της χροιάς μιας οποιασδήποτε χρωματικής διέγερσης ή αλλιώς το πόσο απέχει ένα χρώμα από το γκρι.

4.4.5 CHROMA

Είναι μια ιδιότητα για την οποία δεν υπάρχει αντίστοιχος ελληνικός όρος και συνδέεται με την πληρότητα χρώματος με τον ίδιο ακριβώς τρόπο που συνδέεται η φωτεινότητα με τη λάμψη: Είναι η δηλαδή η πληρότητα χρώματος ενός τυχόντος αντικειμένου σε σχέση με αυτήν ενός δεύτερου αντικειμένου που υπό τις ίδιες ακριβώς συνθήκες γίνεται αντιληπτό ως λευκό.

4.4.6 SATURATION

Ο όρος *κορεσμός* περιγράφει και αυτός μια ψυχοσωματική αντίληψη: Αυτήν του πόσο "καθαρό" είναι ένα χρώμα, πόσο δηλαδή έχει "αναμειχθεί" με λευκό ή γκρι. Συχνά, χρησιμοποιείται σε εκατοστιαία κλίμακα θεωρώντας πως το 100% είναι ένα καθαρό χρώμα και το 0% το γκρι χρώμα (υπό σταθερές συνθήκες φωτεινότητας) .

4.5. ΧΡΩΜΑΤΙΚΑ ΜΟΝΤΕΛΑ ΚΑΙ ΧΡΩΜΑΤΙΚΟΙ ΧΩΡΟΙ

Για την περιγραφή των χρωμάτων και τη μοντελοποίηση τους για χρήση σε ηλεκτρονικά συστήματα υπάρχουν διάφορες προσεγγίσεις. Μία από αυτές είναι η χρήση *χρωματικών μοντέλων* , η προσπάθεια περιγραφής δηλαδή όλων των χρωμάτων ως συνδυασμούς τριών ή τεσσάρων βασικών. Τα χρώματα που προκύπτουν ως συνδυασμοί των *βασικών* (ή *πρωτογενών*) ονομάζονται *δευτερογενή.*

Εκτός αυτού όμως χρειάζεται και λεπτομερής περιγραφή των συνιστωσών ενός χρωματικού μοντέλου.Οι υποομάδες χρωμάτων που προκύπτουν είναι οι λεγόμενοι *χρωματικοί χώροι*, οι οποίοι όμοια με τα χρωματικά μοντέλα μπορούν να βασίζονται σε βασικά χρώματα ώστε να παράξουν τα υπόλοιπα ως συνδυασμούς τους ή σε μεταβλητές που καλούνται να διαχειριστούν τα αντίστοιχα μοντέλα.

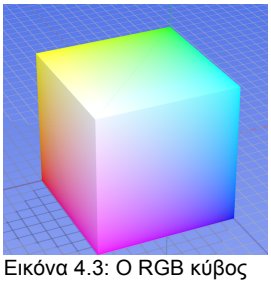
Η ανθρώπινη όραση αντιλαμβάνεται το χρώμα ως αποτέλεσμα σύνθεσης τριών βασικών ομάδων ανάλογα με το μήκος κύματος. Έτσι, οποιοδήποτε χρώμα θα μπορεί να αναπαρασταθεί ως συνδυασμός τριών βασικών χρωμάτων. Μάλιστα ο συνδυασμός αυτός θα είναι γραμμικός αφού έχει αποδειχτεί ότι για μια οπτική διέγερση S η χρωματική απεικόνιση θα είναι

$$C_S(S) = C_1(I_1) + C_2(I_2) + C_3(I_3)$$

όπου τα C_i είναι σταθεροί αριθμοί ενώ τα I_i είναι τα βασικά χρώματα: *μπλε,πράσινο* και *κόκκινο*. Ακριβώς αντίστοιχα μπορούν να αναπαρασταθούν και τα χρώματα σε μαθηματική μορφή ώστε να αναπαρασταθούν και να επεξεργαστούν σε ψηφιακά μέσα. Πιο συγκεκριμένα, στους ηλεκτρονικούς υπολογιστές, κάθε εικόνα αναλύεται σε ένα συγκεκριμένο αριθμό παρόμοιων στοιχείων, τα *εικονοστοιχεία* ή *pixel*. Με βάση τα χρωματικά μοντέλα, το χρώμα κάθε pixel μπορεί να κωδικοποιηθεί κατάλληλα ώστε να προκύψει τελικά ο συνολικός χρωματισμός της εικόνας.

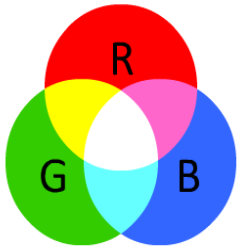
4.5.1 ΧΡΩΜΑΤΙΚΟ ΜΟΝΤΕΛΟ ΚΑΙ ΧΩΡΟΣ RGB

Το όνομα RGB προκύπτει από τα αρχικά Red,Green,Blue και προφανώς στηρίζεται στα βασικά χρώματα κόκκινο,πράσινο και μπλε. Ο χώρος αυτός μπορεί να αναπαραστήσει ένα μεγάλο μέρος του ορατού φάσματος και για την "υλοποίηση" του κάθε βασικό χρώμα αντιστοιχίζεται σε μία μεταβλητή (R,G,B). Ανάλογα με την ένταση της κάθε συνιστώσας η εκάστοτε μεταβλητή παίρνει τιμές από 0 έως και 255 και το τελικό χρώμα είναι η προσθετική μίξη των τριών συνιστωσών.Αφού μάλιστα $2^8 = 256$ για την ψηφιακή αναπαράσταση κάθε μεταβλητής απαιτούνται 8 δυαδικά ψηφία (bits). Απεικονίζοντας μάλιστα τις τριάδες (R,G,B) στον τρισδιάστατο χώρο μπορεί να προκύψει μια σχηματική αναπαράσταση του μοντέλου,όπως φαίνεται παρακάτω.



Εικόνα 4.3: Ο RGB κύβος

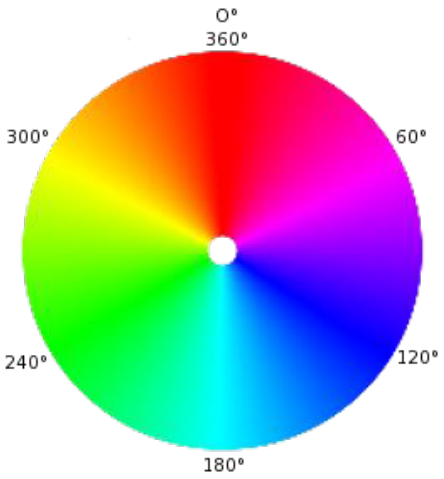
Οι ακραίες τιμές θα είναι η τριάδα (0,0,0) για το μαύρο και η τριάδα (255,255,255) για το άσπρο χρώμα.



Εικόνα 4.2: Χρώματα στο RGB

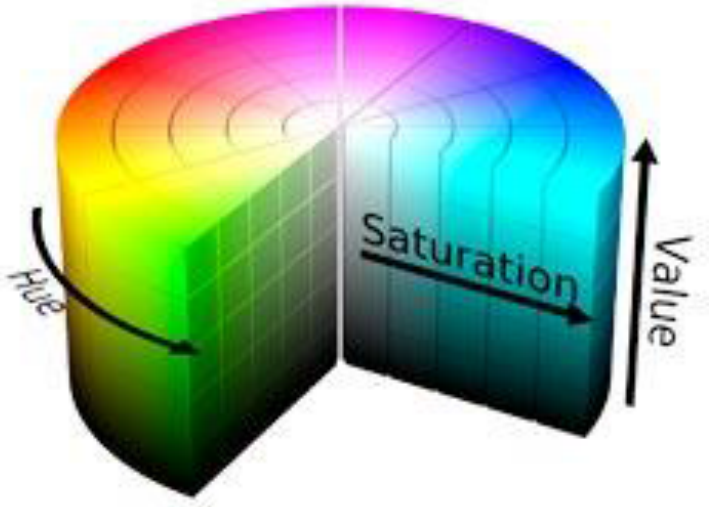
4.5.2 ΧΡΩΜΑΤΙΚΟ ΜΟΝΤΕΛΟ ΚΑΙ ΧΩΡΟΣ HSV/HSL

Παρουσιάζονται μαζί καθώς είναι παρόμοια. Τα μοντέλα αυτά είναι πιο κοντά στην περιγραφή που θα έκανε ο καθένας μας σε φυσική γλώσσα για ένα χρώμα: Τι απόχρωση, πόσο έντονο, πόσο σκούρο, πόσο ανοιχτό είναι και τα λοιπά. Αντίστοιχα με το μοντέλο RGB, οι μεταβλητές που καθορίζουν ένα χρώμα είναι αυτές που καθορίζουν με τα αρχικά τους το όνομα: η χροιά που προανεφέραμε (Hue), ο κορεσμός (Saturation) και αντίστοιχα λάμψη (Value αλλά χρησιμοποιείται με την έννοια του Brightness) και φωτεινότητα (Lightness). Από καθαρά μαθηματική άποψη, τα δύο μοντέλα είναι μη γραμμικοί μετασχηματισμοί του μοντέλου RGB σε πολικές συντεταγμένες.Η μεταβλητή Hue έχει ακριβώς τον ίδιο ορισμό και στα δύο μοντέλα και παίρνει τιμές από $0°$ έως $360°$. Στο σχήμα φαίνεται η σχέση της τιμής της με την απόχρωση του χρώματος.



Εικόνα 4.4: Τιμή & Απόχρωση

Ο κορεσμός όπως αναφέρθηκε και προηγουμένως στην περιγραφή των χαρατηριστικών των χρωμάτων παίρνει τιμές στην εκατοστιαία κλίμακα 0-100% όπως και τα Value και Lightness και η κλιμάκωση τους φαίνεται παρακάτω:

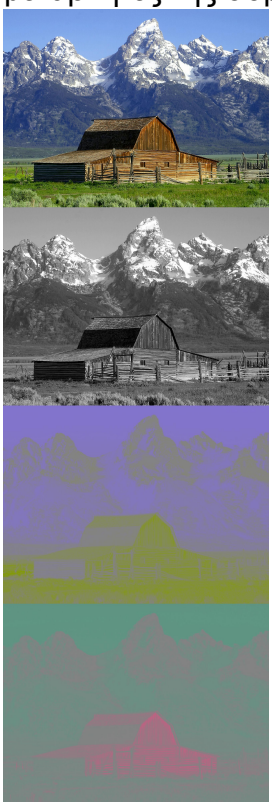


Εικόνα 4.5: Hue, Saturation, Value

4.5.3 ΧΡΩΜΑΤΙΚΟ ΜΟΝΤΕΛΟ YCbCr

Το συγκεκριμένο μοντέλο παραμετρικοποιεί το χρώμα με βάση τις εξής μεταβλητές: Το αρχικό Y αντιστοιχεί στο χαρακτηριστικό *Luma* , το οποίο στα ελληνικά θα αποδιδόταν ως *φωτεινότητα* αλλά πρακτικά έχει πιο άμεση σχέση με το *Brightness* (λάμψη) και αφορά δηλαδή το πόσο άσπρο και μαύρο "περιλαμβάνει" το χρώμα.Τα αρχικά Cb αντιστοιχεί στη διαφορά του χρώματος από τη μπλε συνιστώσα (blue-diffrence) και τα αρχικά Cr στη διαφορά από την κόκκινη συνιστώσα (red-diffrence). Το μοντέλο αυτό δεν αποτελεί ακριβώς ένα χρωματικό χώρο παρά περισσότερο έναν τρόπο κωδικοποίησης του RGB μοντέλου.

Παρακάτω φαίνεται μια εικόνα και το πως διακρίνεται στις τρεις συνιστώσες μεταβλητές της σύμφωνα με το μοντέλο YCbCr:



Πρώτα φαίνεται η αρχική εικόνα ,

έπειτα η Y-συνιστώσα της,

η Cb συνιστώσα,

και τέλος η Cr.

Εικόνα 4.5: YCbCr συνιστώσες

5. ΚΕΦΑΛΑΙΟ 5 :
ΡΥΘΜΟΣ ΒΙΝΤΕΟ

5.1. ΑΝΑΣΚΟΠΗΣΗ ΤΑΣΕΩΝ ΠΡΟΗΓΟΥΜΕΝΗΣ ΕΡΕΥΝΑΣ

Ο όρος "ρυθμός βίντεο" δεν έχει οριστεί μονοσήμαντα ακόμη στη βιβλιογραφία. Στα αγγλικά χρησιμοποιείται ο όρος video tempo ισοδύναμα με το video pace ενώ πιο σπάνια και με σχετικά διαφορετική έννοια χρησιμοποιείται και ο όρος video rhythm. Επίσης, ανάλογα με το περιεχόμενο και τη διάρκεια του βίντεο, οι όροι tempo/pace ,στη βιβλιογραφία ορίζονται πολύ διαφορετικά:

Σε μεγαλύτερα βίντεο (συνήθως σε κινηματογραφικές ταινίες) αφορά στο περιεχόμενο και στο "χαρακτήρα" των διάφορων σκηνών ως προς την πλοκή ενώ σε συντομότερα βίντεο (τα οποία συνήθως αποτελούνται από μία ενιαία σκηνή) αφορούν στην χρονική εξέλιξη της κίνησης μέσα σε αυτά. Οι δύο αυτοί ορισμοί είναι που διαμορφώνουν και τις διαφορετικές τάσεις του προσανατολισμού της έρευνας για την εξαγωγή του ρυθμού ενός βίντεο. Θα εξετάσουμε και τους δύο ορισμούς ανεξάρτητα και αναλυτικά.

5.2. ΠΡΩΤΗ ΤΑΣΗ - Ο ΡΥΘΜΟΣ ΣΤΟΝ ΚΙΝΗΜΑΤΟΓΡΑΦΟ

Στο βιβλίο τους του 1987 οι Thomas & Vivian Sobchack [29],ορίζουν το tempo ενός βίντεο ως το ρυθμό της επεξεργασίας του φίλμ (του μοντάζ δηλαδή) σε συνδυασμό με την ταχύτητα της κίνησης σε κάθε καρέ. Επιπρόσθετα, η εγκυκλοπαίδεια Britannica αναφέρει πως το tempo αφορά σε τρία βασικά πράγματα : την ταχύτητα της κίνησης και της αλλαγής κίνησης στο βίντεο, στη συνοδευτική μουσική αλλά και στο περιεχόμενο της πλοκής [5].

5.2.1 ADAMS, DORAI, VENKATESH

Σε μια προσπάθεια να ποσοτικοποιηθούν οι παραπάνω ορισμοί, οι Adams, Dorai και Venkatesh [30] εισήγαγαν το μέγεθος Pace

$$P(n) = \alpha \frac{med_s - s(n)}{\sigma_s} + \beta \frac{m(n) - \mu_m}{\sigma_m}$$

όπου:

α, β είναι δύο σταθερές αρχικοποιημένες αμφότερες στο 0.5 ή στο 1,
s είναι το μήκος μιας σκηνής μετρημένο σε καρέ,
m είναι η "ποσότητα" της κίνησης, δηλαδή μια μέση τιμή για κάθε σκηνή για την απόλυτη τιμή του αθροίσματος της κίνησης της κάμερας σε όλους τους άξονες για κάθε καρέ,
μ είναι η μέση τιμή του μεγέθους που σημειώνεται ως δείκτης,
$median$ ο διάμεσος,
σ η τυπική απόκλιση του μεγέθους-δείκτη ενώ
n είναι ο αύξων αριθμός της κάθε σκηνής.

Αξίζει να σημειώσουμε πως οι σταθερές α και β λειτουργούν ουσιαστικά ως βάρη για την επιρροή στο ρυθμό του μήκους της σκηνής και της ποσότητας της κίνησης αντίστοιχα. Για αυτό και υπάρχουν εφαρμογές όπου ανάλογα με την ταινία (το είδος της και το περιεχόμενο της) ή το σκηνοθέτη της έχουν δοθεί διαφορετικές και διακριτές μεταξύ τους τιμές.

Έχοντας ως αφετηρία τον ορισμό αυτό, οι Adams et al μελέτησαν πολλαπλά αποσπάσματα από διαφορετικές ταινίες. Στόχος τους ήταν εκτός των άλλων, μέσω της τιμής που θα υπολογιζόταν για το *Pace* να μπορούν να διακρίνουν τις ταινίες από τις οποίες προέρχονταν τα αποσπάσματα σε είδη, κάνοντας συγκρίσεις μεταξύ τους. Ακόμη, ανέλυσαν αποσπάσματα από τις ίδιες ταινίες με στόχο να συνδέσουν το περιεχόμενο της κάθε σκηνής με το Pace της. Έτσι, ανέπτυξαν δείκτες για το ύφος μια τυχούσας σκηνής μιας ταινίας ανάλογα με την τιμή που εξαγόταν: Σκηνές δράσης, χιούμορ και συγκινητικές σκηνές παρουσίαζαν όλες διαφορές στο μέγεθος του *ρυθμού* τους.

5.2.2 BATES, JHALA

Στον ορισμό των Adams,Dorai,Venkatesh έχουν στηριχτεί και άλλοι ερευνητές. Για παράδειγμα, οι Bates και Jhala του University of California Santa Cruz [31] χρησιμοποίησαν το μέγεθος Pace και το επέκτειναν για το ηχητικό σήμα που συνοδεύει μια ταινία. Θεωρώντας πως ο ήχος που "δένει" με ένα βίντεο θα πρέπει να επηρεάζει το ρυθμό του, όρισαν το μέγεθος Auditory Pace P_{audio} απόλυτα αντίστοιχα με τον ορισμό των Adams et al :

$$P_{audio}(n) = \alpha \frac{B(n)-med_B}{\sigma_B} + \beta \frac{L(n)-med_L}{\sigma_L}$$

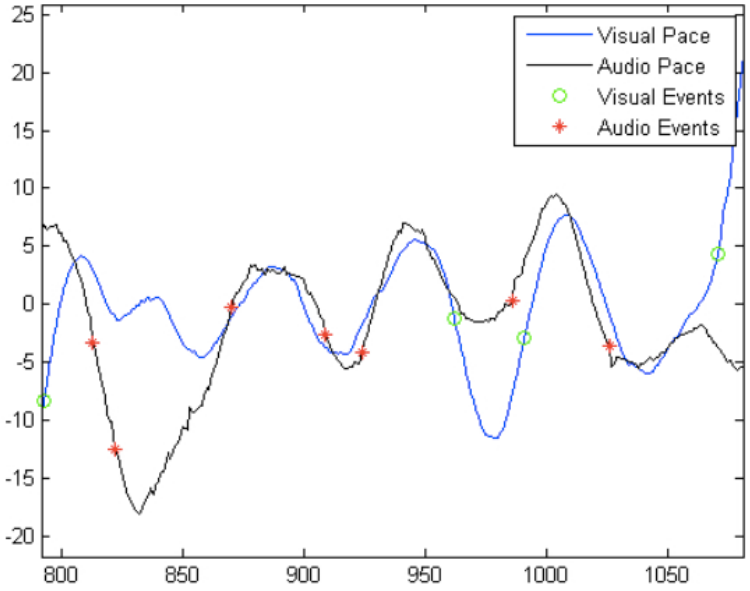
όπου πλέον χρησιμοποιούνται τα νέα μεγέθη
$L\,(Loudness)$ δηλαδή η μέση τιμή της ενέργειας του σήματος σε db (decibel) καθ'όλη τη διάρκεια μιας σκηνής και
$B(Brightness)$ δηλαδή το φασματικό κεντροειδές του σήματος στο πεδίο των συχνοτήτων.

Σημειώνεται ότι το *φασματικό κεντροειδές* είναι ένα μέγεθος που χρησιμοποιείται στην ανάλυση ηχητικών σημάτων και ορίζεται ως το κέντρο βάρους του φάσματος ενός σήματος. Τυπικά, είναι ο σταθμισμένος μέσος των συχνοτήτων που υπάρχουν στο φάσμα, όπως αυτό υπολογίζεται μέσω διακριτού Μετασχηματισμού Fourier και τα μέτρα των συχνοτήτων χρησιμοποιούνται ως βάρη.Έστω λοιπόν ότι έχουμε ένα διακριτό σήμα χωρισμένο σε i "παράθυρα" μήκους N δειγμάτων το καθένα. Το φασματικό κεντροειδές θα είναι

$$C_i = \frac{\sum_{k=1}^{N}(k+1)X_i(k)}{\sum_{k=1}^{N}X_i(k)}$$

με τα $X_i(k)$ να είναι οι συντελεστές του DFT (Discrete Fourier Transform) στο παράθυρο k.

Συγκρίνοντας μάλιστα τις τιμές των δύο μεγεθών (οπτικού και ηχητικού pace) για διάφορες σκηνές κατέληξαν σε όμοια αποτελέσματα, επικυρώνοντας την ευστάθεια και τη χρησιμότητα του ορισμού των Adams, Dorai και Venkatesh και καταδεικνύοντας την ανάγκη να συμπεριληφθεί και η ηχητική πληροφορία μιας σκηνής υπ' όψιν.



Εικόνα 5.1: Σύγκριση Audio και Visual Pace

5.2.3 LIU, YANG, WU, ZHANG, LI

Στην κατεύθυνση της προέκτασης του αρχικού ορισμού έχουν εργαστεί και οι Liu, Yang, Wu, Zhang και Li [32] . Αυτοί επιχειρηματολογούν πως ο ρυθμός θα πρέπει να εμπεριέχει εκτός της ταχύτητας της τεχνικής εξέλιξης μιας ταινίας και την ταχύτητα της ανθρώπινης αντίληψης του περιεχομένου αυτής. Βασίζονται λοιπον στα εξαχθέντα των Adams et al και μετονομάζουν το μέγεθος Pace σε "*γραμματική του φιλμ*" ("*film grammar*") .
Περιλαμβάνοντας στο μέγεθος m και το zoom in-zoom out αλλά και την ακινησία της κάμερας ο όρισμος της γραμματικής του φιλμ (FG) είναι όπως πριν :

$$FG(n) = \alpha \frac{med_s - s(n)}{\sigma_s} + \beta \frac{m(n) - \mu_m}{\sigma_m}$$

Κάπως όμως θα πρέπει να ποσοτικοποιηθεί και ο τρόπος αντίληψης της πληροφορίας του βίντεο από τον άνθρωπο,τόσο του οπτικού περιεχομένου όσο και του ηχητικού.

Το οπτικό περιεχόμενο μοντελοποιείται μέσω της κατηγοριοποίησης της κίνησης που μπορεί να διακριθεί. Όσο πιο "έντονη" και "περίπλοκη" είναι η κίνηση τόσο πιο ταχύς είναι ο ρυθμός αντίληψης της από τον θεατή.

Βασιζόμενοι στο πρότυπο MPEG-7 μπορούμε να ορίσουμε την "ένταση κίνησης" (MI_{MV}) ενός καρέ (καρέ τύπου P, που χρησιμοποιούν δηλαδή πληροφορίες από προηγούμενα καρέ και είναι εύκολα συμπιέσιμα) χωρίζοντας το σε macroblocks. Στο τυχόν (i,j) macroblock έχουμε

$$MI_{MV}(i,j) = \sqrt{x_{i,j}^2 + y_{i,j}^2}$$

όπου $(x_{i,j}, y_{i,j})$ είναι το διάνυσμα κίνησης αυτού, και επομένως συνολικά σε ένα καρέ P θα έχουμε ένταση κίνησης στο υπόβαθρο ίση με

$$MI_{MV}(P) = \sum_{(i,j) \in \upsilon\pi\acute{o}\beta\alpha\theta\rho o} MI_{MV}(i,j)$$

Τέλος, σε μία ολόκληρη σκηνή που υποθέτουμε ότι περιλαμβάνει Q συνολικά καρέ (τύπου P) θα έχουμε κατά μέσο όρο

$$MI_{MV}^{AVE} = \frac{1}{Q} \sum_{n=1}^{Q} MI_{MV}(P_n)$$

Ορίζουμε επίσης τη πολυπλοκότητα της κίνησης: 'Εχοντας υπολογίσει ένα *ιστόγραμμα προσανατολισμού* με N συνολικά bins (μπλοκ δηλαδή τιμών) για την κατανομή της φάσης των διανυσμάτων κίνησης ενός καρέ τύπου P, η πολυπλοκότητα MC είναι ουσιαστικά η εντροπία της διεύθυνσης της κίνησης:

$$MC(P) = -\sum_{n=1}^{N} h(n) \cdot log\, h(n)$$

όπου $h(n)$ είναι το n-οστό bin του ιστογράμματος.
'Ομοια λοιπόν για ολόκληρη τη σκηνή (με Q καρέ συνολικά) :

$$MC^{AVE} = \frac{1}{Q}\sum_{n=1}^{Q} MC(P_n)\ .$$

Λαμβάνοντας λοιπόν υπόψιν και τα δύο μεγέθη, η συνολική οπτική πληροφορία συνοψίζεται ως εξής:

$$VI = \sigma_1 \cdot MI\ + \sigma_2 \cdot MC$$

με τα σ_1, σ_2 να είναι και πάλι βάρη αρχικά ίσα με 0.5 αλλά να μπορούν να μεταβληθούν ανάλογα με τη βαρύτητα που επιθυμούμε για τον κάθε όρο.

Για το ηχητικό περιεχόμενο του βίντεο στηριζόμαστε στο [6]. Μπορούμε πάλι να υπολογίσουμε δύο μεγέθη: την ενέργεια του ήχου (Audio Energy, AE) και το ρυθμό του ήχου (Audio Pace, AP) .

Για το Audio Energy στηριζόμαστε στην ενέργεια κάθε καρέ κι έπειτα από τη συνολική ενέργεια εξάγεται ένας μέσος όρος για το σύνολο μιας σκηνής. Έχοντας υπολογίσει την τιμή του, μπορούμε να βρούμε το σύνολο N_{AP} των κορυφών του ηχητικού σήματος που ξεπερνούν μία συγκεκριμένη, εμπειρική συνήθως τιμή κατωφλίου Th_{AP} . Έτσι, η συχνότητα αλλαγής της ενέργειας του ήχου σε μια σκηνή P με N_{total} συνολικά frames ήχου, θα είναι

$$P = \frac{N_{AP}}{N_{total}}$$

Αφού πάλι βγάλουμε μέσο όρο για όλες τις σκηνές, συνθέτουμε τα δύο μεγέθη, AE και AP ώστε να προκύψει η ηχητική πληροφορία (Audio Information) :

$$AI = \omega_1 \cdot AE + \omega_2 \cdot AP$$

όπου τα βάρη ω_1, ω_2 ισούνται με 0.5 .

Απομένει τώρα να εισάγουμε το τελικό μέγεθος για το ρυθμό της ανθρώπινης αντίληψης (Human Perception, HP) του περιεχομένου ενός βίντεο ή μιας ταινίας.

$$HP(n) = \varphi \frac{VI(n) - \mu_{VI}}{\sigma_{VI}} + \psi \frac{AI(n) - \mu_{AI}}{\sigma_{AI}}$$

Το μ συμβολίζει τη μέση τιμή και το σ την τυπική απόκλιση.Τα φ, ψ είναι για άλλη μια φορά βάρη αρχικοποιημένα στο 0.5, και τυχούσα μεταβολή των τιμών αυτών θα μπορούσε να υποδηλώσει μεγαλύτερη ή μικρότερη βαρύτητα του οπτικού ή του ηχητικού σήματος αντίστοιχα σε μια ταινία.

Ο ορισμός του ρυθμού μιας ταινίας μπορεί πλέον να επεκταθεί ώστε να συμπεριλάβει τόσο την περιγραφή της "τεχνικής" ταχύτητας δηλαδή το Film Grammar όσο και την ανθρώπινη αντίληψη. Πλέον για κάθε σκηνή με αύξοντα αριθμό n ισχύει

$$Tempo(n) = \varepsilon \cdot FG(n) + \gamma \cdot HP(n) \quad , \quad \varepsilon = \gamma = 0.5$$

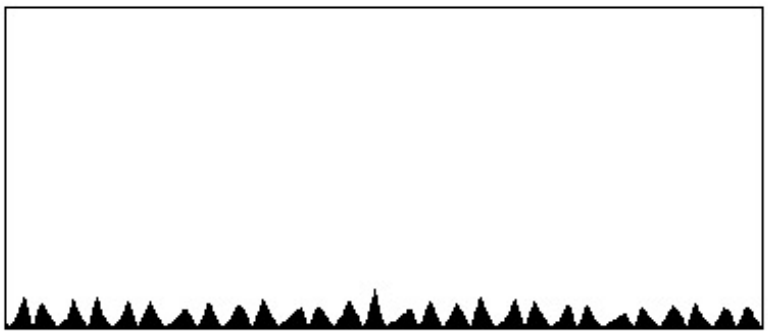
Γίνεται εμφανές λοιπόν πως η συγκεκριμένη προσέγγιση ως προς το ρυθμό ενός βίντεο στηρίζεται σε στοιχεία που χαρακτηρίζουν μία ολόκληρη ταινία. Δεν θα είχε νόημα να εφαρμοστούν σε μικρότερες διάρκειες ή σε βίντεο ενός πλάνου αφού το αποτέλεσμα δε θα είχε ιδιαίτερη θεωτηρική αξία. Αντιθέτως, εφαρμόζοντας την παραπάνω θεωρία σε ολόκληρα κινηματογραφικά έργα είναι δυνατόν να αναπτυχθούν εφαρμογές για διάκριση διαφορετικών σκηνών μιας ταινίας μεταξύ τους, για περιγραφή τους ως προς το περιεχόμενο ανάλογα με τον υψηλό/χαμηλό "ρυθμό", για κατηγοριοποίηση ολόκληρων ταινιών καθώς και για την εξαγωγή συνόψεων και διαφημιστικών αποσπασμάτων (trailers) μιας ταινίας περιλαμβάνοντας κομμάτια με διαφορετικούς ρυθμούς μέσα από ένα φιλμ.

5.3. ΔΕΥΤΕΡΗ ΤΑΣΗ - Ο ΡΥΘΜΟΣ ΣΕ ΣΥΝΤΟΜΑ ΒΙΝΤΕΟ

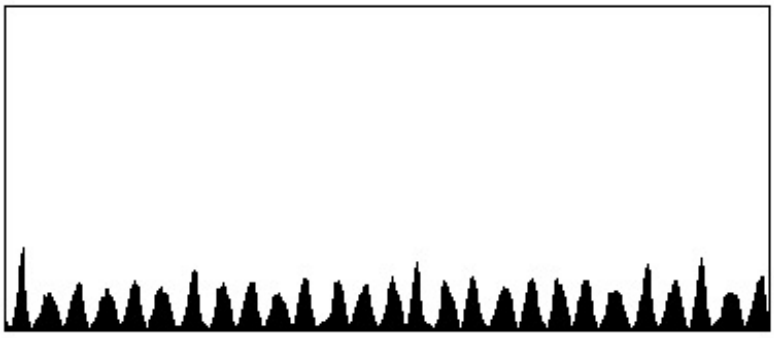
Πέρα από τα παραπάνω, έχουν υπάρξει και διαφορετικές προσεγγίσεις ως προς τι θα μπορούσε να θεωρηθεί ως ρυθμός ενός βίντεο. Στην περίπτωση που το βίντεο είναι μικρό σε διάρκεια και κατά κύριο λόγο το περιεχόμενο του δεν διαφοροποιείται ριζικά από καρέ σε καρέ, ως ρυθμός μπορεί να θεωρηθεί η ταχύτητα και γενικότερα η χρονική εξέλιξη της κίνησης ενός ανθρώπινου σώματος εάν και εφόσον υπάρχει μέσα στο βίντεο. Επίσης, ρυθμός μπορεί να εξαχθεί και από τη χρονική εξέλιξη της φωτεινότητας στο σύνολο των pixels από καρέ σε καρέ.

5.3.1 GUEDES, BRANCO

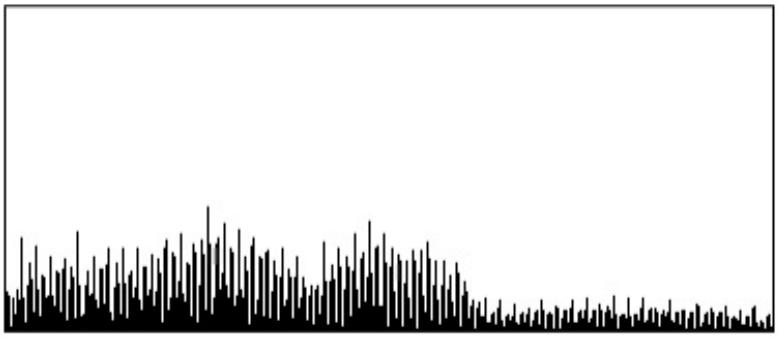
Πιο συγκεκριμένα οι Guedes και Branco [33] ακολούθησαν την εξής προσέγγιση: Για να υπολογίσουμε τη συνολική κίνηση ("*amount of motion*") ενός βίντεο μπορούμε ανά διαδοχικά καρέ να υπολογίζουμε τη συνολική μεταβολή στη *φωτεινότητα* κάθε pixel. Στις περιοχές των δύο διαδοχικών καρέ όπου δεν έχει υπάρξει καμία σημαντική κίνηση, η μεταβολή στις φωτεινότητας θα τείνει στο μηδέν ενώ σε περιοχές με έντονη κίνηση και άρα μεγάλο ενδιαφέρον, η απόλυτη τιμή της μεταβολής θα είναι σίγουρα μεγαλύτερη του μηδενός. Αθροίζοντας λοιπόν τις διαφορές στη φωτεινότητα κάθε pixel για διαδοχικά καρέ, μπορεί να εξαχθεί μια απεικόνιση της "ποσότητας της κίνησης" . Προφανώς όσο περισσότερο κινείται ένα σώμα μέσα στο βίντεο, τόσα περισσότερα pixel θα παρουσιάζουν μεταβολές στη φωτεινότητά τους. Επίσης, αλλαγές στη διεύθυνση κίνησης (π.χ. στροφές, σταματήματα κλπ) θα προκαλούν μηδενισμό στη μεταβολή της φωτεινότητας της "γειτονιάς" των pixel όπου αυτές παρατηρούνται, αφού τέτοιες μεταβολές απαιτούν ένα στιγμιαίο σταμάτημα.Παρατηρώντας περιοδικές κινήσεις μπορούμε να βρούμε περιοδικότητες και στη μεταβολή της φωτεινότητας των pixel στα οποία παρατηρούνται οι κινήσεις. Κινήσεις με μεγαλύτερο χωρικό εύρος αντιστοιχούν σε μεγαλύτερο πλάτος ενώ ταχύτερες κινήσεις αντιστοιχούν σε μεγαλύτερη συχνότητα.



Εικόνα 5.2: Πρώτη Απεικόνιση Φωτεινότητας



Εικόνα 5.3: Δεύτερη Απεικόνιση Φωτεινότητας



Εικόνα 5.4: Τρίτη Απεικόνιση Φωτεινότητας

Στις εικόνες 5.2 και 5.3 φαίνονται οι απεικονίσεις της διαφοράς φωτεινότητας που προέκυψαν από βίντεο με χέρια να χαιρετάνε. Και στις δύο περιπτώσεις η κίνηση έχει ίδια συχνότητα αλλά διαφορετικό πλάτος. Αντίθετα, η 5.4 προέκυψε από αντίστοιχο βίντεο όπου η συχνότητα είναι μεγαλύτερη ενώ το πλάτος μεταβάλλεται.

Οι Guedes και Branco επίσης παρατήρησαν πως αφαιρώντας την DC συνιστώσα από ένα σήμα που προκύπτει από την παραπάνω διαδικασία, όπως για παράδειγμα αυτά των σχημάτων 2-4, προκύπτουν ομοιότητες με ηχητικά σήματα. Επομένως εάν γίνει χρήση ενός αλγορίθμου ικανού να εντοπίσει την περιοδικότητα ενός σήματος, όπως ο γρήγορος μετασχηματισμός Fourier (Fast Fourier Transform-FFT), μπορούμε να εντοπίσουμε ένα ρυθμό που υπάρχει μέσα στο σήμα.

Για την υλοποίηση της διαδικασίας που περιγράφηκε ακολουθείται η εξής διαδικασία:Αρχικά χρησιμοποιούνται 150 αναδρομικά ζωνοπερατά φίλτρα δεύτερης τάξης. Αναδρομικά ονομάζονται τα φίλτρα που επαναχρησιμοποιούν μία ή περισσότερες από τις εξόδους τους ως είσοδο. Η κεντρική συχνότητα των φίλτρων αυτών κυμαίνεται από 0.5Hz έως την περιοχή των 12.5-15HZ και το εύρος ζώνης τους είναι περίπου το 10% της κεντρικής συχνότητας. Στην έξοδο αυτών των φίλτρων εφαρμόζεται ο αλγορίθμος Goertzel για την εξαγωγή μιας απεικόνισης του πλάτους και της φάσης για κάθε κεντρική συχνότητα. Ο αλγόριθμος Goertzel προτιμάται έναντι του FFT που αναφέρθηκε πιο πάνω λόγω χαμηλότερης πολυπλοκότητας.

Για τη συνάρτηση μεταφοράς των ζωνοπερατών φίλτρων έχουμε τα εξής:

$$Y(n) = X(n) + 2 \cdot C \cdot R \times Y(n-1) - R^2 \times Y(n-2)$$

όπου $C = cos(\frac{2\pi f}{sr})$,

$R = e^{\frac{-\pi\ bw}{sr}}$,

cf είναι η κεντρική συχνότητα,

bw είναι το εύρος ζώνης,

sr είναι ο ρυθμός δειγματοληψίας (sampling rate),

και $e = 2.7182818\ldots$ είναι ο αριθμός του Euler.

Εφαρμόζοντας μια μορφή του αλγορίθμου του Goertzel, υπολογίζεται το πραγματικό και φανταστικό μέρος της εξόδου των φίλτρων:

$$Y_{real} = Y(n) - R\ cos(\frac{2\pi\ bw}{sr}) \times Y(n-1)$$

$$Y_{imag} = -R\ sin(\frac{2\pi\ bw}{sr}) \times Y(n-1)$$

και επομένως το πλάτος και η φάση είναι κατά τα γνωστά

$$Y_{πλάτος} = \sqrt{Y_{real}^2 + Y_{imag}^2}$$

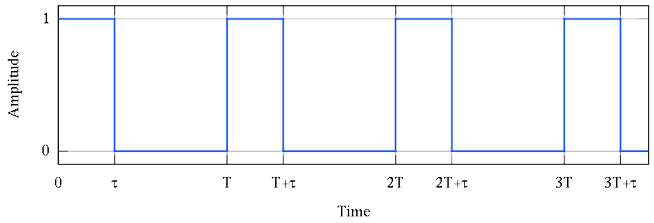
$$Y_{φάση} = arctan(\frac{Y_{real}}{Y_{imag}})$$

Οι Guedes και Branco στηρίζονται στα εξαχθέντα της διαδικασίας αυτής για να προσδιορίσουν τον "παλμό" ενός σύντομου βίντεο (το οποίο συγκεκριμένα περιλαμβάνει την κίνηση ενός χορευτή). Σύμφωνα με το [8] η αίσθηση του "παλμού" ή αίσθηση "ρυθμικότητας" είναι μια εγγενής ιδιότητα του μουσικού ρυθμού που προκαλείται από τη συνύπαρξη και αλληλεπίδραση των χαμηλών συχνοτήτων σε ένα μουσικό σήμα. Γι' αυτό και η εξαγωγή του, δοθέντος ενός μουσικού ηχητικού σήματος, έγκειται στην ανεύρεση της πιο "ισχυρής" χαμηλής συχνότητας καθώς και της φάσης αυτής. Για το λόγο αυτό η διαδικασία του εντοπισμού του ρυθμού (*beat tracking*) συχνά θεωρείται ως εξειδικευμένη εφαρμογή της διαδικασίας εντοπισμού του τόνου (*pitch tracking*). Οι "παλμοί" που θα εξαχθούν θα πρέπει να έχουν διάρκεια παρόμοια με αυτή ενός μουσικού ρυθμού και να είναι συμμετρικοί ο ένας ως προς τον άλλον ώστε να ενισχύουν την αίσθηση που αναφέρθηκε πριν. Πρέπει λοιπόν να υπάρχει ενα κριτήριο "αρμονικότητας" ώστε να ελέγχεται ότι η αίσθηση ενισχύεται από περιοδικότητες που είναι ακέραια πολλαπλάσια του παλμού.

Έτσι, χρησιμοποιώντας το προγραμματιστικό περιβάλλον Max/MSP,οι Guedes, Branco παράλληλα με τη χρησιμοποίηση των ζωνοπερατών φίλτρων και την εφαρμογή του αλγορίθμου Goertzel, εξάγουν από τη φασματική αναπαράσταση του βίντεο και την πιο "κυρίαρχη" στιγμιαία συχνότητα του. Αυτό το πετυχαίνουν εφαρμόζοντας ετεροσυσχέτιση την έξοδο της προηγούμενης διαδικασίας σε μία δοσμένη χρονική στιγμή με ένα τετραγωνικό παλμό 1Hz. Υπενθυμίζεται ότι για δύο συνεχείς συναρτήσεις f, g η πράξη της ετεροσυσχέτισης (cross-correlation) ορίζεται ως:

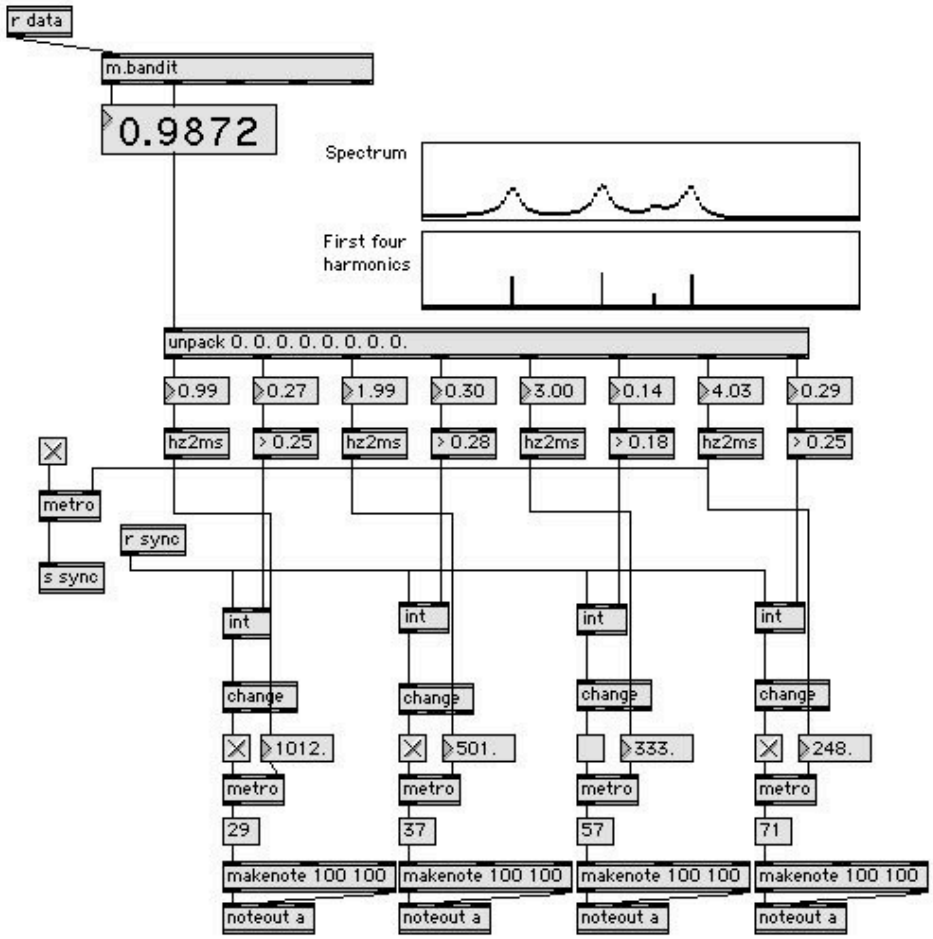
$$f \star g = \int_{-\infty}^{+\infty} f^*(t)\, g(t+\tau)\, dt \quad ,$$ όπου f^* η συζυγής της f

Ο τετραγωνικός παλμός (της εικόνας 5 για παράδειγμα) στο πεδίο των συχνοτήτων περιέχει αρμονικές κορυφές, και έτσι ετεροσυσχετίζοντας τον με το σήμα μας θα "ευνοεί" τις συχνότητες που επίσης εμπεριέχουν αρμονικές κορυφές, κάνοντας τις συχνότητες αυτές υποψήφιες να αποτελούν τον "παλμό".



Εικόνα 5.5: Τετραγωνικός Παλμός

Εντέλει λοιπόν, με τη χρήση ενός αντικειμένου της γλώσσας Max ονόματι **m.bandit** [7] το οποίο διαμορφώνεται ώστε να πραγματοποιεί και την πράξη της ετεροσυσχέτισης, να εκτιμά τη θεμελιώδη συχνότητα του σήματος και να εξάγει τιμές που μπορούν να θεωρηθούν ως μουσικοί χτύποι. Αν αυτοί δοθούν σε ένα άλλο αντικείμενο που να προσομοιάζει τη λειτουργία του μετρονόμου (εν προκειμένω ονομάζεται **m.clock** [7]), αυτό μπορεί να παράγει αυτό που πιο πριν περιγράψαμε ως ρυθμό του αρχικού βίντεο με καθαρά μουσικούς όρους.



Εικόνα 5.6: Συνολική Εφαρμογή στο Max/MSP

Στην εικόνα 5.6, φαίνεται μια ολόκληρη εφαρμογή των Guedes και Branco στο περιβάλλον Max/MSP όπου από τα αρχικά δεδομένα εξάγεται μία "θεμελιώδης" συχνότητα κοντά στο 1Hz καθώς και τέσσερις αρμονικές διαφορετικού πλάτους η καθεμία. Οι συχνότητες αυτών μετατρέπονται σε milliseconds και συγκρίνονται με ένα κατώφλι τιμών ώστε να ελέγχουν τα αντικείμενα **metro** που μπορούν να "παίξουν" MIDI νότες. Έτσι, βασιζόμενοι στην θεωρητική ανάλυση, ουσιαστικά το βίντεο που δίνεται ως είσοδος γίνεται το "όργανο" που θα παίξει συγκεκριμένες νότες στο ρυθμό που αναγνωρίστηκε από την εξέταση διαδοχικών καρέ .

5.3.2 CHU, TSAI

Άλλη μια προσέγγιση όσον αφορά το ρυθμό ενός βίντεο μικρής διάρκειας είναι αυτή που ακολούθησαν οι Chu, και Tsai [34]. Εστιάζοντας και αυτοί σε βίντεο που περιέχουν χορευτές, βασίζονται στην κίνηση του σώματος του χορευτή ώστε να εξάγουν μια ρυθμική δομή που να μπορεί να μετατραπεί κατάλληλα σε ηχητικό σήμα που να μπορεί να συνοδεύσει και να "ντύσει" μουσικά τον χορό.

Οι Chu,Tsai εισάγουν την έννοια του "ρυθμού της κίνησης" (rhythm of motion, ROM) ως την χρονική εξέλιξη της, που προκύπτει από τις κινήσεις που επαναλαμβάνονται περιοδικά αλλά και τις αλλαγές στην κίνηση που παρατηρούνται περιοδικά, όπως στροφές, άλματα, παύσεις και άλλα. Οι κινήσεις ενός χορευτή μάλιστα συχνά δεν παρουσιάζουν αυστηρή επαναληπτικότητα ωστόσο ο θεατής τις αντιλαμβάνεται ως έναν ενιαίο "ρυθμό κίνησης" λόγω της ροής τους. Επίσης, παρ'όλη την υποκειμενικότητα της αντίληψης της μουσικής και του τρόπου "ερμηνείας" της από κάθε χορευτή, οι περισσότεροι αντιλαμβάνονται το ίδιο μουσικό σήμα με παρόμοιο τρόπο και το χορεύουν αρκετά όμοια καθιστώντας το ρυθμό της κίνησης που του αντιστοιχεί λίγο ή πολύ συγκρίσιμο για διαφορετικούς ερμηνευτές.

Γίνεται λοιπόν αντιληπτό ότι βασικό κομμάτι της εξαγωγής του ρυθμού ενός τέτοιου βίντεο είναι η εύρεση και η παρακολούθηση της τροχίας της κίνησης του σώματος. Αυτή θα βασιστεί σε ορισμένα χαρακτηριστικά σημεία (feature points) αφού η παρακολούθηση όλων των pixels του βίντεο, εκτός του τεράστιου κόστους της θα ήταν και άσκοπα αναλυτική. Για το λόγο αυτό χρησιμοποιείται ένας ανιχνευτής γωνιών για τα feature points ενώ για την πρόβλεψη της κίνησης τους εφαρμόζεται η μέθοδος των Lucas-Kanade για την ανίχνευση της οπτικής ροής (Pyramid Lucas-Kanade (PLK) optical flow detection method). Περισσότερα για την οπτική ροή θα δούμε και παρακάτω, στο επόμενο κεφάλαιο, ωστόσο αξίζει να εξετάσουμε συνοπτικά τη μέθοδο των Bruce D. Lucas και Takeo Kanade [35].

Ο αλγόριθμος αυτός υποθέτει ότι η μετατόπιση των οπτικών περιεχομένων μεταξύ δύο γειτονικών στιγμιοτύπων - frames είναι ελάχιστη και κατά προσέγγιση μηδενική σε μια περιοχή του υπό εξέταση σημείου p . Έτσι, η εξίσωση του optical flow μπορούμε να υποθέσουμε ότι ισχύει για όλα τα pixels σε ένα "παράθυρο" με κέντρο το p . Το διάνυσμα της ροής της εικόνας λοιπόν (ουσιαστικά της ταχύτητας) (V_x, V_y) θα πρέπει να ικανοποιεί τα εξής:

$$
\begin{aligned}
I_x(q_1)V_x + I_y(q_1)V_y &= -I_t(q_1) \\
I_x(q_2)V_x + I_y(q_2)V_y &= -I_t(q_2) \\
&\vdots \\
I_x(q_n)V_x + I_y(q_n)V_y &= -I_t(q_n)
\end{aligned}
$$

όπου τα $q_1, q_2, .., q_n$ είναι τα pixels μέσα στο παράθυρο και $I_x(q_i), I_y(q_i), I_t(q_i)$ είναι οι μερικές παράγωγοι της εικόνας I ως προς τις διαστάσεις x, y και το χρόνο t στο q_i .

Οι παραπάνω εξισώσεις γράφονται στη μητρική μορφή

$$Av = b$$

με τους πίνακες

$$A = \begin{bmatrix} I_x(q_1) & I_y(q_1) \\ I_x(q_2) & I_y(q_2) \\ & \cdot \\ & \cdot \\ I_x(q_n) & I_y(q_n) \end{bmatrix}, \qquad v = \begin{bmatrix} V_x \\ V_y \end{bmatrix} \qquad \text{και} \qquad b = \begin{bmatrix} -I_t(q_1) \\ -I_t(q_2) \\ \vdots \\ -I_t(q_n) \end{bmatrix}$$

Το σύστημα αυτό έχει συνήθως περισσότερες εξισώσεις από αγνώστους και η μέθοδος χρησιμοποιεί μια λύση μέσω της αρχής των ελαχίστων τετραγώνων λύνοντας ως εξής

$$A^T A\, v = A^T b \Rightarrow v = (A^T A)^{-1} A^T b$$

Έτσι τελικά

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum_i I_x(q_i)^2 & \sum_i I_x(q_i) I_y(q_i) \\ \sum_i I_y(q_i) I_x(q_i) & \sum_i I_y(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_x(q_i) I_t(q_i) \\ -\sum_i I_y(q_i) I_t(q_i) \end{bmatrix}$$

Επανερχόμενοι λοιπόν στη διαδικασία των Chu και Tsai, εάν η θέση ενός χαρακτηριστικού σημείου s_i σε ένα καρέ t είναι $s_i(x, y)$ τότε στο επόμενο καρέ $t+1$ θα είναι $s_i'(x', y') = PLK(\, s_i(x, y)\,)$.

Για τη συνολική τροχιά κάθε feature point θα πρέπει να ενώσουμε τις εκτιμώμενες θέσεις του σε όλα τα διαδοχικά καρέ. Για ένα σημείο s_i στο καρέ t το πλέον κατάλληλο σημείο προς "ένωση" από το επόμενο καρέ $t+1$ θα είναι εκείνο το s_{j^*} για το οποίο ισχύει πως

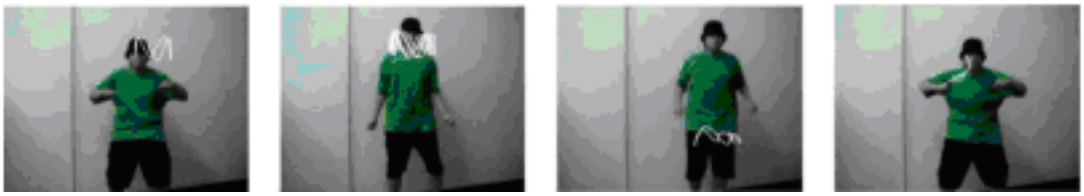
$$j^* = arg\ min_{s_j \in N(s_i')} d(s_i, s_j)$$

όπου:

$N(s_i')$ είναι η "γειτονιά" της εκτιμώμενης θέσης s_i' .
Ως γειτονιά του pixel θεωρούμε όλα τα pixels στον κύκλο με κέντρο το s_i' και μια ορισμένη ακτίνα r

$d(s_i, s_j)$ είναι η απόσταση των δύο σημείων s_i, s_j . Με τη σειρά της, η απόσταση ορίζεται ως $d(s_i, s_j) = \sum_{m=0}^{M} |h_i(m) - h_j(m)|$
όπου h_i, h_j τα χρωματικά ιστογράμματα στο χρωματικό χώρο HSV των 9×9 γειτονιών γύρω από τα s_i, s_j

Για να αποφύγουμε μικρές αποστάσεις που προκαλούνται από χαρακτηριστικά σημεία που αποτελούν θόρυβο, κατωφλιώνουμε τις αποστάσεις που υπολογίζουμε με ένα προκαθορισμένο κατώτατο όριο.



Εικόνα 5.7: Τροχιές με διαφορετικά feature points

Στην εικόνα 5.7 φαίνονται διάφορες τροχιές που υπολογίστηκαν για μια ακολουθία χορευτικών κινήσεων στηριζόμενοι σε διαφορετικά feature points κάθε φορά.

Προκύπτει λοιπόν μια τροχιά ενός τυχόντος σημείου που συμβολίζουμε με $j = \{ s, (x_0, y_0), (x_1, y_1), .. (x_m, y_m) \}$, με s να είναι το καρέ στο οποίο στο οποίο ξεκινά η τροχιά και με τα ζεύγη (x_k, y_k) να είναι οι συντεταγμένες του σημείου στο $s + k$ καρέ.

Οι θεμελιώδεις "χτύποι" (beats) ενός ρυθμού της κίνησης που προέκυψε θα είναι οι παύσεις, οι στροφές και οποιεσδήποτε άλλες σημαντικές αλλαγές στο πλάτος της κίνησης και στην κατεύθυνση της. Για να τους διακρίνουμε, αρχικά θα επιχειρήσουμε να ελαχιστοποίησουμε τους θορύβους στην τροχιά που υπολογίστηκε, οι οποίοι θεωρούμε ότι ακολουθούν την Γκαουσιανή κατανομή.

Περνάμε λοιπόν την τροχιά από ένα βαθυπερατό φίλτρο, υπολογίζοντας τη συνέλιξη της με τη συνάρτηση του Gauss

$$G(t) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{t^2}{2\sigma^2}}$$

όπου σ είναι η τυπική απόκλιση και t είναι η διαφορά ως προς τον αριθμό ενός τυχαίου καρέ από το κεντρικό της κατανομής Gauss.
Η οριζόντια κίνηση φιλτράρεται ως

$$\widehat{j_x}(x) = \sum_{u=0}^{m} j_x(u) \cdot G(i-u)$$

με $j_x(i)$ να είναι η οριζόντια μετατόπιση στο καρέ $s + i$.
Αντίστοιχα φιλτράρεται και η κατακόρυφη κίνηση, ώστε συνολικά η τροχιά j να είναι πιο ομαλή.

Για να εντοπίσουμε τις παύσεις μέσα στην τροχιά της κίνησης θα εξετάσουμε τη χρονική εξέλιξη του "μεγέθους" της κίνησης (motion magnitude)
$Hg = \{g_0, g_1, .., g_{m-1}\}$ με τα g_i να είναι τα πλάτη της κίνησης μεταξύ των καρέ $i,\ i+1$ δηλαδή:

$$g_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

Το g θα μειώνεται καθώς το σώμα επιβραδύνει και θα παρουσιάζει τοπικό ελάχιστο κατά τη στιγμή μιας παύσης. Για να εντοπίσουμε τοπικά ελάχιστα αυτού του μεγέθους θα υιοθετήσουμε μια τροποποίηση του γνωστού αλγορίθμου Hill Climbing [36] .

Γενικά, ο αλγόριθμος χρησιμοποιείται για τη μεγιστοποίηση (ή την ελαχιστοποίηση ανάλογα με την περίσταση) μιας δοθείσας συνάρτησης $f(X)$ όπου το X είναι ένα διάνυσμα συνεχών ή διακριτών τιμών. Σε κάθε επανάληψη (iteration), ο αλγόριθμος τροποποιεί ένα στοιχείο του X και ελέγχει εάν η τροποποίηση βελτίωσε την τιμή της συνάρτησης σύμφωνα με το αρχικό κριτήριο (αύξηση ή μείωση). Εάν όντως τη βελτίωσε, η τροποποίηση γίνεται αποδεκτή και ο αλγόριθμος συνεχίζεται μέχρι το σημείο που δεν μπορεί να βρεθεί τρόπος να βελτιωθεί η αρχική συνάρτηση.

Στην προκειμένη περίπτωση οι Chu και Tsai δίνουν ως είσοδο το μέγεθος Hg που υπολογίστηκε πιο πάνω και δουλεύουν ως εξής: Με σταθερό σε κάθε επανάληψη ένα δεδομένο καρέ, έστω `cIdx`, συγκρίνουμε το μέγεθος της κίνησης του με τα γειτονικά καρέ. Εάν κάποιο από αυτά , έστω `nIdx`, παρουσιάζει μικρότερη κίνηση τότε αντικαθιστά το αρχικό καρέ. Η διαδικασία αυτή συνεχίζεται έως ότου το $Hg(cIdx)$ να είναι το ελάχιστο στη γειτονιά αυτή, η οποία έχει προκαθοριστεί να περιλαμβάνει τα καρέ $cIdx+1,..,cIdx+\Delta$ με $\Delta = 7$. Μετά από αυτή την αναζήτηση τοπικού ελαχίστου, χρησιμοποιείται ο αλγόριθμος Hill Climbing, για να βρεθεί ένα τοπικό μέγιστο. Από εκεί θα ξεκινήσει η διαδικασία για να βρεθεί το επόμενο στη σειρά τοπικό ελάχιστο. Επαναλαμβάνουμε έως ότου ελεγχθεί όλο το μέγεθος Hg και πλέον τα τοπικά ελάχιστα μπορούν να θεωρήθουν ως υποψήφια "χτύποι" (beats) του ρυθμού.

Η παραπάνω περιγραφή του αλγορίθμου γράφεται σε ψευδοκώδικα με είσοδο το Hg και έξοδο το αρχικά κενό σύνολο L (που περιλαμβάνει τα τοπικά ελάχιστα) ως εξής:

```
L ← ∅
cIdx← 0
decreaseFlag ← True
while cIdx ≤ m-1
 if decreaseFlag
        nIdx ← arg min_i H_g[i] , i ϵ 𝒩(cIdx)
        if H_g[nIdx] ≤ H_g[cIdx]
        cIdx ← nIdx
 else
        L = L∪{cIdx}
        decreaseFlag ← False
 else
        nIdx ← arg max_i H_g[i] , i ϵ 𝒩(cIdx)
        if H_g[cIdx] ≤ H_g[nIdx]
         cIdx ← nIdx
 else
        decreaseFlag ← True
end while
```

Αντίστοιχα, θα αναζητήσουμε υποψήφιους χτύπους στις στροφές του χορευτή υπολογίζοντας τον προσανατολισμό του στα διάφορα καρέ και "χτίζοντας" το μέγεθος H_o (ιστορικό προσανατολισμού, Orientation History) ως $H_o = \{o_0, o_1, ., o_{m-1}\}$ όπου τα o_i είναι τα διανύσματα κίνησης από το καρέ ανάμεσα στα καρέ $i, i+1$ και τα οποία μάλιστα έχουν γίνει μοναδιαία διαιρώντας με το εκάστοτε motion magnitude:

$$o_i = \frac{1}{g_i}(x_{i+1} - x_i, y_{i+1} - y_i)$$

Βασιζόμενοι σε αυτά τα δεδομένα, οι συγγραφείς παρατήρησαν ότι ανάμεσα σε καρέ όπου η κίνηση διατηρείται σε σταθερή γενικά κατεύθυνση το εσωτερικό γινόμενο των o_i, o_{i+1} προσεγγίζει το 1. Αντίθετα,στα καρέ που παρουσιάζονται στροφές, το εσωτερικό γινόμενο μειώνεται και σε κάποιες περιπτώσεις φτάνει να αλλάξει πρόσημο. Επομένως, αρκεί να υπολογιστούν τα εσωτερικά γινόμενα των διανυσμάτων κίνησης σε μια ακολουθία από καρέ και να υπολογιστεί ένας μέσος όρος τους. Εάν αυτός είναι μικρότερος από ένα ορισμένο κατώφλι τιμών, βρίσκουμε τη στιγμή κατά την οποία αλλάζει περισσότερο. Η στιγμή αυτή αποθηκεύεται και γίνεται το νέο σημείο.
Η διαδικασία επαναλαμβάνεται μέχρι να ελεγχθεί όλο το ιστορικό προσανατολισμού και όμοια με πριν τα σημεία που εξάγονται είναι πάλι υποψήφια ως χτύποι του ρυθμού της κίνησης.

```
U ← ∅
start ← 0
while start ≤ m-1
 history← 0
 diff ← ∅
 avg ← ∅
 for j= start + 1 to m-1
        history← history + H_o[start] · H_o[j]
        avg[j]← history/j− start
        diff[j] ← avg[j] − avg[j-1]
        if avg[j]≤ threshold
        i* = arg max start <I<j diff [i]
        U ← U u {i*}
        start = j
        break
end while.
```

Σε αυτό το κομμάτι θα επιχειρήσουμε να ανακαλύψουμε την "κυρίαρχη" περίοδο (dominant period) από τους υποψήφιους χτύπους του ρυθμού της κίνησης ώστε να εξάγουμε τους χτύπους αναφοράς (reference beats). Με βάση αυτούς, οι πραγματικοί χτύποι θα είναι όλοι οι υποψήφιοι που δεν απέχουν πολύ τους χτύπους αναφοράς.

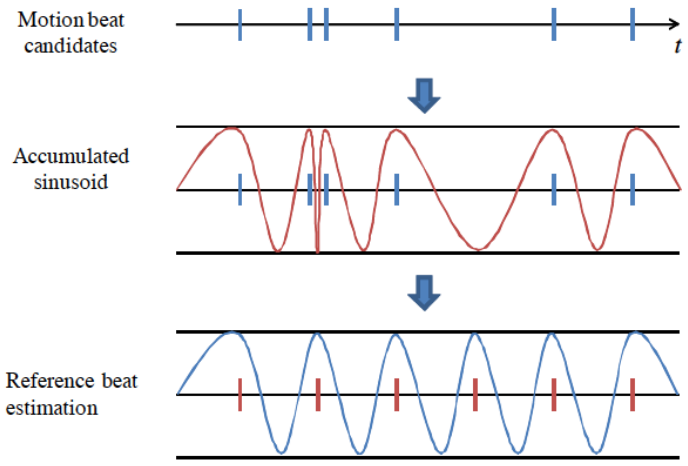
ΑΠΛΗ ΤΡΟΧΙΑ

Γνωρίζοντας τα υποψήφια beats της κίνησης, θα υπολογίσουμε το διάστημα επανάληψης του παλμού (pulse repetition interval, PRI) βασιζόμενοι σε ένα σήμα που προέκυψε από τα χρονικές στιγμές των χτύπων. Δεδομένης μιας τροχιάς, θα συμβολίσουμε μια ακολουθία από χτύπους-beats ως $S = \{b_0, b_1, .., b_n\}$ όπου τα b_i είναι ο αριθμός του καρέ στο οποίο λαμβάνει χώρα το i −οστό υποψήφιο beat.

Η παραγωγή τους θα μπορούσε να μοντελοποιηθεί ως $b_i = \varphi + k_i T + \eta_i$, όπου T είναι η άγνωστη περίοδος, φ είναι μια φάση που κυμαίνεται στο διάστημα $[0, T)$, η_i είναι θόρυβος που προκαλείται από την κίνηση του χορευτή και k_i είναι ένας θετικός ακέραιος- δείκτης του χτύπου. Αντίστοιχα, οι χτύποι αναφοράς μοντελοποιούνται ως $r_j = \varphi + jT$. Για να υπολογίσουμε τους χτύπους αναφοράς εργαζόμαστε ως εξής. Αρχικά η ακολουθία των υποψήφιων χτύπων μετατρέπεται σε ένα συνεχές σήμα με εξίσωση

$$y = \begin{cases} cos\left(\ 2\pi \cdot \frac{t - b_{k-1}}{b_k - b_{k-1}}\ \right) & , \varepsilon\acute{\alpha}\nu\ b_{k-1} < t < b_k \\ 1 & , \ \varepsilon\acute{\alpha}\nu\ t = b_k \end{cases} \quad \text{με} \ \ k = 2,3,\ldots,n$$

Προφανώς,το σήμα παρουσιάζει μέγιστο ίσο με 1 όταν $t = b_k$, δηλαδή όταν εμφανίζεται ένα υποψήφιο beat. Όταν το t είναι ανάμεσα σε δύο υποψήφια beat το σήμα παίρνει την τιμή του μέσω ενός συνημιτόνου. Για κάθε υποψήφιο χτύπο εφαρμόζεται ένα συνημίτονο "κεντραρισμένο" στο b_i και όλα τα συνημιτονοειδή μαζί συνθέτουν το τελικό σήμα όπως φαίνεται και στην παρακάτω απεικόνιση (εικόνα 5.8).



Εικόνα 5.8: Εξαγωγή χτύπων και τελικό σήμα

Γνωρίζοντας το σήμα $y(t)$, οι Chu και Tsai στηρίζονται στο βιβλίο Digital Signal Processing: Principles, Algorithms and Applications [37] για να υπολογίσουν την πυκνότητα του φάσματος ισχύος (power spectrum density,PSD) και από εκεί να εξάγουν την κυρίαρχη περίοδο. Η μέθοδος που προτείνεται στο [11] βρίσκει την ενέργεια του συνολικού συνημιτονοειδούς σε διαφορετικές ζώνες συχνότητας και από το θεώρημα Nyquist περί δειγματοληψίας ξέρουμε πως η μέγιστη συχνότητα που θα μπορούμε να διακρίνουμε θα είναι η μισή από το ρυθμό δειγματοληψίας.

Ωστόσο, είμαστε σίγουροι ότι η συχνότητα των χτύπως της κίνησης (motion beats) θα είναι χαμηλότεροι από το μισό του ρυθμού δειγματοληψίας, που στην προκειμένη περίπτωση είναι το frame rate που ισούται με 30fps αφού είναι φυσικώς αδύνατο για το ανθρώπινο σώμα να κινηθεί τόσο γρήγορα.

Έτσι έχουμε πως

$$PSD_y(f) = |\sum_{j=1}^{M} y(t)e^{i2\pi\, f t_j}|^2$$

όπου M είναι το μήκος του συνολικού συνημιτονοειδούς και f είναι ο δείκτης μιας ζώνης συχνότητας.

Η κυρίαρχη συχνότητα θα είναι εκείνη που δίνει τη μέγιστη PSD_y , δηλαδή

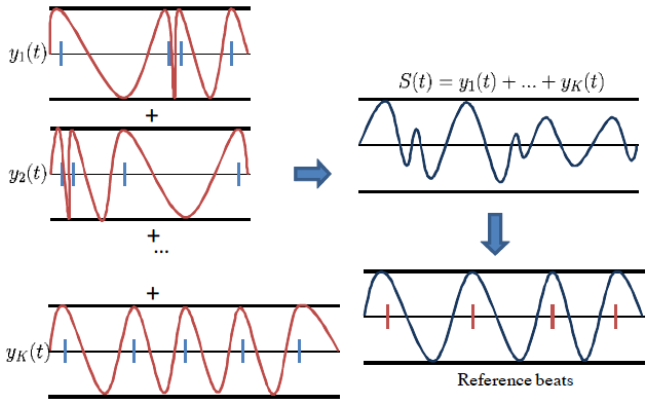
$$f_d = arg\, max_f\, PSD_y(f)$$ και προφανώς $T_d = \frac{1}{f_d}$.

Έπειτα, μπορούμε να υπολογίσουμε και τη φάση φ που εισάγαμε πιο πάνω ως τη φάση εκείνη που προκαλεί το το μέγιστο άθροισμα περιοδικών θετικών κορυφών.
Με το $\varphi\, \epsilon\, [0, T)$ λοιπόν θα είναι:

$$\bar{\varphi} = arg\, max_{\varphi} \sum_{j=1}^{M-1} y\,(jT + \varphi)$$

ΠΟΛΛΑΠΛΕΣ ΤΡΟΧΙΕΣ

Στην περίπτωση που υπάρχουν περισσότερες από μία τροχιές από τις οποίες πρέπει να εξαχθεί μια ακολουθία από beats για το ρυθμό της κίνησης, η προηγούμενη διαδικασία πρέπει να επεκταθεί κατάλληλα: Από κάθε τροχιά θα πρέπει να διαμορφωθεί ένα σήμα $y_i(t)$,ώστε εάν έχουμε K τροχιές το τελικό σήμα να προκύψει από την υπέρθεση των $y_1(t), .., y_i(t), .. y_K(t)$: $S(t) = y_1(t) + y_2(t) + .. + y_K(t)$. Αυτό το σήμα S θα χρησιμοποιήσουμε για να υπολογίσουμε και πάλι την PSD που όμοια με πριν θα είναι $PSD_S(f) = |\sum_{j=1}^{M} S(t)e^{i2\pi\, f t_j}|^2$



Εικόνα 5.9: Σύνθεση τροχιών

Ακόμα λοιπόν και στην περίπτωση που διαφορετικά σημεία του σώματος του χορευτή κινούνται σε διαφορετικές τροχιές μεταξύ τους (κάτι που παρατηρείται πολύ συχνά) είδαμε πώς να εξάγουμε τη θέση των χτύπων αναφοράς του

ρυθμού της κίνησής του.Μένει πλέον να βρούμε τους πραγματικούς χτύπους και να φιλτράρουμε τις όποιες παρεκκλίσεις ή ενδείξεις θορύβου. 'Ενας υποψήφιος χτύπος b_i θεωρούμε πως είναι στη γειτονιά ενός χτύπου αναφοράς r_j εάν $r_j - \alpha\frac{T}{2} \leq b_i \leq r_j + \alpha\frac{T}{2}$ όπου το α θα καθορίζει κάθε φορά το εύρος της γειτονιάς. Εάν είναι πολύ μεγάλο η ανίχνευση δεν θα είναι ακριβής, ενώ εάν είναι πολύ μικρό είναι πιθανό να φιλτράρονται λανθασμένα και στοιχεία της κίνησης. Αφαιρώντας λοιπόν τα στοιχεία που θεωρούμε ότι δεν ανήκουν στην κίνηση και το ρυθμό που την διέπει βρίσκουμε τα beats της κίνησης ως :

$$B = \{b_j^* \mid b_j^* = arg\ min_{b \in N(r_j)} |b - r_j| , j = 1, .., N\}$$

με b_j^* να είναι οι χτύποι που ελέγχθηκε ότι ανήκουν στη γειτονιά $N(r_j)$ των beats αναφοράς.
Έχοντας λοιπόν το σύνολο των χτύπων-beats της κίνησης και έχοντας τη δυνατότητα μέσω της γνωστής διαδικασίας του beat tracking να εξάγουμε και τους χτύπους ηχητικού σήματος μπορούμε να επιχειρήσουμε μια ευθυγράμμιση των δύο στοιχείων. Για διευκόλυνση της διαδικασίας αυτής υιοθετούμε τον εξής συμβολισμό: Τα beats κίνησης και ήχου θα είναι αντίστοιχα $B_{mt} = \{b_0, b_1, .., b_{M-1}\}$ και $B_{mu} = \{b_0, b_1, .., b_{N-1}\}$ όπου τα b_i θα είναι δυαδικές τιμές (0 ή 1) ανάλογα με το εάν υπάρχει χτύπος στο $i-$οστό χιλιοστό του δευτερολέπτου του βίντεο ή όχι: 1 εάν υπάρχει και 0 εάν δεν υπάρχει. Έτσι, μια απλή εφαρμογή ευθυγράμμισης θα ήταν η αναζήτηση δοθέντων για παράδειγμα δύο ακολουθιών $B_{mt} = 1001001$, $B_{mu} = 101010101011$ των μακρύτερων ίδιων υποακολουθιών των εισόδων. Το πρόβλημα των ίδιων υποακολουθιών είναι πολύ κοινό στο χώρο της Πληροφορικής με πολλές και πολύ αποδοτικές επιλύσεις, ωστόσο στην προκειμένη περίπτωση θα πρέπει να λάβουμε υπ'όψιν πως στο χορό (όπως και στο αντίστοιχο μουσικό σήμα) δίνεται πολύ μεγαλύτερη βαρύτητα στους χτύπους της ακολουθίας, τα ψηφία 1 δηλαδή, παρά στα μηδενικά. Για το λόγο αυτό, ακολουθείται άλλη προσέγγιση.

Προκειμένου, να μετρήσουμε το βαθμό της ταύτισης των ακολουθιών ορίζουμε την απόσταση ενός beat κίνησης $B_{mt}[i]$ και του κοντινότερου ηχητικού του beat στην ακολουθία με τη μετάθεση Δ ως

$$d(\Delta, i) = min_{0 \leq j \leq N-1} |i - j| , \forall B_{mu}[j + \Delta] = 1,$$

όπου N είναι το μήκος της ακολουθίας B_{mu} .

Εισάγουμε και την έννοια της συνοχής ως $C(\Delta) = \frac{1}{M}\sum_{i=0}^{M-1} \frac{B_{mt}[i]}{d(\Delta,i)+1}$ και καταλαβαίνουμε ότι αυξάνεται όσο μικρότερες είναι οι χρονικές αποστάσεις μεταξύ των δύο ειδών χτύπων.Επίσης, η διαφορά $D(\Delta)$ υπολογίζεται ως $D(\Delta) = \frac{1}{M}\sum_{i=0}^{M-1} B_{mt}[i] \cdot d(\Delta, i)$ και πλέον μπορούμε να βρούμε τον οριστικό βαθμό ταύτισης (degree of matching) :

$$DOM(\Delta) = \frac{C(\Delta)}{D(\Delta)}$$

Η καταλληλότερη μετάθεση της δοθείσας ακολουθίας χτύπων κίνησης θα είναι αυτή που μεγιστοποιεί το βαθμό ταύτισης, δηλαδή η Δ^* για την οποία θα ισχύει ότι:

$$\Delta^* = arg\ max_{0 \le k \le N-M} DOM(k)$$

Έχοντας λοιπόν βρει τη βέλτιση μετάθεση Δ^*, οι Chu και Tsai καταλήγουν πως η υποακολουθία $\{b_{\Delta^*}, b_{\Delta^*+1}, \ldots, b_{\Delta^*+M-1}\}$ της αρχικής B_{mu} μπορεί να χρησιμοποιηθεί για να αντικαταστήσει το αρχικό μουσικό υπόβαθρο του βίντεο του χορού στα κατάλληλα καρέ.

5.4. ΣΥΜΠΕΡΑΣΜΑΤΑ

Συνοπτικά, και στις δύο προσεγγίσεις του τι είναι ο ρυθμός για ένα οπτικοακουστικό μέσο, βλέπουμε ότι γίνεται μια προσπάθεια να ποσοτικοποιηθεί η εξέλιξη της δράσης που παρατηρείται. Στην τάση που παρουσιάστηκε πρώτη, μπορούμε να πούμε ότι ο ρυθμός (*Pace*) δεν έχει ιδιαίτερο μουσικό περιεχόμενο αφού ορίζεται ως αριθμητικό μέγεθος και η μόνη σύνδεση που μπορεί να γίνει είναι με το *tempo* ενός μουσικού έργου. Υπό αυτό το πρίσμα, είναι όντως χρήσιμο στον ορισμό του όπως προφανώς και το μουσικού tempo, για την κατηγοριοποίηση και τη διάκριση των διαφορετικών έργων αλλά και του χαρακτήρα που αυτά εμπεριέχουν. Όπως ακριβώς άλλη αίσθηση προκαλεί στον ακροατή ένα έργο με *allegro* tempo κι άλλη αίσθηση ένα με *adagio,* έτσι ακριβώς σκηνές με μεγάλες αποκλίσεις στα εξαχθέντα *Pace* θα προκαλούν στο θεατή πολύ διαφορετικές αντιδράσεις.

Στην δεύτερη τάση, είναι προφανές ότι το πεδίο εφαρμογής είναι πιο σύντομα βίντεο (ή ίσως και μεγαλύτερα αρκεί να μην υπάρχει μεγάλη μεταβολή στο περιεχόμενο τους). Όπως και στη διαδικασία υπολογισμού του *Pace,* σημαίνουσα σημασία έχει η κίνηση και το πόσο έντονη είναι. Πλέον, όμως η προσπάθεια μετατοπίζεται στην εύρεση των χρονικών εκείνων σημείων που μπορούμε να διακρίνουμε ως πιο σημαντικά. Είτε αυτά είναι χαμηλές συχνότητες ενός υποθετικού ηχητικού σήματος που δείχνουν να το χωρίζουν σε επαναλαμβανόμενα μέρη, είτε είναι τα μέρη μιας ανθρώπινης κίνησης που το μάτι αντιλαμβάνεται ως πιο έντονα (παύσεις, στροφές κ.ά.) αυτό που προσπαθούμε ουσιαστικά να βρούμε σε ποια ακριβώς σημεία θα έπαιζε νότες ή θα χρησιμοποιούσε παύσεις ένας παίχτης κρουστού οργάνου στην προσπάθεια του να συνοδεύσει το βίντεο που παρακολουθεί. Για το λόγο αυτό, οι προσεγγίσεις αυτής της "κατηγορίας" μπορούν να συζευκτούν πιο ομαλά με έναν ολοκληρωμένο μουσικό ρυθμό και να θεωρηθούν μέρος μιας αλγοριθμικής σύνθεσης όπως την ορίσαμε προηγουμένως, ενώ αυτές της πρώτης κατηγορίας να χρησιμοποιηθούν περισσότερο για κινηματογραφικές εφαρμογές.

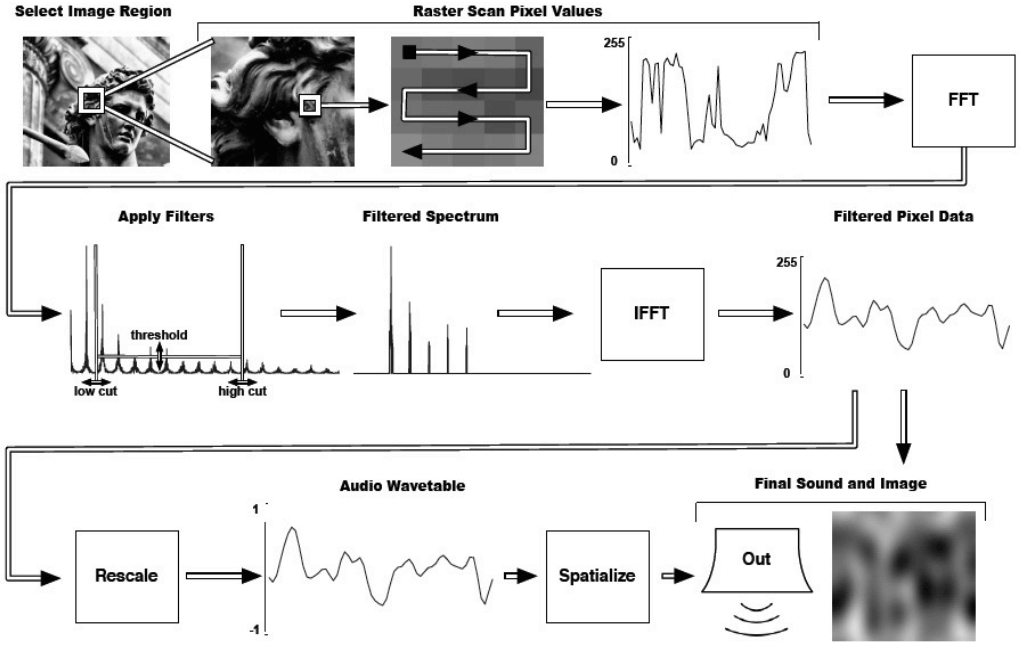
5.5. RYAN McGEE - VOSIS

Ακόμα μια εφαρμογή του sonification που αξίζει να αναφέρουμε, ακόμα κι αν δεν έχει άμεση σχέση με το ρυθμό, είναι το σύστημα VOSIS. Το VOSIS είναι μια διαδραστική διαπροσωπεία για υπολογιστές και tablets που ανέπτυξε ο Ryan McGee του πανεπιστημίου της California (University of California, Santa Barbara) για το sonification εικόνων και βίντεο σε grayscale μορφή [38]. Το

σύστημα στηρίζεται στην προσέγγιση χρόνου-συχνότητας (time-frequency approach) κατά την οποία η εικόνα ή ένα καρέ του βίντεο λειτουργεί ως φασματογράφος (spectograph) για τον ήχο. Ο χρήστης μπορεί να επιλέξει την περιοχή την οποία επιθυμεί να "ηχοποιήσει" και να μεταβάλλει μόνος του παραμέτρους όπως για παράδειγμα ο θόρυβος που θα εισαχθεί στην εικόνα. Με το που επιλεχθεί μια περιοχή της εικόνας, παράγεται αυτόματα ένας ήχος αντίστοιχος των πίξελ της συγκεκριμένης περιοχής κι έτσι η όλη εικόνα μετατρέπεται σε ένα πρωτότυπο μουσικό όργανο.

Η διαδικασία σύνθεσης του ήχου είναι η εξής: Όταν επιλεχθεί μια περιοχή σαρώνεται "σπειροειδώς" από αριστερά προς τα δεξιά κι από πάνω προς τα κάτω ώστε να διαβαστούν όλες οι τιμές των pixel από 0 έως 255. Αυτές οι τιμές μετατρέπονται σε μια ενιαία κυματομορφή στην οποία εφαρμόζεται ο γρήγορος μετασχηματισμός
Fourier (FFT) για να εξαχθούν τα πλάτη και οι φάσεις των συχνοτήτων που περιέχονται σε εκείνη την περιοχή. Το φάσμα που προκύπτει "περνάει" από διάφορα φίλτρα ανάλογα με τις ρυθμίσεις που επιλέγει ο χρήστης κι αφού λάβει την τελική του μορφή, πρέπει να επανέλθουμε σε τιμές pixel. Έτσι, εφαρμόζεται αντίστροφος γρήγορος μετασχηματισμός Fourier (IFFT) και προκύπτουν τα τελικά φιλτραρισμένα pixels. Τέλος, βασιζόμενοι στη μέθοδο της scanned synthesis που εισήγαγαν οι Verplank, Mathews και Shaw,τα pixels αυτά μετατρέπονται σε κυματομορφή ήχου και παράγεται το τελικό αποτέλεσμα [39].

Αξίζει να αναφέρουμε το τελικό αποτέλεσμα προκύπτει αυστηρά και μόνο από την επεξεργασία των οπτικών δεδομένων: Οποιαδήποτε μεταβολή στον ήχο-έξοδο, προκύπτει από μεταβολή στην περιοχή που επιλέγει ο χρήστης και επομένως η προαναφερθείσα διαδικασία επαναλαμβάνεται από την αρχή για να προκύψουν οι τιμές των pixel πριν και μετά την εφαρμογή των μετασχηματισμών Fourier από τις οποίες εξαρτάται ο εξαγόμενος ήχος. Παρακάτω φαίνεται σχηματικά η όλη διαδικασία όπως την έχει σχεδιάσει ο ίδιος ο McGee.



Εικόνα 5.10: Το σύστημα VOSIS σχηματικά

6. ΚΕΦΑΛΑΙΟ 6 : ΥΛΟΠΟΙΗΣΗ

6.1. ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΕΡΕΥΝΗΤΙΚΗΣ ΔΙΑΔΙΚΑΣΙΑΣ

Στα πλαίσια της παρούσας διπλωματικής εργασίας, επιδιώκουμε να προσεγγίσουμε τη ρυθμική υπόσταση ενός βίντεο συσχετίζοντας τον οπτικό και τον ηχητικό ρυθμό. Θα επιχειρήσουμε αρχικά να υπολογίσουμε ένα μέγεθος το οποίο να είναι χαρακτηριστικό του ρυθμού του βίντεο, σύμφωνα με την έρευνα των Adams et al και ύστερα θα διαμορφώσουμε μια συστηματοποιημένη διαδικασία για να αντιστοιχούμε ένα απόσπασμα βίντεο σε ένα κατάλληλο ρυθμικό, μουσικό μέρος.

Για όλο το πρακτικό μέρος της εργασίας, ως βίντεο εισόδου χρησιμοποιήθηκαν διαφορετικά αποσπάσματα του φιλμ μικρής διάρκειας του 1937, "*An Optical Poem*" του *Oskar Fischinger* που αναφέρθηκε και στην εισαγωγή. Ο λόγος που επιλέχθηκε το συγκεκριμένο βίντεο είναι το γενονός ότι το οπτικό περιεχόμενο του, οι εναλλαγές γεωμετρικών μορφών, καθιστούν πιο σαφή τη μεθοδολογία και την πρόθεση της υλοποίησης σε MATLAB. Επίσης, η φύση της ίδιας της ταινίας και το κίνητρο της δημιουργίας της από τον Fischinger, διευκολύνουν τη σύνδεση με το ηχητικό περιεχόμενο και την εξαγωγή συμπερασμάτων από τα αποτελέσματα που θα προκύψουν.

Αφετηρία της διαδικασίας διαμόρφωσης του αλγορίθμου επεξεργασίας ήταν το μουσικό περιεχόμενο του βίντεο, η 2η Ούγγρικη Ραψωδία του Franz Liszt. Στηριζόμενοι στη σύνδεση της Ραψωδίας με την εξέλιξη των μορφών του βίντεο, επιχειρήθηκε η εύρεση "φράσεων" μέσα από το όλο έργο οι οποίες επαναλαμβανόμενες να μπορούν να παράγουν μια ρυθμική δομή. Οι φράσεις αυτές ήταν όλες της διάρκειας των 3-4 sec και όντως η αναπαραγωγή τους σε ένα βρόχο (loop) κάνει εμφανή τον παλμό που τις διέπει λόγω της σχετικά απλής μελωδικής τους εξέλιξης. Έπειτα, "πάνω" από τις φράσεις αυτές παίχτηκαν απλοί ρυθμοί στα τύμπανα, μέσω του λογισμικού Ableton Live, μετατρέποντας ουσιαστικά το μουσικό σήμα σε ένα υποτυπώδες beat και ενισχύοντας τη ρυθμική φύση της φράσης. Τα οπτικά αποσπάσματα που συνοδεύουν τις φράσεις αυτές, θα χρησιμοποιηθούν παρακάτω για την δευτερεύουσα εφαρμογή, τον υπολογισμό δηλαδή του Video Pace με βάση την παλαιότερη έρευνα που προαναφέρθηκε: Τα διαφορετικά μέρη της Ραψωδίας του Liszt που "ντύνουν" τα διαφορετικά αποσπάσματα, παρουσιάζουν παρόμοια μετρική δομή αλλά διαφορετικά τέμπο (αν απομονωθούν και εξεταστούν ανεξάρτητα) και οι διαφορές αυτές ενυπάρχουν και στα διαφορετικά οπτικά μέρη. Όπως γίνεται και πιο εύκολα κατανοητό με τη συνοδεία των ντραμς, τα μέρη με πιο "ζωηρές" συνοδευτικές μελωδίες παρουσιάζουν πιο γρήγορη κίνηση και πιο έντονες εναλλαγές. Έτσι, η σύγκριση των εξαχθέντων μετρήσεων για το ρυθμό του βίντεο με το μουσικό ρυθμό των εκάστοτε μουσικών αποσπασμάτων μπορεί να ποσοτικοποιηθεί πιο αυστηρά και να συμπεράνουμε κατά πόσο έχει λογική βάση η εφαρμογή της προυπάρχουσας έρευνας σε βίντεο όπως αυτά που επιλέξαμε.

Ωστόσο, είναι προφανές ότι η επιλογή του έργου του Liszt ως συνοδευτικό των οπτικών μορφών ήταν αυστηρά υποκειμενική και προϊόν της κρίσης του καλλιτέχνη. Έτσι, δεν αποτελεί ασφαλή μέθοδο το να στηριχθούμε στο ηχητικό μέρος για να εξάγουμε το οποιοδήποτε συμπέρασμα. Αντ' αυτού, είναι σαφώς πιο ορθό να χρησιμοποιηθεί η ίδια η εικόνα και πως αυτή εξελίσσεται στο χρόνο για να βρούμε ένα ρυθμό του βίντεο. Παρατηρώντας τα αποσπάσματα-φράσεις που προαναφέρθηκαν, αν απομονώσουμε μόνο την εικόνα και τα συνοδευτικά τύμπανα που προστέθηκαν, φαντάζει πιο "φυσικό" όλα τα χτυπήματα να συμπίπτουν με τα στοιχεία της εικόνας που το μάτι αντιλαμβάνεται ως πιο έντονα: Είτε απότομες στιγμιαίες αλλαγές που διαφέρουν κατά πολύ από τη ροή των προηγούμενων καρέ, είτε στιγμές όπου οι απεικονιζόμενες μορφές είναι πιο έντονες, σε χρώμα ή μέγεθος. Έτσι λοιπόν, αρκεί να ανιχνεύουμε αυτό που μένει να ορίσουμε ως απότομες αλλαγές στην εικόνα, και τις στιγμές που εντοπίζονται, το σύστημα να "παίζει" αυτόματα μία νότα στα τύμπανα: Ένα χτύπημα στην μπότα (bass drum) ή στο ταμπούρο (snare drum) ή ακόμα και σε ένα από τα πιατίνια.

6.2. ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ ΕΠΕΞΕΡΓΑΣΙΑΣ

6.2.1 ΠΡΩΤΗ ΕΦΑΡΜΟΓΗ: ΥΠΟΛΟΓΙΣΜΟΣ ΤΟΥ VIDEO PACE

Όπως είδαμε και προηγουμένως, σύμφωνα με τους Adams, Dorai και Venkatesh, το Pace ενός φιλμ είναι

$$P(n) = \alpha \frac{med_s - s(n)}{\sigma_s} + \beta \frac{m(n) - \mu_m}{\sigma_m}$$

με τα α, β να είναι σταθερές ίσες με 0.5, το s να είναι το μήκος μιας σκηνής μετρημένο σε καρέ,το m η "ποσότητα" της κίνησης, το μ η μέση τιμή του μεγέθους που σημειώνεται ως δείκτης, $median$ η διάμεσος, σ η τυπική απόκλιση του μεγέθους-δείκτη, ενώ n να είναι ο αύξων αριθμός της κάθε σκηνής. Γίνεται σαφής λοιπόν η σημαίνουσα σημασία του μεγέθους του m, motion magnitude για την εξαγωγή του Pace. Στα πλαίσια της δικιάς μας εφαρμογής ωστόσο δεν είναι υλοποιήσιμος ο υπολογισμός της κάθε κίνησης της κάμερας λόγω βασικά των υπολογιστικών περιορισμών της CPU. Για το λόγο αυτό, ορίζουμε την ποσότητα της κίνησης με βάση το μέγεθος της οπτικής ροής - optical flow.

Ως optical flow κατανοούμε το μοτίβο μιας φαινόμενης κίνησης αντικειμένων, επιφανειών και ακμών σε μια οπτική σκηνή. Για την εύρεση του, επιχειρείται συνήθως ο υπολογισμός της κίνησης μεταξύ δύο διαδοχικών καρέ που ενεργοποιούνται στις στιγμές t και $t + \Delta t$. Το περιβάλλον MATLAB περιέχει ενσωματωμένη την κλάση `vision.OpticalFlow` που επιτρέπει διάφορες λειτουργίες με βάση την εξαγωγή του Optical Flow ενός βίντεο. Μία από αυτές, είναι ο υπολογισμός της ταχύτητας κάθε pixel ενός βίντεο σε όλη τη διάρκεια του. Με βάση αυτή τη δυνατότητα, θα υπολογίσουμε το προσαρμοσμένο motion magnitude,βρίσκοντας ένα μέσο μέτρο ταχύτητας συνολικά για κάθε ένα από τα σύντομα βίντεο που θα συγκρίνουμε.

Το MATLAB υπολογίζει το Optical Flow κάθε φορά κάνοντας χρήση της μεθόδου Horn-Schunck. Ο αλγόριθμος υποθέτει ότι η ροή είναι λεία σε όλη την εικόνα και ορίζει το Optical Flow ως μία συνάρτηση ενέργειας που πρέπει να ελαχιστοποιηθεί. Για δισδιάστατες εικόνες, η συνάρτηση αυτή είναι:

$$E = \iint [\ (I_x u + I_y v + I_t)^2\ +\ a^2(\|\nabla u\|^2 + \|\nabla v\|^2)\ ]\ dxdy$$

όπου:
I_x, I_y, I_z είναι οι μερικές παράγωγοι της έντασης της εικόνας ως προς τις διαστάσεις x,y και το χρόνο, $\vec{V} = [u(x,y), v(x,y)]^{\mathrm{T}}$ είναι το διάνυσμα της "οπτικής ροής" και α είναι μια σταθερά κανονικοποιήσης (όσο μεγαλύτερη η τιμή του, τόσο πιο λεία είναι η ροή). Για την ελαχιστοποίηση αυτού του μεγέθους, η μέθοδος λύνει τις σχετικές πολυδιάστατες εξισώσεις Euler-Lagrange [40] . Αυτές είναι οι εξής:

$$\frac{\partial L}{\partial u} - \frac{\partial}{\partial x}\frac{\partial L}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial u_y} = 0$$

$$\frac{\partial L}{\partial v} - \frac{\partial}{\partial x}\frac{\partial L}{\partial v_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial v_y} = 0$$

όπου L είναι η παράγωγος της έκφρασης ενέργειας, δίνοντας

$$I_x(I_x u + I_y v + I_t) - a^2 \Delta u = 0$$

$$I_y(I_x u + I_y v + I_t) - a^2 \Delta v = 0$$

όπου $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y}$ είναι ο Λαπλασιανός τελεστής.

Στην πράξη, ο τελεστής (αναφέρεται και ως "η Λαπλασιανή") προσεγγίζεται με αριθμητικές μεθόδους, χρησιμοποιώντας τους σταθμισμένους μέσους των u, v σε γειτονιές των εκάστοτε pixel. Έτσι, τελικά οι παραπάνω εξισώσεις γράφονται ως:

$$({I_x}^2 + a^2)\ u + I_x I_y\ v = a^2 \bar{u} - I_t I_x$$

$$({I_y}^2 + a^2)\ v + I_x I_y\ u = a^2 \bar{v} - I_t I_y$$

δίνοντας τελικά τις λύσεις:

$$u^{k+1} = \bar{u}^k - \frac{I_x\ (I_x \bar{u}^k + I_y \bar{v}^k + I_t)}{{I_x}^2 + {I_y}^2 + a^2}$$

$$v^{k+1} = \bar{v}^k - \frac{I_y\ (I_x \bar{u}^k + I_y \bar{v}^k + I_t)}{{I_x}^2 + {I_y}^2 + a^2}$$

όπου με $k+1$ συμβολίζεται η επόμενη επανάληψη ενώ με k η τελευταία που έχει υπολογιστεί.

Βασιζόμενοι λοιπόν σε αυτή τη μέθοδο, θα εξάγουμε την ταχύτητα κάθε pixel σε μια ακολουθία από καρέ. Παρακάτω φαίνονται οι απαραίτητες αρχικοποιήσεις του κώδικα για τον υπολογισμό των ζητούμενων με τα απαραίτητα σχόλια:

```
% Αρχικοποίηση του αντικειμένου που θα διαβάσει το βίντεο
videoReader=vision.VideoFileReader('exampleVideo.mp4','ImageColorSpace',
                                'Intensity','VideoOutputDataType','uint8');
%Ορίζουμε αντικείμενο για τη μετατροπή των καρέ
%στο επιθυμητό format
converter=vision.ImageDataTypeConverter;

%Ορίζουμε ένα νέο αντικείμενο τύπου vision.OpticalFlow
opticalFlow=vision.OpticalFlow;

%Τα εξαχθέντα από το παραπάνω αντικείμενο θέλουμε να είναι
%μιγαδική μορφή ώστε να έχουμε τς συνιστώσες της ταχύτητας στους
%άξονες x,y
opticalFlow.OutputValue='Horizontal and vertical components
                                                    in complex form';

%Στο βίντεο εξόδου που συνοδεύει την επεξεργασία,
%σχεδιάζουμε και τα μη μηδενικά διανύσματα κίνησης
shapeInserter=vision.ShapeInserter('Shape','Lines','BorderColor',
                                  'Custom', 'CustomBorderColor',255);

%Ορίζουμε το αντικείμενο που θα αναπαράγει το βίντεο εξόδου
videoPlayer=vision.VideoPlayer('Name','Optical Poem Processing');
```

Παραθέτουμε επίσης και το βασικό loop επεξεργασίας, για τον υπολογισμό των ταχυτήτων. Το loop αυτό θα φανεί και αργότερα, εμπλουτισμένο με περισσότερες γραμμές κώδικα, απαραίτητες για την τελική εφαμοργή.

```
%Όσο υπάρχουν καρέ προς επεξεργασία, συνεχίζουμε το βρόχο
while ~isDone(videoReader)

    %Κρατάμε ένα καρέ κάνοντας τη συνέλιξη του βίντεο με τη
    %βηματική συνάρτηση
    frame=step(videoReader);

    %Αυξάνουμε τον έως τώρα υπολογισθέντα αριθμό καρέ
     numberOfFrames=numberOfFrames+1;

    %Μετατρέπουμε παρόμοια το καρέ στον κατάλληλο τύπο
    %Βρίσκουμε τη συνέλιξη του αντικείμενου μετά
    im=step(converter,frame);

    %Ο πίνακας των ταχυτήτων ανανεώνεται με το νέο καρέ μέσω
    %του αντικειμένου opticalFlow
    velocities=step(opticalFlow,im);

    %Αρχικοποιούμε τη μορφή των διανυσμάτων που θα σχεδιαστούν
    lines=videooptflowlines(velocities,20);

    %Και τα σχεδιάζουμε στην έξοδο αρκεί να είναι μη μηδενικά
    if ~isempty(lines)
        out=step(shapeInserter,im,lines)
        step(videoPlayer,out)
    end

end
```

Πλέον, στον δισδιάστατο πίνακα `velocities`, έχουμε στη μορφή $v_x + j\, v_y$ τις ταχύτητες όλων των pixels. Για τον ορισμό του motion magnitude, θα χρησιμοποιήσουμε το μέτρο των ταχυτήτων αυτών.

Ορίζουμε λοιπόν ως *ποσότητα της κίνησης - motion magnitude* το μέσο όρο του μέτρου της ταχύτητας της εκάστοτε ακολουθίας καρέ. Ωστόσο, οι Adams, Dorai και Venkatesh ορίζουν το motion magnitude ουσιαστικά ως μέγεθος μετατόπισης. Για το λόγο αυτό, ο ορισμός μέσω της ταχύτητας δεν συνάδει ιδιαίτερα με την παλιότερη έρευνα. Για να μεταβούμε στην έννοια της μετατόπισης, αρκεί να πολλαπλασιάσουμε τη μέση ταχύτητα των pixels με τη συνολική χρονική διάρκεια της εξέλιξης της κίνησης τους. Επειδή όμως στην επεξεργασία που κάνουμε, τα χρονικά στιγμιότυπα που μεταχειριζόμαστε είναι τα καρέ, δεν είναι εκ των προτέρων γνωστή. Γνωρίζοντας όμως το συνολικό αριθμό των καρέ και το Frame Rate μπορούμε εύκολα να υπολογίσουμε τη χρονική διάρκεια ενός βίντεο αποσπάσματος. Στα αντικείμενα `vision.VideoFileReader` του MATLAB, το Frame Rate έχει default τιμή ίση με 30 frames per second.Ο υπολογισμός του μέσου όρου του motion magnitude δεν παρουσιάζει δυσκολία. Βρίσκοντας τις διαστάσεις του video, υπολογίζουμε μέσω ενός διπλού βρόχου το συνολικό άθροισμα των μέτρων της ποσότητας κίνησης.

Έπειτα, αρκεί μία διαίρεση με το συνολικό αριθμό των pixels και το αποτέλεσμα σώζεται σε έναν array `motion_magnitude(i)` όπου θα κρατήσουμε τα αποτελέσματα από τα διαφορετικά αποσπάσματα. Η αποθήκευση σε πίνακα θα χρειαστεί και στον υπολογισμό των στατιστικών μεγεθών επί του motion magnitude που περιέχει ο ορισμός (μέση τιμή και απόκλιση).

```
%Διατηρούμε έναν αύξοντα δείκτη για τα διαφορετικά αποσπάσματα
videoCounter=1;

%Στη δομή videoInfo υπάρχουν πληροφορίες για τις διαστάσεις,
%το frame rate και άλλες ιδιότητες του video
videoInfo=info(videoReader);

videoWidth=videoInfo.VideoSize(1);
videoHeight=videoInfo.VideoSize(2);

frameRate=videoInfo.VideoFrameRate;

motionMagnitude=0;

for i=1:videoWidth
    for j=1:videoHeight

         %Αθροίζουμε όλα τα μέτρα ταχυτήτων
         %μέσω της συνάρτησης abs
        motionMagnitude=motionMagnitude + abs (velocities(j,i));

    end
end

%Ο συνολικός αριθμός pixels θα είναι όσο το "εμβαδόν" του βίντεο
NumberOfPixels=videoWidth*videoHeight;

%Υπολογίζουμε το μέσο όρο
motionMagnitude=motionMagnitude/NumberOfPixels;

%Η χρονική διάρκεια του βίντεο είναι τα συνολικά καρέ
%διά του frame rate
time=NumberOfFrames/frameRate;

%Η τελική ποσότητα κίνησης του τρέχοντος αποσπάσματος εκφρασμένη
%με τρόπο που να προσομοιάζει μετατόπιση,όχι ταχύτητα
motion_magnitude(videoCounter)=motionMagnitude/time;

%Τέλος, αυξάνουμε το δείκτη για το επόμενο απόσπασμα
videoCounter= videoCounter + 1;
```

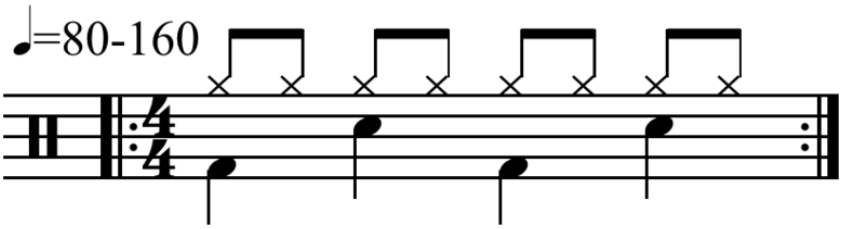
Ενοποιώντας λοιπόν τα παραπάνω βασικά αποσπάσματα κώδικα, και δημιουργώντας κατάλληλες δομές για τη συνολική επεξεργασία όλων των αποσπασμάτων, δεν μας μένουν παρά μόνο οι συνολικοί υπολογισμοί.

Θα φανεί παρακάτω στη συνολική παράθεση του κώδικα και προς το παρόν παραλείπεται αλλά δημιουργούμε πίνακες τόσο για την ποσότητα κίνησης όσο και για τον αριθμό των καρέ των αποσπασμάτων. Καθώς έχει αποφασιστεί εκ των προτέρων εφαρμογή σε τέσσερα αποσπάσματα, επιλέγουμε στατική δέσμευση των πινάκων αυτών ώστε να έχουν από τέσσερα στοιχεία.
Έπειτα, μπορούμε να υπολογίσουμε όλα τα μεγέθη που χρειάζονται για την εφαρμογή του ορισμού των Adams, Dorai και Venkatesh. Παρακάτω θεωρούμε ότι έχουν δημιουργηθεί οι πίνακες τεσσάρων στοιχείων `shotLength` με τους τέσσερις αριθμούς καρέ και `motion_magnitude` με τα τέσσερα motion magnitudes.

```
%Υπολογίζουμε τη διάμεσο του shot length για τα αποσπάσματα
lengthMedian=median(shotLength);

%Υπολογίζουμε την τυπική απόκλιση του shot length
%για όλα αποσπάσματα
lengthDeviation=std(shotLength);


%Υπολογίζουμε τη μέση τιμή του motion magnitude
motionMagnitudeMean=mean(motion_magnitude);

%Υπολογίζουμε την τυπική απόκλιση του motion magnitude
motionMagnitudeDeviation=std(motion_magnitude);



%Αρχικοποιούμε τις σταθερές μας
a=0.5;
b=0.5;



%Υπολογίζουμε τις τέσσερις τιμές του μεγέθους Pace
%σύμφωνα με τον ορισμό:

for i=1:4
 pace(i) = a*((lengthMedian-shotLength(i)) / lengthDeviation ) +
 b*((motion_magnitude(i)-motionMagnitudeMean)/motionMagnitudeDeviation)
end
```

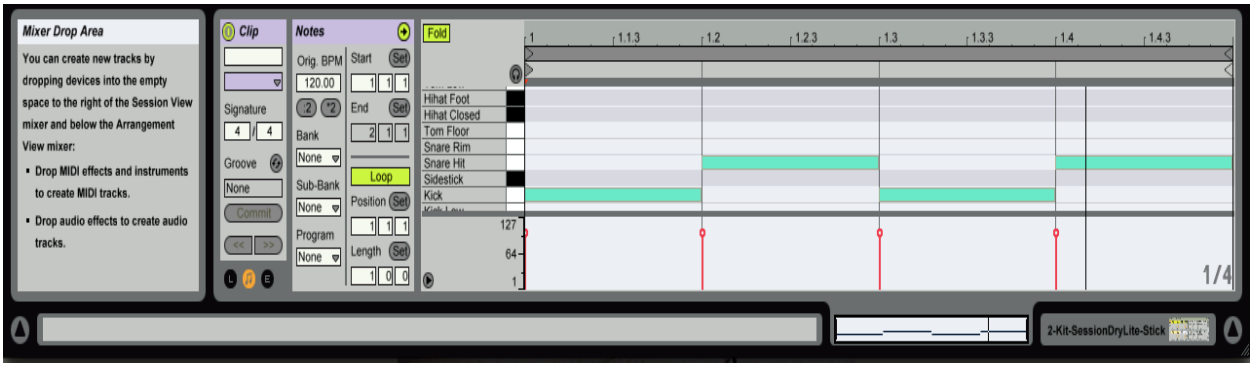
Πριν παραθέσουμε το συνολικό κώδικα της πρώτης και εισαγωγικής εφαρμογής, καλό είναι να επεξηγηθούν οι λόγοι για τους οποίους επιλέχθηκαν τα συγκεκριμένα τέσσερα αποσπάσματα για την εξαγωγή του μεγέθους Pace.

Όπως προαναφέρθηκε, ως βάση για τη συνολική εφαρμογή χρησιμοποιήθηκε το φιλμ μικρού μήκους *An Optical Poem* για λόγους σημασιολογίας. Σύμφωνα πάλι με την παραπάνω περιγραφή, η αναζήτηση "φράσεων" μέσα στο ηχητικό περιεχόμενο του φιλμ που να μπορούν να επαναληφθούν ώστε να δημιουργήσουν μια ρυθμική δομή, μας οδήγησε στην απομόνωση ορισμένων αποσπασμάτων διάρκειας κατά μέσο όρο τριών με τεσσάρων δευτερολέπτων. Αυτά συνοδεύτηκαν από έναν στοιχειωδώς απλό ρυθμό παιγμένο στα τύμπανα, ώστε να γίνει εμφανής η μετρική τους δομή και να ενισχυθεί ο παλμός τους. Από τα διάφορα αποσπάσματα που απομονώθηκαν, θα προτιμήσουμε εκείνα που παρουσιάζουν την ίδια, απλή δομή: Και στα τέσσερα μέρη όπου θα εφαρμόσουμε τον κώδικα που περιγράψαμε, τα τύμπανα παίζουν έναν ρυθμό τεσσάρων τετάρτων ($^4/_4$), που είναι από τους πλέον διαδεδομένους στη σύγχρονη μουσική παραγωγή. Παρακάτω φαίνεται η γραφή του:



Εικόνα 6.1: Ρυθμός 4/4

Η μπότα (bass drum) "παίζει" στο πρώτο και τρίτο τέταρτο και το ταμπούρο στο δεύτερο και το τέταρτο. Οι νότες που σημειώνονται με x είναι οι νότες στα πιατίνια που έχουν αξίες του ενός όγδοου. Στη δική μας περίπτωση, για λόγους απλότητας παραλείπουμε και αυτά τα όγδοα αφήνοντας μόνο τις νότες στην μπότα και το ταμπούρο. Σε μορφή κλιπ σε όργανο MIDI, ο ρυθμός φαίνεται σχηματικά κάπως έτσι:



Εικόνα 6.2: MIDI clip στο Ableton Live

Το κλιπ είναι χωρισμένο σε τέσσερα μέρη και η μπότα (Kick) παίζει μια νότα στο πρώτο και τρίτο μέρος ενώ το ταμπούρο (Snare Hit) παίζει στο δεύτερο και τέταρτο μέρος.

Έχοντας λοιπόν "ντύσει" τα διαφορετικά αποσπάσματα με τον ίδιο ουσιαστικά ρυθμό αλλά σε διαφορετικά tempo, είναι πολύ εύκολη μια σημασιολογική σύγκριση. Τα κομμάτια που παρουσιάζουν πιο έντονη εναλλαγή και κίνηση στο οπτικό περιεχόμενο συνοδεύτονται από πιο "ζωηρά" μέρη της Ούγγρικης Ραψωδίας του Liszt. Όταν αυτά τα μέρη επενδυθούν με τις κρουστές ρυθμικές δομές που περιγράψαμε, παρατηρούμε όμοιες διαφοροποιήσεις και μεταξύ αυτών των δομών. Το ερώτημα που προκύπτει και μένει πλέον να εξετάσουμε είναι κατά πόσο τα διαφορετικά αυτά αποσπάσματα θα παρουσιάζουν αντίστοιχες διαφοροποιήσεις και στις τιμές του μεγέθους *Pace* που θα υπολογίσουμε.

Πιο κάτω φαίνεται ο συνολικός κώδικας που χρησιμοποιήθηκε για την εφαρμογή αυτή σε MATLAB.

```
clear all; clc;
% θα χρειαστεί να επαναλάβουμε τέσσερις φορές τον ίδιο
% κώδικα λόγω των διαφορετικών ονομάτων των αποσπασμάτων
% Πρώτο απόσπασμα:
videoCounter=1;

videoReader=vision.VideoFileReader('segment1.mp4','ImageColorSpace',
                        'Intensity','VideoOutputDataType','uint8');
converter=vision.ImageDataTypeConverter;
opticalFlow=vision.OpticalFlow;
opticalFlow.OutputValue='Horizontal and vertical components in
                                                    complex form';
shapeInserter=vision.ShapeInserter('Shape','Lines','BorderColor',
                                'Custom','CustomBorderColor',255);
videoPlayer=vision.VideoPlayer('Name','optical poem looped');

NumberofFrames=0;

while ~isDone(videoReader)
    frame=step(videoReader);
    NumberofFrames=NumberofFrames+1;
    im=step(converter,frame);
    velocities=step(opticalFlow,im);
    lines=videooptflowlines(velocities,20);
    if ~isempty(lines)
        out = step(shapeInserter,im,lines)
        step(videoPlayer,out)
    end
end

%Κάθε φορά θα καταργούμε τα αντικείμενα ανάγνωσης και
%προβολής του βίντεο για αποφυγή σφαλμάτων με τα άλλα
%αποσπάσματα.
release(videoPlayer);
release(videoReader);
```

```
videoInfo = info(videoReader);
videoWidth = videoInfo.VideoSize(1);
videoHeight = videoInfo.VideoSize(2);

frameRate = videoInfo.VideoFrameRate;

motionMagnitude=0;
for i=1:videoWidth
    for j=1:videoHeight
        motionMagnitude=motionMagnitude + abs (velocities(j,i));
    end
end

NumberOfPixels=videoWidth*videoHeight;
motionMagnitude=motionMagnitude/NumberOfPixels;
time=NumberOfFrames/frameRate;
motion_magnitude(videoCounter)=motionMagnitude/time;
shotLength(videoCounter)=NumberofFrames;
videoCounter=videoCounter+1;

%Δεύτερο Απόσπασμα

videoReader=vision.VideoFileReader('segment2.mp4','ImageColorSpace',
                        'Intensity','VideoOutputDataType','uint8');
converter=vision.ImageDataTypeConverter;
opticalFlow=vision.OpticalFlow;
opticalFlow.OutputValue='Horizontal and vertical components in
                                                    complex form';
shapeInserter=vision.ShapeInserter('Shape','Lines','BorderColor',
                                'Custom', 'CustomBorderColor',255);
videoPlayer=vision.VideoPlayer('Name','optical poem looped');

NumberofFrames=0;


while ~isDone(videoReader)
    frame=step(videoReader);
    NumberofFrames=NumberofFrames+1;
    im=step(converter,frame);
    velocities=step(opticalFlow,im);
    lines=videooptflowlines(velocities,20);
    if ~isempty(lines)
        out = step(shapeInserter,im,lines)
        step(videoPlayer,out)
    end
end


release(videoPlayer);
release(videoReader);


videoInfo = info(videoReader);
videoWidth = videoInfo.VideoSize(1);
videoHeight = videoInfo.VideoSize(2);

frameRate = videoInfo.VideoFrameRate;
```

```
motionMagnitude=0;

for i=1:videoWidth
    for j=1:videoHeight
        motionMagnitude=motionMagnitude + abs (velocities(j,i));
    end
end

NumberOfPixels=videoWidth*videoHeight;
motionMagnitude=motionMagnitude/NumberOfPixels;

time=NumberOfFrames/frameRate;

motion_magnitude(videoCounter)=motionMagnitude/time;
shotLength(videoCounter)=NumberofFrames;
videoCounter=videoCounter+1;


%Τρίτο απόσπασμα


videoReader=vision.VideoFileReader('segment3.mp4','ImageColorSpace',
                        'Intensity','VideoOutputDataType','uint8');
converter=vision.ImageDataTypeConverter;
opticalFlow=vision.OpticalFlow;
opticalFlow.OutputValue='Horizontal and vertical components in
                                                    complex form';
shapeInserter=vision.ShapeInserter('Shape','Lines','BorderColor',
                                'Custom', 'CustomBorderColor',255);
videoPlayer=vision.VideoPlayer('Name','optical poem looped');

NumberofFrames=0;


while ~isDone(videoReader)
    frame=step(videoReader);
    NumberofFrames=NumberofFrames+1;
    im=step(converter,frame);
    velocities=step(opticalFlow,im);
    lines=videooptflowlines(velocities,20);
    if ~isempty(lines)
        out = step(shapeInserter,im,lines)
        step(videoPlayer,out)
    end
end


release(videoPlayer);
release(videoReader);


videoInfo = info(videoReader);
videoWidth = videoInfo.VideoSize(1);
videoHeight = videoInfo.VideoSize(2);

frameRate = videoInfo.VideoFrameRate;

motionMagnitude=0;
```

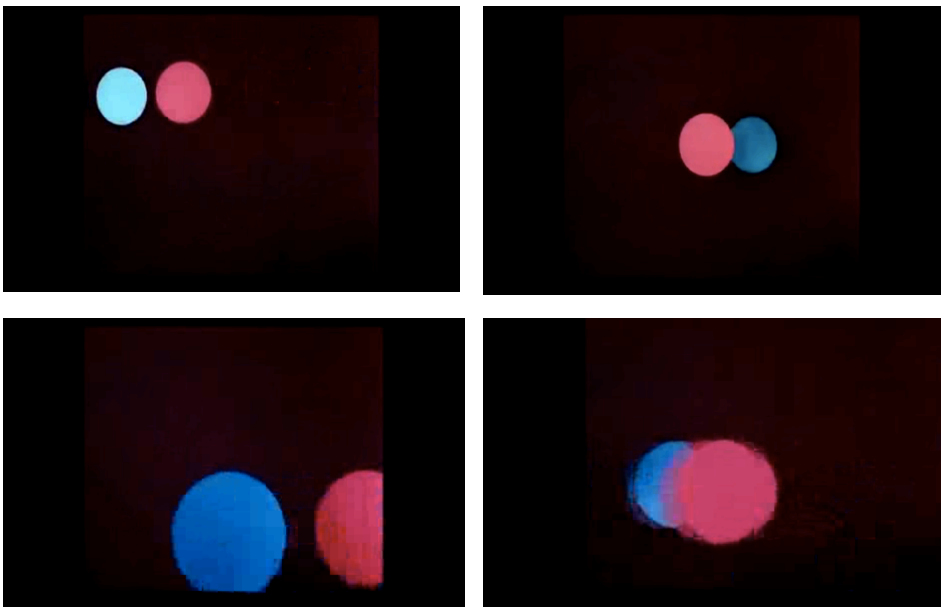
```
for i=1:videoWidth
    for j=1:videoHeight
        motionMagnitude=motionMagnitude + abs (velocities(j,i));
    end
end

NumberOfPixels=videoWidth*videoHeight;
motionMagnitude=motionMagnitude/NumberOfPixels;

time=NumberOfFrames/frameRate;

motion_magnitude(videoCounter)=motionMagnitude/time;
shotLength(videoCounter)=NumberofFrames;
videoCounter=videoCounter+1;



%Τέταρτο Απόσπασμα

videoReader=vision.VideoFileReader('segment4.mp4','ImageColorSpace',
                        'Intensity','VideoOutputDataType','uint8');
converter=vision.ImageDataTypeConverter;
opticalFlow=vision.OpticalFlow;
opticalFlow.OutputValue='Horizontal and vertical components in
                                                    complex form';
shapeInserter=vision.ShapeInserter('Shape','Lines','BorderColor',
                                'Custom','CustomBorderColor',255);
videoPlayer=vision.VideoPlayer('Name','optical poem looped');

NumberofFrames=0;


while ~isDone(videoReader)
    frame=step(videoReader);
    NumberofFrames=NumberofFrames+1;
    im=step(converter,frame);
    velocities=step(opticalFlow,im);
    lines=videooptflowlines(velocities,20);
    if ~isempty(lines)
        out = step(shapeInserter,im,lines)
        step(videoPlayer,out)
    end
end


release(videoPlayer);
release(videoReader);


videoInfo = info(videoReader);
videoWidth = videoInfo.VideoSize(1);
videoHeight = videoInfo.VideoSize(2);

frameRate = videoInfo.VideoFrameRate;

motionMagnitude=0;
```

```
for i=1:videoWidth
    for j=1:videoHeight
        motionMagnitude=motionMagnitude + abs (velocities(j,i));
    end
end

NumberOfPixels=videoWidth*videoHeight;
motionMagnitude=motionMagnitude/NumberOfPixels;

time=NumberOfFrames/frameRate;

motion_magnitude(videoCounter)=motionMagnitude/time;

shotLength(videoCounter)=NumberofFrames;




%Τελικό μέρος της επεξεργασίας

%Στατιστικά μεγέθη

lengthMedian=median(shotLength);

lengthDeviation=std(shotLength);

motionMagnitudeMean=mean(motion_magnitude);

motionMagnitudeDeviation=std(motion_magnitude);

%Υπολογισμός του ζητούμενου πίνακα Pace
for i=1:4
    pace(i)= a * ( (lengthMedian-shotLength(i)) / lengthDeviation )+
      b* ( (motion_magnitude(i)-
motionMagnitudeMean)/motionMagnitudeDeviation)
end;
```

6.2.3 ΔΕΥΤΕΡΗ ΕΦΑΡΜΟΓΗ: ΑΥΤΟΜΑΤΗ ΕΞΑΓΩΓΗ ΡΥΘΜΙΚΩΝ ΑΠΟΣΠΑΣΜΑΤΩΝ

Απομακρύνοντας τη μουσική επένδυση του *An Optical Poem* και παρατηρώντας τα αποσπάσματα που απομονώθηκαν για την παραπάνω εφαρμογή μαζί με τα κρουστά μέρη που συντέθηκαν, γίνεται άμεσα αντιληπτό ότι σε πολλές περιπτώσεις το παίξιμο των τυμπάνων δε συγχρονίζεται ακριβώς με αυτό που το μάτι μας αντιλαμβάνεται ως εξέλιξη του οπτικού περιεχομένου. Για να υπάρξει όντως μια καλά ορισμένη σύνδεση του βίντεο με μια ρυθμική μουσική συνοδεία, θα πρέπει να επινοήσουμε δικά μας κριτήρια εξαγωγής των ρυθμικών παλμών και όχι να στηριχτούμε στο μουσικό κομμάτι που για καλλιτεχνικούς λόγους επιλέχτηκε από τον σκηνοθέτη ως κατάλληλο για συνοδεία των απεικονιζόμενων μορφών.Πιο συγκεκριμένα, οφείλουμε να συνδέσουμε την εξέλιξη των γεωμετρικών μορφών με ένα ρυθμό, ο οποίος να παρουσιάζει τα ασθενή και ισχυρά στοιχεία του (τις άρσεις και τις θέσεις του) στα χρονικά σημεία που ένας θεατής του βίντεο αντιλαμβάνεται ότι υπάρχει σημαντική αλλαγή σε αυτό που παρακολουθεί. Όσο υποκειμενική κι αν φαντάζει εκ πρώτης όψης αυτή η αντίληψη των αλλαγών και της σημαντικότητας τους, στις περισσότερες των περιπτώσεων είναι αρκετά κοινή για πολλούς θεατές-παρατηρητές.

Ένα συγκεκριμένο απόσμασμα μέσα από το φιλμ, ενίσχυσε αυτή την ιδέα και υπήρξε το πρώτο ερέθισμα για τη διαμόρφωση του αλγορίθμου εξαγωγής του ρυθμού. Το συγκεκριμένο απόσμασμα εντοπίζεται στο πέμπτο περίπου λεπτό του φιλμ και διαρκεί περίπου πέντε με έξι δευτερόλεπτα. Απεικονίζει ουσιαστικά δύο κύκλους να περιστρέφονται με σταθερή ταχύτητα ο ένας ως προς τον άλλον και ανά ένα περίπου δευτερόλεπτο να αλλάζουν ταυτόχρονα και στιγμιαία τη θέση τους στην οθόνη. Παρακάτω φαίνονται μερικά διαδοχικά καρέ όπου διακρίνονται οι διαφορετικές φάσεις των περιστροφών και οι μετατοπίσεις τους ως ζεύγος αντικειμένων.



Εικόνα 6.3: Τέσσερα στιγμιότυπα του αποσπάσματος

Εάν το απόσπασμα αυτό λοιπόν τοποθετηθεί εντός ενός βρόχου ώστε να επαναλαμβάνεται για ορισμένες φορές, παρατηρούμε ότι αντιλαμβανόμαστε πιο έντονα τις απότομες αλλαγές της θέσης του ζεύγους επί της οθόνης παρά τη συνεχόμενη περιστροφή του καθενός. Για το λόγο αυτό, επιχειρώντας να "παίξουμε" ένα ρυθμό που να συνοδεύει την ακολουθία αυτή από καρέ, τονίζουμε ως παλμούς τις στιγμές εκείνες που αναμένουμε να υπάρξει μια μετατόπιση του ζεύγους σε μια τυχαία θέση επί του μαύρου φόντου.Μένει λοιπόν να διαμορφωθεί ένας μηχανισμός που να διακρίνει όλες εκείνες τις στιγμές που εμφανίζουν πιο έντονες μεταβολές κατά τη διάρκεια ενός αποσπάσματος με κοινό περίπου περιεχόμενο: Κοινό φόντο, κοινή νοηματικά κίνηση και εξέλιξη. Σημειωτέον ότι ο εντοπισμός των πιο "έντονων" χρονικών στιγμών γίνεται πάντα συγκριτικά με το σύνολο των υπόλοιπων στιγμών ενός αποσπάσματος και θα ήταν κενό περιεχομένου να αναπτυχθεί ένα κριτήριο καθολικό για κάθε είδους βίντεο.

Κατανοούμε λοιπόν πως ουσιαστικά αναζητούμε ένα κριτήριο για να εκτιμούμε το μέγεθος της "αλλαγής" στο οπτικό περιεχόμενο του βίντεο διαμέσω του χρόνου. Τέτοια κριτήρια έχουν κατά καιρούς διαμορφωθεί με διάφορους τρόπους. Οι Guedes και Branco για παράδειγμα που αναφέρθηκαν σε προηγούμενο κεφάλαιο, ουσιαστικά χρησιμοποιούν τη μέτρηση της διαφοράς της φωτεινότητας των pixels ενός βίντεο στην εξέλιξη του χρόνου ως ένα τέτοιο κριτήριο. Ένα άλλο κριτήριο που είναι πολύ διαδεδομένο για τη σύγκριση δύο εικόνων και επομένως θα μπορούσε να χρησιμοποιηθεί για διαδοχικά καρέ είναι η *Ευκλίδεια Απόσταση* συγκεκριμένων σημείων. Υπενθυμίζεται ότι ως ευκλίδεια απόσταση δύο σημείων ορίζουμε ουσιαστικά το μέτρο του διανύσματος που ορίζεται στο δισδιάστατο χώρο από αυτά τα δύο σημεία. Αν $A(x_A, y_A)$ είναι το ένα σημείο λοιπόν και $B(x_B, y_B)$ το άλλο, η ευκλίδεια απόσταση τους είναι

$$d_{AB} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

Για την ομοιότητα δύο εικόνων λοιπόν, συχνά αναφερόμαστε σε γειτονιές εντός μίας εικόνας και στην ευκλίδεια απόσταση που έχουν από μία δεύτερη εικόνα κάποια σημεία ενδιαφέροντος (feature points) εντός μιας γειτονιάς.

Αντ' αυτών εμείς θα χρησιμοποιήσουμε κάτι σαφώς πιο στοιχειώδες
Ως γνωστόν, ένας πολύ καλός δείκτης της ομοιότητας δύο σημάτων αποτελεί η *ετεροσυχέτιση* τους. Υπενθυμίζεται ότι για δύο συνεχή σήματα f, g , η ετεροσυσχέτιση ορίζεται ως εξής:

$$(f \star g)(\tau) \triangleq \int_{-\infty}^{+\infty} f^*(t)\, g(t+\tau)\, dt$$

όπου f^* η συζυγής συνάρτηση της f και τ η χρονική υστέρηση των δύο σημάτων [41] .

Αντίστοιχα, για διακριτά σήματα ορίζουμε:

$$(f \star g)[n] \triangleq \sum_{m=-\infty}^{+\infty} f^*[m]\, g[m+n]$$

Η ετεροσυσχέτιση συνήθως χρησιμοποιείται για σήματα τα οποία παρουσιάζουν μια χρονική μετατόπιση το ένα ως προς το άλλο. Αν για παράδειγμα οι f, g διαφέρουν κατά μία άγνωση μετατόπιση στον άξονα x, η ετεροσυσχέτιση βρίσκει πόσο πρέπει να μετακινηθεί το σήμα g ώστε να συμπίπτει με το σήμα f. Από τον ορισμό, ουσιαστικά η συνάρτηση g "σύρεται" επί του άξονα x, υπολογίζοντας το ολοκλήρωμα του γινομένου με την f σε κάθε θέση. Όταν οι δύο συναρτήσεις ταυτιστούν, τότε η ετεροσυσχέτιση μεγιστοποιείται, αφού οι θετικές κορυφές των σημάτων έχουν τη μέγιστη συνεισφορά όπως επίσης και οι αρνητικές καθώς το γινόμενο τους προκύπτει θετικό.

Συχνά, χρησιμοποιείται η κανονικοποιημένη μορφή τους μεγέθους αυτού. Αυτή, τυπικά προκύπτει αφαιρώντας σε κάθε βήμα τη μέση τιμή και αφαιρώντας την τυπική απόκλιση. Έτσι, η κανονικοποιημένη ετεροσυσχέτιση μιας εικόνας $t(x,y)$ με μία άλλη $f(x,y)$ υπολογίζεται ως :

$$\frac{1}{n}\sum_{x,y}\frac{(f(x,y)-\bar{f})(t(x,y)-\bar{t})}{\sigma_f \sigma_t}$$

όπου n είναι το σύνολο των pixels στις $t(x,y)$ και $f(x,y)$ [42].

Αξίζει να σημειώσουμε πως αν θεωρήσουμε δύο κανονικοποιημένα μεγέθη

$$F(x,y) = f(x,y) - \bar{f} \quad \text{και} \quad T(x,y) = t(x,y) - \bar{t}\ ,$$

τότε το παραπάνω είναι στην πράξη το μέγεθος

$$\langle \frac{F}{\|F\|}, \frac{T}{\|T\|} \rangle$$

όπου με $\langle .. \rangle$ συμβολίζουμε το εσωτερικό γινόμενο και με $\|..\|$ την L^2 νόρμα του εκάστοτε μεγέθους.

Στον τομέα της επεξεργασίας εικόνας, ο υπολογισμός της κανονικοποιημένης ετεροσυσχέτισης συχνά χρησιμοποιείται σε εφαρμογές ανίχνευσης κίνησης (motion tracking): Η υποπεριοχή μιας αρχικής εικόνας της οποίας η κίνηση πρέπει να ανιχνευτεί, μετατοπίζεται σε διαφορετικά μέρη των επόμενων εικόνων-καρέ. Στα μέρη όπου όντως έχει υπάρξει μετατόπιση της ζητούμενης υποπεριοχής, η ετεροσυσχέτιση θα μεγιστοποιείται και έτσι είναι αρκετά απλή η παρακολούθηση της κίνησης.

Μια ιδιότητα μάλιστα της κανονικοποιημένης μορφής που ευνοεί τη χρήση της, είναι ότι οι τιμές της είναι φραγμένες στο διάστημα $[-1.0\ , 1.0]$, κάνοντας πολύ απλές τις συγκρίσεις μεταξύ αποτελεσμάτων και το χαρακτηρισμό διαφορετικών εφαρμογών.

Στο περιβάλλον MATLAB υπάρχει ενσωματωμένη η συνάρτηση `normxcorr2(A,B)` που υπολογίζει την κανονικοποιημένη μορφή της ετεροσυσχέτισης των δισδιάστατων πινάκων A και B [43]. Αφού μάλιστα γνωρίζουμε ότι οι εικόνες αναπαρίστανται ως δισδιάστατοι πίνακες στο περιβάλλον αυτό, η ευκολία της χρήσης της συνάρτησης είναι εμφανής.
Η συνάρτηση λειτουργεί ως εξής:

- Αρχικά, ανάλογα με το μέγεθος των εικόνων, υπολογίζει την ετεροσυσχέτιση τους είτε στο πεδίο του χώρου είτε των συχνοτήτων,

- υπολογίζει τοπικά αθροίσματα, έχοντας προϋπολογίσει ορισμένα τρέχοντα αθροίσματα
  (κατά τη διάρκεια ανάγνωσης δηλαδή των πινάκων A,B) και

- χρησιμοποιεί τα τοπικά αυτά αθροίσματα για να κανονικοποιήσει την ετεροσυσχέτιση και να εξάγει τους συντελεστές της συσχέτισης.

- Η τελική υλοποίηση ακολουθεί τη σχέση

$$\gamma(u,v) = \frac{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}]\,[t(x-u,y-v) - \bar{t}]}{\left\{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}]^{\,2} \sum_{x,y}[t(x,y) - \bar{t}_{u,v}]^{\,2}\right\}^{1/2}}$$

  όπου με f αναφέρεται η εικόνα B , με t η εικόνα A που ονομάζεται και πρότυπο και με $\bar{f}_{u,v}$ η μέση τιμή της f στην περιοχή κάτω από το πρότυπο.

Το τελικό αποτέλεσμα μιας κλήσης της μορφής `c = normxcorr2(A,B);` επιστρέφει στη μεταβλητή c ένα δισδιάστατο πίνακα με τιμές τύπου double που είναι οι συντελεστές της συσχέτισης και οι τιμές τους μεταβάλλονται από -1.0 έως 1.0.

Οι περιοχές όπου παρατηρούνται τιμές να πλησιάζουν το 1.0 πειρέχουν ομοιότητες των δύο εικόνων ενώ προφανώς, όσο μειώνονται οι τιμές των συντελεστών τόσο οι δύο εικόνες θα διαφέρουν ως προς το περιεχόμενο.

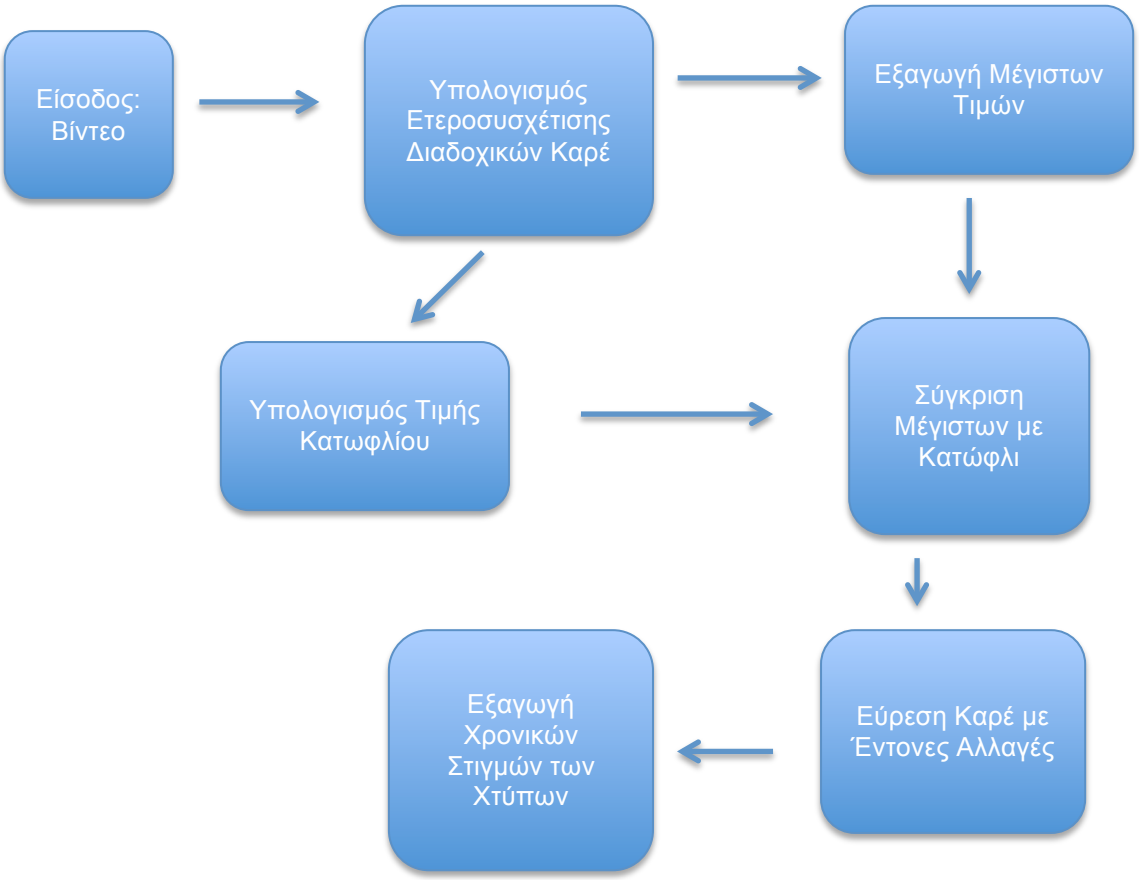
Έχοντας κατανοήσει το υπόβαθρο και τη χρησιμότητα της ετεροσυσχέτισης λοιπόν, μπορούμε πλέον να εξετάσουμε τον αλγόριθμο που διαμορφώθηκε για την εξαγωγή ενός ρυθμικού αποσπάσματος με βάση τα βίντεο μας.

Ο αλγόριθμος που ακολουθήθηκε στηρίζεται στον υπολογισμό της κανονικοποιημένης ετεροσυσχέτισης για κάθε ζεύγος διαδοχικών καρέ του εκάστοτε αποσπάσματος βίντεο και της σύγκρισης των αποτελεσμάτων. Παρατηρώντας τις τιμές που προκύπτουν, βλέπουμε πως ανάμεσα σε καρέ τα οποία δεν παρουσιάζουν μεγάλες αλλαγές, το αποτέλεσμα μιας κλήσης της `normxcorr2` όπως προαναφέρθηκε, επιστρέφει πληθώρα τιμών κοντά στο +1.0, δηλαδή από 0.95 έως και 0.99. Αντίθετα, ανάμεσα σε καρέ όπου εκ των προτέρων έχουμε παρατηρήσει έντονες αλλαγές στο περιεχόμενο, η συνάρτηση δεν φτάνει σε τόσο μεγάλες τιμές.

Παίρνοντας ως παράδειγμα το παραπάνω απόσπασμα, υπολογίζουμε την `normxcorr2` ανάμεσα σε δύο καρέ που περιέχουν περιστροφή των δύο κύκλων στην ίδια περιοχή του φόντου. Εκεί, έχουμε μια μεταβολή των αποτελεσμάτων από αρνητικές τιμές έως και πάνω από 0.9 . Αντίθετα, εάν υπολογίσουμε τη συνάρτηση ανάμεσα σε δύο καρέ όπου υπάρχει μετατόπιση του ζεύγους των κύκλων, οι μέγιστες τιμές που εξάγονται φτάνουν έως το 0.6 με 0.7 . Ως πρώτο βήμα λοιπόν, φαίνεται πως μια σύγκριση των μέγιστων τιμών που λαμβάνει η ετεροσυσχέτιση είναι ένας αρκετά καλός δείκτης των χρονικών σημείων που θα παρατηρηθούν έντοντες αλλαγές στο οπτικό περιεχόμενο.

Ωστόσο, για την πρώτη αυτή δοκιμαστική σύγκριση είχαμε την ευχέρεια να γνωρίζουμε το σύνολο των τιμών και αφού τις εξετάσουμε να αποφασίσουμε ποιες θα είναι αυτές που θα εξασφαλίζουν το επιθυμητό κριτήριο. Για μια πιο γενική εφαρμογή, ποια θα έπρεπε να είναι η τιμή του μεγίστου της `normxcorr2` ώστε να διαπιστώνεται ότι όντως συντελείται μια σημαντική μεταβολή στα όσα βλέπουμε; Η απάντηση θα πρέπει να είναι μια τιμή κατωφλίου η οποία να προκύπτει κάθε φορά, για κάθε απόσπασμα μέσα από μία παρόμοια διαδικασία επεξεργασίας. Οι τιμές που ικανοποιούν αυτή την κατάλληλα διαμορφωμένη συνθήκη θα μας υποδεικνύουν τα καρέ εκείνα που θα πρέπει συνοδεύονται από τους χτύπους του ρυθμικού μέρους που θα "ντύνει" το βίντεο.

Αν δούμε λοιπόν σχηματικά τη διαδικασία αυτή θα έχουμε:



Πώς όμως θα πρέπει να προκύψει η κατάλληλη τιμή κατωφλίου που να δίνει αποδεκτά αποτελέσματα; Για να βρούμε μια ικανοποιητική απάντηση θα "τρέξουμε" τον παρακάτω κώδικα για διάφορα αποσπάσματα στα οποία έχουμε διακρίνει ήδη ορισμένες αλλαγές τις οποίες διασθητικά ένας παίχτης κρουστών οργάνων θα έδινε έμφαση στο παίξιμο του.

```
%Αρχικοποιούμε τα αντικείμενα εισόδου όπως και στην πρώτη
%εφαρμογή.Τα διάφορα αποσπάσματα ονομάζονται
%chop_1,chop_2 και ούτω καθεξής
videoReader=vision.VideoFileReader('chop_1.mp4','ImageColorSpace',

'Intensity','VideoOutputDataType','uint8');

converter=vision.ImageDataTypeConverter;
%Όσο υπάρχουν καρέ προς επεξεργασία, συνεχίζουμε το βρόχο
while ~isDone(videoReader)

%Κρατάμε ένα καρέ κάνοντας τη συνέλιξη του βίντεο με τη
%βηματική συνάρτηση
    frame=step(videoReader);

%Αυξάνουμε τον έως τώρα υπολογισθέντα αριθμό καρέ
     numberOfFrames=numberOfFrames+1;

%Για κάθε ζεύγος, το τρέχον καρέ σώζεται ως frame2 και το
%προηγούμενο προϋπάρχει ως frame1. Υπολογίζουμε την
%κανονικοποιημένη ετεροσυσχέτιση του ζεύγους και η μέγιστη
%τιμή σώζεται σε έναν πίνακα max_array.
     if NumberofFrames == 1
         frame1=frame;
    else
         frame2=frame;
         k = findMax(normxcorr2(frame1,frame2));
         max_array(counter)=k;
         counter=counter+1;
         frame1=frame;
    end
end

release(videoPlayer);
release(videoReader);
```

Η συνάρτηση `findMax` είναι μια στοιχειώδης συνάρτηση εύρεσης της μέγιστης τιμής ενός δισδιάστατου πίνακα A:

```
function Max = findMax(A)

Max = A(1,1);

k=size(A);
ii=k(1);
jj=k(2);

for i=1:ii
    for j=1:jj
        if A(i,j)> Max
            Max = A(i,j);
        end
    end
end

end
```

Δοκιμάζουμε λοιπόν τον κώδικα αυτόν στα διάφορα αποσπάσματα βίντεο και εξετάζουμε τους διαφορετικούς πίνακες `max_array` που προκύπτουν. Πρώτα δοκιμάσαμε το απόσπασμα που περιγράψαμε προηγουμένως και για αυτό, ο πίνακας παίρνει τιμές από **0.4625** έως **0.9993**. Επιλέγοντας μάλιστα το βίντεο να φαίνεται σε ένα αντικείμενο `videoPlayer` όπως αυτό που χρησιμοποιήθηκε στην πρώτη εφαρμογή, εντοπίζουμε πως οι πιο έντονες αλλαγές παρατηρούνται στα καρέ με αύξοντα αριθμό **13, 23, 32** και **44**. Εξετάζοντας τα μέγιστα της ετεροσυσχέτισης στα συγκεκριμένα καρέ (ή καλύτερα στις μεταβάσεις προς αυτά) βρίσκουμε τις τιμές **0.7578, 0.5318, 07236** και **0.7960**. Η μέση τιμή των μεγίστων είναι **0.9386** άρα είναι προφανές το πόσο μεγάλη είναι η απόκλιση των καρέ αυτών από το μέσο όρο.

Για το επόμενο απόσπασμα χρησιμοποιήθηκε ως κατώφλι ένα αυθαίρετο άνω όριο των παραπάνω τιμών ώστε να πειραματιστούμε για το αν αυτό θα "δούλευε" και για διαφορετικό οπτικό περιεχόμενο. Το όριο που επελέγη ήταν **0.8** και οι στιγμές που διακρίνουμε διασθητικά είναι με μια πρώτη εξέταση τα καρέ **6, 19 ,42, 61, 78, 96** και **128** . Σε αυτά τα στιγμιότυπα, ο πίνακας των μέγιστων έχει τις τιμές **0.5669, 0.8847, 0.6321, 0.8987, 0.7814, 0.6166** και **0.5914** . Άρα λοιπόν, και πάλι σε όλες τις αλλαγές με την εξαίρεση μόνο μίας, τα μέγιστα της ετεροσυσχέτισης είναι κάτω από το αυθαίρετα επιλεγμένο κατώφλι.

Κάνουμε άλλες δύο δοκιμές για αυτό το πρώτο στάδιο: Στο τρίτο κατά σειρά δοκιμαστικό απόσπασμα, οι αλλαγές που διακρίνουμε με το μάτι, παρατηρούνται στα καρέ **1, 31** και **39** και εκεί ο `max_array` έχει τιμές **0.91, 0.6744** και **0.6663** . Στο τελευταίο απόσπασμα διακρίνουμε τέσσερις αλλαγές στα καρέ **2, 36, 42** και **47** όπου υπολογίζονται οι τιμές **0.7226, 0.6409, 0.8294** και **0.8046** .

Άρα λοιπόν και πάλι το άνω όριο που επιλέχτηκε από το πρώτο απόσπασμα φαίνεται να "πιάνει" την πλειοψηφία των μεταβολών. Η αυθαίρετη επίλογη του ωστόσο δεν είναι θεμιτή τόσο λόγω της αδυναμίας εξήγησης της όσο και λόγω του ότι δεν υπολογίζουμε λιγοστές αλλά σημαντικές αλλαγές στο περιεχόμενο. Εξετάζοντας τα δεδομένα, φαίνεται πως η τελευταία αδυναμία οφείλεται στις σημαντικές διαφορές των αποσπασμάτων ως προς το περιεχόμενο, κυρίως δηλαδή τα χρώματα και τις γεωμετρικές μορφές. Για αυτό, το κατώφλι θα πρέπει να διαμορφώνεται κάθε φορά μέσα από το ίδιο το απόσπασμα και όχι καθολικά για κάθε βίντεο.

Παρατηρούμε πως όλα τα μέγιστα της αυτοσυσχέτισης στα καρέ αλλαγών είναι σαφώς μικρότερα από το μέσο όρο του εκάστοτε πίνακα `max_array`. Στο πρώτο απόσπασμα η μέση τιμή είδαμε ότι είναι **0.9386**, στο δεύτερο υπολογίζεται στο **0.9665**, στο τρίτο **0.9742** και στο τελευταίο **0.9082** . Πόσο "κάτω" όμως από το μέσο όρο πρέπει να αναζητήσουμε το κατάλληλο κατώφλι; Η σκέψη αυτή μας οδηγεί προφανώς στην έννοια της απόκλισης. Η τυπική απόκλιση των πινάκων στα τέσσερα βίντεο είναι **0.0676 ,0.0686, 0.0585** και **0.0738** . Μια τιμή κατωφλίου λοιπόν που εξάγεται ως
μέση τιμή - τυπική απόκλιση θα δώσει αντίστοιχα τιμές: **0.8709, 0.8979, 0.9157** και **0.8344** που όντως λειτουργούν ως εκάστοτε άνω όρια για τις ετεροσυσχετίσεις στα καρέ που εκ των προτέρων είχαμε διακρίνει.

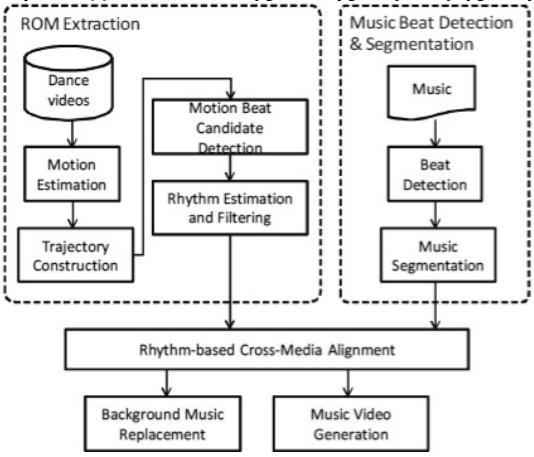
Ακολουθώντας τη διαμόρφωση στατιστικών κατανομών, δοκιμάστηκε και ως κατώφλι μία τιμή $\boldsymbol{\mu\acute{\varepsilon}\sigma\eta\ \tau\iota\mu\acute{\eta}\ -\ 2\ \times\ \tau\upsilon\pi\iota\kappa\acute{\eta}\ \alpha\pi\acute{o}\kappa\lambda\iota\sigma\eta}$ δίνοντας : **0.8033, 0.8293, 0.8572** και **0.7606** . Ενώ λοιπόν εξάγεται ικανοποιητική τιμή για το πρώτο απόσπασμα, στα υπόλοιπα κόβονται αρκετές αλλαγές που θα πρέπει να τονιστούν ως ρυθμικά στοιχεία.

Επομένως λοιπόν η διαμόρφωση της τιμής του κατωφλίου ως η διαφορά της μέσης τιμής και της τυπικής απόκλισης των μέγιστων τιμών της υπολογισθείσας ετεροσυσχέτισης φαίνεται να δίνει ικανοποιητικά αποτελέσματα. Πρέπει λοιπόν να την εντάξουμε στον κώδικα μας ώστε να εξάγεται αυτόματα σε κάθε απόσπασμα και να βρίσκονται όλα τα καρέ στα οποία όντως πρέπει να δώσουμε έμφαση. Το μέρος του κώδικα που αντιστοιχεί στη διαδικασία που περιγράψαμε φαίνεται παρακάτω:

```
%Βρίσκουμε μέσο όρο και απόκλιση του max_array
%και υπολογίζουμε το κατώφλι
megethos=size(max_array);
changeCounter=1;
mesos_oros=mean(max_array);
apoklisi=std(max_array);
katwfli=mesos_oros-apoklisi;

%Ελέγχουμε όλα τα μέγιστα και εάν είναι μικρότερα της τιμής
%κατωφλίου,τοποθετούμε το frame τους σε έναν πίνακα
%changes_array
for i=1:megethos
    if max_array(i)<katwfli
        changes_array(changeCounter)=i;
        changeCounter=changeCounter+1;
    end
end

%Υπολογίζουμε το μέγεθος του πίνακα, δηλαδή το πλήθος των
%μεταβολών που διακρίναμε
megethos2=size(changes_array);
megethos_allagwn=megethos2.(2);

videoInfo=info(videoReader);
frameRate=videoInfo.VideoFrameRate;

%Διαιρούμε τον πίνακα των καρέ αλλαγών με το frame rate
%για να βρούμε τις χρονικές στιγμές τους
time_array=changes_array/frameRate;
```

Πως όμως μπορούμε να τεκμηριώσουμε ακόμη πιο ισχυρά την επινόηση της ερευνητικής διαδικασίας; Στην ουσία μπορούμε να πούμε πως η διαδικασία που ακολουθείται είναι μια μορφή Beat Tracking η οποία αφορά ένα οπτικό περιεχόμενο: Γνωρίζοντας το σύνολο της μορφής, αναζητούμε τα στιγμιότυπα εκείνα τα οποία οι αισθήσεις μας επισημαίνουν ως πιο σημαντικά. Απώτερος στόχος είναι προφανώς να δημιουργήσουμε μια μουσική οντότητα, μια ακολουθία ισχυρών και ασθενών στοιχείων που να συνοδεύει και να ενισχύει τα όσα μας επέδειξε η αισθητική μας αντίληψη.

Έτσι λοιπόν, η ιδέα που ακολουθούμε μοιάζει σημαντικά με τη διαδικασία των Chu και Tsai όπως περιγράφηκε και παραπάνω [34] . Χρησιμοποιώντας ως οπτικό περιεχόμενο βίντεο χορευτών, αντίστοιχα με εμάς αναζητούν τις "θέσεις" της κίνησης, τα ισχυρά δηλαδή στιγμιότυπα της όλης εξέλιξης της μορφής. Όπως φαίνεται και στην εικόνα 6.4 η οποία είναι παρμένη από το έργο των Chu και Tsai, θεμελιώδες στάδιο της μεθόδου τους είναι η ανίχνευση των beats, τόσο της κίνησης όσο και του ήχου που θα χρησιμοποιήσουν στην τελική τους εφαρμογή. Η πρακτική που επιλέγουν όμως για το "Motion Beat Candidate Detection", καθορίζεται εν μέρει και από το πεδίο της εφαρμογής τους: Η κίνηση ενός χορευτή παρουσιάζει στοιχεία που μπορούμε σε κάθε περίπτωση να αναμένουμε (στροφές, παύσεις και λοιπά) ενώ είμαστε σίγουροι πως το συνολικό περιεχόμενο του βίντεο δεν πρόκειται να αλλάξει ριζικά (για παράδειγμα, δεν πρόκειται να υπάρξει έντονη αλλαγή στό φόντο ή στην περιοχή της οθόνης στην οποία παρατηρείται κίνηση) .



Εικόνα 6.4 : Σχηματικά η διαδικασία των Chu,Tsai

Συμπερασματικά, δε θα ήταν δόκιμο να ακολουθήσουμε κατά γράμμα την ιδέα των Chu et al. Αντίθετα, η επιστράτευση της κανονικοποιημένης ετεροσυσχέτισης, επιτρέπει την ανίχνευση των ζητούμενων στιγμών σε ένα μεγαλύτερο εύρος εφαρμογών ενώ παρά τους όποιους περιορισμούς μπορεί να παρουσιάζει, είναι ιδαίτερα εύχρηστη με χρήση του MATLAB.

Ένας βασικός τέτοιος περιορισμός είναι δυστυχώς η αδυναμία εφαρμογής σε πραγματικό χρόνο (real time). Αυτό συμβαίνει καθώς για να ανιχνεύσουμε τις υποψήφιες θέσεις χτυπημάτων σε ένα κρουστό όργανο, πρέπει να έχουμε "διαβάσει" το σύνολο του οπτικού περιεχομένου του βίντεο ώστε να είμαστε σε θέση να εξάγουμε τους πίνακες `max_array` και `changes_array` μέσω του υπολογισμού των κατάλληλων τιμών κατωφλίου. Σε βελτιώσεις και επεκτάσεις ωστόσο της διαδικασίας θα αναφερθούμε σε παρακάτω κεφάλαιο.

Έτσι λοιπόν, απομένει να δημιουργήσουμε ένα αρχείο MIDI το οποίο να αντιστοιχίζει νότες στις χρονικές στιγμές που εισάγαμε στον `time_array`. Για τη διαχείριση του πρωτοκόλλου MIDI μέσα από το MATLAB, θα χρησιμοποιήσουμε κάποια έτοιμα κομμάτια κώδικα του Dr. Ken Schutte [44]. Πρακτικά, η μορφή της εξόδου θα είναι ένα απλοποιημένο MIDI κλιπ,αφού δεν παρέχουμε πληροφορίες για παραμέτρους όπως π.χ. το velocity, και στηριζόμαστε μόνο στους χρόνους ενεργοποίησης και τερματισμού μιας νότας και στη συχνότητα της, δηλαδή τη θέση της σε μια οκτάβα. Ουσιαστικά, αντιστοιχούμε σε κάθε μεταβολή που έχουμε βρει από την προηγούμενη διαδικασία, τη χρονική ενεργοποίηση μιας νότας η οποία τερματίζεται μετά από ένα σταθερό χρονικό διάστημα. Το διάστημα αυτό, προκύπτει ως η ελάχιστη χρονική διάρκεια που υπάρχει ανάμεσα σε δύο διαδοχικά συμβάντα του πίνακα `time_array`. Το πρώτο χρονικά τέτοιο συμβάν αντιστοιχίζεται στο μεσαίο Ντο και τα επόμενα διαδοχικά στις αμέσως ψηλότερες νότες. Η επιλογή των νοτών δεν έχει ουσιαστική σημασία αφού το όργανο το οποίο θα κληθεί να "ερμηνεύσει" το αρχείο MIDI, θα αποτελείται κατ' επιλογήν μας από νότες σε κρουστά όργανα που ανά δύο ή ανά τρεις θα επαναλαμβάνονται, π.χ. μπότα-ταμπούρο-μπότα-ταμπούρο κ.ο.κ ή μπότα-ταμπούρο-πιατίνι- μπότα-ταμπούρο-πιατίνι κ.ο.κ . Παρακάτω φαίνεται το μέρος του κώδικα που υλοποιεί τη δημιουργία του αρχείου. Οι συναρτήσεις `matrix2midi` και `writemidi` είναι δημιουργήματα του Dr Schutte και εξηγούνται στο παράρτημα.

```
%Συνολικός αριθμός νοτών
N = megethos_allagwn;
%Η θέση της τελευταίας νότας, έχοντας ξεκινήσει από το μεσαίο Ντο
final_note = 60 + N -1;
%Αρχικοποίηση του πίνακα που μετατραπεί σε .mid
M = zeros(N,6);

%Όλες οι νότες στο track 1
M(:,1) = 1 ;

%Όλες οι νότες στο κανάλι 1
M(:,2) = 1;

%Η θέση κάθε νότας στην οκτάβα της
M(:,3) = (60:final_note)';

%Οι εντάσεις αυξάνονται ομοιόμορφα και ανεπαίσθητα από 80 έως 120
M(:,4) = round(linspace(80,120,N));

%Οι χρόνοι ενεργοποίησης
M(:,5) = (time_array)';

%και απενεργοποίησης κάθε νότας
M(:,6) = M(:,5) + onset;

midi_new = matrix2midi(M);

%Δημιουργία της εξόδου
writemidi(midi_new,'Rhythm_Track.mid');
```

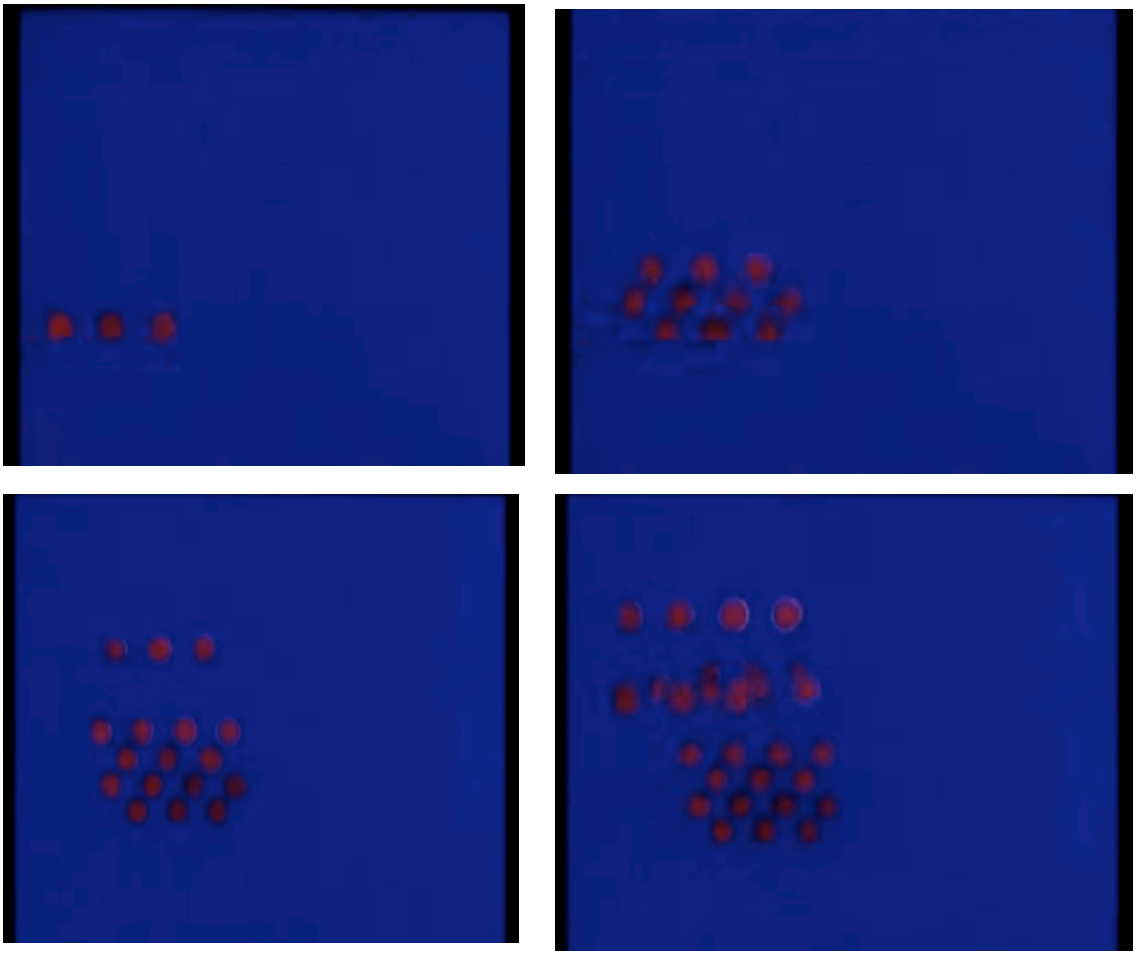
7. ΚΕΦΑΛΑΙΟ 7:
ΑΠΟΤΕΛΕΣΜΑΤΑ - ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΕΠΕΚΤΑΣΗ

Στο έβδομο και τελευταίο κεφάλαιο θα παρουσιαστούν τα αποτελέσματα της πειραματικής διαδικασίας που περιγράφηκαν στο προηγούμενο κεφάλαιο και ένας σχολιασμός τους. Έπειτα, θα παρατεθούν ορισμένες τροποποιήσεις και βελτιώσεις που έγιναν στους κώδικες για τη βελτιστοποίηση των εφαρμογών και τέλος θα προταθούν ορισμένοι τρόποι επέκτασης αυτών.

7.1. ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΩΤΗΣ ΕΦΑΡΜΟΓΗΣ

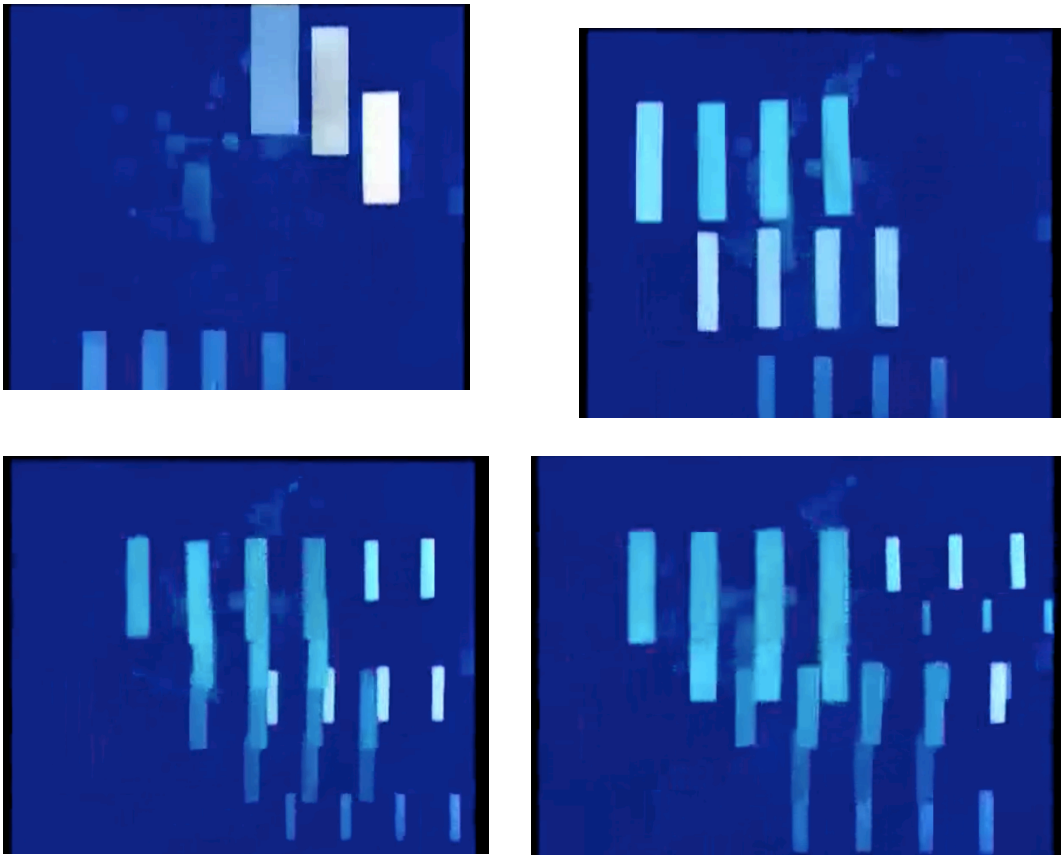
Όπως εξηγήθηκε προηγουμένως τα αποσπάσματα που χρησιμοποιήθηκαν σε αυτό το μέρος συνοδεύονται από ηχητικά σήματα παρόμοιας μετρικής δομής αλλά διαφορετικών τέμπο. Εφαρμόσαμε τον κώδικα για τέσσερα λοιπόν κομμάτια από το *Optical Poem* τα οποία παρουσιάζουν μεγάλες διαφορές μεταξύ τους ως προς το περιεχόμενο:

- Το πρώτο απόσπασμα περιλαμβάνει κινούμενες μορφές σε ένα μόνο μέρος της οθόνης, που καταλαμβάνει περίπου το μισό του συνολικού φόντου. Η κίνηση των αντικειμένων που περιλαμβάνει (σύνολα κύκλων), δεν παρουσιάζει πολύ έντονες μεταβολές. Παρακάτω φαίνονται ορισμένα καρέ από την εξέλιξη του:



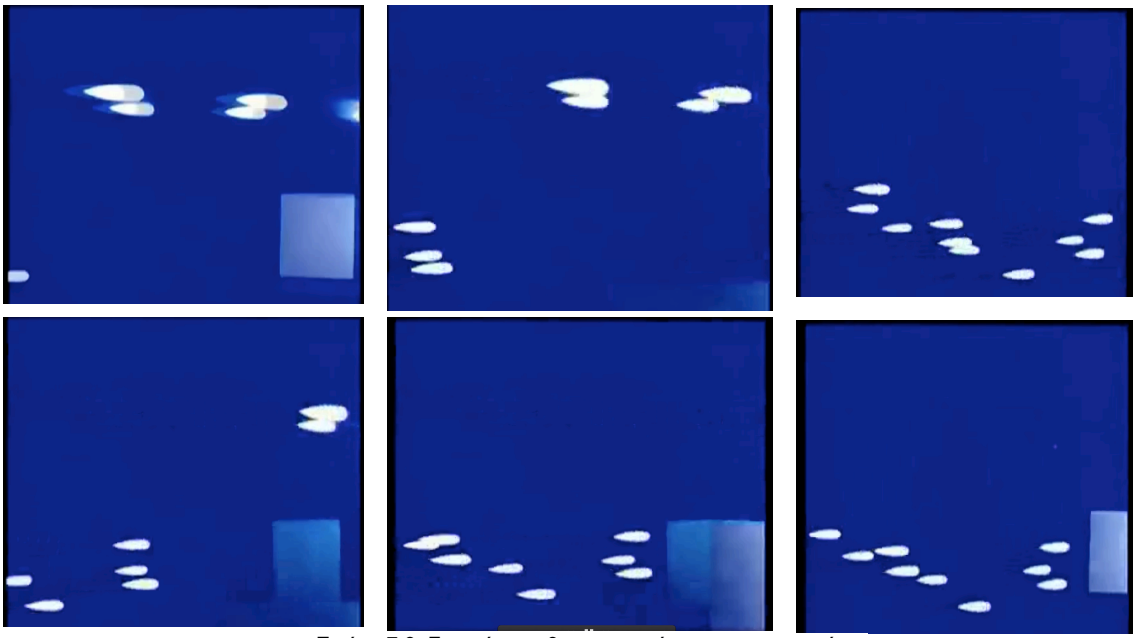
Εικόνα 7.1: Στιγμιότυπα 1ου αποσπάσματος εφαρμογής

- Το δεύτερο απόσπασμα παρουσιάζει κίνηση στο σύνολο της οθόνης, με μικρότερη ωστόσο ταχύτητα κι έχοντας μάλιστα αρκετά μεγάλη διάρκεια. Είναι μάλιστα το απόσπασμα εκείνο που εμπεριέχει την ελάχιστες χρωματικές μεταβολές σε σχέση με τα υπόλοιπα, με τις απεικονιζόμενες μορφές να ομοιάζουν χρωματικά και με το φόντο.



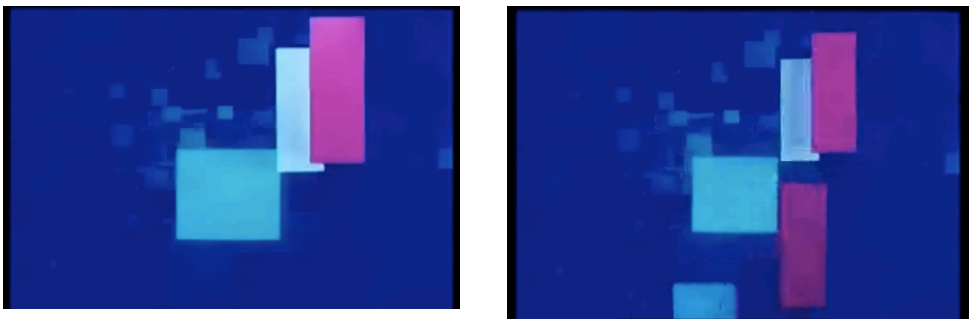
Εικόνα 7.2: Στιγμιότυπα 2ου αποσπάσματος εφαρμογής

- Το τρίτο βίντεο είναι αυτό που προκαταβολικά αναμένουμε να έχει το υψηλότερο Pace. Παρουσιάζει τη γρηγορότερη εναλλαγή μορφών με ταχέως κινούμενες ελλειψοδειδείς μορφές και τετράγωνα που "αναβοσβήνουν" στην οθόνη. Λόγω και της έντονης δραστηριότητας που περικλείει, συνοδεύεται και από το μουσικό μέρος με το υψηλότερο tempo.



Εικόνα 7.3: Στιγμιότυπα 3ου αποσπάσματος εφαρμογής

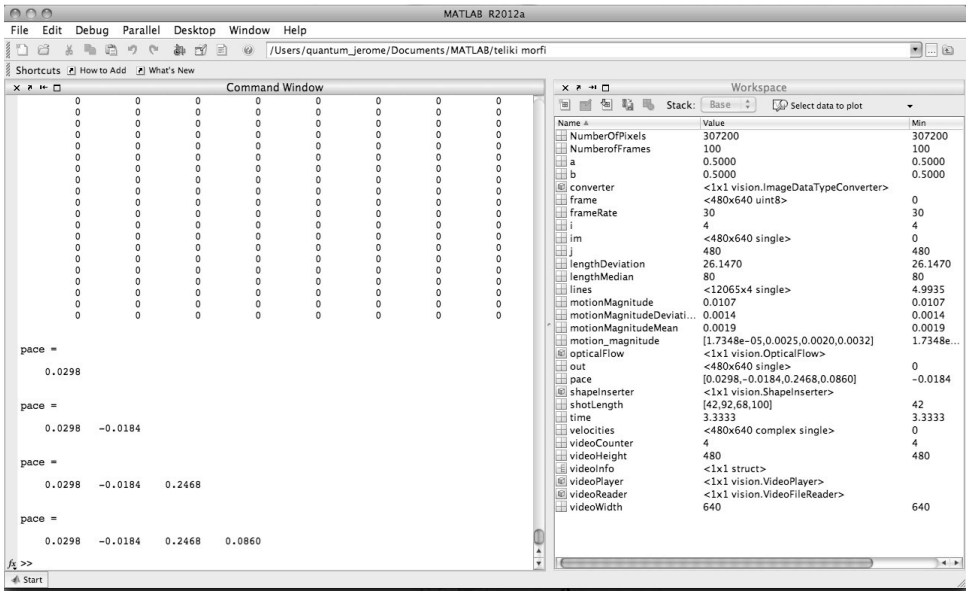
- Τέλος, το τέταρτο απόσπασμα παρουσιάζει κι αυτό κίνηση στο σύνολο της οθόνης με αρκετά γρήγορες μεταβολές μάλιστα. Είναι ακόμα το απόσπασμα με τις μεγαλύτερες μεταβολές στο χρωματικό περιεχόμενο με τις μορφές να αλλάζουν χρώματα από μπλε σε αποχρώσεις του κόκκινου έως και λευκό.



Εικόνα 7.4: Στιγμιότυπα 4ου αποσπάσματος εφαρμογής

Τρέχοντας λοιπόν τον κώδικα σε MATLAB για τα αποσπάσματα αυτά λαμβάνουμε τα εξής αποτελέσματα. Ο πίνακας `pace(i)` έχει τις εξής τιμές:

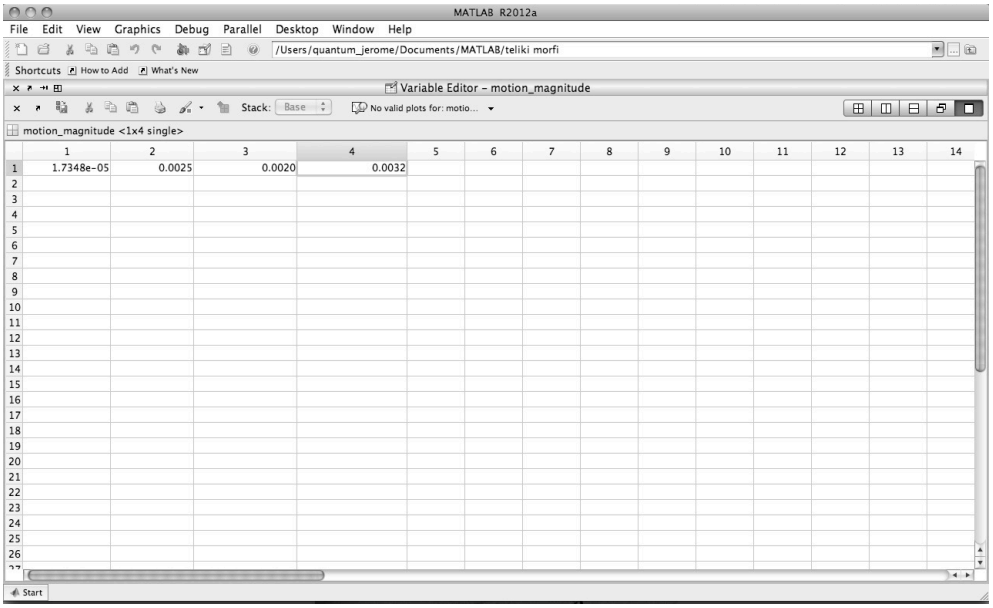
```
[0.0298    -0.0184    0.2468    0.0860]
```



Εικόνα 7.5: Screen shot αποτελέσματος για ***pace***

Είναι εμφανές λοιπόν πως τα αποτελέσματα είναι όντως κοντά στα όσα αναμέναμε. Θα εξετάσουμε και τους πίνακες `motion_magnitude` και `shotLength` που υπολογίστηκαν για να μπορέσουμε να σχολιάσουμε αναλυτικότερα. Ο πρώτος λοιπόν έχεις τις τιμές :

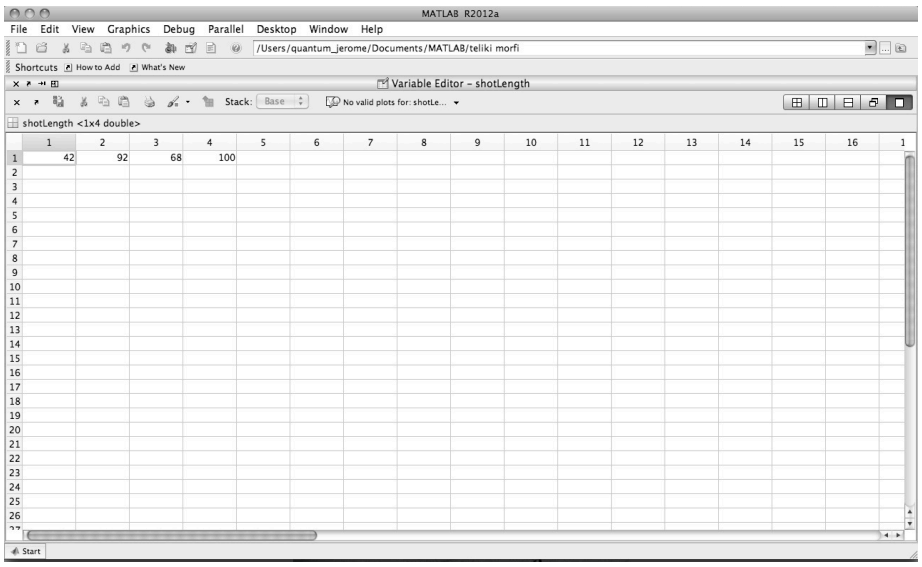
```
[1.7348e-05    0.0025    0.0020    0.0032]
```



Εικόνα 7.6: Screen shot αποτελέσματος για ***motion_magnitude***

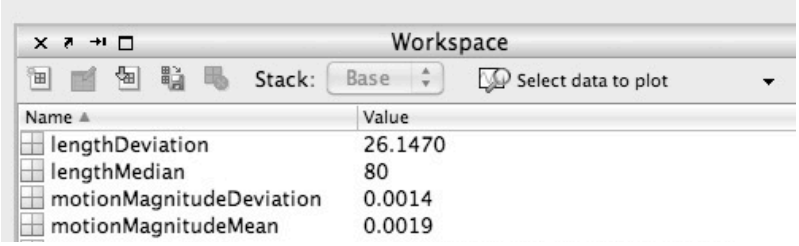
ενώ ο δεύτερος τις παρακάτω τιμές, μετρημένες σε καρέ:

```
[42    92    68    100]
```



Εικόνα 7.7: Screen shot αποτελέσματος για ***shotLength***

Τέλος τα στατιστικά μεγέθη που χρειαζόμαστε και αφορούν την ποσότητα της κίνησης και το μήκος κάθε αποσπάσματος υπολογίστηκαν ως εξής: Η τυπική απόκλιση του μήκους βρέθηκε 26.1470 καρέ, η κεντρική τιμή 80 καρέ, η μέση τιμή της ποσότητας κίνησης 0.0019 και η τυπική της απόκλιση 0.0014.



Εικόνα 7.8: Screen shot αποτελεσμάτων για στατιστικά μεγέθη

Συμπεραίνουμε λοιπόν τα εξής:

- Το `pace(3)` είναι η μέγιστη τιμή του αποτελέσματος ακόμα και εάν η τιμή του motion magnitude στο συγκεκριμένο απόσπασμα είναι μικρότερη απ' ότι στο δεύτερο και το τέταρτο. Η μικρότερη διάρκεια του ωστόσο, είναι αυτή που οδηγεί στη μεγιστοποίηση του συνολικού όρου αφού η διαφορά από την κεντρική τιμή είναι θετική.

- Δεύτερη μεγαλύτερη τιμή είναι αυτή του τέταρτου αποσπάσματος. Σε αυτό, ενώ παρατηρείται η μέγιστη ποσότητα κίνησης συγκριτικά με όλα τα υπόλοιπα αποσπάσματα, η μεγάλη διάρκεια του ρίχνει την τελική τιμή του Pace. Αυτό προκύπτει ποσοτικά από το ότι το shot length εδώ είναι πάνω από την κεντρική τιμή και η διαφορά τους προκύπτει αρνητική, ενώ έχει και διαισθητική επιβεβαίωση: Όσο περισσότερη ώρα παρακολουθούμε ένα απόσπασμα, τόσο περισσότερο "συνηθίζουμε" το περιεχόμενό του και ο ρυθμός της εξέλιξης του φαίνεται λιγότερο ταχύς.

- Τρίτο στη σειρά τοποθετείται το πρώτο απόσπασμα. Αυτό, εάν και παρουσιάζει την ελάχιστη ποσότητα στην κίνηση, ευνοείται από τη μικρή του διάρκεια η οποία είναι και πάλι η ελάχιστη ανάμεσα στα τέσσετα αποσπάσματα.

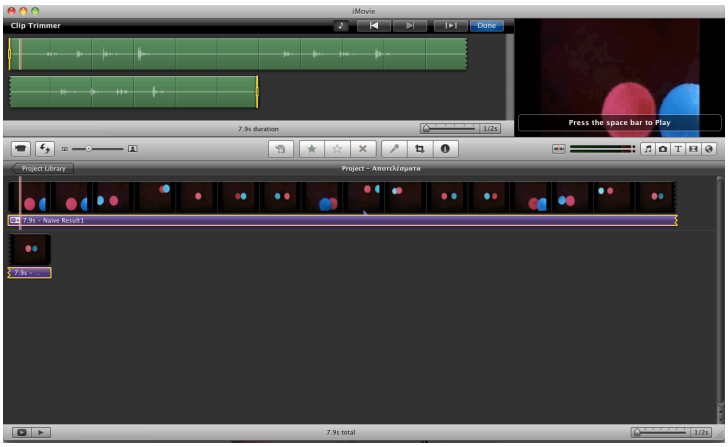
- Τελευταίο ταξινομείται το δεύτερο απόσπασμα. Αυτό παρουσιάζει και τις μεγαλύτερες ιδιαιτερότητες καθώς είναι το μοναδικό απόσπασμα που χαρακτηρίζεται από αρνητική τιμή του μεγέθους Pace. Αν και η κίνηση που περικλείει είναι αρκετά γρήγορα μεταβαλλόμενη, υπολογίζεται ως αρκετά μικρότερη στο μέγεθό της σε σχέση με το τέταρτου απόσπασμα που έχει παρόμοια διάρκεια με αυτό. Έτσι λοιπόν, και καθώς τα βάρη a και b για το ρόλο της διάρκειας και της κίνησης στο τελικό μέγεθος είναι ίσα, το motion magnitude δεν είναι αρκετά μεγάλο ώστε να εξισορροπήσει τον πρώτο όρο του αθροίσματος στον ορισμό που αποτιμάται αρνητικό.

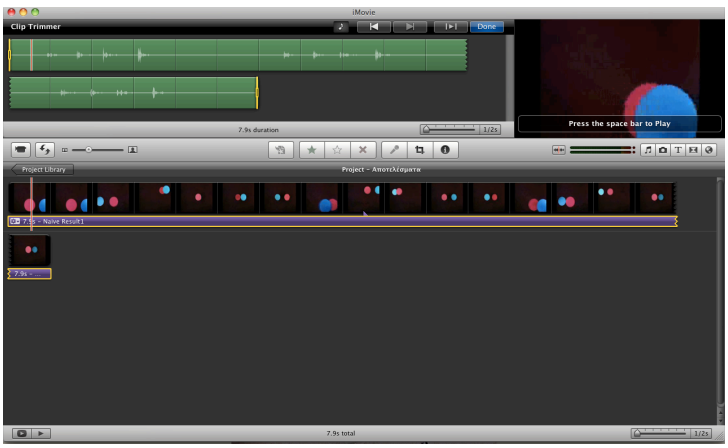
Συνολικά, καταλαβαίνουμε πως η οπτική και διαισθητική μας αντίληψη σχετικά με το ρυθμό ενός αποσπάσματος βίντεο κατά βάση συμφωνεί με τον ορισμό των Adams, Dorai και Venkatesh για το Video Pace. Βλέπουμε ωστόσο πως σε κάποιες περιπτώσεις, η ποσοτικοποίηση του ρόλου της διάρκειας ενός αποσπάσματος οδηγεί σε αποτελέσματα τα οποία να απέχουν ως ένα βαθμό από τη διαισθητική αντίληψη. Το γεγονός μπορεί να εξηγείται εν μέρει και από τη φύση των αποσπασμάτων εφαρμογής του ορισμού. Οι Adams et al στόχευαν όλη την έρευνα τους σε μεγάλης διάρκειας βίντεο ή και ολόκληρες κινηματογραφικές ταινίες. Στις περιπτώσεις αυτές, η χρονική διάρκεια όντως θα επηρεάζει διαφορετικά αυτό που ο θεατής αντιλαμβάνεται ως ρυθμό του οπτικού περιεχομένου: Στα δικά μας παραδείγματα εφαρμογής, διαφορές στη διάρκεια της τάξης του ενός δευτερολέπτου οδηγούν σε πολύ διαφορετικά αποτελέσματα.

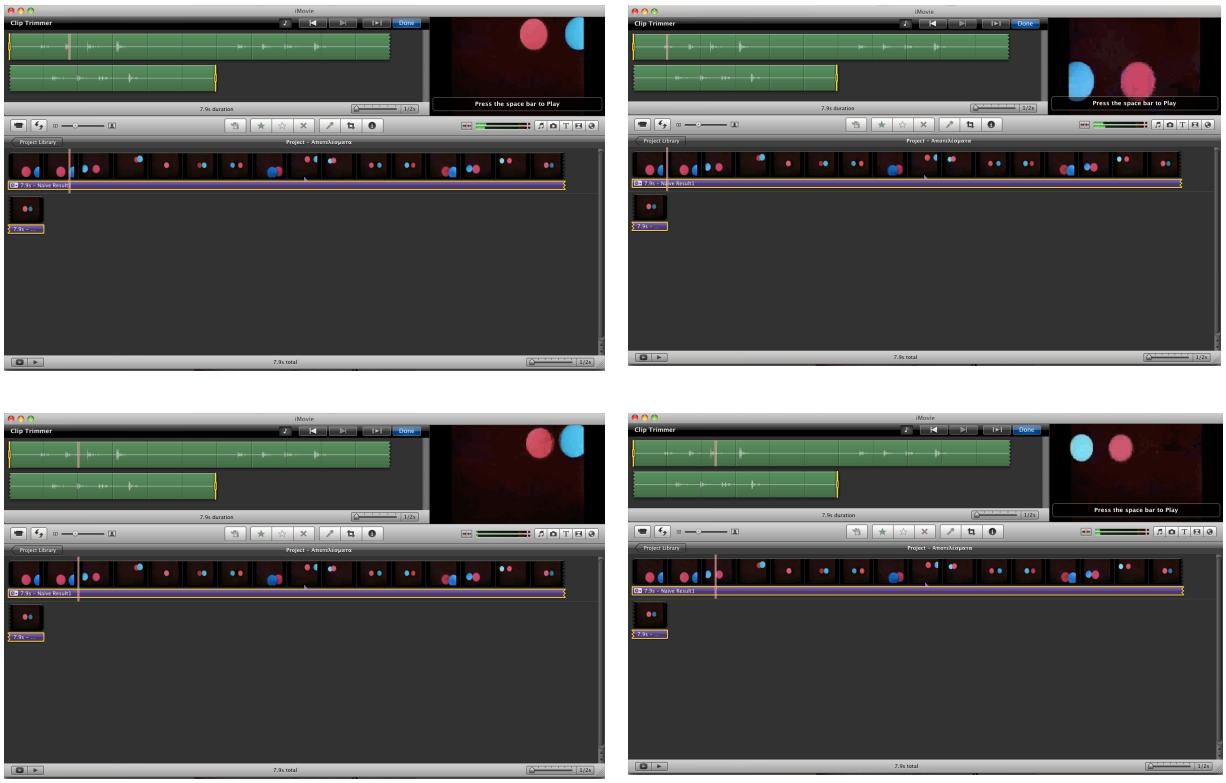
7.2. ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΕΥΤΕΡΗΣ ΕΦΑΡΜΟΓΗΣ

Για τη δεύτερη και βασική εφαρμογή, ο κώδικας που περιγράφηκε στο προηγούμενο κεφάλαιο εφαρμόστηκε αρχικά στο απόσπασμα του *An Optical Poem* που αναφέρθηκε ως το πρωταρχικό ερέθισμα για τη διαμόρφωση της διαδικασίας. Το MIDI κλιπ που εξάχθηκε τροφοδοτήθηκε στο πρόγραμμα Ableton Live, σε ένα MIDI όργανο (ουσιαστικά σε ένα ξεχωριστό κανάλι) το οποίο ορίστηκε να είναι ένα σετ από ντραμς. Στις νότες που εκ προοιμίου γνωρίζαμε ότι θα ξεκινάει το κλιπ, ορίσαμε τον ήχο από ένα χτύπημα σε μπότα, στην επόμενη τον ήχο από ένα χτύπημα σε ταμπούρο και επαναλάβαμε το μοτίβο (μπότα-ταμπούρο) για συνολικά 10 νότες. Το κλιπ εξόδου περιείχε 4 μόνο νότες (όσες οι αλλαγές που ανιχνεύθηκαν) και επομένως δεν θα χρειαστούμε όλους τους ήχους που αναθέσαμε.

Το πρώτο αποτέλεσμα είναι ιδιαίτερα ικανοποιητικό. Το οπτικό περιεχόμενο έχει συνοδευτεί από χτύπους ακριβώς τις χρονικές στιγμές που παρατηρούνται οι έντονες αλλαγές στις θέσεις του ζεύγους των σφαιρών (βλέπε και κεφάλαιο 6). Για να παρουσιάσουμε το αποτέλεσμα αυτό, επαναλάβαμε το βίντεο και το μουσικό ρυθμό που πλέον το συνοδεύει, τρεις φορές συνολικά και το εισάγαμε στο πρόγραμμα *iMovie*. Στα screen shots που φαίνονται παρακάτω, στο δεξί μέρος της οθόνης φαίνεται η εξέλιξη των οπτικών μορφών και αριστερά οι κόκκινες γραμμές υποδεικνύουν το χρονικό σημείο στο οποίο βρισκόμαστε στην κυματομορφή του ηχητικού μέρους (πάνω) και σε ποιο καρέ βρίσκεται το οπτικό μέρος (κάτω).







Εικόνες 7.9 - 7.14: Screen shots από παρουσίαση αποτελέσματος στο iMovie

Στα screen shots εξετάζουμε τις κυματομορφές που αναπαριστούν τον ήχο εξόδου: Όλες οι διαταραχές λαμβάνουν χώρα τις στιγμές που το περιστρεφόμενο ζεύγος των κύκλων αλλάζει τη θέση του στην οθόνη. Ενδιάμεσα, αφού οι μόνοι ήχοι είναι τα χτυπήματα στα τύμπανα, η κυματομορφή είναι μηδενική και οι κύκλοι παραμένουν στο ίδιο περίπου μήκος και πλάτος της οθόνης.

Η πρώτη μορφή αυτή του κώδικα, εφαρμόστηκε και σε άλλα πέντε αποσπάσματα με πολύ ικανοποιητικά αποτελέσματα. Με την εξαίρεση ορισμένων χρονικών σημείων, ο αλγόριθμος ανίχνευσε όλες τις καίριες μεταβολές στο οπτικό περιεχόμενο των βίντεο και τις "έντυσε" με ρυθμικά μέρη τα οποία πήραν την τελική τους μορφή όπως και προηγουμένως στο λογισμικό Ableton Live. Με την κατάλληλη μάλιστα επανάληψη ορισμένων αποσπασμάτων, δημιουργούνται και ρυθμοί που θα μπορούσαν να υφίστανται και ανεξάρτητα του οπτικού περιεχόμενου σε ένα καθαρά μουσικό περιβάλλον.

Σε κάποια άλλα αποσπάσματα ωστόσο ο αλγόριθμος δεν παρουσιάζει τα επιθυμητά αποτελέσματα. Σε ένα παράδειγμα ιδιαίτερα, στο οποίο στην οθόνη αναβοσβήνουν ταχύτατα άσπρες τριγωνικές μορφές, το κλιπ που εξάγεται έχει πάρα πολλές νότες, παιγμένες πάρα πολύ γρήγορα. Ενώ λοιπόν ανιχνεύονται οπτικές μεταβολές, το αποτέλεσμα δεν έχει κάποιο συγκεκριμένο ρυθμικό περιεχόμενο. Εάν μειώσουμε την ταχύτητα αναπαραγωγής του βίντεο και του συνοδευτικού ήχου, φαίνεται ότι τότε το αποτέλεσμα είναι όντως ρυθμικό, αυτό όμως δεν μας επαρκεί. Στο σημείο αυτό λοιπόν αξίζει να παρατεθούν ορισμένες βελτιστοποιήσεις που έγιναν στην πειραματική διαδικασία, τόσο στο επίπεδο της υλοποίησης όσο και στο επίπεδο της διαμόρφωσης της ιδέας.

7.3. ΒΕΛΤΙΩΣΕΙΣ ΤΗΣ ΔΕΥΤΕΡΗΣ ΕΦΑΡΜΟΓΗΣ

Η πρώτη στοιχειώδης βελτιστοποίηση του αλγορίθμου προέκυψε έχοντας το τελευταίο απόσπασμα που αναφέρθηκε ως ερέθισμα. Παρατηρώντας αυτό λοιπόν, φάνηκε ότι ορισμένες νότες της εξόδου είναι πολύ κοντά χρονικά, αφού μια αλλαγή που το μάτι αντιλαμβάνεται ως ενιαία μπορεί να γίνεται σε βάθος ορισμένων διαδοχικών καρέ. Μια προφανής λοιπόν βελτίωση θα ήταν για ένα προκαθορισμένο αριθμό καρέ, ο αλγόριθμος να επιτρέπει την αναγνώριση μίας μόνο μεταβολής. Μετακινούμενοι λοιπόν εντός του πίνακα των αλλαγών που ανιχνεύθηκαν, εάν βρεθούν συμβάντα που απέχουν μεταξύ τους μικρότερο από αυτό το ελάχιστο επιτρεπτό διάστημα καρέ, το σύνολο τους θα αντικαθίσταται από την πιο έντονη μεταβολή, το καρέ δηλαδή όπου σημειώθηκε η ελάχιστη τιμή για την ετεροσυσχέτιση. Παρακάτω, φαίνεται οι κώδικας για αυτή την τροποποίηση της διαδικασίας:

```
%Αρχικοποίηση του μετρητή,του διαστήματος αναζήτησης και του
%νέου πίνακα αλλαγών
counterNew=1;
margin=6;
changesNew(1)=changes_array(1);

%Αναζητούμε στο σύνολο του αρχικού πίνακα
for nn:2=megethos_allagwn

    if (changes_array(nn)-changesNew(counterNew)< margin )
%Εάν βρεθούν διαδοχικές τιμές με διαφορά εντός του μη επιτρεπτού
%διαστήματος, θα πρέπει να ανανεώσουμε το νέο πίνακα με την
%τιμή στην οποία λαμβάνει χώρα η εντονότερη αλλαγή

        if max_array(changes_array(nn))<max_array(changesNew(counterNew))
            %Ανανεώνεται το περιθώριο αναζήτησης
            margin=margin-(changes_array(nn)-changesNew(counterNew));
            %και η τιμή του πίνακα
            changesNew(counterNew)=changes_array(nn)
        end

    else
%Διαφορετικά, το παράθυρο στο οποίο κάνουμε την αναζήτηση
%αρχικοποιείται ξανά και ο νέος πίνακας θα πάρει ως νέα τιμή,
%προσωρινά αυτήν του αρχικού

        margin=6;
        counterNew=counterNew+1;
        changesNew(counterNew)=changes_array(nn);
    end
end
%Ο αριθμός των συνολικών μεταβολών που ανιχνεύθηκαν ανανεώνεται
%με βάση το νέο πίνακα changesNew
megethos_al=size(changesNew);
megethos_allagwn=megethos_al(2);
```

Η βασική ιδέα αυτού του αλγορίθμου βελτιστοποίησης είναι για κάθε στοιχείο του αρχικά εξαχθέντος πίνακα αλλαγών, να αναζητούνται επόμενα στοιχεία των οποίων η τιμή να απέχει από τη δικιά του, λιγότερο από μια καθορισμένη τιμή. Εάν βρεθεί μια τέτοια τιμή, πρέπει να εξεταστούν οι μέγιστες τιμές που έχουν βρεθεί για την ετεροσυσχέτιση σε κάθε στοιχείο του υπό εξέταση ζεύγους.
Εάν έχουμε μικρότερη τιμή για το καρέ που είναι το αρχικό στοιχείο στον πίνακα αλλαγών, τότε δεν απαιτείται κάποια ενέργεια: Ο νέος πίνακας θα περιλαμβάνει αυτό ακριβώς το καρέ ενώ το διάστημα τιμών δεν θα αλλάξει. Εάν ωστόσο συμβαίνει το αντίστροφο, τότε ο πίνακας που επιθυμούμε να διαμορφώσουμε να περιλαμβάνει το νέο καρέ μόνο. Επίσης, το διάστημα αναζήτησης θα μειωθεί κατά την απόσταση των δύο τιμών αφού δεν επιθυμούμε ένα κυλιόμενο "παράθυρο" αναζήτησης με σταθερό μήκος αλλά ένα διάστημα με προκαθορισμένο μήκος και σταθερό αρχικό σημείο το εκάστοτε αρχικό στοιχείο. Στην περίπτωση τώρα που δεν βρεθεί μία τιμή του αρχικού πίνακα που να ικανοποιεί τη συνθήκη που ορίσαμε, τότε ο changesNew θα είναι ουσιαστικά ένα αντίγραφο του αρχικού μέχρι να βρεθούν στοιχεία που απέχουν μεταξύ τους λιγότερο από το ορισθέν διάστημα.

Δοκιμάζοντας αυτό τη βελτίωση του κώδικα στο απόσπασμα που παρήγαγε μη ικανοποιητικά αποτελέσματα, είναι εύκολα κατανοητό ότι όντως η έξοδος είναι βελτιωμένη. Οι συνολικές νότες του κλιπ που εξάγεται μειώνονται από 51 για διάρκεια 9.6 δευτερολέπτων σε 18. Έτσι, η έξοδος έχει πιο συγκεκριμένο και πιο εύκολα αντιληπτό ρυθμικό υπόβαθρο. Παράλληλα, ο νέος κώδικας δεν επηρεάζει τα θετικά αποτελέσματα που είχαμε στις πρώτες δοκιμές αφού το διάστημα των 5 frames είναι αρκετά μικρό ώστε να "φιλτράρει" μόνο μικρότερες μεταβολές του περιεχομένου τις οποίες το μάτι δεν είναι πάντα ικανό να αντιληφθεί.

Η δεύτερη βελτιστοποίηση που επιχειρήθηκε αφορά κυρίως στη διαμόρφωση της αρχικής ιδέας. Όπως προαναφέρθηκε, η αντιστοίχιση των διαφόρων χρονικών στιγμών όπου βρέθηκε ένα σημαντικό συμβάν σε μία νότα, γίνεται αυθαίρετα ανάλογα με το κανάλι που θα επιλεχθεί να "ερμηνεύσει" το MIDI κλιπ στο Ableton Live. Μήπως λοιπόν θα ήταν προτιμότερο να προκαθορίζεται ποιες μεταβολές θα αντιστοιχούν σε ποιους ήχους;

Η υλοποίηση μιας τέτοιας προέκτασης στηρίζεται σε μια διαδικασία διαχωρισμού των μεταβολών σε δύο διακριτά σύνολα. Γνωρίζοντας λοιπόν τα καρέ στα οποία σημειώθηκαν οι πιο αξιοσημείωτες μεταβολές, μπορούμε να κάνουμε τη διάκριση αυτή και πάλι μέσω του πίνακα με τα μέγιστα της ετεροσυσχέτισης. Θα ορίσουμε λοιπόν οι μισές από τις ανιχνευθείσες χρονικές στιγμές να αντιστοιχούν σε ήχους ταμπούρου και οι άλλες μισές σε ήχους μπότας. Καθώς ένα χτύπημα στο ταμπούρο παράγει πολύ πιο οξύ (υψίσυχνο) ήχο απ' ότι ένα χτύπημα στη μπότα, ορίζεται οι μισές πιο έντονες μεταβολές να αντιστοιχούν σε νότες σε ταμπούρο και οι υπόλοιπες σε μπότα. Το διαχωριστικό κριτήριο θα είναι προφανώς ο μέσος όρος της μέγιστης ετεροσυσχέτισης στα καρέ που υπάρχουν είτε στον πίνακα changes_array

είτε στον changesNew ανάλογα με το αν επιλέξουμε να εντάξουμε την προηγούμενη βελτίωση ή όχι. Θα χρειαστεί επίσης, στο βρόχο επεξεργασίας όπου δημιουργείται ο changes_array μέσω της σύγκρισης με το αρχικό κατώφλι, να κρατάμε επίσης σε έναν νέο πίνακα τις τιμές του max_array στα καρέ που εξάγονται. Αρκεί λοιπόν το παρακάτω loop να ακολουθεί τους υπολογισμούς των mesos_oros, apoklisi και katwfli στη θέση του ήδη υπάρχοντος.

```
for i=1:megethos
    if max_array(i)<katwfli
        changes_array(changeCounter)=i;
        values(changeCounter)=max_array(i);
        changeCounter=changeCounter+1;
    end
end
```

Σε αυτό το σημείο θα χρειαστεί να δημιουργήσουμε δύο διαφορετικούς πίνακες οι οποίοι με τη σειρά τους θα χρησιμοποιηθούν για τη δημιουργία δύο ξεχωριστών αρχείων .mid . Το ένα από αυτά θα τροφοδοτηθεί σε ένα κανάλι που θα "παίζει" μόνο νότες σε μπότα και το άλλο μόνο σε ταμπούρο.

```
%Υπολογίζουμε το μέσο όρο και αρχικοποιούμε τους
%μετρητές για κάθε σύνολο από νότες
mesosOrosAllagwn=mean(values);
snareCounter=1;
bassCounter=1;


for i=1:megethos

    if values(i)>mesosOrosAllagwn

        snareDrum(snareCounter)=changes_array(i);
        snareCounter=snareCounter+1;

    else

        bassDrum(snareCounter)=changes_array(i);
        bassCounter=bassCounter+1;
    end
end



%Ο time_array θα χρειαστεί για να ορίσουμε μια καθολική
%διάρκεια νοτών και έπειτα μετατρέπουμε τα frames για
%κάθε "όργανο" σε πραγματικές χρονικές στιγμές
time_array=changes_array/30;
time_snare=snareDrum/30;
time_bass=bassDrum/30;
```

Πλέον, αρκεί να ξαναγίνει η διαδικασία υπολογισμού του `onset` όπως και προηγουμένως και να επαναληφθεί η διαδικασία δημιουργίας αρχείου .mid δύο φορές ώστε να προκύψουν οι διακριτές έξοδοι, ***snareDrum.mid*** και ***bassDrum.mid*** .

```
m=size(snareDrum);
megethos_snare=m(2);
mm=size(bassDrum);
megethos_bass=mm(2);

%Ταμπούρο
N1 = megethos_snare;
final_note1 = 60 + N1 -1;

M1 = zeros(N1,6);
M1(:,1) = 1 ;

M1(:,2) = 1;

M1(:,3) = (60:final_note1)';

M1(:,4) = round(linspace(80,120,N1));

M1(:,5) = (time_snare)';

M1(:,6) = M1(:,5) + onset;

midi_new1 = matrix2midi(M1);
writemidi(midi_new1,'snareDrum.mid');

%Μπότα
N2 = megethos_bass;
final_note2 = 60 + N2 -1;

M2 = zeros(N2,6);
M2(:,1) = 1 ;

M2(:,2) = 1;

M2(:,3) = (60:final_note2)';

M2(:,4) = round(linspace(80,120,N2));

M2(:,5) = (time_bass)';

M2(:,6) = M2(:,5) + onset;

midi_new2 = matrix2midi(M2);

%Δημιουργία της δεύτερης εξόδου
writemidi(midi_new2,'bassDrum.mid');
```

7.4. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΠΕΡΑΙΤΕΡΩ ΕΠΕΚΤΑΣΕΙΣ

Αντί επιλόγου, αξίζει να κάνουμε μια ανασκόπηση του συστήματος που διαμορφώθηκε. Το επιστέγασμα της παρούσας εργασίας είναι μια κατανοητή και εύχρηστη εφαρμογή που επιτρέπει σε οποιονδήποτε χρήστη, μουσικό ή όχι, να πειραματιστεί με τη σύνδεση οπτικών περιεχομένων με το μουσικό ρυθμό. Δίνοντας στο σύστημα ένα βίντεο της επιλογής του, ο χρήστης έχει τη δυνατότητα να χρησιμοποιήσει τις εξόδους τόσο για την κατανόηση ενός πιθανού ρυθμού του οπτικού αντικειμένου όσο και για τη μουσική επένδυση του βίντεο. Τα αρχεία .mid που εξάγονται μπορούν είτε να χρησιμοποιηθούν ανεξάρτητα, όπως φάνηκε κι από τα αποτελέσματα ορισμένων αποσπασμάτων, είτε να αποτελέσουν τη ρυθμική βάση για μια πιο πλούσια μουσική επένδυση. Τέλος, εφόσον εντάσσουμε όλη τη διαδικασία στα πλαίσια του sonification θα πρέπει τα αποτελέσματα να μπορούν να εξεταστούν και ως ανεξάρτητα ηχητικά μέρη που προέκυψαν μέσω της επεξεργασίας της εισόδου. Στην προκειμένη περίπτωση λοιπόν, χρησιμοποιώντας ως "έμπνευση" τα βίντεο εισόδου, συστηματοποιούμε τη μετατροπή τους σε ήχο ώστε όντως να παράγουμε αυστηρά ορισμένα μουσικά μέρη που θα μπορούσαν να παίζονται από το ρυθμικό μέρος ενός οποιουδήποτε μουσικού συνόλου.

Ωστόσο, το σύστημα που διαμορφώθηκε παρουσιάζει και ορισμένους περιορισμούς. Είδαμε ότι όλη η επεξεργασία των στοιχείων του βίντεο πρέπει να προηγηθεί και έπειτα να χρησιμοποιηθεί η ενδιάμεση έξοδος για το τελικό αποτέλεσμα. Εάν η διαδικασία μπορούσε να γίνει σε πραγματικό χρόνο (real time), θα είχαμε πολύ πιο ευπαρουσίαστα αποτελέσματα αλλά και δυνατότητα επέκτασης των χρήσεων της. Για παράδειγμα, θα μπορούσαμε να χρησιμοποιήσουμε το σύστημα σε συνδυασμό με βίντεο που προκύπτουν σε ζωντανό χρόνο από μία webcam: Αντιστοιχώντας για παράδειγμα τις πιο έντονες χειρονομίες ενός ατόμου στους ήχους των τυμπάνων όπως προηγουμένως, μία απλή κάμερα μπορεί να μετατραπεί σε ένα ψηφιακό, "κρουστό" μουσικό όργανο. Η ιδέα αυτή μπορεί μάλιστα να συμπεριλάβει πολλούς εμπλουτισμούς χρησιμοποιώντας για παράδειγμα, απλές εφαρμογές *ανίχνευσης αντικειμένων (object tracking).* Για παράδειγμα, με μια απλή ένδειξη με τα δάχτυλα, ο χρήστης θα μπορούσε να ορίζει τη δομή του ρυθμού που θα παιχτεί ώστε το σύστημα να παίζει αυτόματα και με αντίστοιχο τρόπο τις νότες για παράδειγμα στα πιατίνια. Ακόμη, χωρίζοντας την επιφάνεια που σαρώνει η κάμερα σε υποπεριοχές μπορούμε να αντιστοιχήσουμε μονοσήμαντα μία χειρονομία σε έναν ήχο: Γροθιά στο κάτω δεξιά τεταρτημόριο χτύπημα στην μπότα, γροθιά στο κάτω αριστερά χτύπημα στο ταμπούρο και ανοιχτή παλάμη στο πάνω δεξιά χτύπημα στο πιατίνι crash (crash cymbal) κ.ό.κ.

Εικόνα 7.15: Πιθανή Ένδειξη για ρυθμό 4/4

Για μια τέτοια επέκταση ωστόσο, το περιβάλλον MATLAB δεν προσφέρεται ιδιαίτερα λόγω των υψηλών του απαιτήσεων σε υπολογιστική ισχύ. Μια απλή μείωση των απαιτήσεων από την CPU χωρίς να εγκαταλείψουμε τελείως το MATLAB, θα ήταν η ενσωμάτωση βιβλιοθηκών όπως η *OpenCV* που περιλαμβάνουν κώδικες και σε άλλες γλώσσες (κυρίως C++) .

Τέλος, μία ακόμη επέκταση της εφαρμογής θα ήταν η ενσωμάτωση και άλλων "ρυθμικών" οργάνων στο τελικό ηχητικό αποτέλεσμα. Λόγω της διαδικασίας που ακολουθήθηκε, δεν μπορούμε να περιμένουμε ιδιαίτερο μελωδικό περιεχόμενο. Για μικρά αποσπάσματα ωστόσο όπως αυτά που εξετάστηκαν, η τροφοδότηση των αρχείων .mid σε ένα MIDI όργανο ηλεκτρικού μπάσου για παράδειγμα, ακόμα και εάν απλά "έπαιζε" μία κλίμακα νοτών, θα βελτίωνε αισθητά το ηχητικό αποτέλεσμα.

Σε κάθε περίπτωση, θεωρούμε πάντως πως η προηγούμενη ανάλυση είναι ιδιαίτερα χρήσιμη για την κατανόηση της έννοιας του sonification γενικότερα αλλά και την παραγωγή εύχρηστων αποτελεσμάτων, εστιάζοντας πάντα στο βίντεο ως πηγή της διαδικασίας και στο ρυθμό ως επίκεντρο των αποτελεσμάτων της.

ΠΑΡΑΡΤΗΜΑ 1
ΤΟ ΠΡΩΤΟΚΟΛΛΟ MIDI

Ο όρος MIDI (Musical Instrument Digital Interface) χρησιμοποιείται για τόσο πρωτόκολλο όσο για τις διεπαφές και τις διάφορες συσκευές που επιτρέπουν την επικοινωνία μεταξύ διαφόρων ηλεκτρονικών μουσικών οργάνων και ηλεκτρονικών υπολογιστών. Ένα "μήνυμα" MIDI μπορεί να περιλαμβάνει έως και 16 κανάλια πληροφορίας, το καθένα από τα οποία μπορεί να δοθεί ως είσοδος σε διαφορετική συσκευή. Οι πληροφορίες αυτές αφορούν στην αναπαράσταση του μουσικού κομματιού, στο pitch, σε σήματα ελέγχου για παραμέτρους όπως η ένταση και το vibrato και για σήματα ρολογιού για τον ορισμό και το συγχρονισμό του τέμπο μεταξύ διαφόρων τερματικών. Η πληροφορία σχετικά με το τονικό ύψος (pitch) κωδικοποιείται με τιμές στο διάστημα 0-127, αντιστοιχώντας κάθε τιμή σε ένα φθόγγο ξεκινώντας από τη νότα C1 για το 0 [45].

Με αντίστοιχο τρόπο (συνήθως σε δεκαεξαδική μορφή) κωδικοποιούνται επίσης πληροφορίες για την ενεργοποίηση κάθε νότας, την απενεργοποίηση της ή και το velocity δηλαδή την ταχύτητα με την οποία ένα πλήκτρο ενός εικονικού πιάνο μετακινείται από τη θέση ηρεμίας του στην πλέον συμπιεσμένη. Συνήθως, βρίσκουμε αρχικά ένα byte κατάστασης που καθορίζει την παράμετρο την οποία αφορά το εκάστοτε μήνυμα και έπειτα δύο byte δεδομένων αναφορικά με την παράμετρο αυτή.

Στην παρούσα εργασία έγινε χρήση του έργου του Dr. Ken Schutte για τη χρήση του πρωτοκόλλου MIDI μέσα από το MATLAB, υπό την GNU General Public License [44]. Στους κώδικές του ο Schutte ορίζει πως το μήνυμα MIDI θα περιέχει 6 κανάλια πληροφορίας τα οποία ελέγχουν τα εξής: Πόσα διαφορετικά κανάλια θα έχει το αρχείο, εάν θα έχουμε στερεοφωνικό ήχο ή όχι, πόσες νότες περιλαμβάνονται, ποιες θα είναι οι εντάσεις τους, πότε ενεργοποιείται κάθε νότα και πότε απενεργοποιείται. Τα δύο scripts που χρησιμοποιήθηκαν φαίνονται παρακάτω:

writemidi.m

```
function rawbytes=writemidi(midi,filename,do_run_mode)
% rawbytes=writemidi(midi,filename,do_run_mode)
%
% writes to a midi file
%
% midi is a structure like that created by readmidi.m
%
% do_run_mode: flag - use running mode when possible.
% if given, will override the msg.used_running_mode
% default==0. (1 may not work...)
%
% TODO: use note-on for note-off... (for other function...)
%
% Copyright (c) 2009 Ken Schutte
% more info at: http://www.kenschutte.com/midi
%if (nargin<3)
do_run_mode = 0;
%end
% do each track:
Ntracks = length(midi.track);
```

```
for i=1:Ntracks
databytes_track{i} = [];
for j=1:length(midi.track(i).messages)
msg = midi.track(i).messages(j);
msg_bytes = encode_var_length(msg.deltatime);
if (msg.midimeta==1)
% check for doing running mode
run_mode = 0;
run_mode = msg.used_running_mode;
% should check that prev msg has same type to allow run
% mode...
% if (j>1 && do_run_mode && msg.type == midi.track(i).messages(j-1).type)
% run_mode = 1;
% end
msg_bytes = [msg_bytes; encode_midi_msg(msg, run_mode)];
else
msg_bytes = [msg_bytes; encode_meta_msg(msg)];
end
% disp(msg_bytes')
       %if (msg_bytes ~= msg.rawbytes)
       % error('rawbytes mismatch');
       %end
       databytes_track{i} = [databytes_track{i}; msg_bytes];
       end
       end
       % HEADER
       % double('MThd') = [77 84 104 100]
       rawbytes = [77 84 104 100 ...
       0 0 0 6 ...
       encode_int(midi.format,2) ...
       encode_int(Ntracks,2) ...
       encode_int(midi.ticks_per_quarter_note,2) ...
       ]';
       % TRACK_CHUCKS
       for i=1:Ntracks
       a = length(databytes_track{i});
       % double('MTrk') = [77 84 114 107]
       tmp = [77 84 114 107 ...
              encode_int(length(databytes_track{i}),4) ...
              databytes_track{i}']';
              rawbytes(end+1:end+length(tmp)) = tmp;
              end
              % write to file
              fid = fopen(filename,'w');
              %fwrite(fid,rawbytes,'char');
              fwrite(fid,rawbytes,'uint8');
              fclose(fid);
              % return a _column_ vector
              function A=encode_int(val,Nbytes)
              for i=1:Nbytes
              A(i) = bitand(bitshift(val, -8*(Nbytes-i)), 255);
              end
              function bytes=encode_var_length(val)
              % What should be done for fractional deltatime values?
              % Need to do this round() before anything else, including that
              % first check for val<128 (or results in bug for some fractional values)
              % Probably should do rounding elsewhere and require
              % this function to take an integer.
              val = round(val)
              if val<128 % covers 99% cases!
              bytes = val;
              return
              end
              binStr = dec2base(round(val),2);
              Nbytes = ceil(length(binStr)/7);
              binStr = ['00000000' binStr];
              bytes = [];
              for i=1:Nbytes
              if (i==1)
              lastbit = '0';
              else
              lastbit = '1';
              end
              B = bin2dec([lastbit binStr(end-i*7+1:end-(i-1)*7)]);
```

```
bytes = [B; bytes];
end
function bytes=encode_midi_msg(msg, run_mode)
bytes = [];
if (run_mode ~= 1)
bytes = msg.type;
% channel:
bytes = bytes + msg.chan; % lower nibble should be chan
end
bytes = [bytes; msg.data];
function bytes=encode_meta_msg(msg)
bytes = 255;
bytes = [bytes; msg.type];
bytes = [bytes; encode_var_length(length(msg.data))];
 bytes = [bytes; msg.data];
```

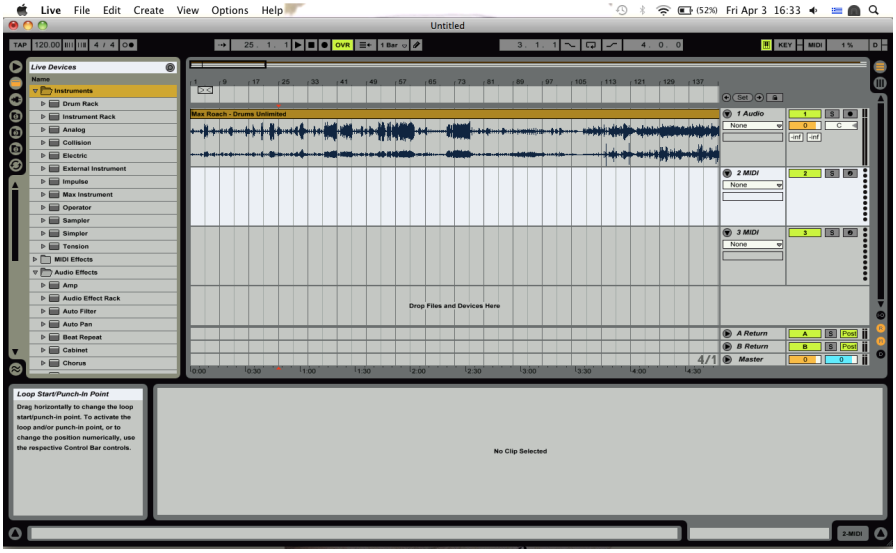
matrix2midi.m

```
function midi=matrix2midi(M,ticks_per_quarter_note,timesig)
% midi=matrix2midi(M,ticks_per_quarter_note)
%
% generates a midi matlab structure from a matrix
% specifying a list of notes. The structure output
% can then be used by writemidi.m
%
% M: input matrix:
% 1 2 3 4 5 6
% [track chan nn vel t1 t2] (any more cols ignored...)
%
% optional arguments:
% - ticks_per_quarter_note: integer (default 300)
% - timesig: a vector of len 4 (default [4,2,24,8])
%
% Copyright (c) 2009 Ken Schutte
% more info at: http://www.kenschutte.com/midi
% TODO options:
% - note-off vs vel=0
% - tempo, ticks, etc
if nargin < 2
ticks_per_quarter_note = 300;
end
if nargin < 3
timesig = [4,2,24,8];
end
tracks = unique(M(:,1));
Ntracks = length(tracks);
% start building 'midi' struct
if (Ntracks==1)
midi.format = 0;
else
midi.format = 1;
end
midi.ticks_per_quarter_note = ticks_per_quarter_note;
tempo = 500000; % could be set by user, etc...
% (microsec per quarter note)
for i=1:Ntracks
trM = M(tracks(i)==M(:,1),:);
note_events_onoff = [];
note_events_n = [];
note_events_ticktime = [];
% gather all the notes:
for j=1:size(trM,1)
% note on event:
note_events_onoff(end+1) = 1;
note_events_n(end+1) = j;
note_events_ticktime(end+1) = 1e6 * trM(j,5) * ticks_per_quarter_note / tempo;
% note off event:
note_events_onoff(end+1) = 0;
note_events_n(end+1) = j;
note_events_ticktime(end+1) = 1e6 * trM(j,6) * ticks_per_quarter_note / tempo;
end
msgCtr = 1;
% set tempo...
```

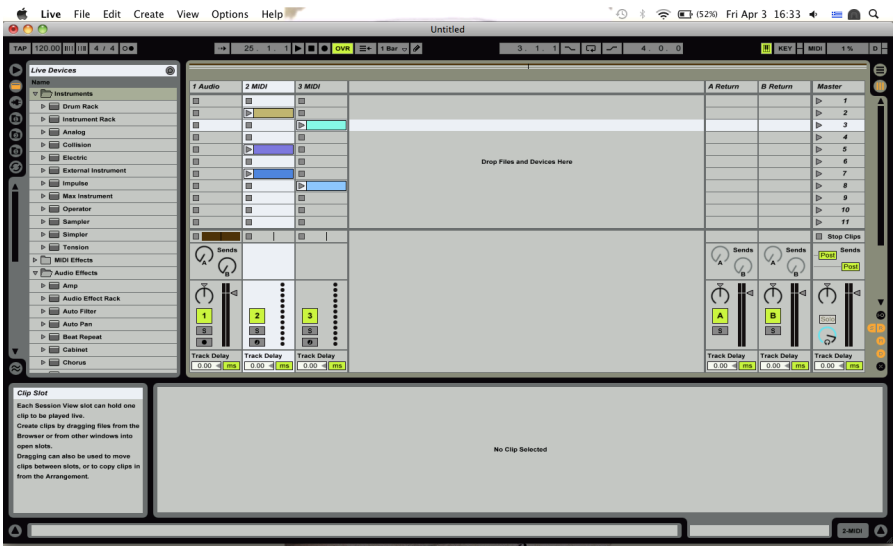
```
midi.track(i).messages(msgCtr).deltatime = 0;
midi.track(i).messages(msgCtr).type = 81;
midi.track(i).messages(msgCtr).midimeta = 0;
midi.track(i).messages(msgCtr).data = encode_int(tempo,3);
midi.track(i).messages(msgCtr).chan = [];
msgCtr = msgCtr + 1;
% set time sig...
midi.track(i).messages(msgCtr).deltatime = 0;
midi.track(i).messages(msgCtr).type = 88;
midi.track(i).messages(msgCtr).midimeta = 0;
midi.track(i).messages(msgCtr).data = timesig(:);
midi.track(i).messages(msgCtr).chan = [];
msgCtr = msgCtr + 1;
[junk,ord] = sort(note_events_ticktime);
prevtick = 0;

for j=1:length(ord)
n = note_events_n(ord(j));
cumticks = note_events_ticktime(ord(j));
midi.track(i).messages(msgCtr).deltatime = cumticks - prevtick;
midi.track(i).messages(msgCtr).midimeta = 1;
midi.track(i).messages(msgCtr).chan = trM(n,2);
midi.track(i).messages(msgCtr).used_running_mode = 0;
if (note_events_onoff(ord(j))==1)
% note on:
midi.track(i).messages(msgCtr).type = 144;
midi.track(i).messages(msgCtr).data = [trM(n,3); trM(n,4)];
else
%-- note off msg:
%midi.track(i).messages(msgCtr).type = 128;
%midi.track(i).messages(msgCtr).data = [trM(n,3); trM(n,4)];
%-- note on vel=0:
midi.track(i).messages(msgCtr).type = 144;
midi.track(i).messages(msgCtr).data = [trM(n,3); 0];
end
msgCtr = msgCtr + 1;
prevtick = cumticks;
end
% end of track:
midi.track(i).messages(msgCtr).deltatime = 0;
midi.track(i).messages(msgCtr).type = 47;
midi.track(i).messages(msgCtr).midimeta = 0;
midi.track(i).messages(msgCtr).data = [];
midi.track(i).messages(msgCtr).chan = [];
msgCtr = msgCtr + 1;
end
% return a _column_ vector
% (copied from writemidi.m)
function A=encode_int(val,Nbytes)
A = zeros(Nbytes,1); %ensure col vector (diff from writemidi.m...)
for i=1:Nbytes
A(i) = bitand(bitshift(val, -8*(Nbytes-i)), 255);
end
```

ΠΑΡΑΡΤΗΜΑ 2
ΤΟ ΛΟΓΙΣΜΙΚΟ ABLETON LIVE

Το Ableton Live είναι ένα λογισμικό για Windows και OS X που λειτουργεί ως sequencer και ως digital audio workstation (DAW), παρέχει δηλαδή τη δυνατότητα για ηχογράφηση, επεξεργασία και παραγωγή μουσικών αρχείων. Έχει σχεδιαστεί τόσο για τη σύνθεση, τη μίξη ή την ηχογράφηση μουσικής όσο και για την χρήση ως όργανο σε ζωντανές παραστάσεις. Το πρόγραμμα είναι γραμμένο στη γλώσσα C++ και περιλαμβάνει ενσωματωμένα προσομοιωμένα μουσικά όργανα κάθε οικογένειας όπως και μια πληθώρα από ηχητικά εφέ που μπορούν να εφαρμοστούν στα ηχογραφηθέντα μέρη. Περιλαμβάνει δύο interfaces εργασίας, το Arrangement Mode και το Session Mode. Το πρώτο, που φαίνεται στην πρώτη εικόνα, προορίζεται κυρίως για την ενορχήστρωση και τη μίξη ενώ το δεύτερο, στη δεύτερη εικόνα, για την ηχογράφηση κλιπ, προκαθορισμένου μήκους ή όχι, σε διάφορα όργανα ή για τη ζωντανή αναπαραγωγή τους. Τέλος, όπως είδαμε και στη διαδικασία παραπάνω, το Ableton Live υποστηρίζει και ορισμένους τύπους βίντεο, στα οποία μπορεί να εφαρμόσει μια μερική επεξεργασία. Ένα βίντεο μπορεί να εισαχθεί στο πρόγραμμα ως αρχείο audio, ώστε ο χρήστης να μπορεί να μεταχειριστεί όπως επιθυμεί το ηχητικό μέρος του και να εφαρμόσει ορισμένες λειτουργίες στο οπτικό [46].



Arrangement Mode



Session Mode

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] T. Hermann, “TAXONOMY AND DEFINITIONS FOR SONIFICATION AND AUDITORY DISPLAY,” in *icad.org*, 2008, pp. 1–8.

[2] R. L. Alexander, S. O’Modhrain, D. A. Roberts, J. A. Gilbert, and T. H. Zurbuchen, “The bird’s ear view of space physics: Audification as a tool for the spectral analysis of time series data,” *J. Geophys. Res. Sp. Phys.*, vol. 119, no. 7, pp. 5259–5271, 2014.

[3] Wikipedia, “Visual music,” *In Wikipedia, The Free Encyclopedia*, 2015. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Visual_music&oldid=646819071.

[4] O. E. D. Online, “Oxford English Dictionary Online,” *Oxford English Dictionary*, 2010. [Online]. Available: http://dictionary.oed.com.

[5] E. Britannica, “Encyclopedia - Britannica Online Encyclopedia,” *EBU*, 2011. [Online]. Available: http://www.school.eb.com.au/all/comptons/article-9275557?query=martin luther&ct=null.

[6] J. Molino, J. Underwood, and C. Ayrey, “Musical fact and the semiology of music,” *Music Anal.*, vol. 9, pp. 105–111, 1990.

[7] M. Bradshaw and I. Xenakis, “Formalized Music: Thought and Mathematics in Composition,” *Music Educators Journal*, vol. 59. p. 85, 1973.

[8] K. Muscutt, “Composing with Algorithms: An Interview with David Cope,” *Computer Music Journal*, vol. 31. pp. 10–22, 2007.

[9] Roger B Dannenberg, *Algorithmic Composition: A Guide to Composing Music with Nyquist*, University. University of Michigan Press, 2013, p. 262.

[10] H. C. Longuet-Higgins, “Perception of melodies,” *Nature*, vol. 263. pp. 646–653, 1976.

[11] P. E. Allen and R. B. Dannenberg, “Tracking Musical Beats in Real Time,” in *International Computer Music Conference Glasgow 1990*, 1990, pp. 140 – 143.

[12] A. T. Cemgil, B. Kappen, P. Desain, and H. Honing, “On tempo tracking: Tempogram Representation and Kalman filtering,” *Journal of New Music Research*, vol. 29. pp. 259–273, 2000.

[13] C. Raphael, “Automated Rhythm Transcription,” in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2001, pp. 99–107.

[14] W. A. Sethares, R. D. Morris, and J. C. Sethares, “Beat tracking of musical performances using low-level audio features,” *IEEE Trans. Speech Audio Process.*, vol. 13, pp. 275–285, 2005.

[15] D. P. W. Ellis, “Beat Tracking by Dynamic Programming,” *Journal of New Music Research*, vol. 36. pp. 51–60, 2007.

[16] P. Cooper, *Perspectives in music theory: an historical-analytical approach*. Dodd, Mead, 1973.

[17] J. D. Kramer, *The time of music: new meanings, new temporalities, new listening strategies*. Schirmer/Mosel Verlag GmbH, 1988.

[18] M. Yeston, *The Stratification of Musical Rhythm*. Yale University Press, 1976.

[19] M. J. Moravcsik and D. Rosenbluth, *Musical Sound: An Introduction to the Physics of Music*. Springer US, 2001.

[20] S. M. Boker, “The Perception of Structure in Simple Auditory Rhythmic Patterns,” 1994.

[21] T. L. Bolton, “Rhythm.,” *Am. J. Psychol.*, 1894.

[22] P. Fraisse, *Les structures rythmiques: {é}tude psychologique*. Publications universitaires de Louvain, 1956.

[23] H. Woodrow, “Time Perception,” in *Handbook of experimental psychology*, .

[24] D.-J. Povel and H. Okkerman, “Accents in equitone sequences,” *Percept. Psychophys.*, no. 30, 1981.

[25] J. M. Thomassen, “Melodic Accents: experiments and a tentative model,” *J. Acoust. Soc. Am.*, vol. 71, 1982.

[26] M. H. Thaut, P. D. Trimarchi, and L. M. Parsons, “Human brain basis of musical rhythm perception: common and distinct neural substrates for meter, tempo, and pattern.,” *Brain Sci.*, vol. 4, pp. 428–52, 2014.

[27] A. Moles, *Information Theory and Esthetic Perception*. University of Illinois Press, 1968.

[28] Wikipedia, “Video,” *In Wikipedia, The Free Encyclopedia*, 2015. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Video&oldid=655267524.

[29] T. Sobchack and V. C. Sobchack, *An Introduction to Film*. Little, Brown, 1987.

[30] B. Adams, C. Dorai, and S. Venkatesh, “Study of shot length and motion as contributing factors to movie tempo (poster session),” *Proc. eighth ACM Int. Conf. Multimed. - Multimed. ’00*, pp. 353–355, 2000.

[31] D. Bates and A. Jhala, “Multi-Modal Analysis of Movies for Rhythm Extraction,” pp. 14–17.

[32] A. Liu, Z. Yang, J. Wu, Y. Zhang, and J. Li, “An innovative tempo model for movie content analysis,” *Proc. 2008 IEEE Int. Conf. Networking, Sens. Control. ICNSC*, pp. 1348–1352, 2008.

[33] C. Guedes and C. Guedes, “Extracting Musically-Relevant Rhythmic Information from Dance Movement by Applying Pitch Tracking Techniques to a Video Signal,” *Proc. 2006 Sound Music Comput. Conf.*, pp. 25–33, 2006.

[34] W. T. Chu and S. Y. Tsai, “Rhythm of motion extraction and rhythm-based cross-media alignment for dance videos,” *IEEE Trans. Multimed.*, vol. 14, pp. 129–141, 2012.

[35] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” *Proc. 7th Int. Jt. Conf. Artif. Intell.*, pp. 674–679, 1981.

[36] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach, Third edition*. 2014.

[37] J. Proakis and D. Manolakis, *Digital Signal Processing : Principles, Algorithms and Applications*. Pearson Prentice Hall, 2007.

[38] R. Mcgee, “VOSIS : a Multi-touch Image Sonification Interface.”

[39] B. Verplank, M. Mathews, and R. Shaw, “2) Scan 1) Manipulate.”

[40] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” *Artificial Intelligence*, vol. 17. pp. 185–203, 1981.

[41] E. W. Weisstein, “Cross-Correlation,” *From MathWorld--A Wolfram Web Resource*. .

[42] Wikipedia, “Cross-correlation,” *In Wikipedia, The Free Encyclopedia*. .

[43] MathWorks, “normxcorr2 : Normalized 2-D cross correlation.” [Online]. Available: http://www.mathworks.com/help/images/ref/normxcorr2.html?refresh=true#zmw57dd0e84821.

[44] K. Schutte, “MATLAB and MIDI.” [Online]. Available: http://www.kenschutte.com/midi.

[45] P. J. Hass, “Chapter Three : Midi How does the MIDI system work?,” *Introduction to Computer Music: Volume One, Indiana University*, 2013. [Online]. Available: http://www.indiana.edu/~emusic/etext/MIDI/chapter3_MIDI.shtml.

[46] Ableton, “Ableton Live.” [Online]. Available: https://www.ableton.com/en/live/new-in-9/.