



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Αλγοριθμική Σύνθεση με την Χρήση Εικόνων Βίντεο

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ξενοφών Καρακωνσταντής

Επιβλέπων: Γεώργιος Καμπουράκης Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2017



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Αλγοριθμική Σύνθεση με την Χρήση Εικόνων Βίντεο

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ξενοφών Καρακωνσταντής

Επιβλέπων: Γεώργιος Καμπουράκης Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2017

Εγκρίθηκε από την τριμελή επιτροπή την --/--/2017

.....

Καμπουράκης Γεώργιος
Καθηγητής Ε.Μ.Π.

.....

Λούμος Βασίλειος
Καθηγητής Ε.Μ.Π.

.....

Παπαοδυσσεύς Κωνσταντίνος
Καθηγητής Ε.Μ.Π.

.....

Ξενοφών Καρακωνσταντής

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

Copyright © Ξενοφών Καρακωνσταντής 2017

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

ΠΕΡΙΛΗΨΗ

Από την μουσική των δεδομένων του σωματιδίου Higgs boson μέχρι και την μουσική της διάταξης του γαλαξία, η ηχητικοποίηση έχει ευρύ φάσμα εφαρμογής που συνεχώς μεγαλώνει. Η ανάγκη της εναλλακτικής μελέτης και ανάλυσης των δεδομένων αυξάνεται τόσο όσο και ο όγκος και η περιπλοκότητα τους στον σύγχρονο επιστημονικό κόσμο. Παράλληλα, η πληροφορία των περιεχομένων της σύγχρονης μουσικής συνεχώς πληθαίνει και οι συνθέτες μουσικής καθώς και το κοινό τους, αναζητούν συνεχώς νέα μέσα διάδοσης της. Η αλγοριθμική σύνθεση πρόκειται για την πλέον πρωτοπόρα μέθοδο της σύλληψης και δημιουργίας της μουσικής αυτής.

Στα πλαίσια της παρούσης διπλωματικής εργασίας, εξετάζεται η διαδικασία και οι τεχνικές της αλγοριθμικής σύνθεσης, όπως επίσης και ο τρόπος κατά τον οποίο εφαρμόζεται με την χρήση των σύγχρονων μεθόδων επεξεργασίας και ανάλυσης της εικόνας. Επιπρόσθετα, παρουσιάζονται ορισμένες προσεγγιστικές μέθοδοι υλοποίησης της παραπάνω λογικής με την χρήση δεδομένων εικόνας του βίντεο.

Λέξεις-κλειδιά

Ηχητικοποίηση, βίντεο, αλγοριθμική σύνθεση, χρώματα, κίνηση, οπτικοακουστικό, οπτική αντίληψη, ακουστική αντίληψη, υπολογιστική όραση

ABSTRACT

From the music of the Higgs boson particle data to the music of the galaxy's structure, sonification has a wide range of applications that are constantly growing. The necessity for alternative means of studying and analysing data increases as much as the volume and complexity of the data itself in the modern scientific world. At the same time, the information and content of contemporary music is constantly increasing and music composers as well as their counterpart audience are constantly seeking new means of dissemination. Algorithmic synthesis is the most pioneering method of capturing and creating this music.

In the context of this diploma thesis, the algorithmic synthesis process and techniques are examined, as well as the way they are applied with the use of modern methods of image processing and analysis. Additionally, an assessment is made following the implementation and definition of the above method and the video data used.

Keywords

Sonification, video, algorithmic composition, colours, motion, audio-visual, visual perception, acoustic perception, computer vision

Ευχαριστώ θερμά τον επιβλέποντα καθηγητή κ. Γεώργιο Καμπουράκη και τον υποψήφιο διδάκτορα Κωνσταντίνο Μπακογιάννη για την πολύτιμη βοήθεια τους, καθώς και τους Δημήτρη Αποστολακίδη, Δημήτρη Κρητικό και Χριστίνα Σπαή για την καθοριστική συμβολή τους για την ολοκλήρωση της εργασίας αυτής.

ΠΕΡΙΕΧΟΜΕΝΑ

1.	ΕΙΣΑΓΩΓΗ	12
1.1	Αλγόριθμοι στην μουσική	12
1.2	Η μουσική ως πνευματική οντότητα.....	13
1.3	Ηχητικοποίηση Δεδομένων (Sonification)	14
1.4	Ιστορική Αναδρομή στην Αλγοριθμική Μουσική.....	14
1.4.1	Σειριακή Μουσική	17
1.4.2	Στοχαστική Μουσική	18
1.4.3	Αλεατορική Μουσική	20
1.4.4	Σύγχρονη Αλγοριθμική Σύνθεση	21
1.5	ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ	22
2.	ΗΧΟΣ & ΑΚΟΗ.....	24
2.1	ΠΩΣ ΑΝΤΙΛΑΜΒΑΝΕΤΑΙ Ο ΑΝΘΡΩΠΟΣ ΤΟΝ ΗΧΟ	24
2.2	Η ΦΥΣΙΚΗ ΣΗΜΑΣΙΑ ΤΟΥ ΗΧΟΥ.....	26
2.3	Η ΜΟΥΣΙΚΗ ΩΣ ΗΧΟΣ.....	28
2.3.1	Ηχόχρωμα.....	28
2.3.2	Τονικό ύψος.....	29
2.3.3	Ακουστότητα – Ένταση του Ήχου.....	30
2.3.4	Δυναμικές.....	31
2.3.5	Διάρκεια – Τέμπο – Ρυθμός	32
3.	ΟΡΑΣΗ & ΟΠΤΙΚΗ ΑΝΤΙΛΗΨΗ.....	34
3.1	Η ΛΕΙΤΟΥΡΓΙΑ ΤΗΣ ΟΡΑΣΗΣ ΣΤΟΝ ΑΝΘΡΩΠΟ	34
3.2	ΤΟ ΧΡΩΜΑ, ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥ ΚΑΙ ΟΙ ΧΡΩΜΑΤΙΚΟΙ ΧΩΡΟΙ	35
3.2.1.	Χαρακτηριστικά του χρώματος	36
3.3	ΧΡΩΜΑΤΙΚΟΙ ΧΩΡΟΙ	37
3.3.1	Το μοντέλο CIEXYZ.....	38
3.3.2	Το μοντέλο CIELUV	39
3.3.3	Το μοντέλο CIELAB.....	40
3.3.4	Το μοντέλο RGB.....	41
3.3.5	Το μοντέλο YCbCr	43
3.3.6	Το μοντέλο HSV/HSL.....	44
3.3.7	Το μοντέλο CMY(K).....	45
3.4	ΧΡΩΜΑΤΙΚΟΣ ΚΥΚΛΟΣ	46
4.	ΕΙΚΟΝΑ & ΒΙΝΤΕΟ.....	48
4.1	ΨΗΦΙΑΚΗ ΕΙΚΟΝΑ	48

4.1.1	Γραφικά Raster	49
4.1.2	Γραφικά Vector.....	49
4.2	ΨΗΦΙΑΚΟ ΒΙΝΤΕΟ	50
4.3	ΥΠΟΛΟΓΙΣΤΙΚΗ ΟΡΑΣΗ.....	51
4.3.1	Ιστορία της Υπολογιστικής Όρασης.....	51
4.4	ΛΕΙΤΟΥΡΓΙΕΣ ΥΠΟΛΟΓΙΣΤΙΚΗΣ ΟΡΑΣΗΣ.....	52
4.4.1	Αναγνώριση Προτύπων	52
4.4.2	Κινητική Ανάλυση.....	53
4.4.3	Ανακατασκευή πλάνου (Scene reconstruction).....	53
4.4.4	Επαναφορά εικόνας (Image Restoration)	54
4.5	ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΑΣ.....	55
4.6	ΤΕΧΝΙΚΕΣ ΚΑΤΑΤΜΗΣΗΣ ΕΙΚΟΝΑΣ.....	55
4.6.1	Μέθοδος Κατωφλίου (Thresholding)	55
4.6.2	K-Means Clustering.....	56
4.6.3	Ανίχνευση Ακμών	57
4.6.4	Median.....	58
4.6.5	Μέθοδοι μερικής διαφόρισης κατάτμησης εικόνας.....	59
4.6.6	Watershed	61
4.6.7	Άλλες μέθοδοι κατάτμησης.....	61
5.	ΥΛΟΠΟΙΗΣΗ ΑΛΓΟΡΙΘΜΟΥ.....	63
5.1	ΕΡΓΑΛΕΙΑ ΓΙΑ ΤΗΝ ΥΛΟΠΟΙΗΣΗ	63
5.1.1	Python.....	63
5.1.2	OpenCV.....	64
5.1.3	Mingus	64
5.1.4	Scikit-Learn	64
5.1.5	Matplotlib	65
5.1.6	Numpy	65
5.1.7	Imutils	65
5.1.8	Πρωτόκολλο MIDI.....	66
5.2	ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ ΑΛΓΟΡΙΘΜΩΝ ΕΠΕΞΕΡΓΑΣΙΑΣ.....	66
5.3	ΛΕΙΤΟΥΡΓΙΑ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ	70
5.3.1	Αλγόριθμος δειγματοληψίας και καταμερισμού (Πρώτο στάδιο επεξεργασίας)	70
5.3.2	Επεξήγηση αλγορίθμου δειγματοληψίας και καταμερισμού.....	73
5.3.3	Αλγόριθμος χρωματικής και κινητικής παραμετροποίησης (Δεύτερο στάδιο επεξεργασίας)	74

5.3.4	Επεξήγησης αλγορίθμου χρωματικής και κινητικής παραμετροποίησης	81
5.3.5	Αλγόριθμος αρμονίας.....	83
6.	ΣΥΜΠΕΡΑΣΜΑΤΑ	88
6.1	ΑΠΟΤΕΛΕΣΜΑΤΑ ΥΛΟΠΟΙΗΣΗΣ	88
6.2	ΠΡΟΣΘΗΚΕΣ ΑΛΓΟΡΙΘΜΟΥ & ΕΠΙΠΡΟΣΘΕΤΕΣ ΛΕΙΤΟΥΡΓΙΕΣ	91
6.2.1	Χρωματική αρμονία	92
6.2.2	Θερμά & ψυχρά χρώματα	94
6.3	ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΕΠΕΚΤΑΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ	96
	Βιβλιογραφία	98

1. ΕΙΣΑΓΩΓΗ

1.1 ΑΛΓΟΡΙΘΜΟΙ ΣΤΗΝ ΜΟΥΣΙΚΗ

Οι αλγόριθμοι, ως πεπερασμένα σύνολα διαδικασιών και κανόνων, χρησιμοποιούνται για την εκπόνηση βασικών λειτουργιών των ανθρώπων εδώ και αιώνες. Δεν πρόκειται μόνο για την υλοποίηση αλγορίθμων με την σύγχρονη μέθοδο υπολογιστικών συστημάτων, αλλά γενικότερα, για την μεθοδευμένη διαδικασία επίλυσης προβλημάτων μέσω συγκεκριμένων πεπερασμένων οδηγιών, με λογική σειρά και σε πεπερασμένο χρόνο. Πιο συγκεκριμένα, κάθε ακολουθία οδηγιών στην οποία μπορεί να υπακούσει ένα ρομπότ, ονομάζεται αλγόριθμος. (Stone, 1971) Συχνά αποκτά πολύπλοκη μορφή και δομή όπως και επιτελεί πολύπλοκο έργο, σε ένα ευρύ φάσμα εφαρμογής. Επομένως και εννοιολογικά, ο αλγόριθμος ερμηνεύεται ανάλογα με την εφαρμογή του. Στον 21^ο αιώνα, παρατηρείται μια έξαρση στην εφαρμογή της αλγοριθμικής λογικής σε τέτοιο βαθμό, ώστε να αποτελεί πλέον βασικός πυλώνας της σύγχρονης βιομηχανίας και επιστήμης.

Ούτε για την διαδικασία σύνθεσης της μουσικής είναι άγνωστη η έννοια του αλγορίθμου. Εν τούτοις, η μουσική αποτελεί μία διαδικασία που ο αλγόριθμος καλείται να εκτελέσει, η αλλιώς, το πρόβλημα προς επίλυση. Η μορφή που θα αποκτήσει το μουσικό έργο, εξαρτάται από τις οδηγίες που δόθηκαν στον αλγόριθμο και η σειρά με τις οποίες αυτές εκτελούνται. Ενδεικτικά της σύνθεσης της μουσικής μέσω αλγορίθμου είναι τα διαφορά αρμονικά μοντέλα που κυριαρχούν και σήμερα. Είτε αυτό του 'δυτικού' πολιτισμού με το συγκεκριμένο σύστημα διδασκαλίας και εκμάθησης, η αντίστιξη και το σύνολο της αρμονίας, είτε τα διαφορά τροπικά συστήματα ή μικροτονικά συστήματα διδασκαλίας της μουσικής (Τουρκία, Μέση Ανατολή κ.α.) που ακολουθούν επί των πλείστων μια συγκεκριμένη ακολουθία κανόνων και οδηγιών για να μπορέσει να αποκτήσει υπόσταση το μουσικό κομμάτι. Όμως, η μουσική είναι επίσης μια ρευστή έννοια και ενδεχομένως να μην αποτελεί ανάγκη να ακολουθεί κανείς συγκεκριμένους κανόνες, αλλά εμπειρικά και βιωματικά, ενδεχομένως λόγω πολιτιστικής κληρονομιάς, να κατασκευάσει ο συνθέτης το εν λόγω έργο.

1.2 Η ΜΟΥΣΙΚΗ ΩΣ ΠΝΕΥΜΑΤΙΚΗ ΟΝΤΟΤΗΤΑ

«Υπάρχει σίγουρα μια καθολική και ασυνείδητη τάση να επιβάλλεται ένας ρυθμός, ακόμη και όταν ακούγεται μια σειρά πανομοιότυπων ήχων σε σταθερά διαστήματα ... Έχουμε την τάση να ακούμε τον ήχο ενός ψηφιακού ρολογιού, για παράδειγμα, ως "τικ-τοκ, τικ-τοκ"- παρόλο που στην πραγματικότητα είναι " τικ τικ, τικ τικ".»

- (Sacks, 2008)

Όπως και στον αλγόριθμο, δύσκολα μπορεί κανείς να δώσει μια συνοπτική και σαφή ερμηνεία στην μουσική, διότι πρόκειται για μια σειρά τονικών μοτίβων και κρότων σε συγκεκριμένο συχνοτικό φάσμα το οποίο αντιλαμβάνεται σχεδόν αποκλειστικά το ανθρώπινο είδος με μια σχετική συνάφεια. Δεν έχει υλική υπόσταση, δεν εμπεριέχει κάποια φυσική σημασία πέρα από αυτήν που του δίνουμε εμείς. Παρόλα αυτά, από την γέννηση του ανθρωπίνου πολιτισμού, τον συνοδεύει, τον θεμελιώνει και αποτελεί κεντρικό στοιχείο αυτού.

Πλέον είναι βέβαιο πως η πλειοψηφία των ανθρώπων, έχει μια προδιάθεση στο να αντιλαμβάνεται την μουσική, να την απολαμβάνει και να την δημιουργεί. Ο άνθρωπος μπορεί να αντιληφθεί την τονικότητα, την χροιά, το τέμπο, τα αρμονικά διαστήματα, τις μελωδικές γραμμές και τον ρυθμό. Ακούγοντας μουσική, του προκαλεί συναισθήματα και μηχανικές ή ρυθμικές κινήσεις, επηρεάζει την διάθεση του και αποτελεί ερέθισμα της μνήμης του, σε κάποιους δε, δημιουργεί την ικανότητα αντίληψης χρωμάτων με κάθε νέα τονική μετάβαση (συναισθησία). Πρόκειται για μία τέχνη τόσο βαθιά συνδεδεμένη με την ανθρώπινη ύπαρξη που δύσκολα μπορεί να εξηγηθεί η υπόστασή της. Σύμφωνα με τον Σοπενχάουερ η μουσική είναι η μόνη τέχνη η οποία δεν αντιγράφει απλά ιδέες, αλλά αποτελεί την ίδια την ενσάρκωση της βούλησής της. (Schopenhauer, 1966) Όμως είναι αναγκαία η εκλογίκευση της εννοίας της μουσικής για τον ανθρώπινο νου, αρά και της αντίστοιχης ερμηνείας της.

Ως παραδοχή λοιπόν, δεχόμαστε τον ορισμό του Oxford Universal Dictionary πως *«μουσική είναι οι φωνητικοί ή οργανικοί ήχοι (ή ο συνδυασμός τους) που κατά την ένωση τους παράγουν την ομορφιά της φόρμας, της αρμονίας και της έκφρασης συναισθημάτων»*. (Oxford University, 2010) Πως ξεχωρίζει όμως η μουσική από τον απλό ήχο ή τον θόρυβο; Πρόκειται για μία ερώτηση που δεν έχει ξεκάθαρη απάντηση, πάρα μόνο σε ατομικές εκτιμήσεις. Δηλαδή αν κάποιος ήχος θεωρείται θόρυβος ή όχι συχνά εξαρτάται από το ποιος τον ακούει. Το να χτυπάς με σταθερό βήμα ένα μολύβι σε ένα γραφείο μπορεί να είναι ρυθμικό και μουσικό για κάποιον, αλλά ενοχλητικός θόρυβος για κάποιον άλλον. Σε πολλές περιπτώσεις η καθοριστική διαφορά ενός θορύβου με κάτι μουσικό είναι η κυματομορφή του εν λόγω ήχου. Στη μια περίπτωση υπάρχουν πολλά ανώμαλα σημεία (εξαιρετικά πολλές αρμονικές), με συχνές αλλαγές στο μήκος του κύματος, ή κάποια χαμηλής

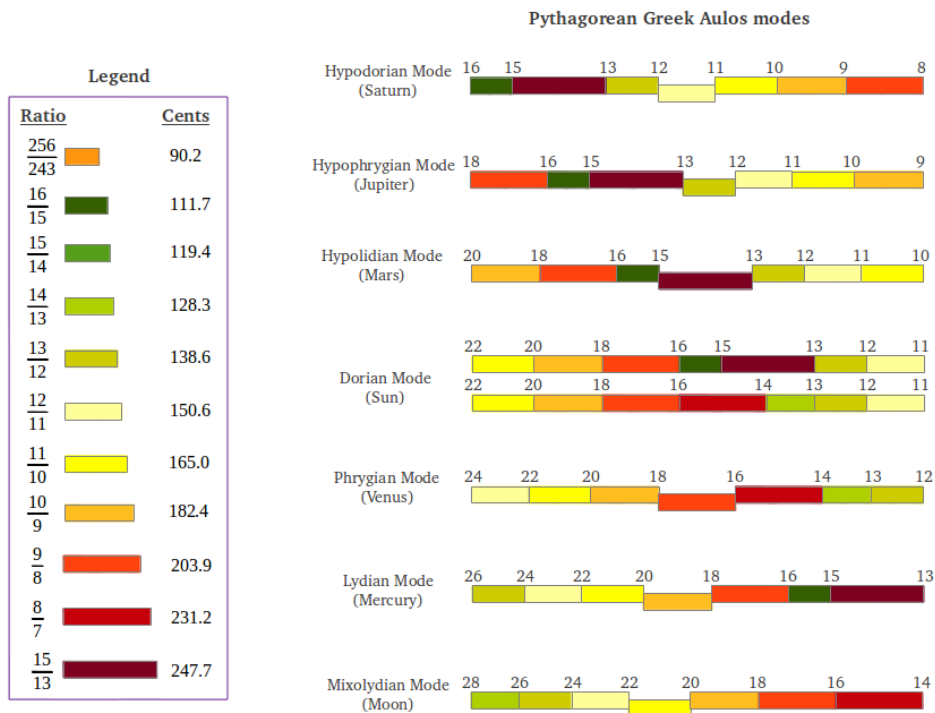
συχνότητας κυματομορφή, ενώ στην άλλη περίπτωση κάτι πιο σταθερό με σχετικά λίγες αρμονικές και σε 'μεσαίο' επίπεδο στο ακουστικό πεδίο.

1.3 ΗΧΗΤΙΚΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ (SONIFICATION)

Η Ηχητικοποίηση είναι η χρήση ήχου εκτός ομιλίας για τη μετάδοση πληροφοριών ή την αντίληψη δεδομένων. (Kramer, 1994) Η ακουστική μορφή των δεδομένων έχει πλεονεκτήματα σε χρονική και χωρική ανάλυση αλλά και σε ανάλυση πλάτους και συχνότητας. Πρόκειται για μια εναλλακτική λύση αλλά και συμπλήρωμα στις τεχνικές απεικόνισης. Ένα παράδειγμα ηχητικοποίησης είναι ο ήχος του καρδιακού παλμογράφου, ο οποίος για κάθε παλμό της καρδιάς απηχεί παράλληλα έναν συγκεκριμένο τόνο. Μπορεί να αντιληφθεί κανείς την χρησιμότητα της συγκεκριμένης εφαρμογής, μιας και τον καρδιακό παλμό δεν τον ακούμε σε συνθήκες θορυβώδους περιβάλλοντος.

1.4 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ ΣΤΗΝ ΑΛΓΟΡΙΘΜΙΚΗ ΜΟΥΣΙΚΗ

Όπως προαναφέρθηκε, η χρήση της αλγοριθμικής σκέψης έχει αρχέγονες ρίζες. Η ιδέα της χρήσης οδηγιών και διαδικασιών για τη δημιουργία μουσικής χρονολογείται ήδη από την εποχή των αρχαίων Ελλήνων. Ο Πυθαγόρας πίστευε σε μια άμεση σχέση μεταξύ των νόμων της φύσης και της αρμονίας των ήχων που εκφράζονται στη μουσική. Πιο συγκεκριμένα, συνέδεσε τις αριθμητικές αναλογίες που προκύπταν από την μελέτη του γνωστού για την εποχή σύμπαντος, με τα αρμονικά διαστήματα που προκύπταν μετά την μουσική εφαρμογή τους, (Philip, 1966) αντίληψη την οποία αποκόμισαν αργότερα ο Πτολεμαίος και ο Πλάτωνας. Προφανώς το μουσικό σύστημα της εποχής αποδίδεται στα μαθηματικά των πυθαγορείων, αλλά σε καμία περίπτωση δεν υπήρξε ο μουσικός απολύτως μονωμένος από την διαδικασία λήψης αποφάσεων όσον αφορά την σύνθεση.



Εικόνα 1: Τα πυθαγόρεια μονόχορδα (σκάλες) και οι λόγοι που προκύπτουν από τα διαστήματα μεταξύ των θέσεων τους.

Κατά τον μεσαίωνα, όταν πλέον ο πυθαγόρειος συσχετισμός αναλογίας-διαστήματος είναι λόγια και απόκρυφη γνώση, η εκκλησιαστική μουσική, ως γνωστόν πρωτοπόρα της εποχής, έρχεται να κατασκευάσει την έννοια της πολυφωνίας. Από τον 9^ο αιώνα συναντάμε το Οργκάνουμ, την πρώτη απόπειρα πολυφωνικής μουσικής των Γρηγοριανών μοναχών κατά την ψαλμωδία. Συνήθως δύο ή περισσότερες φωνές ξεκινούσαν από την ίδια νότα, στη συνέχεια απομακρύνονταν η μία από την άλλη κινούμενες παράλληλα (συνήθως στην 4^η ή 5^η βαθμίδα) και τέλος κατέληγαν πάλι στην ίδια νότα. Η λογική εστιαζόταν στο πως με βάση ένα κυρίαρχο φωνητικό κομμάτι (Γρηγοριανή ψαλμωδία) μπορούν να δοθούν οδηγίες (κανόνες) στις επιπρόσθετες φωνές για να προκύψει τελικά ένα αρμονικό προϊόν.



Early organum; from *Musica enchiriadis* (c. 859).

Εικόνα 2: Ένα δείγμα σημειογραφίας του Organum όπου φαίνονται οι δύο φωνές αρχικά να αποκλίνουν και στην συνέχεια συγκλίνουν.

μουσικών κομματιών. Συνθέτες όπως ο Γιόχαν Φίλιπ Κίρνμπεργκερ, ο Μαξιμιλιανός Στάντλερ, Βόλφγκανγκ Αμαντέους Μότσαρτ και ο Ιωσήφ Χάυντν εισήγαγαν την έννοια του *Musikalisches Würfelspiel* (Μουσικά Ζάρια), στο οποίο συνέθεταν **έντεκα** διαφορετικές μορφές του μέτρου ενός μενουέτου (αριστοκρατικός χορός, ανεξάρτητο μέρος μιας σονάτας). (Alpern, 1995) Το κάθε μέτρο μπορούσε να συνδεθεί με το επόμενο με οποιονδήποτε συνδυασμό του αθροίσματος δύο ζαριών (**έντεκα** πιθανά αποτελέσματα). Με αυτόν τον τρόπο η αρμονία του κομματιού αποκτούσε διάφορες μορφές, στηριζόμενο φυσικά από τον κανόνα των ζαριών.

WOLFGANG AMADEUS MOZART

Musikalisches Würfelspiel

Table of Measure Numbers

	Part One								Part Two								
	I	II	III	IV	V	VI	VII	VIII	I	II	III	IV	V	VI	VII	VIII	
2	96	22	141	41	105	122	11	30	2	70	121	26	9	112	49	109	14
3	32	6	128	63	146	46	134	81	3	117	39	126	56	174	18	116	83
4	69	95	158	13	153	55	110	24	4	66	139	15	132	73	58	145	79
5	40	17	113	85	161	2	159	100	5	90	176	7	34	67	160	52	170
6	148	74	163	45	80	97	36	107	6	25	143	64	125	76	136	1	93
7	104	157	27	167	154	68	118	91	7	138	71	150	29	101	162	23	151
8	152	60	171	53	99	133	21	127	8	16	155	57	175	43	168	89	172
9	119	84	114	50	140	86	169	94	9	120	88	48	166	51	115	72	111
10	98	142	42	156	75	129	62	123	10	65	77	19	82	137	38	149	8
11	3	87	165	61	135	47	147	33	11	102	4	31	164	144	59	173	78
12	54	130	10	103	28	37	106	5	12	35	20	108	92	12	124	44	131

Table of Measures



Εικόνα 4: Οι πίνακες αντιστοίχισης με την ρίψη ζαριών στο *Musikalisches Würfelspiel* του Μότσαρτ.

1.4.1 Σειριακή Μουσική

Από τον 20^ο αιώνα και μετά, η οργάνωση της μουσικής με τον δωδεκαθογγισμό/ισοσυγκερασμό, σύστημα το οποίο προκαθορίζει την τονικότητα και εξαφανίζει την ιεραρχική σχέση τόνου και κλειδιού, είναι πλέον ευρέως διαδεδομένο και κυρίαρχο ενοποιητικό στοιχείο της δυτικής μουσικής. Η μουσική αποκτάει αναγνωρίσιμες παραμέτρους όπως η διάρκεια, οι δυναμικές, το

ηχώχρωμα και πλέον υπάρχει μια καθολική συνάφεια στις μουσικές συνθέσεις. Συνεπώς η σύνθεση κατ' αυτόν τον τρόπο, δηλαδή η συστηματική οργάνωση των φθόγγων, αποτελεί αλγοριθμική προσέγγιση της μουσικής. Ο συνθέτης, για να έχει στη διάθεσή του περισσότερο υλικό, μπορεί να επεξεργαστεί τη δωδεκάφθογγη σειρά με τις τεχνικές της παραδοσιακής αντίστιξης όπως το τρανσπόρτο (μεταφορά σε άλλο κλειδί), την αναστροφή (μίμηση της πρώτης φωνής σε άλλον τόνο της ίδιας σκάλας), την μετατροπία (αλλαγή σκάλας) κ.α., οι οποίες τεχνικές εφαρμόζονται η κάθε μία αλγοριθμικά στο κυρίαρχο μοτίβο της μελωδίας.

Με την σειριακή αυτή μορφή που απέκτησε η δυτική μουσική, οι συνθέτες της απέκτησαν μια ευκολία χειρισμού και δημιουργίας λόγω της αλγοριθμικής προσέγγισης της μουσικής και τον εν μέρει προκαθορισμό της πορείας που ακολουθούσε ένα κομμάτι, γεγονός που επέτρεψε την εισαγωγή του ηλεκτρονικού υπολογιστή στην διαδικασία σύλληψης ενός μουσικού κομματιού.

1.4.2 Στοχαστική Μουσική

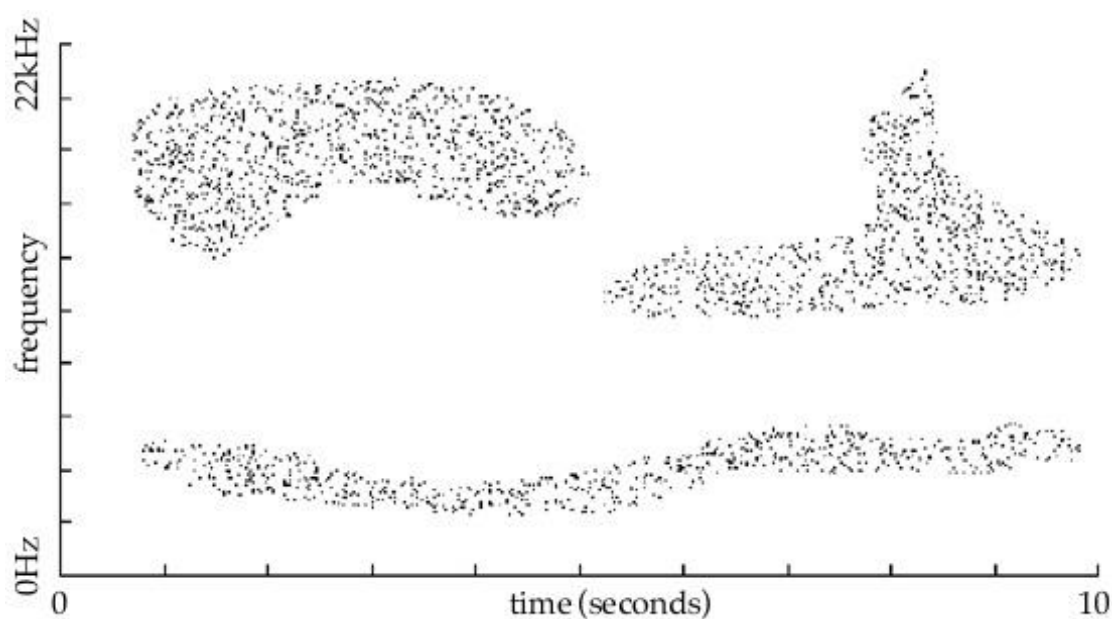
Μετά τον μεσοπόλεμο, αρκετοί νέοι συνθέτες, νιώθοντας ότι ο δωδεκαφθογγισμός είχε οδηγήσει σε ένα μουσικό αδιέξοδο λόγω των επιβλητικών θεμελιωδών αρχών της σειράς και του τονικού κέντρου, προσπάθησαν να ξεφύγουν από αυτόν. Ένας εκ των βασικών επικριτών υπήρξε ο Ιάννης Ξενάκης, ο οποίος στην έκθεση του με τίτλο 'Η κρίση της σειριακής μουσικής' αναφέρει χαρακτηριστικά:

«Η γραμμική πολυφωνία καταστρέφεται από την πολυπλοκότητά της. Αυτό που ακούει κάνεις στην πραγματικότητα δεν είναι τίποτα παρά μια μάζα νοτών σε διάφορες εκτάσεις. Η τεράστια αυτή πολυπλοκότητα εμποδίζει το κοινό από το να ακολουθήσει το την σύνδεση των μελωδικών γραμμών και έχει ως μακροσκοπικό αποτέλεσμα μια παράλογη και τυχαία διασπορά των ήχων σε όλη την έκταση του ηχητικού φάσματος. Υπάρχει επομένως μια αντίφαση μεταξύ του πολυφωνικού γραμμικού συστήματος και το ακουστικού αποτελέσματος, το οποίο πρόκειται για την επιφάνεια ή μια μάζα. Αυτή η αντίφαση, εγγενής στην πολυφωνία, θα εξαφανιστεί όταν υπάρξει καθολική ανεξαρτησία των ήχων. Στην πραγματικότητα, όταν γραμμικοί συνδυασμοί και οι πολυφωνικές υπερθέσεις δεν λειτουργούν πλέον, αυτό που θα έχει σημασία θα είναι ο στατιστικός μέσος των απομονωμένων καταστάσεων των μετασχηματισμών του κάθε ηχητικού στοιχείου σε μια δεδομένη στιγμή . . . Επομένως, εισάγεται η έννοια της πιθανότητας, η οποία, παρεμπιπτόντως, συνεπάγεται συνδυαστικός λογισμός σε αυτή τη συγκεκριμένη περίπτωση.» - (Xenakis, 1955)

Επί της ουσίας, ο Ξενάκης πρότεινε την αντικατάσταση της ντετερμινιστικής λογικής της σειριακής μουσικής με την γενικότερη έννοια της πιθανολογίας, η οποία θα μπορούσε να εμπεριέχει την αυστηρή σειριακή προσέγγιση ως μία ιδιαίτερη

περίπτωση.

Η στοχαστική μουσική συνεπώς βασίζεται σε τυχαίες πράξεις μέσα σε χρονικά μεταβλητούς περιορισμούς. Η λογική αυτή οδήγησε τον Ξενάκη στην εισαγωγή της έννοιας της κοκκώδους σύνθεσης (granular synthesis), δηλαδή η μέθοδος σύνθεσης ήχου που εφαρμόζεται στην κλίμακα μικρόηχων¹ του χρόνου. (Roads, 2004) Βασίζεται στην ίδια αρχή με τη δειγματοληψία (sampling). Ωστόσο, τα δείγματα δεν αναπαράγονται συμβατικά, αλλά χωρίζονται σε μικρά κομμάτια περίπου 1 έως 50 ms. Αυτά τα μικρά κομμάτια ονομάζονται κόκκοι. Οι πολλαπλοί κόκκοι συνήθως είναι στοιβαγμένοι το ένα πάνω στο άλλο και μπορούν να παίζουν με διαφορετικές ταχύτητες, φάσεις, ένταση και συχνότητα, μεταξύ άλλων παραμέτρων. Σε χαμηλές ταχύτητες αναπαραγωγής, το αποτέλεσμα είναι ένα είδος ηχοτοπίου (soundscape). Σε υψηλές ταχύτητες, το αποτέλεσμα παίρνει την μορφή νότας ή νοτών ενός νέου ηχοχρώματος. Με την παραμετροποίηση της διάρκειας, του φίλτρου, την χωρική τοποθέτηση και την 'πυκνότητα' των κόκκων, παράγονται πολλοί και διαφορετικοί ήχοι. Ο Ξενάκης δημιούργησε πρώτος κοκκώδης ήχους χρησιμοποιώντας αναλογικές γεννήτριες τόνων και με την συγκόλληση μαγνητικών ταινιών. Τα παραπάνω εμφανίζονται στο κομμάτι του Ξενάκη Analogique A-B (1959).



Εικόνα 5: Οπτικοποίηση της κοκκώδους σύνθεσης. Κάθε κουκίδα αντιπροσωπεύει έναν κόκκο (μικρόηχος) σε μια συγκεκριμένη συχνότητα και χρονική στιγμή.

¹ Οι μικρόηχοι περιλαμβάνουν όλους τους ήχους της χρονικής κλίμακας μικρότερες από τις μουσικές νότες, τους ήχους της κλίμακας χρόνου ηχοαντικειμένων και τους αυτούς με αξία μεγαλύτερη από την κλίμακα χρόνου δειγματοληψίας. Συγκεκριμένα, είναι μικρότερης αξίας από το ένα δέκατο του δευτερολέπτου και μεγαλύτερης αξίας από τα δέκα χιλιοστά του δευτερολέπτου, συμπεριλαμβανομένης της περιοχής ακουστικών συχνοτήτων (20 Hz έως 20 kHz) και της περιοχής φασματικών συχνοτήτων (κάτω από 20 Hz).

Το 1962, ενώ μέχρι τότε ο Ξενάκης συνέθετε μουσική με την χρήση αλγορίθμων χωρίς όμως τη βοήθεια ηλεκτρονικών υπολογιστών, εφάρμοσε την πιθανοτική λογική της στοχαστικής μουσικής στις συνθέσεις του μέσω της γλώσσας FORTRAN και συνεπώς, ενός ηλεκτρονικού υπολογιστή (IBM-7090). (Nick Collins, Julio d'Escrivan, 2007) Έτσι συνέθεσε την σειρά έργων ονόματι *ST-xxx* για κουαρτέτο εγχόρδων αλλά και ορχηστρικά, όπου το 'ST' προκύπτει από την λέξη *stochastic*. Στην προκειμένη ο Ξενάκης σχεδίασε πρώτα διαγράμματα ροής για τις διάφορες υπορουτίνες οι οποίες στη συνέχεια μεταφράστηκαν σε γλώσσα μηχανής. Το αποτέλεσμα εκτυπώθηκε ως μία λίστα η οποία στην συνέχεια έπρεπε να μεταγλωττιστεί σε μουσική σημειογραφία.

1.4.3 Αλεατορική Μουσική

Αλεατορική μουσική (από το λατινικό *alea*²) ονομάζεται η μουσική στην οποία κάποια στοιχεία της σύνθεσης έχουν αφηθεί στην τύχη ή στην κρίση του ερμηνευτή την εκάστοτε στιγμή. Ουσιαστικά, ο συνθέτης δεν έχει τον πλήρη έλεγχο της σύλληψης του κομματιού και δεν προκαθορίζει την μορφή που θα αποκτήσει. Συνεπώς στοιχεία όπως ο ρυθμός, το τονικό εύρος, η αξία του φθόγγου κτλ., αλλάζουν διαρκώς σχετικά με την κρίση του ερμηνευτή ή κάποιον παράγοντα της τύχης όπως την ρίψη ζαριών ή την παραγωγή τυχαίων αριθμών από έναν υπολογιστή.

Μεταξύ των αξιοσημείωτων αλεατορικών έργων είναι το *Music of Changes* (1951) γραμμένο για πιάνο και το *Concert for Piano and Orchestra* (1958) από τον Αμερικανό συνθέτη John Cage και το *Klavierstück XI* (1956, *Keyboard Piece XI*), από τον Karlheinz Stockhausen της Γερμανίας. Το τελευταίο έργο αποτελείται από 19 σημεία σε μια μεγάλη ενιαία σελίδα. Ο ερμηνευτής μπορεί να ξεκινήσει από οποιοδήποτε σημείο και να συνεχίσει σε οποιοδήποτε άλλο, προχωρώντας με τέτοιο τρόπο ώστε όταν φτάσει σε ένα σημείο για τρίτη φορά, τελειώνει η ερμηνεία. Οι ενδείξεις για το ρυθμό, τη δυναμική κ.λπ. στο τέλος κάθε σημείου εφαρμόζονται στο αμέσως επόμενο σημείο που θα επιλέξει ο ερμηνευτής. Αν και αποτελείται από μία περίπλοκη **σειριακή** δομή, οι τονικότητες δεν έχουν καμία σχέση με την δωδεκάφθογγη μέθοδο αλλά αντλούνται από τις αναλογίες των ρυθμικών συνθέσεων που προκύπτουν από τις προηγουμένως παιγμένες ενότητες.

² Η λέξη *alea* σημαίνει 'ζάρια', γεγονός που οδήγησε στην ονομασία *aleatoric music*, δηλαδή η μουσική της τύχης/ζαριών.



Εικόνα 6: Το έργο του Karlheinz Stockhausen, *Klavierstück XI* που αποτελείται από 19 αυτοτελή μέρη.

Αντίστοιχα, το 1963 ο Γερμανός συνθέτης Gottfried Michael Koenig άρχισε να συνθέτει ένα κομμάτι βασισμένο σε αλγοριθμικό μοντέλο το οποίο υλοποίησε στην συνέχεια ως πρόγραμμα για ηλεκτρονικό υπολογιστή, το επονομαζόμενο Projekt I (PR1). (Nick Collins, Julio d'Escrivan, 2007) Χρησιμοποίησε λίστες με τυχαίες τιμές παραμέτρων και στοχαστικές λειτουργίες, οι οποίες καθόριζαν τις τιμές για τον τόνο, το ρυθμό, τη δυναμική και το ηχόχρωμα. Οι τυχαίες διεργασίες ελέγχονταν από έναν ομαδικό μηχανισμό που καθόριζε τα διαφορά αποτελέσματα συναρτήσεως του χρόνου.

1.4.4 Σύγχρονη Αλγοριθμική Σύνθεση

Στο Ινστιτούτο Ήχου της Ουτρέχτης, με επικεφαλής τον Koenig (1964-1986), δημιουργήθηκε ένα δομοστοιχειωτό αναλογικό στούντιο βασισμένο στον έλεγχο της τάσης. Αποτελούταν από ανεξάρτητες μονάδες (όπως ταλαντωτές, φίλτρα, γεννήτριες, λογικά κυκλώματα κλπ.) που συνδεόταν μεταξύ τους και να ελέγχονταν παραμετρικά από την τάση. Αυτό θεωρείται μία από τις πρώτες εφαρμογές παραγωγής ήχου σε πραγματικό χρόνο χρησιμοποιώντας αλγοριθμικά δομημένο hardware. Ο πυρήνας αυτού του συστήματος ονομαζόταν «γεννήτρια μεταβλητών λειτουργιών» (Nick Collins, Julio d'Escrivan, 2007), ο οποίος ήταν ουσιαστικά ένας προγραμματιζόμενος ανιχνευτής ακολουθιακών ελέγχων (sequencer) στον οποίο αποθηκευόταν μια μεταβλητή συνάρτηση της τάσης που διαβαζόταν συναρτήσεως μιας μεταβλητής ταχύτητας και κατά συνέπεια προκίπτανε ποίκιλα αποτελέσματα στην πράξη. Η βασική διαφορά με προηγουμένως είναι πως η παραπάνω τεχνολογία επέτρεψε στον Koenig να συνθέτει χωρίς το 'μοντάζ' της μαγνητικής ταινίας του ηχογραφημένου αποτελέσματος, αλλά σε πραγματικό χρόνο και με την

χρήση ηλεκτρικού ρεύματος. Η εφαρμογή αυτή σε επίπεδο τεχνολογικών μέσων οδήγησε στην ψηφιακή υλοποίηση μουσικών ακολουθιακών ελέγχων και σύνθεσης σε πραγματικό χρόνο όπως η δημιουργία της αντικειμενοστραφής γλώσσας προγραμματισμού Max-Max/MSP που επιτρέπει στον χρήστη του Η/Υ να “χτίζει” την διαρρύθμιση του hardware ψηφιακά και να παράγει αλγοριθμική μουσική σε πραγματικό χρόνο, σε ένα φιλικό για τον ίδιο περιβάλλον.

1.5 ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ

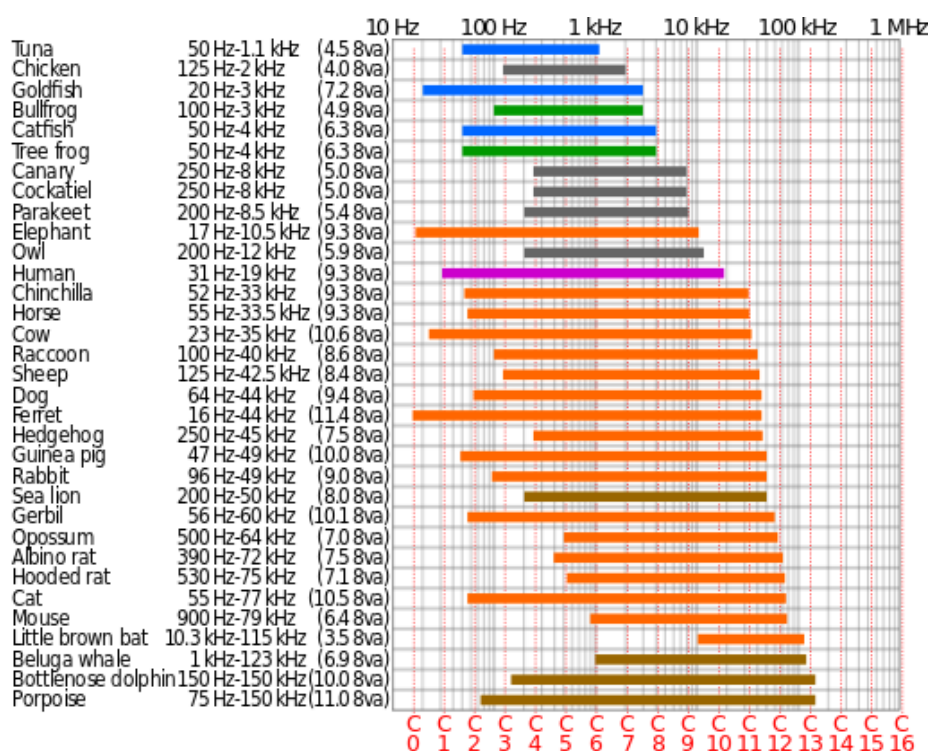
Η συγκεκριμένη διπλωματική εργασία αφορά την μετατροπή εικόνας σε ήχο, σε ένα μουσικό πλαίσιο. Η διαδικασία αυτή γίνεται μέσω της επεξεργασίας συγκεκριμένων παραμέτρων ενός βίντεο. Κατά την είσοδο του βίντεο στο σύστημα, ο υπολογιστής αντιστοιχεί την κάθε παράμετρο σε κάποιο θεμελιώδες μουσικό στοιχείο και στην συνέχεια, το βίντεο αποδίδεται μουσικά με αναλογία κίνησης σε ρυθμό όπως και χρωμάτων σε μελωδία. Για την επίτευξη του σκοπού αυτού, θα γίνει χρήση ενός περιβάλλοντος σε Python με την χρήση βιβλιοθηκών της OpenCV για την ανάλυση και δειγματοληψία της εισόδου και την χρήση βιβλιοθηκών της Mingus για την αντιστοίχιση σε επίπεδο αρμονίας και ρυθμού και την παραγωγή μουσικών αρχείων MIDI (έξοδος) ούτως ώστε να εκφραστεί το ακουστικό αποτέλεσμα. Τα αρχεία MIDI τροφοδοτούνται σε κάποιο πρόγραμμα επεξεργασίας του ήχου (Ableton, Cubase, Reaper κ.α.) ώστε να καθοριστεί το επιθυμητό ηχόχρωμα του «καναλιού» που προέκυψε και στην συνέχεια να αναπαραχθεί συναρτήσει του βίντεο.

2. ΗΧΟΣ & ΑΚΟΗ

Η ακοή του μέσου ανθρώπου τον φέρνει σε επαφή καθημερινά με το περιβάλλον του, καθορίζει την συμπεριφορά και τις αντιδράσεις του σε τέτοιο βαθμό, ώστε να θεμελιώνει την αντίληψη του, να αναγνωρίζει και να επιδρά στο περιβάλλον βάσει αυτού. Έτσι γίνεται βέβαιο πως όλη η ύλη χαρακτηρίζεται από τον εκάστοτε ήχο που παράγει και αποκτάει μια συνοχή αφού γίνει αντιληπτός ο ήχος αυτός.

2.1 ΠΩΣ ΑΝΤΙΛΑΜΒΑΝΕΤΑΙ Ο ΑΝΘΡΩΠΟΣ ΤΟΝ ΗΧΟ

Η ανθρώπινη ακοή γενικά εκτείνεται για ήχους με συχνότητα μεταξύ 20 Hz και 20.000 Hz με μία σχετική απόκλιση από άτομο σε άτομο. Παρατηρείται επίσης μία μείωση της αντίληψης υψηλών συχνοτήτων σε ανθρώπους με μεγάλη ηλικία. Αυτό το εύρος διαφέρει σε άλλους οργανισμούς που έχουν ενδεχομένως, καλύτερη ακοή από τον άνθρωπο, όπως για παράδειγμα ο σκύλος (40 Hz με 60.000 Hz) και η γάτα (55 Hz με 77.000 Hz).

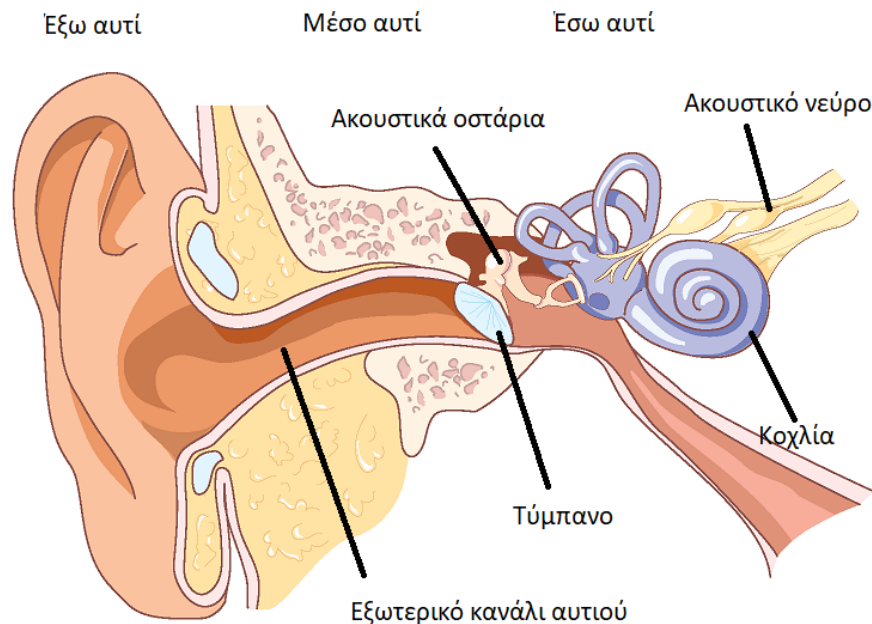


Εικόνα 7: Λογαριθμικός πίνακας συχνοτικού εύρους ακοής ζώων.

Ο ήχος είναι η αίσθηση που προκαλείται λόγω της διέγερσης των αισθητηρίων οργάνων της ακοής από μεταβολές πίεσης του ατμοσφαιρικού αέρα ή κάποιο άλλο

μέσο διάδοσης (π.χ. νερό). Οι μεταβολές αυτές διαδίδονται με τη μορφή ηχητικών κυμάτων, τα οποία δεν είναι ορατά με γυμνό μάτι όμως μπορούν να εμφανισθούν στην επιφάνεια ενός υγρού, και να καταγραφούν από ηλεκτρικά και ηλεκτρονικά μέσα ηχογράφησης.

Στους ανθρώπους, τα ηχητικά κύματα διοχετεύονται στο αυτί μέσω του εξωτερικού καναλιού του αυτιού και φτάνουν στο τύμπανο (τυμπανική μεμβράνη). Η συμπίεση και η αραιώση αυτών των κυμάτων θέτουν αυτή την λεπτή μεμβράνη σε κίνηση, προκαλώντας έτσι μία συμπαθητική ταλάντωση³ στα ακουστικά οστάρια του μέσου αυτιού, στον κοχλία και στα τριχίδια του έσω αυτιού. Τα τριχίδια αυτά ευθυγραμμίζουν τον κοχλία από τη βάση στην κορυφή και το τμήμα που διεγείρεται όπως και η ένταση της διέγερσης δίνουν μια ένδειξη της φύσης του ήχου. Οι πληροφορίες που συλλέγονται από τα κύτταρα των τριχιδίων αποστέλλονται μέσω του ακουστικού νεύρου για επεξεργασία στον εγκέφαλο. (Seashore, 1967)



Εικόνα 8: Ανατομία του ανθρώπινου αυτιού.

³ Ταλάντωση ιδιοσυχνότητας, συγχρονισμού.

2.2 Η ΦΥΣΙΚΗ ΣΗΜΑΣΙΑ ΤΟΥ ΗΧΟΥ

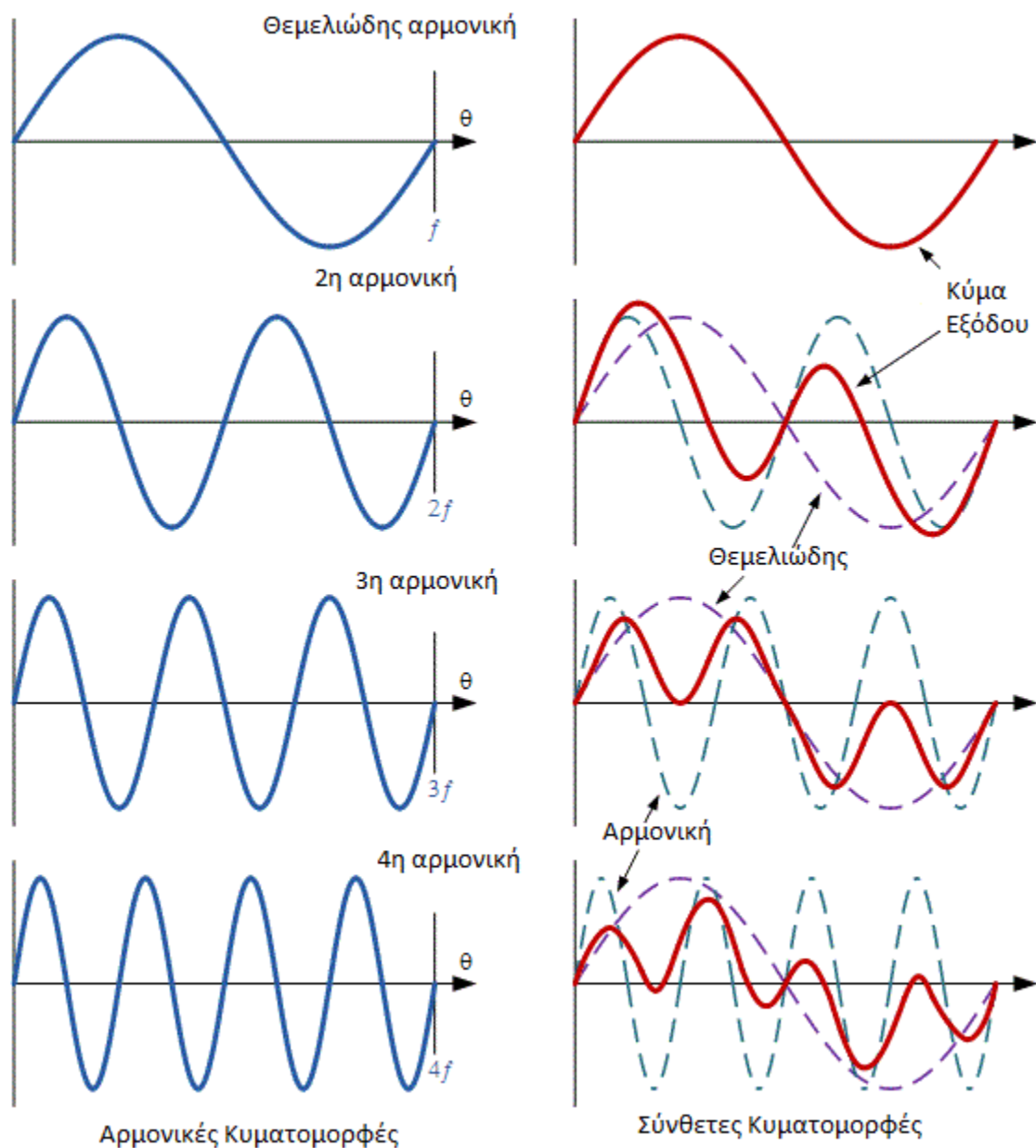
Όπως αναφέρθηκε προηγουμένως, ο ήχος οφείλεται στις μεταβολές πίεσης σε κάποιο μέσο διάδοσης και έχει την μορφή του ηχητικού κύματος. Τα ηχητικά κύματα παράγονται από σώματα που εκτελούν μηχανικές ταλαντώσεις (δονήσεις), και επομένως χαρακτηρίζονται ως μηχανικά κύματα (ελαστικότητας) που μεταφέρουν μηχανική ενέργεια. Ουσιαστικά μία πηγή ήχου δημιουργεί τις δονήσεις στο μέσο (π.χ. ατμοσφαιρικός αέρας) οι οποίες διαδίδονται μακριά από την πηγή με την ταχύτητα του ήχου (σε ατμοσφαιρικό ξηρό αέρα στους 20 °C είναι 343 μέτρα/δευτερόλεπτο), σχηματίζοντας έτσι το ηχητικό κύμα. Στιγμιαία τα μετατοπισμένα μόρια ύλης (του μέσου) ασκούν δυνάμεις στα γειτονικά μόρια, αναγκάζοντάς τα να έλθουν εκτός θέσης ισορροπίας (ταλάντωση) και κατά συνέπεια παρατηρούνται διαδοχικά πυκνώματα και αραιώματα στο μέσο (μεταβολή της πίεσης).

Το ηχητικό κύμα, όπως και οι περισσότερες μορφές κυμάτων, χαρακτηρίζονται από την συχνότητα, την περίοδο, το μήκος κύματος, το πλάτος της ταλάντωσης, την διάδοση στον χρόνο και την μορφή του κύματος. Από αυτές τις ιδιότητες πηγάζουν τέσσερα χαρακτηριστικά που αποσκοπούν στην περιγραφή ενός ήχου από μουσικό ακουστικής προσέγγισης και είναι τα εξής: ύψος, ένταση, διάρκεια και χροιά. Για παράδειγμα, ένα απλό ημιτονοειδές αρμονικό κύμα με περίοδο $T = 2 \text{ ms}$, έχει προφανώς συχνότητα $f = \frac{1}{T} = 500 \text{ Hz}$, η οποία αντιστοιχεί περίπου και στο τονικό ύψος της νότας Σι (B4 : 493.88 Hz). Εάν το μέσο στο οποίο διαδίδεται το κύμα είναι γνωστό, τότε και η ταχύτητα διάδοσης του κύματος είναι επίσης γνωστή (έστω ταχύτητα του ήχου στον ατμοσφαιρικό αέρα), επομένως μπορεί να βρεθεί και το μήκος του κύματος που μελετάται:

$$f = \frac{c}{\lambda} \leftrightarrow \lambda = \frac{c}{f} = \frac{343 \frac{\text{m}}{\text{s}}}{500 \text{ Hz}} = 0,686 \text{ m}.$$

Παρατηρείται λοιπόν μια αντιστρόφως ανάλογη σχέση συχνότητας και μήκους κύματος (π.χ. μήκος χορδής μιας κιθάρας) και κατά συνέπεια τονικού ύψους. Η μέση ένταση του ήχου είναι ανάλογη της πίεσης ή της ενέργειας του κύματος $I = \frac{1}{T} \int_0^T p(t)v(t)dt$ όπου I η ένταση, T ο χρόνος, $p(t)$ η συνάρτηση της πίεσης στον χρόνο και $v(t)$ η συνάρτηση της ταχύτητας των μορίων του μέσου στον χρόνο. Η ένταση εκφράζεται επίσης συναρτήσει του μέγιστου πλάτους ταλάντωσης του κύματος $I = 2\pi^2 f^2 A^2 \rho v$, όπου A το πλάτος και ρ η πυκνότητα του μέσου. Η διάρκεια του ήχου προφανώς εκφράζεται σε σχέση με τον χρόνο διάδοσης του κύματος στο συγκεκριμένο μέσο. Τέλος, η μορφή του κύματος (π.χ. τριγωνικό, τετράγωνο, ημιτονοειδές, κ.λπ.) εκφράζει την χροιά (ηχόχρωμα) του ήχου, πιο συγκεκριμένα το σύνολο των αρμονικών ταλαντώσεων (το σύνολο των διαφόρων αρμονικών συχνοτήτων) που συμβάλουν στην συνολική κυματομορφή καθορίζει το πως θα ακουστεί αυτός ο ήχος.

Όταν παράγεται ένας ήχος από κάποιο μέσο σε μία συγκεκριμένη συχνότητα, ο ανθρώπινος εγκέφαλος αντιλαμβάνεται κυρίως αυτή τη συχνότητα χωρίς να αντιλαμβάνεται και τις επιμέρους συχνότητες από τις οποίες αποτελείται. Οι επιμέρους συχνότητες ονομάζονται αρμονικές ενώ η αντιληπτή συχνότητα ονομάζεται θεμελιώδης. Οι αρμονικές μιας συχνότητας είναι ακέραια πολλαπλάσιά της. Δηλαδή η συχνότητα των 100 Hz έχει σαν δεύτερη αρμονική τα 200 Hz, σαν τρίτη τα 300 Hz, σαν τέταρτη τα 400 Hz και ούτω καθ' εξής. Τον διπλασιασμό μιας συχνότητας, το αυτί μας τον αντιλαμβάνεται ως αύξηση μιας οκτάβας στην διατονική μουσική κλίμακα όπως και την αύξηση κατά το ήμισυ το αντιλαμβάνεται ως διάστημα πέμπτης. Αντίστοιχα προκύπτουν και τα υπόλοιπα διατονικά διαστήματα.



Εικόνα 9: Απεικόνιση των αρμονικών συστατικών του σύνθετου περιοδικού κύματος ήχου.

Όταν η πηγή δεν παράγει περιοδικό ηχητικό κύμα το παραγόμενο αποτέλεσμα συνήθως ορίζεται ως «ηχητικός θόρυβος». Σε αυτήν την περίπτωση ο ήχος δεν εκλαμβάνεται αρμονικά και συμπαγής, δεν είναι ξεκάθαρο το αρμονικό κέντρο ή η θεμελιώδης συχνότητα, έχει περιπλοκή κυματομορφή με πολλές αρμονικές συνιστώσες οι οποίες δεν γίνονται ξεκάθαρες χωρίς κάποια καταγραφή και ανάλυση. Πολλές φορές με τον όρο «ηχητικό θόρυβος» εννοείται και ο «υπόηχος» ο οποίος μπορεί να μην ακούγεται αλλά να επηρεάζει ενδεχομένως τις σωματικές κοιλότητες ενός ανθρώπου και να προκαλεί κάποιου είδους δυσφορία.

2.3 Η ΜΟΥΣΙΚΗ ΩΣ ΗΧΟΣ

Η μουσική πρόκειται για ένα σύνολο ήχων από το οποίο αποτελείται ένα μουσικό κομμάτι. Συνήθως οι ήχοι αυτοί είναι αποτέλεσμα περιοδικών ηχητικών κυμάτων παραγόμενα από μουσικά όργανα. Τα μουσικά όργανα κατηγοριοποιούνται με διάφορους τρόπους, όπως ανάλογα με τη μέθοδο παραγωγής ήχου (χορδόφωνα, μεμβρανόφωνα, αερόφωνα), είτε με τον τρόπο παιξίματος (κρουστά, πνευστά, νυκτά) είτε βάσει του υλικού κατασκευής (ξύλινα, χάλκινα κλπ.). Υπάρχουν πολλές μέθοδοι κατηγοριοποίησης, με πιο διαδεδομένη αυτή των Χόρνμποστελ και Σακς (Erich Moritz von Hornbostel, Curt Sachs, 1961). Κριτήρια της κατηγοριοποίησης των οργάνων με αυτή την μέθοδο, είναι, πρωτίστως, ο τρόπος παραγωγής του ήχου και, δευτερευόντως, ο τρόπος παιξίματος και η κατασκευή του οργάνου.

Στα ιδιόφωνα (αυτόφωνα) (ομάδα κρουστών) ο ήχος παράγεται μέσω της δόνησης του ίδιου του σώματός τους (π.χ. μεταλλόφωνο), τα μεμβρανόφωνα (ομάδα κρουστών) φέρουν στρογγυλές μεμβράνες προς κρούση (π.χ. τύμπανο), στα χορδόφωνα ο ήχος δημιουργείται θέτοντας τεντωμένες χορδές σε ταλάντωση (π.χ. κιθάρα, βιολί), στα αερόφωνα ο ήχος δημιουργείται από αέρα που ταξιδεύει μέσα σε όργανο με σωληνοειδές σχήμα (π.χ. κλαρινέτο) και τα ηλεκτρόφωνα παράγουν ήχο μέσω ηλεκτρονικού κυκλώματος (π.χ. synthesizer, Hammond)

2.3.1 Ηχώχρωμα

Αυτό που ξεχωρίζει το κάθε όργανο ηχητικά είναι η χροιά του ή το ηχώχρωμα του. Σύμφωνα με τον J. F. Schouten η χροιά αναλύεται στα εξής (Erickson, 1975):

- Το εύρος μεταξύ αρμονικού και θορυβώδους χαρακτήρα του οργάνου
- Ο φασματικός φάκελος (φασματικό πλαίσιο συχνοτήτων)
- Το εύρος του χρόνου όσον αφορά την άνοδο, τη διάρκεια και την κάθοδο/decay (ADSR—attack, decay, sustain, release)
- Οι μεταβολές τόσο του φασματικού περιβλήματος (φασματικός φάκελος) όσο και της θεμελιώδους συχνότητας (μικροτονικά)

- Το πρόθεμα ή η εμφάνιση ενός ήχου, αρκετά ανόμοια με την επακόλουθη διαρκή δόνηση του ήχου

Σύμφωνα με τον Erickson η χροιά αναλύεται ακόμα καλύτερα σε δύο τομείς (υποκειμενικά, αντικειμενικά) με βάση την αντίληψη και την φυσική σημασία της κάθε παραμέτρου (Erickson, 1975):

Υποκειμενικά	Αντικειμενικά
Τονικός χαρακτήρας, συνήθως 'κουρδισμένος'	Περιοδικός ήχος
Θορυβώδης, με ή χωρίς κάποιο τονικό χαρακτήρα, συμπεριλαμβανομένου του θορύβου ενός κρότου.	Θόρυβος, συμπεριλαμβανομένου απεριοδικών παλμών ή τυχαίων παλμών με απεριοδικό χαρακτήρα
Χρωματισμός	Φασματικός φάκελος
Έναρξη/λήξη	Φυσικός χρόνος ανόδου/καθόδου
Χρωματική ολίσθηση/ Σχηματική ολίσθηση	Αλλαγή φασματικού φακέλου
Μικροτονικά	Μικρή συχνοτική αλλαγή (πάνω ή κάτω)
Βιμπράτο (Vibrato)	Διαμόρφωση συχνότητας
Τρέμολο (Tremolo)	Διαμόρφωση πλάτους
Ατάκα (Attack)	Πρόθεμα
Τελικός ήχος (τελείωμα)	Κατάληξη

2.3.2 Τονικό ύψος

Όπως προαναφέρθηκε, το τονικό ύψος αφορά την ιδιότητα του ήχου να κατατάσσεται σε μια κλίμακα συχνότητας, μπορεί να ποσοτικοποιηθεί ως συχνότητα δηλαδή, χωρίς όμως να λαμβάνει την καθαρά φυσική ιδιότητα της συχνότητας. Ουσιαστικά πρόκειται για την αξία που έχει ο ήχος να χαρακτηρίζεται ψηλός, χαμηλός ή μεσαίος (πρίμα, μπάσα, ή μεσαία) και διακρίνεται κυρίως σε μουσικούς ήχους και όχι σε θορυβώδεις ήχους. Η έννοια χρησιμοποιείται για να μπορέσει ένας ακροατής να αποδώσει μουσικούς τόνους στις σχετικές θέσεις μιας μουσικής κλίμακας διότι αφορά την υποκειμενική αντίληψη του ακροατή για το ηχητικό κύμα και όχι την μέτρηση ή την καταγραφή του.

Σε μία χορδή, το τονικό ύψος εξαρτάται από τρεις παράγοντες, α) το πάχος της χορδής, β) το μήκος της χορδής και γ) την τάση της χορδής. Η τάση της χορδής καθορίζεται από το πόση δύναμη έχει εφαρμοστεί κατά το χόρδισμά της (κούρδισμα) μέσω των κλειδιών. Χορδές λοιπόν, κατασκευασμένες από το ίδιο υλικό, έχοντας το ίδιο μήκος και το ίδιο πάχος, παράγουν διαφορετικά μουσικά ύψη (μουσικές συχνότητες), ανάλογα με το πόσο τεντωμένες είναι, διότι περιγράφονται

ως στασιμά κύματα όπου αλλάζει το μήκος του κύματος και κατά συνέπεια η συχνότητα. Τα διαφορετικά αυτά ύψη ονομάζονται στην μουσική ορολογία φθόγγοι ή τόνοι. Στην ευρωπαϊκή μουσική (δυτική) η διατονική κλίμακα φθόγγων χωρίζεται σε τόνους και ημιτόνια έτσι ώστε να χωρίζεται σε επτά φθόγγους (π.χ. ντο, ρε, μι, φα, σολ, λα, Σι) ενώ η χρωματική κλίμακα χωρίζεται σε δώδεκα φθόγγους (π.χ. ντο, ντο δίεση, ρε, ρε δίεση, μι, φα, φα δίεση, σολ, σολ δίεση, λα, λα δίεση, Σι) με διαχωρισμό ημιτονίων του ενός φθόγγου από τον επόμενο. Ο συνδυασμός δύο τόνων στη σειρά ονομάζεται μείζων διάστημα τρίτης (χαρούμενος ήχος, φωτεινός) και ο συνδυασμός ενός τόνου με ημιτόνιο ονομάζεται ελάσσων διάστημα τρίτης (σκοτεινός ήχος). Αντίστοιχα προκύπτουν και τα υπόλοιπα αρμονικά διαστήματα.

2.3.3 Ακουστότητα – Ένταση του Ήχου

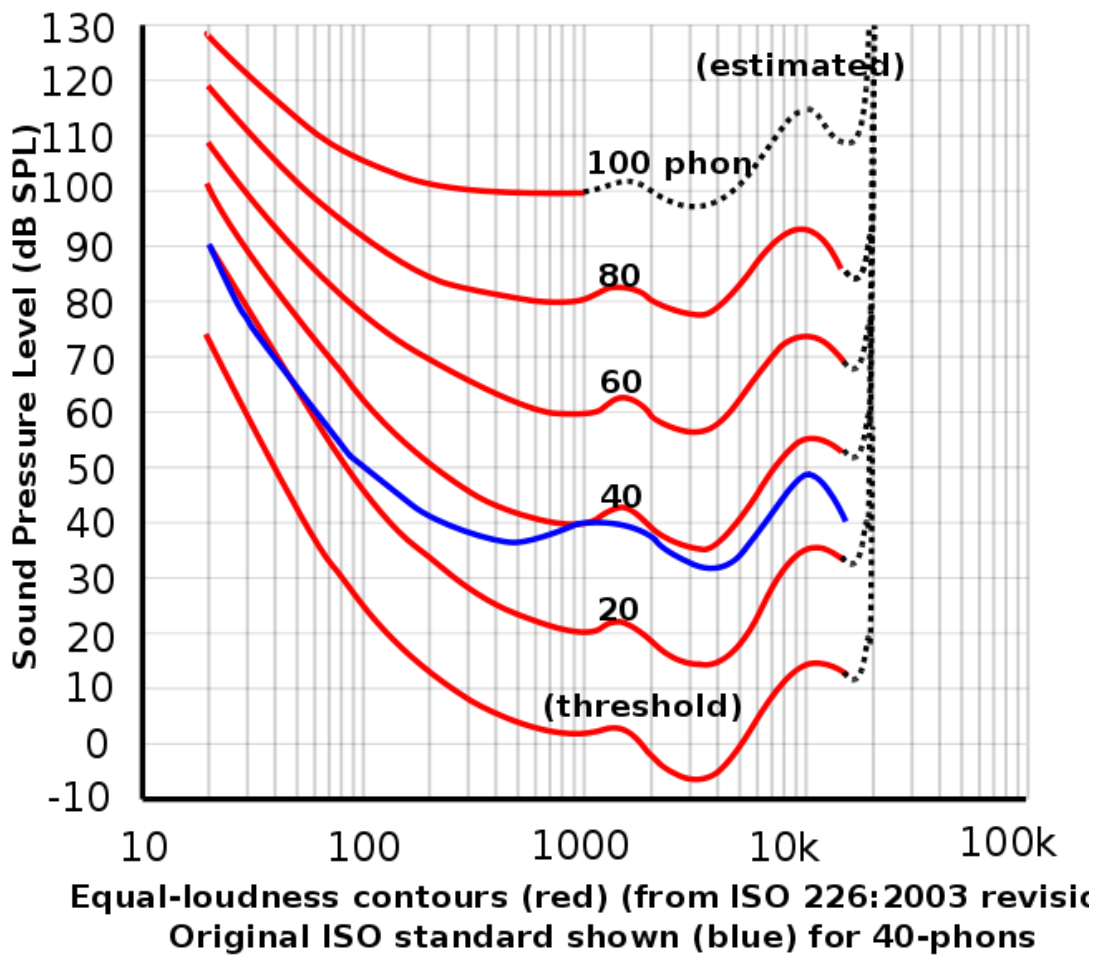
Το πλάτος των ηχητικών κυμάτων συμβατικά αναφέρεται στις μεταβολές της πίεσης του αέρα στο κύμα αλλά κάποιες φορές περιγράφει και τις μετατοπίσεις στο μέσο μετάδοσης του κύματος (κινήσεις του αέρα ή του διαφράγματος ενός ηχείου). Η ακουστότητα του ήχου αναφέρεται στην περιγραφή του ήχου έτσι όπως τον αντιλαμβάνεται το ανθρώπινο αυτί, πιο συγκεκριμένα εάν ένας ήχος περιγράφεται ως δυνατός, ασθενής, απαλός κλπ. Πολλές φορές η ακουστότητα συγγέεται με την ένταση του ήχου η οποία αφορά την ισχύς του ηχητικού κύματος ανά μονάδα επιφανείας ή αλλιώς η ενέργεια που μεταφέρει το ηχητικό κύμα ανά μονάδα επιφανείας και ανά μονάδα χρόνου. Όπως αναφέρθηκε προηγουμένως η ένταση περιγράφεται ως εξής:

$$I = \frac{1}{T} \int_0^T p(t)v(t)dt$$

(όπου p η ακουστική/ηχητική πίεση και v η εγκάρσια ταχύτητα των μορίων του αέρα)

Από την άλλη η ακουστότητα εξαρτάται κυρίως από την ηχητική πίεση και κατά επέκταση από την συχνότητα και την μορφή του κύματος. Η ευαισθησία του ανθρώπινου αυτιού αλλάζει ως συνάρτηση της συχνότητας, όπως φαίνεται στο γράφημα ίσης ακουστότητας παρακάτω. Κάθε γραμμή στο γράφημα δείχνει το μέγεθος της ηχητικής πίεσης (sound pressure level – SPL) που απαιτείται για να γίνουν αντιληπτές οι διάφορες συχνότητες το ίδιο δυνατά. Η κάθε καμπύλη αφορά διαφορετικά επίπεδα ηχητικής πίεσης. Δείχνει επίσης ότι οι άνθρωποι με κανονική ακοή είναι πιο ευαίσθητοι σε ήχους γύρω στα 2-4 kHz, με την ευαισθησία να μειώνεται δεξιά και αριστερά της περιοχής εκείνης. (Olson, 1972)

Έτσι γίνεται αντιληπτός ο λόγος που κάποιες χαμηλές συχνότητες τις ακούμε για παράδειγμα πιο δυνατά από κάποιες μεσαίες, ενώ η ένταση στην πραγματικότητα είναι ίδια.



Εικόνα 10: Καμπύλες ίσης ακουστότητας με τον οριζόντιο άξονα να εκφράζει την συχνότητα και τον κάθετο άξονα να εκφράζει την ηχητική πίεση σε dB.

2.3.4 Δυναμικές

Στην μουσική, με τον όρο δυναμική εννοείται η παραλλαγή στην ακουστότητα μεταξύ νοτών ή φράσεων. Ο σκοπός των δυναμικών είναι να τονίσουν ή να δώσουν έμφαση σε κάποιο λεπτομερές σημείο του μουσικού κομματιού και στην μουσική σημειογραφία περιγράφονται από συγκεκριμένες σημάνσεις που υποδεικνύουν τον τρόπο αυτόν με τον οποίο θα παίξει ο ερμηνευτής. Για παράδειγμα το *p* ή *piano* υποδεικνύει απαλό παίξιμο και όχι δυνατό ενώ το *f* ή *forte* υποδεικνύει δυνατό ή ισχυρό παίξιμο.

2.3.5 Διάρκεια – Τέμπο – Ρυθμός

Ρυθμική αξία ή διάρκεια φθόγγων είναι ο συμβολισμός και το χρονικό διάστημα που χρησιμοποιείται στη μουσική για να εκφραστεί η διάρκεια του ήχου συναρτήσει του μέτρου. Όλες οι ρυθμικές αξίες είναι πολλαπλάσια και διαιρέτες του δυο και η μεγαλύτερη από αυτές είναι το «ολόκληρο» το οποίο διαρκεί τέσσερις χρόνους (4/4). Ουσιαστικά η διάρκεια του ήχου ή του φθογγόσημου διαρκεί τέσερα τέταρτα ή αλλιώς τέσσερις χρόνους. Το επόμενο πολλαπλάσιο είναι το «μισό» στο οποίο μετράτε ο χρόνος δυο τετάρτων μέχρι την λήξη του ήχου. Το «τέταρτο» ουσιαστικά είναι ο χρόνος αναφορά που χρησιμοποιείται και στις προηγούμενες περιπτώσεις. Το «όγδοο» διαρκεί το μισό του χρόνου (πρώτη υποδιαίρεση τέταρτου), άρα αυτό σημαίνει ότι σε έναν χρόνο παίζεται δυο φορές κάποιος φθόγγος (δηλαδή δύο όγδοα). Το «δέκατο-έκτο» διαρκεί το 0.25 του χρόνου (δεύτερη υποδιαίρεση τέταρτου). Υπάρχουν και δύο ακόμα μικρότερες ρυθμικές αξίες, τα «τριακοστά-δεύτερα» και τα «εξηκοστά-τέταρτα» (τρίτη και τέταρτη υποδιαίρεση τέταρτου αντίστοιχα).

Η έννοια του τέμπο αποτελεί την ένδειξη η οποία δηλώνει τον ρυθμικό χαρακτήρα με τον οποίο πρέπει να εκτελεστεί ένα μουσικό έργο. Στην μουσική σημειογραφία το τέμπο ενός έργου συνήθως αναγράφεται στην αρχή του, διπλά από το πρώτο μετρό. Αυτό αλλάζει εάν και εφόσον σε κάποιο μέτρο θελήσει να αποδώσει ο συνθέτης πιο γρηγορά ή πιο αργά το σημείο του έργου, οπότε και αναγράφεται διπλά από το μετρό αυτό. Η σωστή απόδοση του τέμπο είναι υψίστης σημασίας για να αποκτήσει το έργο τον χαρακτήρα και το ύφος που επιθυμεί ο συνθέτης. Το τέμπο, όσον αφορά την ταχύτητα, υπολογίζεται μαθηματικά σε χτύπους ανά λεπτό (bpm). Σαν συμβολισμός αναγράφεται με πόσους χτύπους ανά λεπτό ισούται η βασική ρυθμική αξία στην οποία είναι γραμμένο το μουσικό κείμενο. Αυτή η ένδειξη ονομάζεται μετρονομική. Για παράδειγμα, αν το κείμενο έχει μετρική ένδειξη 4/4, όπου ρυθμική μονάδα είναι το τέταρτο, η μετρονομική ένδειξη θα δηλώνει πόσα τέταρτα χωρούν μέσα σε ένα λεπτό (π.χ. ♩=120).

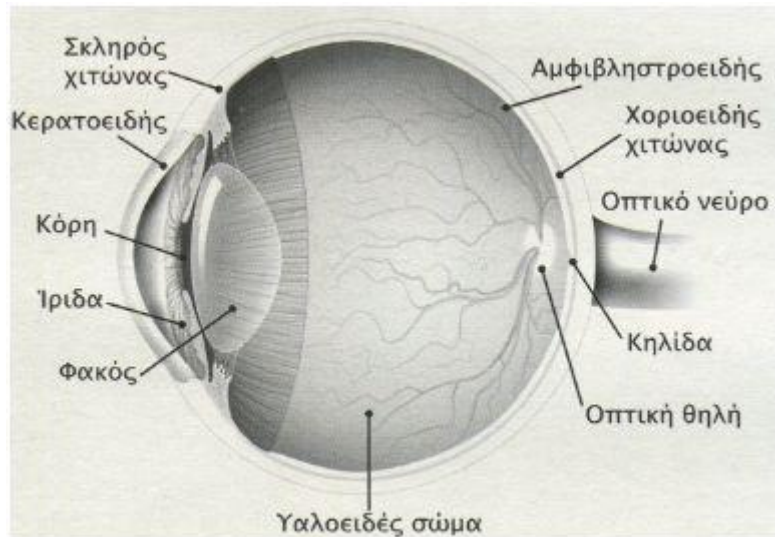
3. ΟΡΑΣΗ & ΟΠΤΙΚΗ ΑΝΤΙΛΗΨΗ

Η οπτική αντίληψη πρόκειται για την ικανότητα που έχει ο άνθρωπος (και τα ζώα) να ερμηνεύει το περιβάλλον του με την χρήση του ορατού φάσματος του φωτός. Το οπτικό σύστημα εκτελεί μια σειρά σύνθετων λειτουργιών, όπως την λήψη φωτός και τον σχηματισμό μονόκλωνων παραστάσεων, την αναγνώριση και την κατηγοριοποίηση οπτικών αντικειμένων, την αξιολόγηση των αποστάσεων προς και μεταξύ των αντικειμένων και την γενικότερη καθοδήγηση των κινήσεων του σώματος σε σχέση με τα αντικείμενα που βρίσκονται στο περιβάλλον.

3.1 Η ΛΕΙΤΟΥΡΓΙΑ ΤΗΣ ΟΡΑΣΗΣ ΣΤΟΝ ΑΝΘΡΩΠΟ

Το φως κάποιας πηγής (ήλιος, λάμπα, φωτιά κλπ.) προσπίπτει σε διάφορα αντικείμενα/εμπόδια και έπειτα ένα μέρος του φτάνει στο μάτι, το οποίο αποτελεί και το βασικότερο όργανο της όρασης. Εκεί, το φως διαθλάται μέσω του φακού του οφθαλμού και του κερατοειδή χιτώνα, ώστε να προβληθεί στον αμφιβληστροειδή χιτώνα η εικόνα. Μεταξύ του φακού και του αμφιβληστροειδούς υπάρχει ένα άχρωμο, παχύρρευστο υλικό που ονομάζεται υαλώδες υγρό, το οποίο απορροφά ορισμένες συχνότητες του εισερχόμενου φωτός. Υπάρχουν δύο τύποι φωτοευαίσθητων κυττάρων στον αμφιβληστροειδή, τα ραβδία και τα κωνία. Κάθε αμφιβληστροειδής περιέχει περίπου 120 εκατομμύρια ραβδία και 8 εκατομμύρια κωνία. Όταν ανταποκρίνονται τα κύτταρα αυτά στη δέσμη φωτός, μεταδίδουν την εικόνα μέσω νευρικών ώσεων (ηλεκτρικών παλμών) στο οπτικό νεύρο και στην συνέχεια μέσω διαφόρων νευρικών δομών στον οπτικό φλοιό του εγκεφάλου.

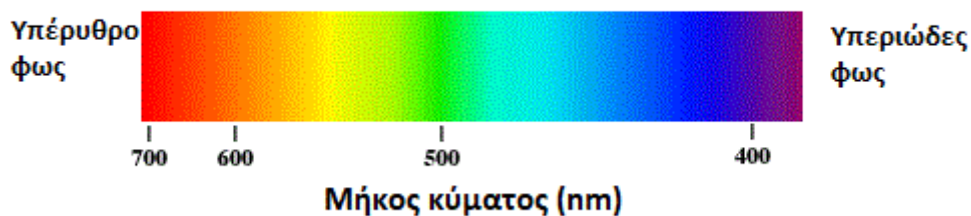
Τα ραβδία είναι συγκεντρωμένα κυρίως στην περιφέρεια του αμφιβληστροειδούς και λειτουργούν συνήθως σε πολύ χαμηλά επίπεδα φωτός. Είναι εξαιρετικά ευαίσθητα στο φως ενώ δεν είναι ευαίσθητα στο χρώμα, παρέχοντας έτσι αχρωματική όραση σε χαμηλά επίπεδα φωτισμού. Αντίστοιχα τα κωνία είναι συγκεντρωμένα κυρίως στο κέντρο του αμφιβληστροειδούς. Είναι λιγότερο ευαίσθητα στο φως από τα ραβδία, αλλά παρέχουν τόσο φωτεινότητα όσο και έγχρωμη όραση με φως της ημέρας.



Εικόνα 11: Ανατομία του οφθαλμού.

3.2 ΤΟ ΧΡΩΜΑ, ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥ ΚΑΙ ΟΙ ΧΡΩΜΑΤΙΚΟΙ ΧΩΡΟΙ

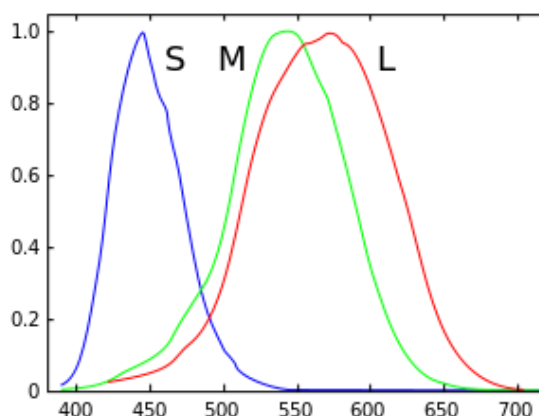
Ως ορατό φάσμα χαρακτηρίζεται το τμήμα εκείνο της ηλεκτρομαγνητικής ακτινοβολίας που μπορεί να αντιληφθεί ο άνθρωπος με το αισθητήριο της όρασης. Το φάσμα αυτό εκτείνεται σε ακτινοβολίες με μήκος κύματος από 390 nm έως 700 nm. Οι συχνότητες με μήκος κύματος κάτω από 390 nm αποτελούν την υπεριώδη ακτινοβολία (Ultra Violet ή UV), ενώ οι συχνότητες με μήκος κύματος πάνω από τα 700 nm αποτελούν την υπέρυθη ακτινοβολία (Infrared).



Εικόνα 12: Το ορατό φάσμα φωτός

Οι ηλεκτρικές ώσεις που αφορούν την πληροφορία του χρώματος προέρχονται από τους υποδοχείς του ματιού, τα κωνία, που αντιδρούν το καθένα στην ανίχνευση φωτός συγκεκριμένου εύρους μήκους κύματος. Ένα χρώμα που παίρνει μορφή στον εγκέφαλο μπορεί να προέρχεται από μία συχνότητα ή συνδυασμό περισσότερων συχνοτήτων του ορατού φάσματος. Για παράδειγμα το κίτρινο χρώμα είναι αποτέλεσμα της επεξεργασίας του σήματος που προέρχεται από την ανίχνευση φωτός από δύο κυρίως είδη κωνίων, τα κωνία που είναι ευαίσθητα σε συχνότητες που αντιστοιχούν στην περιοχή του κόκκινου χρώματος (L-κωνία, μέγιστη ευαισθησία σε μήκος κύματος 560 nm) και τα κωνία που είναι ευαίσθητα στις συχνότητες που αντιστοιχούν στην περιοχή του πράσινου (M-κωνία, μέγιστη ευαισθησία σε μήκος κύματος 530 nm). Αντίστοιχα τα S-κωνία παρουσιάζουν ευαισθησία στην περιοχή του μπλε χρώματος (μέγιστη ευαισθησία σε μήκος κύματος 420 nm).

Η ευαισθησία των κωνίων σε διαφορετικά μήκη κύματος οφείλεται σε φωτοευαίσθητες χρωστικές ουσίες τις φωτοψίνες, οι οποίες εμπεριέχουν τις φωτοευαίσθητες πρωτεΐνες, τις οψίνες. Η οψίνες ουσιαστικά μεσολαβούν για την μετατροπή των φωτονίων σε ηλεκτροχημικά σήματα, που οδηγούνται στη συνέχεια μέσω των οπτικών νευρών στον εγκέφαλο. Κάθε είδος κωνίου περιέχει διαφορετικές φωτοψίνες. Επιπρόσθετα, Το φως, καθώς προσπίπτει στα κωνικά κύτταρα, τα διεγείρει ανάλογα με τα μήκη κύματος που περιλαμβάνει. Συνεπώς, κάθε χρώμα που αντιλαμβανόμαστε οφείλεται στους συνδυασμούς των σημάτων που δίνουν οι φωτοϋποδοχείς.



Εικόνα 13: Κανονικοποιημένη τυπική απόκριση μονοχρωματικών φασματικών ερεθισμάτων του εκάστοτε τύπου κωνικού κυττάρου.

3.2.1. Χαρακτηριστικά του χρώματος

Κάθε χρώμα περιγράφεται γενικότερα από τρία κύρια χαρακτηριστικά: την *απόχρωση* (hue), τον *κορεσμό* (saturation) και την *φωτεινότητα* (lightness). Η απόχρωση προσδιορίζεται ως «ο βαθμός κατά το οποίο ένα (φωτεινό) ερέθισμα μπορεί να περιγραφεί ως παρόμοιο ή διαφορετικό από αντίστοιχα ερεθίσματα που

περιγράφονται ως κόκκινα, πράσινα, μπλε και κίτρινα» (Fairchild, 2004). Συνεπώς, μονοχρωματικά, η απόχρωση είναι ανάλογη του μήκους κύματος. Ο κορεσμός είναι «ο χρωματισμός μιας επιφάνειας προς την φωτεινότητά του» (Fairchild, 2004) που στην πραγματικότητα αφορά κατά πόσο μπορεί να γίνει αντιληπτή ο διαχωρισμός της χρωματισμένης επιφάνειας από το λευκό του φωτός. Ψυχοφυσικά υποδηλώνει το πόσο έντονο ή πόσο 'ζωντανό' είναι το χρώμα. Τέλος, η φωτεινότητα αφορά το πόσο φωτεινή είναι μία επιφάνεια συναρτήσει της πιο φωτεινής λευκής επιφάνειας του συγκεκριμένου περιβάλλοντος. Για παράδειγμα, οι περιοχές ενός ομοιόμορφα χρωματισμένου αντικειμένου σε άμεσο φως έχουν μεγαλύτερη φωτεινότητα από τις περιοχές που βρίσκονται στη σκιά. Η μορφές αυτές ανάλυσης και χαρακτηρισμού των χρωμάτων χρησιμοποιούνται κυρίως για την οργάνωση των χρωμάτων και των αντίστοιχων χώρων τους. Χρησιμοποιούνται και επιπλέον συμβατικές 'παράμετροι' ανάλυσης των χρωμάτων όπως το chroma (όμοιο μέγεθος με τον κορεσμό/saturation), το brightness (ανάλογο μέγεθος με την φωτεινότητα/lightness) και το colourfulness (ανάλογο επίσης του κορεσμού, της έντασης του χρώματος).

3.3 ΧΡΩΜΑΤΙΚΟΙ ΧΩΡΟΙ

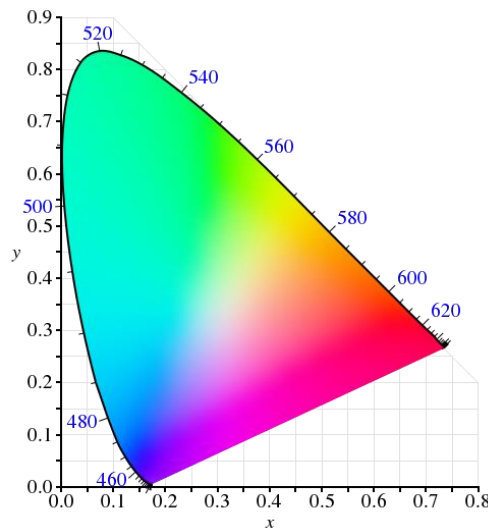
Η έννοια του χρωματικού χώρου αναπτύχθηκε για την ακριβή μαθηματική περιγραφή του χρώματος και των χαρακτηριστικών του σε ψηφιακά και αναλογικά μέσα. Πρόκειται ουσιαστικά για την οργάνωση των διαφόρων χρωμάτων με τέτοιο τρόπο ώστε να γίνεται αντιληπτό όλα τα χαρακτηριστικά και οι διαφορές των χρωμάτων. Επιπλέον, περιγράφουν τον τρόπο με τον οποίο τα χρώματα μπορούν να αναπαρασταθούν ως πλειάδες αριθμών. Σε γενικές γραμμές, τα μοντέλα αυτών χωρίζονται σε τέσσερα βασικά μοντέλα:

- **CIE** (η πρώτη απόπειρα για οργάνωση χρωμάτων, βασισμένη σε ένα σύστημα τριών συντεταγμένων)
- **RGB** (μία προσέγγιση που βασίζεται στο γραμμικό συνδυασμό 3 βασικών χρωμάτων)
- **Luma plus chroma** (ένα μοντέλο στο οποίο η εικόνα προκύπτει από το συνδυασμό της φωτεινότητας και της κατανομής του χρώματος στις περιοχές της εικόνας)
- **Hue and Saturation** (ένα μοντέλο-προέκταση του Luma + Chroma, με βασικές παραμέτρους την απόχρωση, τον χρωματικό κορεσμό και τη φωτεινότητα, που προκύπτει από μετασχηματισμένο του RGB σε κυλινδρικό σύστημα συντεταγμένων)

3.3.1 Το μοντέλο CIEXYZ

Η βασικότερη απόπειρα για αναπαράσταση των χρωμάτων σύμφωνα με την πρόσληψή τους από το ανθρώπινο μάτι είναι το μοντέλο CIE XYZ. Βασίζεται στην αποσύμπλεξη των χρωματικών παραμέτρων από αυτή της φωτεινότητας, με αποτέλεσμα ο κύβος της RGB αναπαράστασης να μετασχηματίζεται σε ένα επίπεδο χωρίο μέσω της κανονικοποίησης:

$$X = \frac{R}{R+G+B} \quad Y = \frac{G}{R+G+B}$$



Εικόνα 14: Το χωρικό διάγραμμα χρώματος CIE 1931 (CIE XYZ). Το εξωτερικό καμπυλωτό όριο είναι ο φασματικός (ή μονοχρωματικός) τόπος, με μήκη κύματος που εμφανίζονται σε νανόμετρα.

Με συντεταγμένες στο χωρίο αυτό ενός χρώματος στο RGB τις (x_r, y_r) , (x_g, y_g) και (x_b, y_b) , και με το λευκό σημείο αναφοράς το (x_w, y_w, z_w) η CIEXYZ αναπαράσταση προκύπτει μέσω του κάτωθι γραμμικού μετασχηματισμού της RGB αναπαράστασης ως εξής:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [M] \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Όπου

$$[M] = \begin{bmatrix} S_r X_r & S_g X_g & S_b X_b \\ S_r Y_r & S_g Y_g & S_b Y_b \\ S_r Z_r & S_g Z_g & S_b Z_b \end{bmatrix}$$

$$\begin{aligned}
X_r &= \frac{x_r}{y_r}, & Y_r &= 1, & Z_r &= \frac{1-x_r-y_r}{y_r} \\
X_g &= \frac{x_g}{y_g}, & Y_g &= 1, & Z_g &= \frac{1-x_g-y_g}{y_g} \\
X_b &= \frac{x_b}{y_b}, & Y_b &= 1, & Z_b &= \frac{1-x_b-y_b}{y_b}
\end{aligned}$$

Και

$$\begin{bmatrix} S_r \\ S_g \\ S_b \end{bmatrix} = \begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix}^{-1} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix}$$

3.3.2 Το μοντέλο CIELUV

Πρόκειται για μετασχηματισμό του χώρου χρώματος CIE XYZ, που όμως επιχειρήσε την αντιληπτική ομοιομορφία. Χρησιμοποιείται εκτενώς για εφαρμογές όπως γραφικά υπολογιστών που ασχολούνται με έγχρωμα φώτα. Για ένα λευκό σημείο με συντεταγμένες (X_n, Y_n, Z_n) , ο μετασχηματισμός σε συντεταγμένες (L^*, u^*, v^*) , όπου L^* η φωτεινότητα (τροποποιημένη έτσι από παλαιότερα μοντέλα ώστε να αυξάνεται με βήματα που είναι ορατά στο ανθρώπινο μάτι), και οι u^* και v^* είναι οι 2 καρτεσιανές διαστάσεις χρώματος που συνδυαζόμενες παράγουν μία απόχρωση, προκύπτει από τις παρακάτω σχέσεις:

$$L^* = \begin{cases} \left(\frac{29}{3}\right)^3 \frac{Y}{Y_n}, & \alpha\nu \frac{Y}{Y_n} \leq \left(\frac{6}{29}\right)^3 \\ 116 \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}}, & \alpha\nu \frac{Y}{Y_n} > \left(\frac{6}{29}\right)^3 \end{cases}$$

$$u^* = 13L^* \times (u' - u'_n)$$

$$v^* = 13L^* \times (v' - v'_n)$$

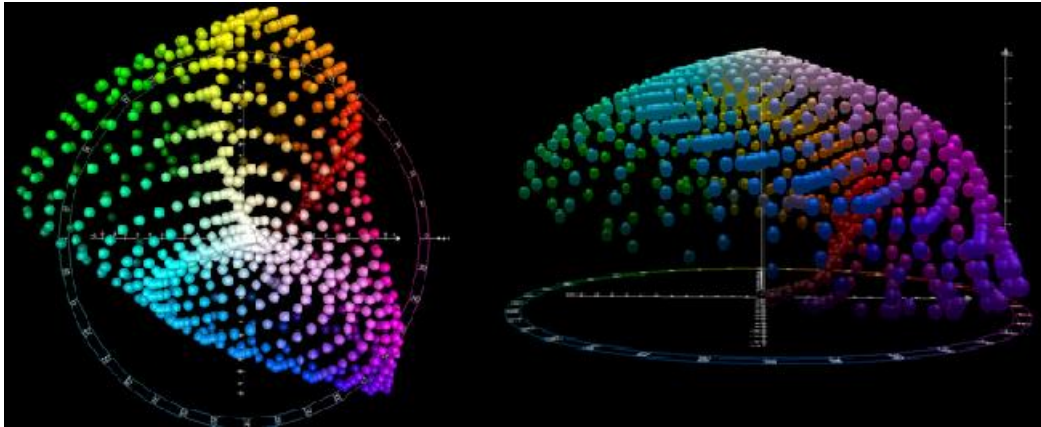
Όπου

$$u' = \frac{4X}{X + 15Y + 3Z} = \frac{4x}{-2x + 12y + 3}$$

$$v' = \frac{9Y}{X + 15Y + 3Z} = \frac{9y}{-2x + 12y + 3}$$

3.3.3 Το μοντέλο CIELAB

Το CIELAB (ή $L^*a^*b^*$ ή *Lab*) παράγει ένα χώρο χρωμάτων που είναι πιο αντιληπτικά γραμμικός από τους άλλους χώρους χρώματος, με την έννοια ότι μια αλλαγή του ίδιου μεγέθους σε μια τιμή χρώματος θα πρέπει να παράγει μια αλλαγή περίπου ίδιας οπτικής σημασίας. Αποδίδει δηλαδή τη διαφορά μεταξύ δύο χρωμάτων (Color Difference) ως μία γραμμική σχέση.



Εικόνα 15: Η άνω και πλάγια όψη αντίστοιχα του CIELAB χωρίου.

Με νέες συντεταγμένες τις ($L^*a^*b^*$), όπου L^* η φωτεινότητα (ο άξονας του κυλίνδρου) και a^*, b^* δύο κάθετες διαστάσεις επί του άξονα στο ύψος μίας τιμής του L^* που παράγουν τις τελικές αποχρώσεις (με το εύρος του a^* να είναι από πράσινο έως κόκκινο και αυτό του b^* από μπλε έως κίτρινο), ο μετασχηματισμός από τον χώρο CIEXYZ είναι ο εξής:

$$L^* = 116f\left(\frac{Y}{Y_n}\right) - 16$$

$$a^* = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right]$$

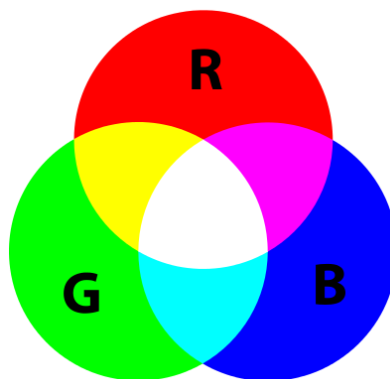
$$b^* = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right]$$

Όπου

$$f(t) = \begin{cases} (t)^{\frac{1}{3}}, & \text{αν } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29}, & \text{αν } t \leq \left(\frac{6}{29}\right)^3 \end{cases}$$

3.3.4 Το μοντέλο RGB

Το RGB (συντομογραφία του red-green-blue) ορίζεται από τις τρεις χρωματικές συνιστώσες του κόκκινου, πράσινου και μπλε και μπορεί να παράγει οποιοδήποτε τριγωνικό συνδυασμό που ορίζεται από τα βασικά αυτά χρώματα. Περιγράφει ουσιαστικά το είδος του φωτός που πρέπει να εκπέμπεται για να παράγει ένα δεδομένο χρώμα. Δηλαδή, για να παραχθεί ένα χρώμα με RGB, πρέπει να επικαλύπτονται τρεις φωτεινές δέσμες (ένα κόκκινο, ένα πράσινο και ένα μπλε) (για παράδειγμα με εκπομπή από μαύρη οθόνη ή με αντανάκλαση από λευκή οθόνη). Το μοντέλο χρώματος RGB είναι προσθετικό με την έννοια ότι οι τρεις φωτεινές δέσμες προστίθενται μαζί και τα φάσματα φωτός τους προστίθενται κατά το αντίστοιχο μήκος κύματος της κάθε συνιστώσας, ούτως ώστε να παράγεται το φάσμα του επιθυμητού χρώματος. Αντίθετα, στο μοντέλο αφαιρετικού χρώματος που ισχύει για χρώματα, μελάνια, χρωστικές ουσίες και κάποιες άλλες ουσίες, το χρώμα εξαρτάται από την ανάκλαση του φωτός κάτω από την οποία το βλέπει το μάτι.



Εικόνα 16: Προσθετική ανάμειξη χρωμάτων: η προσθήκη κόκκινου σε πράσινο αποδίδει κίτρινο. η προσθήκη κόκκινου σε μπλε αποδίδει ματζέντα. Προσθέτοντας πράσινο σε μπλε αποδίδει κυανό. Προσθέτοντας και τα τρία κύρια χρώματα μαζί αποδίδεται το λευκό.

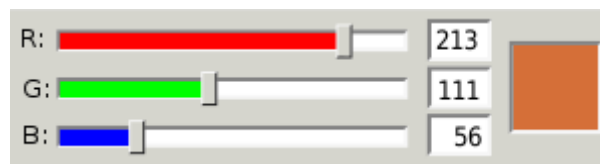
Ένα χρώμα στο μοντέλο χρώματος RGB περιγράφεται με την ποσοστιαία ένδειξη του κόκκινου, πράσινου και μπλε χρώματος. Συνεπώς, το χρώμα εκφράζεται ως πίνακας 1x3, RGB (r, g, b), κάθε συστατικό της οποίας μπορεί να κυμαίνεται από μηδέν έως μια καθορισμένη μέγιστη τιμή. Αν όλα τα στοιχεία είναι στο μηδέν το αποτέλεσμα είναι μαύρο. Αν όλα είναι στο μέγιστο, το αποτέλεσμα είναι το φωτεινότερο αντιπροσωπευτικό λευκό.

Αυτά τα εύρη μπορούν να ποσοτικοποιηθούν με διάφορους τρόπους:

- Από 0 έως 1, με οποιαδήποτε κλασματική τιμή μεταξύ. Αυτή η αναπαράσταση χρησιμοποιείται σε θεωρητικές αναλύσεις, και σε συστήματα που χρησιμοποιούν αναπαραστάσεις κινητής υποδιαστολής.
- Κάθε τιμή συστατικού χρώματος μπορεί επίσης να γραφεί ως ποσοστό, από 0% έως 100%.
- Στους υπολογιστές, οι τιμές συνιστωσών αποθηκεύονται συχνά ως ακέραιοι αριθμοί στην περιοχή 0 έως 255, εύρος το οποίο μπορεί να αποτυπώσει ένα byte δεδομένων. Αυτά συχνά αντιπροσωπεύονται είτε ως δεκαδικοί είτε δεκαεξαδικοί αριθμοί.

Για παράδειγμα, το πιο λαμπερό κορεσμένο κόκκινο γράφεται στις διάφορες αναπαραστάσεις RGB ως:

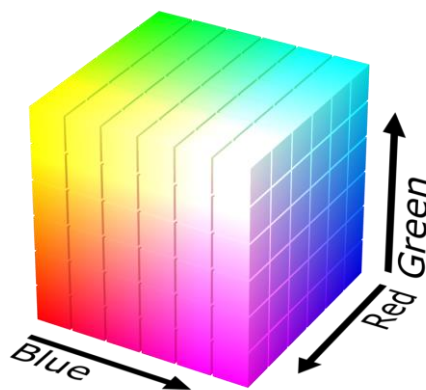
Αναπαράσταση	RGB Πίνακας
Αριθμητική	(1.0, 0.0, 0.0)
Ποσοστιαία	(100%, 0%, 0%)
Ψηφιακή (8-bit ανά κανάλι)	(255, 0, 0) ή #FF0000 (Δεκαεξαδικός)
Ψηφιακή (16-bit ανά κανάλι)	(65535, 0, 0)



Εικόνα 17: Ένας συνηθισμένος επιλογέας χρωμάτων RGB σε γραφικό περιβάλλον. Κάθε ρυθμιστικό κυμαίνεται από 0 έως 255.

Δεδομένου ότι τα χρώματα καθορίζονται από τρεις μεταβλητές, τότε ένας τρισδιάστατος χώρος μπορεί να περιγραφεί από τις τιμές τους ως συνηθισμένες καρτεσιανές συντεταγμένες σε ευκλείδειο χώρο. Στο RGB αυτό αντιπροσωπεύεται ως ένας κύβος που χρησιμοποιεί μη αρνητικές τιμές εύρους 0-1, αποδίδοντας

μαύρο στην αρχή των αξόνων, στην κορυφή (0, 0, 0), με αυξανόμενες τιμές έντασης που εκτείνονται κατά μήκος των τριών αξόνων, καταλήγοντας στο λευκό στην κορυφή (1, 1, 1). Μια τριάδα RGB (r, g, b) αντιπροσωπεύει την τρισδιάστατη συντεταγμένη του σημείου του δεδομένου χρώματος εντός του κύβου ή των επιφανειών του ή κατά μήκος των άκρων του. Αυτή η προσέγγιση επιτρέπει τον υπολογισμό της ομοιότητας χρώματος δύο δεδομένων χρωμάτων RGB με τον απλό υπολογισμό της απόστασης μεταξύ τους: όσο μικρότερη είναι η απόσταση, τόσο μεγαλύτερη είναι η ομοιότητα. Οι υπολογισμοί εκτός φάσματος μπορούν επίσης να εκτελεστούν με αυτό τον τρόπο.



Εικόνα 18: Το μοντέλο RGB που αντιστοιχεί σε κύβο. Ο οριζόντιος άξονας x, καθώς και οι κόκκινες τιμές αυξάνονται προς τα αριστερά, ο άξονας y με μπλε χρώμα αυξάνεται προς τα κάτω και ο κάθετος άξονας z με πράσινο αυξάνεται προς τα πάνω.

3.3.5 Το μοντέλο YCbCr

Το YCbCr πρόκειται για μια οικογένεια χρωμάτων που χρησιμοποιούνται ως τμήμα της σωλήνωσης έγχρωμης εικόνας σε συστήματα βίντεο και ψηφιακής φωτογραφίας. Το Y αφορά την παράμετρο luma (brightness) και τα Cb, Cr είναι παράμετροι που αντιστοιχούν στην διαφορά χρώματος από την μπλε και κόκκινη συνιστώσα αντιστοίχως (blue-difference, red-difference). Ανάλογο με το YCbCr είναι και το YPbPr το οποίο χρησιμοποιείται για την αποτύπωση αναλογικών βίντεο και προέρχεται από γραμμικό συνδυασμό των αναλογικών τιμών R, G και B (που κυμαίνονται από 0 έως 1) χρησιμοποιώντας δύο καθορισμένες σταθερές (KB) και (KR):

$$Y' = K_R R' + K_G G' + K_B B'$$

$$P_B = \frac{1}{2} \frac{B' - Y'}{1 - K_B}$$

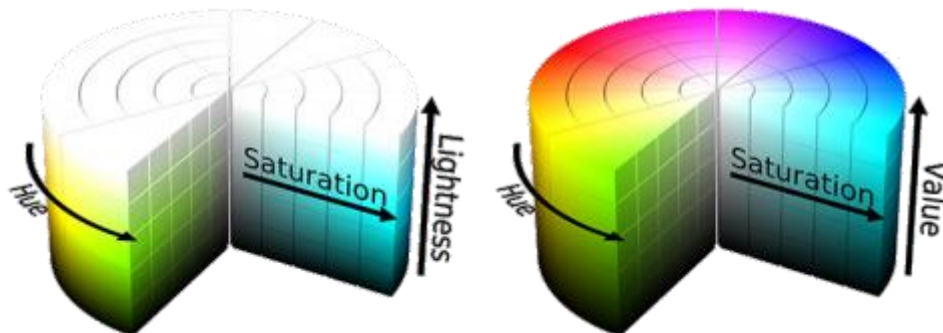
$$P_R = \frac{1}{2} \frac{R' - Y'}{1 - K_R}$$

3.3.6 Το μοντέλο HSV/HSL

Οι HSL και HSV χώροι είναι οι δύο πιο κοινές αναπαραστάσεις συστημάτων κυλινδρικών συντεταγμένων βασισμένες στο RGB μοντέλο. Ουσιαστικά αναδιατάσσουν τη γεωμετρία του RGB σε μια προσπάθεια να γίνουν πιο διαισθητικά αντιληπτές από την αναπαράσταση εκείνη του καρτεσιανού κύβου (RGB). Αναπτύχθηκε στη δεκαετία του 1930 για εφαρμογές έγχρωμης τηλεόρασης. Έτσι, ο χρωματικός χώρος HSL επέτρεψε την προσθήκη πληροφοριών χρώματος στα υπάρχοντα ασπρόμαυρα σήματα, πράγμα που σήμαινε ότι σε αντίθεση με τα προηγούμενα συστήματα χρωμάτων, οι μονόχρωμοι δέκτες θα μπορούσαν να λαμβάνουν έγχρωμα σήματα που εκπέμπονται χρησιμοποιώντας HSL. Επιπρόσθετα, οι HSL και HSV χρησιμοποιούνται σήμερα σε όλα τα συστήματα κωδικοποίησης ψηφιακής τηλεόρασης, τους συλλέκτες χρωμάτων, στα λογισμικά επεξεργασίας εικόνας και ενίοτε στην ανάλυση/επεξεργασία εικόνας και την υπολογιστική όραση (computer vision).

Η ονομασία του HSL προκύπτει από τα hue (απόχρωση), saturation (κορεσμό) και luminance (ή μερικές φορές lightness) (φωτεινότητα), ενώ το HSV προκύπτει από τα hue, saturation και value (τιμή- εδώ με την έννοια της ποιότητας λευκού φωτός) και συχνά ονομάζεται HSB (B για brightness).

Στα κυλινδρικά αυτά συστήματα, η γωνία γύρω από τον κεντρικό κατακόρυφο άξονα αντιστοιχεί στην απόχρωση (hue), η απόσταση από τον άξονα αντιστοιχεί στον κορεσμό (saturation) και η απόσταση κατά μήκος του άξονα αντιστοιχεί σε τιμή (value) ή φωτεινότητα (lightness ή luminance).



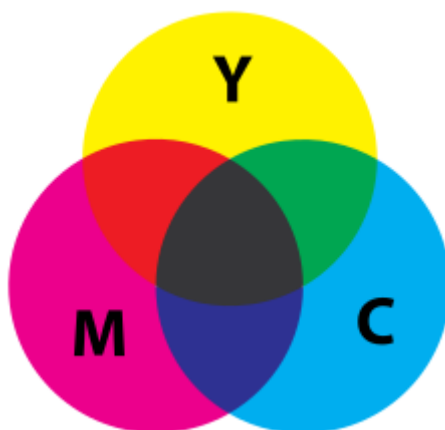
Εικόνα 19: Οι κυλινδρικές αναπαραστάσεις των μοντέλων HSL και HSV αντίστοιχα.

3.3.7 Το μοντέλο CMY(K)

Το CMY είναι επίσης ένας γραμμικός χρωματικός χώρος βασισμένος στην αφαιρετική ανάμειξη. Χρησιμοποιεί ως πρωτεύοντα χρώματα το κυανό (C-Cyan), ματζέντα (M-Magenta) και το κίτρινο (Y-Yellow) τα οποία αντιστοιχούν στα συμπληρωματικά χρώματα των κόκκινο, πράσινο και μπλε αντιστοίχως. Συνεπώς:

$$(C, M, Y) = (1 - R, 1 - G, 1 - B)$$

Το μοντέλο CMY χρησιμοποιείται στις συσκευές έγχρωμης εκτύπωσης, καθώς τα κυανό, ματζέντα και κίτρινο είναι τα κύρια χρώματα των χρωστικών ουσιών και μελανιών. Σε προσθετικά χρωματικά μοντέλα, όπως το RGB, το λευκό είναι ο 'προσθετικός' συνδυασμός όλων των κύριων χρωματισμένων φώτων (R+G+B), ενώ το μαύρο είναι ουσιαστικά η απουσία φωτός. Αντίθετα, στο CMY το λευκό πρόκειται για το φυσικό χρώμα του χαρτιού για παράδειγμα, ή κάποιου άλλου φόντου, ενώ το μαύρο είναι αποτέλεσμα ενός συνδυασμού έγχρωμων μελανιών.



Εικόνα 20: Η αφαιρετική μέθοδος παραγωγής 'μαύρου' χρώματος στο μοντέλο CMY.

Για την εξοικονόμηση μελάνης και προκειμένου να παραχθεί ένα πιο έντονο μαύρο χρώμα, παράγονται ακόρεστα και σκούρα χρώματα χρησιμοποιώντας μαύρο μελάνι (K-key) αντί για τον αποκλειστικό συνδυασμό κυανού, ματζέντα και κίτρινου (CMY).

3.4 ΧΡΩΜΑΤΙΚΟΣ ΚΥΚΛΟΣ



Εικόνα 21: Ο Χρωματικός κύκλος και οι δώδεκα υποδιαιρέσεις του.

Ο χρωματικός κύκλος ή αλλιώς χρωματικός τροχός είναι μια αφηρημένη απεικονιστική οργάνωση των αποχρώσεων των διάφορων χρωμάτων γύρω από έναν κύκλο, η οποία δείχνει τις σχέσεις μεταξύ βασικών χρωμάτων, δευτερευόντων χρωμάτων, τριτογενών χρωμάτων κλπ.

Οι περισσότεροι χρωματικοί κύκλοι βασίζονται σε τρία κύρια χρώματα, σε τρία δευτερεύοντα χρώματα και τα έξι ενδιάμεσα που σχηματίζονται με την ανάμιξη ενός πρωτογενούς με ένα δευτερεύον, γνωστό ως τριτογενή χρώματα, για συνολικά 12 κύριες διαιρέσεις. Μερικοί προσθέτουν περισσότερα ενδιάμεσα προϊόντα, για 24 ονόματα χρωμάτων.

Ο πρώτος χρωματικός κύκλος, του Νεύτωνα, απεικόνιζε μόνο τις φασματικές αποχρώσεις. Είχε ως σκοπό να καταστρώσει έναν μνημονικό κανόνα για το είδος του χρώματος που πρόκυπτε από την ανάμειξη διαφόρων φωτεινών δέσμίων, που θα μπορούσαν να προβλεφθούν κατά προσέγγιση, από το κέντρο βάρους των 'ακτινών' κάθε φασματικού χρώματος. (Newton, 1704) Οι διαιρέσεις του κύκλου του Νεύτωνα είναι άνισου μεγέθους, με βάση τα διαστήματα μιας δωρικής μουσικής κλίμακας. (Briggs, 2013)

Όπως προαναφέρθηκε, οι χώροι χρώματος HSL και HSV είναι απλοί γεωμετρικοί μετασχηματισμοί του κυβικού RGB σε κυλινδρική μορφή. Ο εξωτερικός ανώτερος κύκλος του κυλίνδρου HSV (για μέγιστο saturation και value), ή αντίστοιχα ο εξωτερικός μεσαίος κύκλος του κυλίνδρου HSL (για μέγιστο saturation και το μισό του μεγίστου lightness) μπορεί να θεωρηθεί χρωματικός κύκλος.

4. ΕΙΚΟΝΑ & ΒΙΝΤΕΟ

Ως εικόνα ορίζεται η αποτύπωση ή απομνημόνευση και εμφάνιση ενός ή περισσότερων αντικειμένων ή προσώπων. Η αποτύπωση αυτή μπορεί να γίνεται σε μπορεί να γίνει σε διαφορά μέσα αναλογικά (χαρτί, φιλμ κ.α.), ψηφιακά (υπολογιστής, κάμερα, κινητό κ.α.) ή στην μνήμη ενός ανθρώπου. Οι πληροφορίες μιας εικόνας αποτυπώνονται συνήθως σε δύο ή τρεις διαστάσεις, όπως αντίστοιχα αποτυπώνεται και η συνεχής ροή εικόνων (καρέ) στον χρόνο, κατά κανόνα σε ψηφιακά μέσα που στο πλήθος τους ορίζονται ως βίντεο. Με τις έννοιες εικόνα και βίντεο εννοούνται επίσης και τα ηλεκτρονικά μέσα εγγραφής, καταγραφής και μετάδοσης των πληροφοριών αυτών.

Τα σύγχρονα τεχνολογικά μέσα επιτρέπουν στην συρροή εικόνων και βίντεο στην καθημερινότητα του ανθρώπου. Στην περίπτωση των εικόνων, η χρήση και αποτύπωση τους παρατηρείται από τον πρώιμο ανθρώπινο πολιτισμό για την καταγραφή της καθημερινότητας του, της ζωής και των συναισθημάτων του, γεγονός που ισχύει μέχρι και σήμερα. Το βίντεο ως τέτοιο μέσο, υπάρχει μόνο τον τελευταίο αιώνα όμως η εξέλιξη του ακολουθεί μια εκθετική ανάπτυξη και χρησιμότητα.

4.1 ΨΗΦΙΑΚΗ ΕΙΚΟΝΑ

Η ψηφιακή εικόνα είναι ουσιαστικά μια αριθμητική αναπαράσταση (συνήθως δυαδική) μιας δισδιάστατης εικόνας. Ανάλογα με το εάν η ανάλυση της εικόνας είναι σταθερή, μπορεί να είναι τύπου vector (διανυσματική) ή raster (ένα σύνολο κουκίδων/εικονοστοιχείων). Η βασική διαφορά μεταξύ vector και raster γραφικών είναι ότι τα μεν αποτελούνται διανυσματικά 'μονοπάτια' ενώ τα δε από pixels (εικονοστοιχεία). Επομένως, ένα γραφικό τύπου raster, όπως μια gif ή jpeg εικόνα, είναι μια σειρά από pixels διαφόρων χρωμάτων, τα οποία μαζί σχηματίζουν μια εικόνα. Από την άλλη, ένα γραφικό τύπου vector είναι η μία ακολουθία εντολών ή μαθηματικών κανόνων που τοποθετεί γραμμές και σχήματα σε ένα δεδομένο δισδιάστατο ή τρισδιάστατο χώρο.

4.1.1 Γραφικά Raster

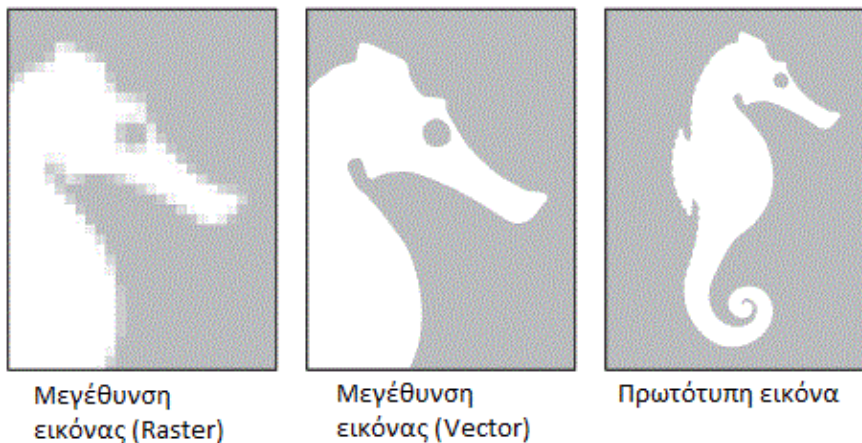
Οι εικόνες raster έχουν ένα πεπερασμένο σύνολο ψηφιακών τιμών, τα pixels (εικονοστοιχεία). Η ψηφιακή εικόνα δηλαδή, περιέχει σταθερό αριθμό γραμμών και στηλών τέτοιων pixels. Τα pixels είναι το μικρότερο μεμονωμένο στοιχείο μιας εικόνας και διατηρούν καθορισμένες τιμές που αντιπροσωπεύουν τη φωτεινότητα και την απόχρωση ενός δεδομένου χρώματος σε οποιοδήποτε συγκεκριμένο σημείο.

Γενικά, τα pixels αποθηκεύονται συλλογικά στη μνήμη του υπολογιστή ως εικόνα raster (ή χάρτη raster), δηλαδή ως δισδιάστατος πίνακας μικρών ακεραίων αριθμών. Αυτές οι τιμές συχνά μεταδίδονται ή αποθηκεύονται σε συμπιεσμένη μορφή.

Τέλος, οι εικόνες raster παράγονται από διάφορες συσκευές και τεχνικές εισόδου, όπως ψηφιακές φωτογραφικές μηχανές, σαρωτές, μηχανήματα μέτρησης συντεταγμένων, σειсмоγραφικό προφίλ, αερομεταφερόμενο ραντάρ και πολλά άλλα. Μπορούν επίσης να συντεθούν από αυθαίρετα δεδομένα, μη συντελεσμένης εικόνας, όπως μαθηματικές λειτουργίες ή τρισδιάστατα γεωμετρικά μοντέλα.

4.1.2 Γραφικά Vector

Τα γραφικά vector βασίζονται σε διανύσματα, τα οποία προσδιορίζονται μέσω συγκεκριμένων τοποθεσιών της εικόνας που ονομάζονται σημεία ελέγχου ή κόμβοι. Κάθε ένα από αυτά τα σημεία έχει μια ορισμένη θέση στους άξονες x και y του καρτεσιανού επιπέδου (επιφάνεια εργασίας) και καθορίζει την κατεύθυνση του διανύσματος (path). Επιπλέον, σε κάθε διάνυσμα μπορούν να αποδοθούν διάφορα χαρακτηριστικά, όπως το χρώμα, το σχήμα, η καμπύλη, το πάχος και η πληρότητα, ανάλογα με την μαθηματική περιγραφή της εικόνας. (Kaufman, 1993)



Εικόνα 22: Μία εικόνα μεγεθυμένη κατά raster και vector γραφικά αντιστοίχως.

4.2 ΨΗΦΙΑΚΟ ΒΙΝΤΕΟ

Το ψηφιακό βίντεο πρόκειται για μια αναπαράσταση κινούμενων οπτικών εικόνων με τη μορφή κωδικοποιημένων ψηφιακών δεδομένων σε αντίθεση με το αναλογικό βίντεο, το οποίο αντιπροσωπεύει κινούμενες οπτικές εικόνες με αναλογικά σήματα.

Επιπρόσθετα, το ψηφιακό βίντεο απαρτίζεται από μια σειρά από ορθογώνιες ψηφιακές εικόνες bitmap (raster), που διαδέχονται η μία την άλλη με σταθερό ρυθμό. Οι εικόνες αυτές, στα πλαίσια του βίντεο, ονομάζονται καρτέ (frames) και ο ρυθμός με τον οποίο προσμετρώνται ονομάζεται Frames Per Second (καρτέ ανά δευτερόλεπτο) ή FPS σε συντομογραφία. Επομένως, η bitmap εικόνα πρόκειται για μία ορθογώνια συστοιχία των pixels που την απαρτίζουν, έτσι, αν έχει πλάτος W pixels και ύψος H pixels, τότε προφανώς το μέγεθος πλαισίου θα είναι $W \times H$. Όπως προαναφέρθηκε, το χρώμα ενός pixel αντιπροσωπεύεται από έναν σταθερό και προκαθορισμένο αριθμό bits. Όσα περισσότερα bits υπάρχουν για να περιγράψουν τα pixels, τόσο μεγαλύτερο εύρος παραλλαγής χρωμάτων μπορεί να αποτυπωθεί στο βίντεο. Αυτό ονομάζεται βάθος χρώματος (CD – Color Depth) του βίντεο.

Για παράδειγμα, ένα βίντεο με διάρκεια μίας ώρας (3600sec), μέγεθος πλαισίου (frame) 640×480 ($W \times H$), με βάθος χρώματος 24bits και ρυθμό πλαισίου 25fps θα έχει τις εξής ιδιότητες:

- **Pixels ανά πλαίσιο** = $640 * 480 = 307.200$
- **Bits ανά πλαίσιο** = $307,200 * 24 = 7,372,800 = 7,37$ Mbits
- **Bit rate (BR)** = $7.37 * 25 = 184.25$ Mbits/sec
- **Μέγεθος του βίντεο (Video Size – VS)** = $184\text{Mbits/sec} * 3600\text{sec} = 662,400\text{Mbits} = 82,800\text{Mbytes} = 82.8\text{Gbytes}$

Όλα τα παραπάνω ισχύουν για βίντεο το οποίο δεν έχει ακολουθήσει την διαδικασία της συμπίεσης. Προφανώς όταν αναφέρεται κανείς στην ποιότητα ενός βίντεο, συνήθως αναφέρεται στο bit rate, δηλαδή την ροή δεδομένων ανά δευτερόλεπτο. Όσο πιο μεγάλο είναι το BR, τόση θα είναι και η οπτοακουστική πληροφορία του βίντεο. Παράλληλα, για συμπιεσμένα βίντεο, το BR είναι ένδειξη της ποιότητας της συμπίεσης και της απόδοσης του αλγορίθμου συμπίεσης που χρησιμοποιήθηκε.

4.3 ΥΠΟΛΟΓΙΣΤΙΚΗ ΟΡΑΣΗ

Η υπολογιστική όραση (Computer Vision) είναι ένα διεπιστημονικό πεδίο που ασχολείται με τον τρόπο κατά τον οποίο οι υπολογιστές μπορούν να χρησιμοποιηθούν για την άντληση δεδομένων και την υψηλού επιπέδου κατανόηση ψηφιακών εικόνων και βίντεο. Ουσιαστικά, επιδιώκει να αυτοματοποιήσει τις λειτουργίες που πραγματοποιεί το ανθρώπινο οπτικό σύστημα. (Ballard & Brown, 1982)

Οι συνηθισμένες λειτουργίες της υπολογιστικής όρασης περιλαμβάνουν μεθόδους επεξεργασίας, ανάλυσης και κατανόησης ψηφιακών εικόνων και την εξαγωγή πολυδιάστατων δεδομένων που υπάρχουν στον φυσικό κόσμο, με σκοπό την παραγωγή αριθμητικών ή συμβολικών πληροφοριών, όπως π.χ. στην μορφή αυτοματισμού ή αποφάσεων. (Klette, 2014) Με την έννοια κατανόηση ψηφιακής εικόνας εννοείται η μετατροπή των εικόνων αυτών (όπως για παράδειγμα η είσοδος του φωτός στον αμφιβληστροειδή) σε περιγραφές του κόσμου που μπορούν να αλληλοεπιδρούν με άλλες διαδικασίες σκέψης και να διεγείρουν τις κατάλληλες ενέργειες. Η κατανόηση αυτή της εικόνας, μπορεί να θεωρηθεί ως απεμπλοκή συμβολικών πληροφοριών από τα δεδομένα της εικόνας, χρησιμοποιώντας μοντέλα κατασκευασμένα με τη βοήθεια της γεωμετρίας, της φυσικής, της στατιστικής και της θεωρίας της μάθησης. (Forsyth & Ponce, 2003)

4.3.1 Ιστορία της Υπολογιστικής Όρασης

Στα τέλη της δεκαετίας του 1960, οι τεχνικές υπολογιστικής όρασης άρχισαν να αναπτύσσονται στα πανεπιστήμια που πρωτοπορούσαν στην τεχνητή νοημοσύνη. Σκοπός της υπολογιστικής όρασης τότε, ήταν να μιμείται το ανθρώπινο οπτικό σύστημα κατά τέτοιο τρόπο ώστε τα ρομπότ να μπορέσουν να προσεγγίσουν την νοητική ικανότητα των ανθρώπων. (Szeliski, 2010) Το 1966, έγινε μία απόπειρα να επιτευχθεί αυτό κατά τη διάρκεια ενός καλοκαιριού εγχειρήματος, καθώς προσαρτήθηκε μια φωτογραφική μηχανή σε έναν υπολογιστή και ο τελευταίος κλήθηκε να «περιγράψει αυτό που είδε». (Boden, 2006)

Αυτό που ξεχώριζε την υπολογιστική όραση από τον υπόλοιπο τομέα της ψηφιακής επεξεργασίας εικόνας εκείνη την εποχή, ήταν η ανάγκη να εξαχθεί τρισδιάστατη δομή των εικόνων, με σκοπό την πλήρη κατανόηση και περιγραφή των πληροφοριών της εικόνας. Οι μελέτες στη δεκαετία του 1970 αποτελούσαν τα πρώιμα θεμέλια για πολλούς από τους αλγόριθμους υπολογιστικής όρασης που υπάρχουν και σήμερα, συμπεριλαμβανομένων τέτοιων όπως την εξαγωγή ακμών των εικόνων, της σήμανσης γραμμών, της μη-πολυεδρικής και πολυεδρικής

μοντελοποίησης, της αναπαράστασης αντικειμένων ως διασυνδέσεις μικρότερων δομών, της οπτικής ροής (optical flow) και την εκτίμηση της κίνησης. (Szeliski, 2010)

Από τη δεκαετία του 1990 κι έπειτα η υπολογιστική όραση έχει γνωρίσει αλματώδη ανάπτυξη, έχει συνδεθεί με το γνωστικό πεδίο της μηχανικής μάθησης και έχει δώσει σημαντικά απτά αποτελέσματα, με αλγορίθμους όρασης πραγματικού χρόνου να υλοποιούνται ακόμα και σε φτηνά κινητά τηλέφωνα εξοπλισμένα με κάμερα. Στο εν λόγω πλαίσιο, η μηχανική όραση έχει διαδραματίσει θεμελιώδη ρόλο στην εξέλιξη της ενισχυμένης πραγματικότητας.

Μετά την ευρύτατη διάδοση του Kinect, ενός καινοτόμου περιφερειακού διασύνδεσης μεταξύ χρηστών και υπολογιστικών συστημάτων, και τη σχετική άνθιση του τρισδιάστατου (στερεοσκοπικού) οπτικού περιεχομένου ύστερα από τη μεγάλη επιτυχία της κινηματογραφικής ταινίας Άβαταρ το 2009, η υπολογιστική όραση έχει αρχίσει να εξετάζει πιο ενδελεχώς και την αξιοποίηση δεδομένων βάθους (π.χ. από στερεοσκοπική κάμερα ή ξεχωριστούς αισθητήρες βάθους) για την επίτευξη των στόχων της.

4.4 ΛΕΙΤΟΥΡΓΙΕΣ ΥΠΟΛΟΓΙΣΤΙΚΗΣ ΟΡΑΣΗΣ

4.4.1 Αναγνώριση Προτύπων

Ίσως η πιο συνηθισμένη λειτουργία στην υπολογιστική όραση είναι η αναγνώριση προτύπων. Πιο συγκεκριμένα αφορά τον καθορισμό των δεδομένων μιας εικόνας και εάν αυτή περιέχει κάποιο συγκεκριμένο αντικείμενο, χαρακτηριστικό ή δραστηριότητα. Παρουσιάζονται οι επιμέρους αυτές λειτουργίες παρακάτω:

- Αναγνώριση αντικειμένων (object recognition) – πρόκειται για ένα ή περισσότερα προκαθορισμένα αντικείμενα ή κατηγορίες αντικειμένων που μπορούν να αναγνωριστούν, συνήθως μαζί με τις καρτεσιανές συντεταγμένες τους στην εικόνα (δύο διαστάσεων) ή την τρισδιάστατη τοποθέτησή τους σε ένα πλάνο. Τα Blippar, Google Goggles και LikeThat παρέχουν αυτόνομα προγράμματα που απεικονίζουν αυτή τη λειτουργικότητα.
- Ταυτοποίηση (identification) – πρόκειται για την μεμονωμένη αναγνώριση και ταυτοποίηση συγκεκριμένου αντικειμένου. Για παράδειγμα, περιλαμβάνει την αναγνώριση ενός προσώπου (face recognition) ή του δακτυλικού αποτυπώματος, την αναγνώριση των χειρόγραφων ψηφίων ή την αναγνώριση και ταυτοποίηση ενός συγκεκριμένου οχήματος.

- Ανίχνευση (detection) – αφορά την σάρωση των δεδομένων της εικόνας για την ανίχνευση μιας συγκεκριμένης κατάστασης. Για παράδειγμα, η ανίχνευση πιθανών μη φυσιολογικών κυττάρων ή ιστών σε ιατρικές εικόνες και η ανίχνευση οχήματος σε αυτοματοποιημένο σύστημα διοδίων. Η ανίχνευση που βασίζεται σε σχετικά απλούς και γρήγορους υπολογισμούς χρησιμοποιείται μερικές φορές για την εύρεση δεδομένων μικρότερων περιοχών της εικόνας, τα οποία μπορούν να αναλυθούν περαιτέρω με πιο απαιτητικές υπολογιστικές τεχνικές για να παραχθεί μια σωστή ερμηνεία της εικόνας.

4.4.2 Κινητική Ανάλυση

Αρκετές λειτουργίες σχετίζονται με την εκτίμηση της κίνησης, όπου επεξεργάζεται μια ακολουθία εικόνων για να παραχθεί μια προσέγγιση της ταχύτητας ενός αντικειμένου, είτε σε κάθε σημείο της εικόνας, είτε σε τρισδιάστατο πλάνο ή ακόμα και στην κάμερα που συλλαμβάνει και παράγει τις εικόνες. Παραδείγματα τέτοιων λειτουργιών είναι:

- Egomotion – ουσιαστικά πρόκειται για τον καθορισμό της άκαμπτης τρισδιάστατης κίνησης (περιστροφή) της κάμερας από μια ακολουθία εικόνων που παράγεται από την κάμερα.
- Tracking (ανίχνευση) – η ανίχνευση και παρακολούθηση των κινήσεων ενός (συνήθως) μικρότερου συνόλου σημείων ή αντικειμένων (π.χ. οχημάτων ή ανθρώπων) στην ακολουθία εικόνων.
- Optical Flow (οπτική ροή) – Ο προσδιορισμός του τρόπου με τον οποίο ένα σημείο κινείται σχετικά με το επίπεδο της εικόνας, για κάθε σημείο της εικόνας. Αυτή η κίνηση είναι αποτέλεσμα τόσο του τρόπου με τον οποίο το αντίστοιχο τρισδιάστατο σημείο κινείται στο πλάνο, όσο και με τον τρόπο που η κάμερα κινείται σε σχέση με το πλάνο.

4.4.3 Ανακατασκευή πλάνου (Scene reconstruction)

Δεδομένης μιας ή περισσότερων εικόνων ενός πλάνου ή ενός βίντεο, η ανακατασκευή πλάνου στοχεύει στον υπολογισμό ενός τρισδιάστατου μοντέλου του πλάνου. Στην απλούστερη περίπτωση, το μοντέλο μπορεί να αποτελείται από ένα σύνολο από τρισδιάστατα σημεία. Οι πιο εξελιγμένες μέθοδοι παράγουν ένα πλήρες μοντέλο τρισδιάστατης επιφάνειας. Η εμφάνιση της τρισδιάστατης απεικόνισης που δεν απαιτεί κίνηση ή σάρωση, όπως και οι σχετικοί αλγόριθμοι επεξεργασίας, επιτρέπουν την ταχεία πρόοδο σε αυτό το πεδίο. Επίσης, η

πλεγματική τρισδιάστατη ανάλυση χρησιμοποιείται για την σύλληψη τρισδιάστατων εικόνων πολλαπλών γωνιών αναφοράς. Πλέον, υπάρχουν αλγόριθμοι για τη συρραφή πολλαπλών τρισδιάστατων εικόνων σε σύννεφα σημείων (point clouds) και τρισδιάστατα μοντέλα.

4.4.4 Επαναφορά εικόνας (Image Restoration)

Η επαναφορά εικόνας έχει ως σκοπό την απομάκρυνση θορύβου (αισθητήριος θόρυβος, θαμπάδα κίνησης κλπ.) από τις διάφορες εικόνες. Η απλούστερη δυνατή προσέγγιση για την απομάκρυνση θορύβου είναι η εφαρμογή των διάφορων τύπων φίλτρου στην εικόνα, όπως τα βαθυπερατά φίλτρα (low-pass filter) ή φίλτρα διάμεσης (median filter). Οι πιο εξελιγμένες μέθοδοι υιοθετούν ένα μοντέλο για το πώς μοιάζουν οι δομές τοπικών εικόνων, ένα μοντέλο που τις διακρίνει από τον θόρυβο. Αρχικά αναλύοντας τα δεδομένα της εικόνας σε επίπεδο δομών της εικόνας, όπως είναι οι γραμμές ή τα άκρα και κατόπιν τον έλεγχο του φιλτραρίσματος με βάση τις τοπικές πληροφορίες από το βήμα της ανάλυσης, επιτυγχάνεται συνήθως καλύτερη απομάκρυνση θορύβου σε σύγκριση με τις απλούστερες προσεγγίσεις.



Εικόνα 23: Παράδειγμα υπολογιστικής όρασης. Η συγκεκριμένη εφαρμογή πρόκειται για μετρητής ανθρώπων σε δημόσιους χώρους όπως π.χ. ένα εμπορικό κέντρο.

4.5 ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΑΣ

Στα πλαίσια της υπολογιστικής όρασης, η κατάτμηση της εικόνας είναι ουσιαστικά η διαδικασία διαίρεσης μιας ψηφιακής εικόνας σε πολλαπλά τμήματα (σύνολα εικονοστοιχείων/pixels, επίσης γνωστά ως υπερ-εικονοστοιχεία). Ο στόχος της τμηματοποίησης είναι να απλοποιήσει ή να αλλάξει την αναπαράσταση μιας εικόνας σε κάτι που είναι πιο ουσιαστικό και ευκολότερο να αναλυθεί. (Shapiro & Stockman, 2001) Η κατάτμηση εικόνων χρησιμοποιείται συνήθως για τον εντοπισμό αντικειμένων και ορίων (γραμμών, καμπυλών κ.λπ.) στις εικόνες. Πιο συγκεκριμένα, η κατάτμηση της εικόνας είναι η διαδικασία εκχώρησης μιας ετικέτας σε κάθε pixel σε μια εικόνα, έτσι ώστε εικονοστοιχεία με την ίδια ετικέτα να μοιράζονται ορισμένα χαρακτηριστικά.

Σε γενικές γραμμές, στην κατάτμηση εικόνας χρησιμοποιούνται τεχνικές κατάτμησης και ομαδοποίησης δεδομένων που ανήκουν χονδρικά στις δυικές κατηγορίες αλγορίθμων των διαχωριστικών και των ιεραρχικών τεχνικών.

Η λογική των διαχωριστικών τεχνικών έγκειται στην υλοποίηση μίας διαδικασίας ομαδοποίησης με τελικό στόχο τον διαχωρισμό και την απομόνωση των ομαδοποιήσεων που θα δημιουργηθούν (K-means, median κλπ.). Αντίθετα, οι ιεραρχικές τεχνικές λειτουργούν στη βάση της συνένωσης των ομαδοποιήσεων που θα προκύψουν (κατάτμηση ιστογράμματος, μέθοδος Watershed κλπ.).

4.6 ΤΕΧΝΙΚΕΣ ΚΑΤΑΤΜΗΣΗΣ ΕΙΚΟΝΑΣ

4.6.1 Μέθοδος Κατωφλίου (Thresholding)

Αυτή η μέθοδος βασίζεται σε μια τιμή κατωφλίου για να μετατραπεί μια grey-scale (κλίμακα απόχρωσης του γκρι) εικόνα σε αντίστοιχη δυαδική εικόνα, όπως επίσης εφαρμόζεται και για την μετατροπή μιας εικόνας σε δυαδική με βάση μια μέγιστη τιμή του ιστορογράμματος της. Συνεπώς, το κλειδί αυτής της μεθόδου είναι στην επιλογή της τιμής κατωφλίου (ή τιμές όταν επιλέγονται πολλαπλά επίπεδα).

Το ιστόγραμμα πρόκειται για την κατανομή χρώματος σε μία εικόνα και αποτελείται από «λόφους» και «κοιλιάδες». Οι λόφοι, ως περιοχές απότομης χρωματικής συμπίκνωσης, μαρτυρούν την παρουσία κάποιου θέματος. Αν ο λόφος είναι πολύ λεπτός, τότε πιθανόν μιλάμε για ένα αντικείμενο, αντίθετα αν είναι πιο φαρδύς, πιθανότατα δηλώνει το φόντο. Αυτό που δε μπορεί να προσδιοριστεί είναι το σε ποιο χωρίο ανήκουν τα εικονοστοιχεία των κοιλιάδων.

Στην περίπτωση της μεθόδου κατωφλίου με ιστοριόγραμμα, η τιμή κατωφλίου ορίζεται κατά προσέγγιση ανάμεσα στους λόφους (δηλαδή μες στην κάθε κοιλάδα) του ιστογράμματος. Στην περίπτωση των δύο λόφων, η τελική εικόνα λαμβάνει μόνο 2 τιμές:

$$t(m, n) = \begin{cases} 1, & \text{αν } f(m, n) > T \\ 0, & \text{αν } f(m, n) \leq T \end{cases}$$

Το αποτέλεσμα είναι μία δυαδική εικόνα, που απεικονίζεται ως μαύρο και άσπρο. Στην περίπτωση όμως περισσότερων αντικειμένων και αποχρώσεων, μπορούν να ορισθούν πολλές τιμές κατωφλίου (μία για κάθε κοιλάδα) και να παραχθεί μία εικόνα σε περισσότερες αποχρώσεις του γκρι. Δεδομένου ότι κάθε χρωματική αλλαγή στην καινούργια εικόνα (δηλαδή μία επιπλέον απόχρωση του γκρι), δηλώνει μία αλλαγή θέματος-αντικειμένου στην εικόνα μας, εύκολα από αυτό προσδιορίζονται οι ακμές τους και αποκόπτονται από το περιβάλλον τους.



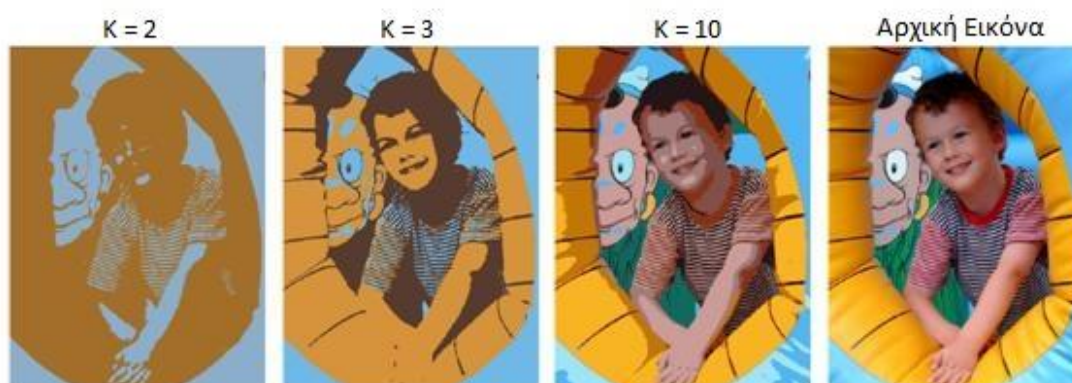
Εικόνα 24: Μία ένδειξη της ανάλυσης/κατάτμησης μιας εικόνας με την μέθοδο κατωφλίου

4.6.2 K-Means Clustering

Μία ιδιαίτερα διαδεδομένη κατηγορία μεθόδων που εντάσσεται σε αυτές των διαχωριστικών τεχνικών είναι αυτή του κέντρου βάρους. Η πιο γνωστή μέθοδος αυτής της κατηγορίας είναι η K-means, η οποία δουλεύει ως εξής: χωρίζεται η

εικόνα (τυχαία) σε N περιοχές με την καθεμία να διαθέτει ένα κέντρο βάρους. Κάθε σημείο γύρω από αυτό εντάσσεται στην αντίστοιχη περιοχή, δεδομένου ότι η απόστασή του (ευκλείδεια, χρωματική κ.λπ.) δεν ξεπερνάει την ακτίνα του, δηλαδή βρίσκεται στο κοντινότερο κέντρο βάρους που είναι διαθέσιμο.

Για κάθε ρixel μίας περιοχής, ελέγχεται το χρωματικό του περιεχόμενο κι έτσι αυτό αποδίδεται σε μία χρωματική περιοχή. Αν το στοιχείο πλησιάζει χρωματικά στην αρχική εκτίμηση του υποβάθρου, η οποία είναι επιλεγμένη ως το χρωματικό ελάχιστο (minimum) της εικόνας, τότε αποδίδεται σε αυτό, ενώ το αντίστοιχο συμβαίνει αν πλησιάζει στην αρχική εκτίμηση του γράμματος, δηλαδή το χρωματικό μέγιστο (maximum). Για κάθε επιπλέον απόδοση ρixel σε μία περιοχή, το χρωματικό περιεχόμενο της περιοχής αυτής ορίζεται και πάλι ως ο χρωματικός μέσος όρος των στοιχείων που έχουν αποδοθεί σε αυτήν μέχρι στιγμής. Η διαδικασία αυτή επαναλαμβάνεται έως ότου υπάρχει μία σύγκλιση μεταξύ του εκάστοτε ρixel και του αντίστοιχου κέντρου βάρους του.



Εικόνα 25: Η κατάτμηση της εικόνας σε $K = 2, 3$ και 10 περιοχές αντίστοιχα.

4.6.3 Ανίχνευση Ακμών

Η ανίχνευση ακμών αποτελεί τοπική μέθοδο επεξεργασίας μίας εικόνας, με στόχο να εντοπίσει τα εικονοστοιχεία μίας γραμμής που μπορεί να θεωρηθεί ακμή, δηλαδή μία γραμμή που οι περιοχές που χωρίζει έχουν μεγάλη διαφορά στη συνάρτηση εικόνας. Για την κατάτμηση ενός αντικειμένου από μια εικόνα, ωστόσο, είναι αναγκαία τα κλειστά περιγράμματα περιοχής. Οι επιθυμητές ακμές συνεπώς, είναι τα όρια μεταξύ τέτοιων αντικειμένων ή χωρικών ταξινομημένων ομάδων.

Μια τυπική ακμή μπορεί για παράδειγμα να είναι το όριο μεταξύ ενός μπλοκ κόκκινου χρώματος και ενός μπλοκ κίτρινου χρώματος. Αντίθετα μια γραμμή (όπως μπορεί να εξαχθεί από έναν ανιχνευτή κορυφογραμμής) μπορεί να είναι ένας μικρός αριθμός εικονοστοιχείων διαφορετικού χρώματος σε ένα κατά τα άλλα

αμετάβλητο υπόβαθρο. Για μια γραμμή, μπορεί συνεπώς να υπάρχει συνήθως μια άκρη σε κάθε πλευρά της γραμμής.

Υπάρχουν πολλές μέθοδοι και αλγόριθμοι για την ανίχνευση ακμών, με τον απλούστερο να είναι αυτός που χρησιμοποιεί την συνάρτηση εξομάλυνσης του Gauss (συνάρτηση σφάλματος) για την εξομάλυνση των ακμών σε πρακτικές εφαρμογές. Έτσι, μια μονοδιάστατη εικόνα f η οποία έχει ακριβώς μία ακμή τοποθετημένη στο $x = 0$ μπορεί να μοντελοποιηθεί ως εξής:

$$f(x) = \frac{I_r - I_l}{2} \left(\operatorname{erf} \left(\frac{x}{\sqrt{2}\sigma} \right) + 1 \right) + I_l$$

Με την συνάρτηση σφάλματος να ορίζεται ως:

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt$$

Στην αριστερή πλευρά της ακμής, η ένταση είναι $I_l = \lim_{x \rightarrow -\infty} f(x)$, και δεξιά της ακμής είναι $I_r = \lim_{x \rightarrow \infty} f(x)$. Η παράμετρος κλίμακας, σ , ονομάζεται κλίμακα θολώματος της ακμής. Στην ιδανική περίπτωση, αυτή η παράμετρος κλίμακας θα πρέπει να προσαρμόζεται με βάση την ποιότητα της εικόνας, ώστε να αποφευχθεί η καταστροφή των πραγματικών ακμών της εικόνας.

4.6.4 Median

Μία εξίσου χρήσιμη μέθοδος είναι αυτή της διαμέσου (median). Με λογική χωρισμού της εικόνας σε K περιοχές, όπως η K -means, διαφέρει ως προς το ότι η αναφορά σε κάθε περιοχή δεν είναι το χρωματικό κέντρο βάρους αλλά η ενδιάμεση χρωματική τιμή που προκύπτει από τα ταξινομημένα pixel της συνολικής περιοχής. Με αναφορά λοιπόν αυτήν, η (χρωματική) ακτίνα απόστασης ενός στοιχείου από το σημείο-αντιπρόσωπο (medoid) είναι εξορισμού μικρότερη, καθότι το στοιχείο δε συγκρίνεται με μία ακραία χρωματική τιμή μίας περιοχής, αλλά με τη μέση τιμή των στοιχείων της. Τέλος, η μέση τιμή παραμένει αυτή που είναι καθ' όλη την ταξινόμηση των στοιχείων και δεν επαναπροσδιορίζεται ως ο καινούργιος μέσος όρος των στοιχείων, όπως στην K -means.

Και οι δύο παραπάνω μέθοδοι λειτουργούν καλύτερα όταν τα σχήματα των περιοχών είναι κυκλικά ή παραπλήσια, με την μέθοδο της διαμέσου να είναι αρκετά πιο αξιόπιστη, όμως και οι δύο μπορούν να γίνουν προβληματικές όταν λειτουργήσουν σε περιοχές με μεγάλη διαφορά μεγεθών μεταξύ τους.

4.6.5 Μέθοδοι μερικής διαφορικής κατάτμησης εικόνας

Χρησιμοποιώντας μια μέθοδο μερικής διαφορικής εξίσωσης (ΜΔΕ) και επιλύοντας την εξίσωση ΜΔΕ με κάποιο αριθμητικό σχήμα, μπορεί να κατατμηθεί μία εικόνα. Η διάδοση καμπύλης είναι μια δημοφιλής τεχνική σε αυτή την κατηγορία, με πολλές εφαρμογές για την ανίχνευση αντικειμένων, την παρακολούθηση αντικειμένων, την στερεοσκοπική ανακατασκευή κλπ. Οι μέθοδοι που βασίζονται σε ΜΔΕ είναι ίσως οι πιο βολικές και αποτελεσματικές μέθοδοι για την κατάτμηση της εικόνας. Το κύριο πλεονέκτημα είναι ότι η θεωρία πίσω από την ιδέα και τις τεχνικές λύσης είναι καλά εδραιωμένα σε άλλους τομείς όπως η φυσική και η μηχανική. Μερικοί τέτοιοι μέθοδοι είναι οι εξής:

- Παραμετρικοί μέθοδοι - Οι τεχνικές Lagrange βασίζονται στην παραμετροποίηση του περιγράμματος ενός αντικειμένου σύμφωνα με κάποια στρατηγική δειγματοληψίας και στη συνέχεια την εξέλιξη κάθε στοιχείου σύμφωνα με την εικόνα και την εσωτερική της δομή. Τέτοιες τεχνικές είναι γρήγορες και αποτελεσματικές, ωστόσο η αρχική, καθαρά παραμετρική διατύπωση (λόγω του Kass, Witkin και Terzopoulos το 1987 και γνωστό ως "snakes") (Kass, Witkin, & Terzopoulos, 1988) επικρίνεται γενικά για τους περιορισμούς της όσον αφορά την επιλογή στρατηγικής δειγματοληψίας, της καμπύλης, των μεταβολών τοπολογίας (διαχωρισμός και συγχώνευση των καμπυλών), αντιμετώπιση προβλημάτων σε υψηλότερες διαστάσεις κλπ. Σήμερα, έχουν αναπτυχθεί αποτελεσματικές "διακριτοποιημένες" διαδικασίες για την αντιμετώπιση αυτών των περιορισμών διατηρώντας παράλληλα υψηλή απόδοση. Σε κάθε περίπτωση, η ελαχιστοποίηση της ενέργειας διεξάγεται γενικά χρησιμοποιώντας κατακόρυφη κλίση, όπου τα παράγωγα υπολογίζονται χρησιμοποιώντας, π.χ., πεπερασμένες διαφορές.
- Μέθοδοι καθορισμού επιπέδου (set level) - αυτές οι μέθοδοι μπορούν να χρησιμοποιηθούν για την αποτελεσματική αντιμετώπιση του προβλήματος διάδοσης της καμπύλης / επιφάνεια / κλπ. με πεπλεγμένο τρόπο. Η κεντρική ιδέα είναι να αντιπροσωπεύει το εξελισσόμενο περίγραμμα χρησιμοποιώντας μια βηματική συνάρτηση της οποίας το μηδέν αντιστοιχεί στο πραγματικό περίγραμμα. Στη συνέχεια, σύμφωνα με την εξίσωση

κίνησης του περιγράμματος, μπορεί κανείς εύκολα να αντλήσει μια παρόμοια ροή για την πεπλεγμένη επιφάνεια που όταν εφαρμοστεί στο μηδενικό επίπεδο θα αντικατοπτρίζει τη διάδοση του περιγράμματος. Η μέθοδος set level παρέχει πολλά πλεονεκτήματα: είναι πεπλεγμένη, είναι ανεξάρτητη από παραμέτρους, παρέχει έναν άμεσο τρόπο εκτίμησης των γεωμετρικών ιδιοτήτων της εξελισσόμενης δομής, επιτρέπει την αλλαγή της τοπολογίας και είναι εγγενής. Μπορεί να χρησιμοποιηθεί για τον καθορισμό ενός πλαισίου βελτιστοποίησης, όπως προτάθηκε από τους Zhao, Merriman και Osher το 1996. (Peng, Merriman, Osher, Zhao, & Kang, 1999) Μπορεί κανείς να καταλήξει στο συμπέρασμα ότι είναι ένα πολύ βολικό πλαίσιο για την αντιμετώπιση πολυάριθμων εφαρμογών υπολογιστικής όρασης και ανάλυσης ιατρικών εικόνων. Η έρευνα στον καθορισμό επιπέδου στις δομές δεδομένων έχει οδηγήσει σε πολύ αποτελεσματικές υλοποιήσεις αυτής της μεθόδου.

- Μέθοδος Ταχύτατης Πορείας (Fast Marching Method) – Η μέθοδος αυτή πρόκειται για μία εξειδικευμένη μορφή της μεθόδου καθορισμού επιπέδου. Ουσιαστικά περιγράφει την εξέλιξη μιας κλειστής επιφάνειας ως συνάρτηση του χρόνου u με ταχύτητα f στην κανονική κατεύθυνση, σε ένα σημείο x επί της διαδιδόμενης επιφάνειας. Η συνάρτηση της ταχύτητας προκαθορίζεται και ο χρόνος κατά τον οποίο το περίγραμμα διέρχεται από ένα σημείο x καθορίζεται με την επίλυση της εξίσωσης. Περιγράφεται επομένως από την μη-γραμμική μερική διαφορική κυματική εξίσωση (Maxwell):

$$|\nabla u(x)| = \frac{1}{f(x)} \text{ για } x \in \Omega$$

$$u(x) = 0 \text{ για } x \in \partial\Omega$$

Αν ο χώρος που μελετάται χωριστεί σε σημεία ενός συνολικότερου πλέγματος, τότε κάθε σημείο του πλέγματος μπορεί να θεωρηθεί κόμβος. Κάθε κόμβος έχει αντίστοιχη τιμή $U_i = U(x_i) \approx u(x_i)$. Ο αλγόριθμος λειτουργεί ακριβώς όπως ο αλγόριθμος του Dijkstra αλλά διαφέρει ως προς τον τρόπο με τον οποίο υπολογίζονται οι τιμές των κόμβων. Στον αλγόριθμο του Dijkstra, η τιμή ενός κόμβου υπολογίζεται χρησιμοποιώντας έναν μόνο από τους γειτονικούς κόμβους. Ωστόσο, κατά την επίλυση της ΜΔΕ στο \mathbb{R}^n , χρησιμοποιούνται μεταξύ 1 και n γειτονικοί κόμβοι.

4.6.6 Watershed

Μία ακόμα διαδομένη μέθοδος είναι η λεγόμενη Watershed. Η συλλογιστική αφητηρία αυτής της μεθόδου προκύπτει από την ιδέα ότι πλημμυρίζοντας ένα γεωγραφικό τοπίο, οι περιοχές του μπορούν ξεχωρίσουν εύκολα χάρη στο νερό που συγκεντρώνεται άνισα στην κάθε μία, εξαιτίας του διαφορετικού τους βάθους. Πρακτικά, προκύπτει ως σύνθεση των 3 βασικότερων ιδεών: της ανίχνευσης ακμών, της κατωφλιοποίησης και των της ανάπτυξης περιοχών.

Ανοίγοντας σπές στη θέση του κάθε ελάχιστου της εικόνας, ξεκινούμε την πλημμυρίδα από αυτά (δηλαδή τις πιο σκοτεινές περιοχές) και αφήνουμε τη στάθμη να ανεβαίνει από κάτω προς τα πάνω με σταθερό ρυθμό. Σε κάποια χρονική στιγμή, η στάθμη θα έχει ανέβει τόσο σε κάθε λεκάνη απορροής (ο χώρος που δημιουργείται γύρω από την οπή) που θα είναι κοντά στο να υπερχειλίσει σε μία διπλανή. Όταν αυτό τείνει να συμβεί, ο αλγόριθμος Watershed κατασκευάζει ένα φράγμα πάνω στην ακμή που εμποδίζει οριακά την υπερχειλίση, κι έτσι δημιουργείται ένα κομμάτι του περιγράμματος της περιοχής. Η εικόνα συνεχίζει να πλημμυρίζεται και η ίδια διαδικασία λαμβάνει χώρα σε κάθε ακμή κάθε περιοχής που τείνει να εξαφανιστεί από την υπερχειλίση της πλημμυρίδας προς διπλανή περιοχή.

Συνήθως ο αλγόριθμος Watershed χρησιμοποιείται στο μέτρο της κλίσης της προς κατάτμηση εικόνας και όχι στην εικόνα αυτή καθαυτή, καθώς με βάση τις ακμές, η κατάτμηση ορίζεται από τις περιοχές μονοτονίας του μέτρου της κλίσης της εικόνας.

4.6.7 Άλλες μέθοδοι κατάτμησης

Με την ραγδαία τεχνολογική ανάπτυξη καθώς και την χρηστικότητα της υπολογιστικής όρασης, αναπτύσσονται και εξελίσσονται συνεχώς νέες μέθοδοι κατάτμησης εικόνας, αναλόγως με την εφαρμογή και την πρακτικότητα τους. Αναφορικά, κάποιοι από αυτές η μέθοδος Ενεργών Περιγραμμάτων/Παραμετρικών Ενεργών Περιγραμμάτων, η μέθοδος Γεωδαιτικών Ενεργών Περιγραμμάτων, οι μέθοδοι κατανομής γραφημάτων, οι μέθοδοι αυξανόμενης περιοχής (region-growing methods), οι μέθοδοι διπλής ομαδοποίησης (dual clustering methods), οι μέθοδοι βασισμένοι στην συμπίεση (compression-based methods) καθώς και οι μέθοδοι κατάτμησης πολλαπλών κλιμάκων (multi-scale segmentation).

5. ΥΛΟΠΟΙΗΣΗ ΑΛΓΟΡΙΘΜΟΥ

5.1 ΕΡΓΑΛΕΙΑ ΓΙΑ ΤΗΝ ΥΛΟΠΟΙΗΣΗ

Για την διεκπεραίωση της αλγοριθμικής σύνθεσης βασισμένη στο βίντεο εισόδου, χρειάστηκε ένας ευρύς συνδυασμός εργαλείων και λογισμικών με το καθένα να έχει συγκεκριμένο ρόλο στην υλοποίηση. Λόγω της πολυδιάστατης λειτουργίας του κώδικα, υπήρχε ανάγκη μία βάση υλοποίησης, κάποιο εργαλείο ή λογισμικό που θα κάλυπτε τις ανάγκες ανάλυσης του βίντεο και των καρτέ, όπως επίσης και κάποιο λογισμικό που θα παρήγαγε το αντίστοιχο μουσικό αποτέλεσμα.

5.1.1 Python

Η Python είναι μια υψηλού επιπέδου γλώσσα προγραμματισμού (O'Grady, 2013) με δημιουργό τον Ολλανδό Γκβίντο βαν Ρόσσουμ (Guido van Rossum) το 1990. Ο κώδικας της χαρακτηρίζεται από την αναγνωσιμότητα, την ευκολία στην χρήση της και το συντακτικό της επιτρέπει στους προγραμματιστές να εκφράσουν έννοιες σε λιγότερες γραμμές κώδικα από ότι θα ήταν δυνατόν σε γλώσσες όπως η C++ ή η Java. (McConnell, 2009) Διακρίνεται λόγω του ότι έχει πολλές βιβλιοθήκες που διευκολύνουν ιδιαίτερα αρκετές συνηθισμένες εργασίες και για την ταχύτητα εκμάθησής της, όπως και για την ευκολία που παρέχει για τον αντικειμενοστραφή προγραμματισμό.

Οι διερμηνευτές της Python είναι διαθέσιμοι για εγκατάσταση σε πολλά λειτουργικά συστήματα, επιτρέποντας στην Python την εκτέλεση κώδικα σε ευρεία γκάμα συστημάτων. Χρησιμοποιώντας εργαλεία τρίτων, όπως το Py2exe ή το Pyinstaller, ο κώδικας της Python μπορεί να πακεταριστεί σε αυτόνομα εκτελέσιμα προγράμματα για μερικά από τα πιο δημοφιλή λειτουργικά συστήματα, επιτρέποντας τη διανομή του βασισμένου σε Python λογισμικού για χρήση σε αυτά τα περιβάλλοντα χωρίς να απαιτείται εγκατάσταση του διερμηνευτή της Python.

Η Python αναπτύσσεται ως ανοιχτό λογισμικό (open source) και η διαχείρισή της γίνεται από τον μη κερδοσκοπικό οργανισμό Python Software Foundation. Ο κώδικας διανέμεται με την άδεια Python Software Foundation License η οποία είναι συμβατή με την GPL. Το όνομα της γλώσσας προέρχεται από την ομάδα Άγγλων κωμικών Μόντυ Πάιθον.

5.1.2 OpenCV

Το OpenCV (Open Source Computer Vision) είναι μια βιβλιοθήκη λειτουργιών προγραμματισμού που χρησιμοποιείται κυρίως σε εφαρμογές υπολογιστικής όρασης σε πραγματικό χρόνο. (Pisarevsky, 2017) Αρχικά αναπτύχθηκε από την Intel, στη συνέχεια υποστηρίχθηκε από το Willow Garage και τώρα διατηρείται από την Itseez.

Το OpenCV είναι γραμμένο σε C++ και το βασικό του περιβάλλον του είναι στην C++, αλλά διατηρεί ακόμα ένα λιγότερο εκτενές αλλά εκτεταμένο παλαιότερο περιβάλλον σε C. Διαθέτει όμως και διασυνδέσεις με περιβάλλοντα Python και Java και υποστηρίζει Windows, Linux, Mac OS, iOS και Android.

5.1.3 Mingus

Το mingus είναι μία πλατφόρμα και βιβλιοθήκη λειτουργιών μουσικής θεωρίας και σημειογραφίας για την Python που υποστηρίζει το αρχείο MIDI και μέσο αναπαραγωγής. (Sraans, 2015) Μπορεί να χρησιμοποιηθεί για εφαρμογές που αφορούν τη θεωρία της μουσικής, να δημιουργήσει πλατφόρμες επεξεργασίας, εκπαιδευτικά εργαλεία και άλλες εφαρμογές που χρειάζεται να επεξεργαστούν ή να παίξουν μουσική. Μπορεί επίσης να χρησιμοποιηθεί για τη δημιουργία παρτιτούρας με το LilyPond.

5.1.4 Scikit-Learn

Το scikit-learn πρόκειται για μία βιβλιοθήκη λειτουργιών με σκοπό το machine learning χωρίς όμως να περιορίζεται σε αυτό απαραίτητα. (scikit-learn developers (BSD License), 2017) Διαθέτει διάφορους αλγόριθμους ταξινόμησης, παλινδρόμησης και ομαδοποίησης, συμπεριλαμβανομένων των μηχανισμών διανυσμάτων υποστήριξης (support vector machines), random forest, αύξησης κλίσης, k-means και DBSCAN και έχει σχεδιαστεί για να συνεργάζεται με τις αριθμητικές και επιστημονικές βιβλιοθήκες της Python NumPy και SciPy.

Το Scikit-learn αναπτύχθηκε αρχικά από τον David Cournapeau στα πλαίσια του καλοκαιρινού προγράμματος κώδικα της Google το 2007. Αργότερα ο Matthieu Brucher προσχώρησε στο έργο και άρχισε να το χρησιμοποιεί ως μέρος της δουλειάς του. Το 2010 η INRIA, το Γαλλικό Ινστιτούτο Έρευνας Πληροφορικής και Αυτοματισμού, συμμετείχε και η πρώτη δημόσια έκδοση (v0.1 beta) δημοσιεύθηκε στα τέλη Ιανουαρίου 2010.

5.1.5 Matplotlib

Το Matplotlib είναι μια βιβλιοθήκη σχεδίασης Python 2D, η οποία παράγει ποιοτικά στοιχεία δημοσίευσης σε διάφορες μορφές έντυπου και διαδραστικά περιβάλλοντα σε διάφορες πλατφόρμες. (Hunter, 2007) Το Matplotlib μπορεί να χρησιμοποιηθεί σε κώδικα Python, το Python shell και το IPython, το jupyter notebook, από διακομιστές εφαρμογών ιστού και τέσσερα εργαλεία γραφικού περιβάλλοντος. Με το Matplotlib μπορεί κανείς να δημιουργήσει ιστογράμματα, φάσματα ισχύος, γραμμικά διαγράμματα, διαγράμματα σφάλματος, scatterplots κλπ., με μερικές μόνο γραμμές κώδικα.

5.1.6 NumPy

Το NumPy είναι το θεμελιώδες πακέτο για επιστημονικό υπολογισμό με την Python. (NumPy Developers, 2017) Περιέχει μεταξύ άλλων:

- Ισχυρό αντικείμενο N-διάστατης συστοιχίας
- Εξελιγμένες λειτουργίες εκπομπής
- Εργαλεία για την ενσωμάτωση κώδικα C / C ++ και Fortran
- Χρήσιμη γραμμική άλγεβρα, μετασχηματισμό Fourier και δυνατότητες τυχαίου αριθμού

Εκτός από τις προφανείς επιστημονικές του χρήσεις, το NumPy μπορεί επίσης να χρησιμοποιηθεί ως αποτελεσματικός πολυδιάστατος χώρος (container) γενικών δεδομένων. Μπορούν να οριστούν αυθαίρετοι τύποι δεδομένων. Αυτό επιτρέπει στο NumPy να ενσωματώνεται απρόσκοπτα και γρήγορα με μια μεγάλη ποικιλία βάσεων δεδομένων.

5.1.7 Imutils

Το imutils πρόκειται για μία σειρά λειτουργικών συναρτήσεων ευκολίας για να πραγματοποιήσει κανείς τις βασικές λειτουργίες επεξεργασίας εικόνας όπως η μετάφραση, η περιστροφή, η αλλαγή μεγέθους, ο σκελετισμός και η ευκολότερη απεικόνιση των εικόνων του Matplotlib με το OpenCV και το Python 2.7 και το Python 3. (Rosebrock, 2017)

5.1.8 Πρωτόκολλο MIDI

Το MIDI (Musical Instrument Digital Interface, ελλ. Ψηφιακή Διασύνδεση Μουσικών Οργάνων) είναι ένα πρωτόκολλο που αποσκοπεί στην επικοινωνία και τον συγχρονισμό μεταξύ ηλεκτρονικών μουσικών οργάνων (όπως synthesizers, drum machines, samplers, συσκευές χρονισμού), υπολογιστών και άλλων ηλεκτρονικών συσκευών, ανεξαρτήτως κατασκευαστή.

Το πρωτόκολλο MIDI δεν μεταδίδει ηχητικό σήμα, αλλά μηνύματα που περιέχουν πληροφορίες σχετικά με το τονικό ύψος και την ένταση μιας νότας, καθώς επίσης και σήμα χρονισμού που προσδιορίζει την ταχύτητα - το tempo - ενός κομματιού.

5.2 ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ ΑΛΓΟΡΙΘΜΩΝ ΕΠΕΞΕΡΓΑΣΙΑΣ

Στην παρούσα διπλωματική εργασία, επιδιώκεται η οπτικοακουστική προσέγγιση των πληροφοριών ενός βίντεο, συσχετίζοντας τα χρώματα και την κίνηση του βίντεο στην αρμονία και τον ρυθμό αντίστοιχα, μιας μουσικής σύνθεσης. Συνεπώς, το προϊόν του αλγορίθμου είναι η ηχητικοποίηση (sonification) του εκάστοτε βίντεο που εισάγεται σε αυτόν, αξιοποιώντας μια παραμετροποίηση των πληροφοριών του βίντεο και εξάγοντας το τελικό αποτέλεσμα ήχου.

Για το πρακτικό μέρος της εργασίας επιλέχθηκε το βίντεο *EYLANDIA | Iceland 4k* των δημιουργών *Marcus Sies & Flo Nick*. Ο λόγος που επιλέχθηκε το βίντεο αυτό είναι ότι εμπεριέχει πολλά 'στατικά' πλάνα τραβηγμένα ως time-lapse⁴, φυσικών και ιδιόμορφων τοπίων στην Ισλανδία. Έτσι υπάρχει η δυνατότητα να εκφραστεί η αρμονία μεταξύ των χρωμάτων της φύσης και η κίνηση σε συγκεκριμένα σημεία του κάθε καρέ.

Η επεξεργασία του βίντεο γίνεται σε πρώτη φάση από ένα υποπρόγραμμα που πραγματοποιεί δειγματοληψία ενός δευτερολέπτου ανά καρέ (1 fps) και 'τεμαχίζει' το κάθε καρέ της δειγματοληψίας αυτής σε οκτώ επιμέρους κομμάτια. Σε κάθε τέτοιο 'τεμάχιο', γίνεται μία ανάλυση χρωματική και κινητική στο δεύτερο στάδιο του αλγορίθμου, το οποίο θα εξηγηθεί στην συνέχεια. Κατόπιν, αποθηκεύονται τα καρέ στο σύνολό τους, καθώς και τα επιμέρους τεμάχια σε ξεχωριστούς φακέλους του υπολογιστή.

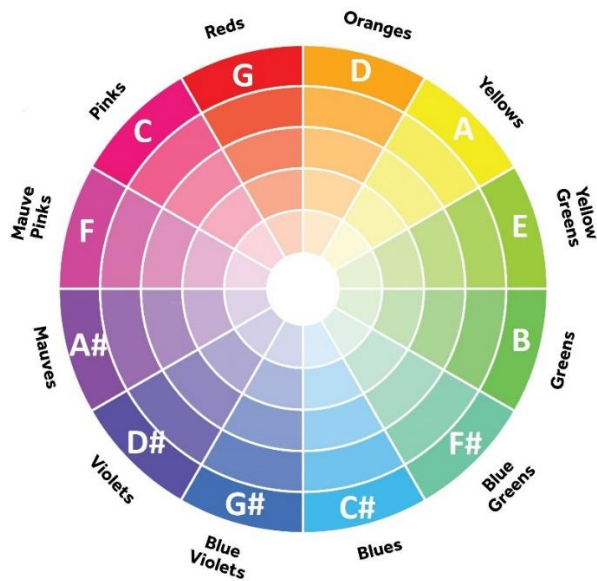
Το κύριο πρόγραμμα, κατά την εκκίνηση του, 'διαβάζει' τα στοιχεία των φακέλων αυτών έτσι ώστε να ξεκινήσει η διαδικασία ανάλυσης των χρωμάτων και της κίνησης. Αρχικοποιούνται οι διάφορες μεταβλητές που χρησιμοποιούνται καθόλη

⁴ Η φωτογραφική τεχνική της λήψης μιας ακολουθίας καρέ σε καθορισμένα χρονικά διαστήματα για την καταγραφή των αλλαγών που λαμβάνουν χώρα με αργό ρυθμό.

τη διάρκεια του προγράμματος και επιλέγεται από τον χρήστη ο αριθμός των clusters (ομάδες) κατά την οποία θα ποσοτικοποιηθεί χρωματικά, αρχικά το κάθε τεμάχιο και στην συνέχεια ολόκληρο το καρέ. Πριν την χρωματική ανάλυση, ο κώδικας αναλύει την κίνηση που υπάρχει σε κάθε τεμάχιο με κάθε αλλαγή καρέ. Αυτό το κάνει αποθηκεύοντας προσωρινά σε μία μεταβλητή buffer καθένα από τα οκτώ τεμάχια του καρέ, αφού το μετατρέψει πρώτα σε greyscale (γκρι ισοδύναμο) και πραγματοποιήσει Gaussian Blur (εξομάλυνση εικόνας, θόλωμα σκληρών ακμών). Έπειτα, συγκρίνει το επεξεργασμένο τεμάχιο αυτό με το αντίστοιχο σε θέση του επόμενου καρέ, ούτως ώστε να εξετάσει το ενδεχόμενο κίνησης στο σημείο αυτό. Αν υποθέσουμε πως το παρόν καρέ είναι το n καρέ και το επόμενο από αυτό το $n+1$ τότε μπορεί να διαπιστωθεί εάν υπάρχει κίνηση σε κάποιο τεμάχιο εάν αφαιρεθεί (κατά pixel) κάποιο τεμάχιο στο n καρέ από το τεμάχιο του ίδιου σημείου στο $n+1$ καρέ (absolute difference). Κατόπιν, το αποτέλεσμα θα εμφανίσει το αρνητικό της διαφοράς μεταξύ των καρέ στο συγκεκριμένο σημείο, δηλαδή μία ή πολλές λευκές επιφάνειες που υποδηλώνουν την κίνηση από τον χρόνο της n δειγματοληψίας μέχρι τον χρόνο της $n+1$ δειγματοληψίας. Τέλος, η επιφάνεια (ή επιφάνειες) που προκύπτει εξετάζεται εάν ξεπερνάει κάποιο προκαθορισμένο μέγεθος που ορίζει ο χρήστης, το οποίο ουσιαστικά πρόκειται για την ελάχιστη επιφάνεια της απόλυτης διαφοράς μεταξύ των αντίστοιχων σημείων των διαδοχικών καρέ (threshold) που υποδηλώνει κάποια εμφανής κίνηση. Προφανώς, η παραπάνω διαδικασία επαναλαμβάνεται για όλα τα τεμάχια των διαδοχικών καρέ.

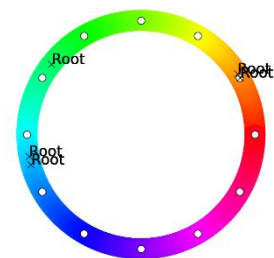
Κάθε τεμάχιο αντιστοιχίζεται σημειογραφικά σε ένα δεκατοέκτο. Εφόσον υπάρχει κίνηση σε διαδοχικά τεμάχια, τότε προσαυξάνεται στην αξία του φθόγγου η αξία του κάθε τεμαχίου. Για παράδειγμα, εάν ο υπολογιστής αντιληφθεί κίνηση σε ένα τεμάχιο και έπειτα στο δεύτερο, τρίτο, κλπ., τότε η αξία του φθόγγου που μεταφράζεται σε δεκατοέκτο, όγδοο, παρεστιγμένο όγδοο, κλπ. αντιστοιχά.

Σε δεύτερη φάση, ο υπολογιστής ανιχνεύει το κυρίαρχο χρώμα που επικρατεί στο τεμάχιο που εξετάζεται με κάθε επανάληψη. Μετατρέπει τον χρωματικό χώρο με τον οποίο αντλεί τα δεδομένα (rgb) σε αντίστοιχο hsv κύλινδρο και εξετάζει σε ποια απόχρωση (hue) αντιστοιχεί το χρώμα για τις μέγιστες τιμές των saturation (κορεσμός) και value (φωτεινότητα). Έτσι ουσιαστικά αντλεί την περίμετρο της άνω επιφάνειας του κυλίνδρου hsv (για μέγιστη ακτίνα-saturation και μέγιστο ύψος-value), που αντιστοιχεί στον χρωματικό κύκλο με χρωματική παράμετρο την γωνία του κύκλου (hue). Ο υπολογιστής χωρίζει τον κύκλο σε δώδεκα ισάξια τόξα (δώδεκα βασικά χρώματα του χρωματικού κύκλου) και αντιστοιχίζονται αυτά σε συγκεκριμένες νότες της χρωματικής κλίμακας. Η αντιστοιχία έχει ως εξής:



Εικόνα 26: Χρωματικός κύκλος με αντιστοίχιση νοτών της χρωματικής κλίμακας, διατεταγμένων έτσι ώστε να αντιστοιχούν στον κύκλο πέμπτων.

Η αντιστοίχιση έχει γίνει έτσι ώστε να προκύπτουν χρωματικές αρμονίες μέσα από τα χρώματα των καρτέ, διότι έτσι προκύπτει και αντίστοιχη μουσική αρμονία. Για παράδειγμα, στο τελευταίο καρτέ του βίντεο που χρησιμοποιήθηκε, φαίνεται η χρωματική ανάλυση του καρτέ, η αναγωγή του στον χρωματικό κύκλο και ο αριθμός και το είδος των clusters κατά το οποίο ομαδοποιήθηκαν τα χρώματα (εικόνα 27). Είναι εμφανές πως οι 'νότες' που εντοπίστηκαν είναι F#, πολλαπλά A και πολλαπλά και G#. Ο συνδυασμός των νοτών αυτών θα μπορούσε να αντιστοιχεί στην πρώτη, δεύτερη και τρίτη βαθμίδα μιας ελάσσονος κλίμακας από F#.



Εικόνα 27: Η ανάλυση του τελευταίου καρτέ του 'EYLEND', στο οποίο διαφαίνεται το καρτέ, τα k-means clusters και η αναγωγή στον χρωματικό κύκλο.

Ο ήχος που θα ακουστεί εξαρτάται από τις τρεις παραμέτρους χρώματος στον hsv χώρο, hue, saturation και value. Παραπάνω εξηγήθηκε πως η απόχρωση (hue) είναι συνδεδεμένη με την χρωματική κλίμακα, δηλαδή ποια από τις δώδεκα νότες θα 'παιχτούν'. Αντίστοιχα, το saturation των κυρίαρχων χρωμάτων του κάθε τεμαχίου και του κάθε καρέ, συσχετίζεται με την οκτάβα της νότας που θα παιχτεί. Επί της ουσίας, γίνεται μία κλιμακοποίηση των τιμών της παραμέτρου saturation που μπορεί να πάρει κάποιο χρώμα και αντιστοιχίζεται με το πόσο υψηλή ή χαμηλή θα ακουστεί η νότα που θα παιχτεί. Συνεπώς, αν η τιμή του κορεσμού είναι χαμηλή στην κλίμακα (απλό μουντό χρώμα, χωρίς πολύ ζωντάνια) τότε θα επιλεχθεί κάποιο χαμηλό εύρος οκτάβας, αν η τιμή του κορεσμού είναι ψηλή τότε θα επιλεχθεί κάποιο ψηλό εύρος οκτάβας αντίστοιχα.

Επιπρόσθετα, η τιμή του value, δηλαδή της φωτεινότητας, κλιμακοποιήθηκε και συσχετίστηκε με την δυναμική των νοτών που θα παιχτούν. Χαμηλό value σημαίνει χαμηλή δυναμική (απαλός ήχος, με μικρή ένταση), ψηλό value σημαίνει υψηλή δυναμική (δυνατός ήχος, με μεγάλη ένταση). Αυτό συμβαίνει διότι η φυσική φωτεινότητα πρόκειται για μέγεθος ανάλογο της ενέργειας του φωτός. Αντίστοιχα και η δυναμική πρόκειται για μέγεθος ανάλογο της ενέργειας του κύματος της ταλάντωσης της πηγής.

Τα αποτελέσματα της ανάλυσης εκφράζονται ως έξοδο του συστήματος με τη μορφή midi αρχείων, τα οποία επεξεργάζονται από τον χρήστη μέσω προγράμματος επεξεργασίας ήχου.

Όλες οι παραπάνω λειτουργίες συμβαίνουν μία φορά για τα τεμάχια του κάθε καρέ και μία επιπλέον φορά καθώς αναλύεται το κάθε καρέ ξεχωριστά. Η μόνη διαφορά είναι πως στην μία περίπτωση (των τεμαχίων αποκλειστικά) το αποτέλεσμα είναι μονοφωνικό, ενώ στην περίπτωση όλου του καρέ, παίζεται μία συγχορδία πέντε νοτών (που αντιστοιχούν στα πέντε clusters χρωμάτων) ως ένα πολυφωνικό 'χαλί', πίσω από την μονοφωνική μελωδία.

Τέλος, το βασικό πρόγραμμα χρησιμοποιεί ένα δεύτερο υποπρόγραμμα/κλάση που κατασκευάστηκε στα πλαίσια της εργασίας, για την εναρμόνιση των δεδομένων που αντλούνται από το βίντεο. Ακολουθεί κάποιους δομημένους κανόνες για να μπορεί να παραχθεί η αρμονία βασισμένη και παραμετροποιημένη πάνω στις πληροφορίες που αντλούνται από το βίντεο.

5.3 ΛΕΙΤΟΥΡΓΙΑ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ

5.3.1 Αλγόριθμος δειγματοληψίας και καταμερισμού (Πρώτο στάδιο επεξεργασίας)

Στο πρώτο στάδιο ανάλυσης και επεξεργασίας του βίντεο-εισόδου γίνεται η δειγματοληψία των καρτέ που πρόκειται να αναλυθούν από το επόμενο στάδιο (βασικό πρόγραμμα). Κατά την εκκίνηση του προγράμματος καλούνται κάποιες βασικές βιβλιοθήκες του OpenCV καθώς και βιβλιοθήκες που προκαθορισμένα έχουν εγκατασταθεί μαζί με την συνολικότερη εγκατάσταση του interpreter python.

```
import cv2
import imutils
import time
import numpy as np
```

Στην συνέχεια ο υπολογιστής καλείται να δείχνει συνεχώς στο βίντεο που πρόκειται να αναλύσει με την μεταβλητή **vidcap** και όσο 'τρέχει' το βίντεο να επιστρέφει μία σχέση αληθείας με την μεταβλητή **success**. Επίσης αρχικοποιούνται οι μεταβλητές που πρόκειται να χρησιμοποιηθούν ως μετρητές χρόνου (**msec**) ή καρτέ (**count**).

```
vidcap = cv2.VideoCapture('iceland.mp4')
success,image = vidcap.read()
count = 0
msec = 0
firstFrame = None
```

Καθορίζεται ο αριθμός στηλών και γραμμών κατά τις οποίες θα καταμεριστεί το κάθε καρτέ.

```
# Numbers of rows
nRows = 2
# Number of columns
mCols = 4
```

Όσο δεν αλλάζει η σχέση αληθείας '**success**', τρέχει ο καταμερισμός των καρτέ και γίνεται δειγματοληψία ένα καρτέ το δευτερόλεπτο, κάτι το οποίο καθορίζεται από την συνάρτηση **vidcap.set(cv2.CAP_PROP_POS_MSEC, msec)** με βάση την είσοδο **msec** που καθορίζεται στο τέλος του κώδικα να προσαυξάνεται 1000 msec μετά από κάθε επανάληψη. Η συνάρτηση **imutils.resize(image, width=500)** πραγματοποιεί μία σμίκρυνση του κάθε καρτέ για να γίνει πιο αποδοτική η δειγματοληψία.

```

while (success):

    vidcap.set(cv2.CAP_PROP_POS_MSEC, msec)
    success,image = vidcap.read()

    if not success:
        break

    image = imutils.resize(image, width=500)

    sizeY = image.shape[0]
    sizeX = image.shape[1]

    #Tam of Y region
    Y=sizeY/nRows
    print Y

    # Tam of X region
    X = sizeX/mCols
    print X

    left = 0
    top = 0
    height = Y
    width = X
    # Total of regions
    cnt = 0

    for i in range(0,mCols):
        top = 0
        for j in range(0, nRows):
            img=image[top:(top+height), left:(left + width)]
            #img = img.reshape((img.shape[0] * img.shape[1], 3))

            cv2.imwrite("/home/xenofon/Pythonstuff/Xen/framez/frame{0}.{1}.jpg"
                .format(count, cnt), img)

```

Παραπάνω αναγνωρίζονται οι διαστάσεις του κάθε καρέ, καθώς και οι διαστάσεις που θα έχουν τα επιμέρους τεμάχια του καρέ με X να αντιπροσωπεύει την διάσταση στον οριζόντιο άξονα και Y στον κατακόρυφο αντίστοιχα. Επειδή γενικά ο υπολογιστής απαριθμεί τις δισδιάστατες εικόνες με το (0,0) στην πάνω αριστερή

γωνία της εικόνας, αρχικοποιούνται οι μεταβλητές **left** και **top** και στην συνέχεια σαρώνεται η εικόνα με βήμα τον αριθμό των στηλών στον οριζόντιο άξονα και με τον αριθμό των γραμμών στον κατακόρυφο αντίστοιχα. Το κάθε νέο τεμάχιο καθορίζεται με βάση τα όρια του προηγούμενου και συναθροίζονται σε αυτά οι διαστάσεις X, Y που υπολογίστηκαν προηγουμένως. Τέλος, οι εικόνες/τεμάχια αποθηκεύονται στον φάκελο *framez* για χρήση στο κύριο πρόγραμμα.

```
top = top + Y
cnt +=1
left = left + X
```

Προφανώς με κάθε σάρωση τα όρια του νέου τεμαχίου πρέπει να καθοριστούν με βάση τις σταθερές διαστάσεις X, Y.

```
print('sec:', msec/float(1000))

print 'Read a new frame: ', success

cv2.imshow('Image', image)
key = cv2.waitKey(1) & 0xFF

# if the `q` key is pressed, break from the loop
if key == ord("q"):
    break

cv2.imwrite("/home/xenofon/Pythonstuff/Xen/wholeframez/whole%d.jpg"
% count, image) # save frame as JPEG file
```

Παραπάνω γίνεται ο σχετικός οπτικός έλεγχος πως η δειγματοληψία και ο καταμερισμός πραγματοποιείται με επιτυχία, καθώς εμφανίζεται η τιμή της μεταβλητής **msec** μετά από κάθε επανάληψη και η μεταβλητή αληθείας **success**. Έτσι επιβεβαιώνεται ο χρήστης πως γίνεται δειγματοληψία κάθε χρονικό διάστημα που ορίζει ο ίδιος. Στην συνέχεια, ολόκληρο το καρέ της δειγματοληψίας αποθηκεύεται στον φάκελο *wholeframez* για την χρήση στο κύριο πρόγραμμα. Τέλος, καθορίζονται οι τιμές του μετρητή των καρέ και το σχετικό χρονικό διάστημα που γίνεται η δειγματοληψία, και απενεργοποιείται η διαδικασία δειγματοληψίας του βίντεο με τις εντολές **vidcap.release()** και **cv2.destroyAllWindows()**.

```
count += 1
msec += 1000

vidcap.release()
cv2.destroyAllWindows()
```


5.3.2 Επεξήγηση αλγορίθμου δειγματοληψίας και καταμερισμού

Η διαδικασία δειγματοληψίας και καταμερισμού παραμετροποιείται ανάλογα με τις προτιμήσεις του χρήστη, καθώς οι βασικοί παράμετροι που καθορίζουν τον 'ρυθμό' τις δειγματοληψίας, όπως και το 'μέγεθος' και τον 'τρόπο' καταμερισμού είναι οι τιμές των **msec** και **nRows, nCols** μεταβλητών αντίστοιχα. Όπως αναφέρθηκε προηγουμένως, η μεταβλητή **msec** καθορίζει το χρονικό διάστημα κατά το οποίο πραγματοποιείται η δειγματοληψία του κάθε καρέ και τα **nRows, nCols** καθορίζουν τον αριθμό των γραμμών και στηλών που επιλέγει ο χρήστης να καταμερίσει το κάθε καρέ.

Στα πλαίσια της εργασίας, θεωρήθηκε πως ο βέλτιστος χρόνος δειγματοληψίας είναι τα 1000 ms ή 1 δευτερόλεπτο (1 fps) διότι με μία συντομότερη δειγματοληψία υπήρχε υπερσυσσώρευση πληροφορίας. Ο συνήθης ρυθμός προβολής των καρέ (frame rate) είναι τα 30 fps. Για το ανθρώπινο μάτι ο ρυθμός αυτός επαρκεί για να αντλήσει την οπτική-κινητική πληροφορία του βίντεο, (Read & Meyer, 2000) όμως για να μπορέσει να μεταφραστεί η πληροφορία αυτή σε αντίστοιχη ακουστική, προϋποθέτει μεγαλύτερα χρονικά διαστήματα για να μπορέσει να γίνει αντιληπτή από το ανθρώπινο αυτί η συχνοτική απόκριση του κάθε ήχου. Επίσης, με ρυθμό δειγματοληψίας 1 fps, γίνεται πιο εύκολα αντιληπτή η ρυθμική διάσταση του βίντεο, διότι:

$$1 \text{ fps} = 60 \text{ frames per minute}$$

Και αν για παράδειγμα καθορίσουμε ένα tempo για το ακουστικό αποτέλεσμα, έτσι ώστε να 'ταιριάξει' στο περιεχόμενο του βίντεο, λ.χ. 120 bpm (beats per minute), τότε το κάθε καρέ έχει αντίστοιχη ακουστική πληροφορία:

$$\frac{120 \text{ bpm}}{60 \text{ fpm}} = 2 \text{ bpf} \text{ ή } \text{beats per frame}$$

Συνεπώς, η αξία του κάθε καρέ αντιστοιχεί σε δύο χτύπους το δευτερόλεπτο, ή σε ένα υποθετικό μέτρο $\frac{4}{4}$, σε δύο τέταρτα εκ των τεσσάρων.

Επιπρόσθετα, γίνεται ο καταμερισμός των καρέ κατά τέτοιον τρόπο, ώστε να χωρίζεται το καθένα σε οκτώ ίσα μέρη (2 σειρές × 4 στήλες). Ο λόγος που συμβαίνει αυτός ο καταμερισμός είναι γιατί εκτελείται μία κινητική ανάλυση στο κύριο πρόγραμμα, σε κάθε τεμάχιο και με τέτοιον τρόπο, ώστε όταν ο υπολογιστής αντιληφθεί κίνηση σε ένα τέτοιο τεμάχιο, να 'πυροδοτεί' τον ήχο που έχει την σχετική χρωματική αντιστοίχιση. Επομένως, εφόσον το καρέ έχει χωριστεί σε 2

χτύπους τότε το κάθε τεμάχιο αντιστοιχεί στο $\frac{1}{4}$ του χτύπου (ή για το παραπάνω παράδειγμα με μέτρο $\frac{4}{4}$, θα αντιστοιχεί σε ένα δεκατοέκτο).

Έτσι, από τον συνδυασμό των παραπάνω, προκύπτει πως κάθε καρτέ διαμορφώνεται σε ένα είδος 'παρτιτούρας' πάνω στο οποίο αποτυπώνεται η 'μουσική' των χρωμάτων και της κίνησης.



Εικόνα 28: Ο καταμερισμός του πρώτου καρτέ του βίντεο 'Iceland' που χρησιμοποιήθηκε για την υλοποίηση.

5.3.3 Αλγόριθμος χρωματικής και κινητικής παραμετροποίησης (Δεύτερο στάδιο επεξεργασίας)

Το δεύτερο στάδιο της ανάλυσης και επεξεργασίας του βίντεο πρόκειται για το κύριο πρόγραμμα και το ουσιαστικότερο μέρος της υλοποίησης. Στο πρόγραμμα αυτό καταστρώνεται η frame by frame ανάλυση του βίντεο, με βάση την δειγματοληψία του προηγούμενου σταδίου αλλά και του καταμερισμού των καρτέ.

Σε πρώτη φάση, καλούνται οι διάφορες βιβλιοθήκες που χρησιμοποιήθηκαν για να 'τρέξει' ο κώδικας:

```
import time, colorsys, StringIO, os, glob
import cv2
from sklearn.cluster import KMeans
import utils
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
from matplotlib import gridspec
```

```
from harmony import Chord
from itertools import groupby
from testharm import *
from mingus.containers import Track
from mingus.containers import Bar
from mingus.midi import midi_file_out
```

```
clusters = 5
beats = 0
```

Στην συνέχεια γίνεται μία αρχικοποίηση (μεταβλητών) τιμών και επιλέγεται από τον χρήστη ο αριθμός των χρωματικών clusters (συστοιχίες) κατά τον οποίο πρόκειται να χωριστούν οι εικόνες μετά την εκτέλεση του προγράμματος.⁵

```
firstFrame0 = None
firstFrame1 = None
firstFrame2 = None
(...)
firstFrame8 = None
beat_end = False
count = 0
```

Καλείται συναρτησιακά με την σειρά του το υποπρόγραμμα εναρμόνισης με την μορφή **play = Instrument()** και **play2 = Instrument()** για τα τεμάχια (μελωδία) και τα καρέ (χαλί) αντίστοιχα και καθορίζονται οι αρχικές παραμέτρους κατά το οποίο θα κατασκευαστεί το μουσικό κομμάτι όπως το tempo και η εκκίνηση του μέτρου.

```
play = Instrument()
play2 = Instrument()
play2.tempo = 120
play.tempo = 120
play.whole = False
play.sixteenth = False
play.eighth = False
play.quater = False
play.half = False
play.rest = False
play.done = False
play.newbar = True
t = Track()
t1 = Track()
b0 = Bar()
b0.place_notes(None, 1)
```

⁵ Θα επεξηγηθεί παρακάτω ο λόγος που επιλέχθηκε αυτός ο αριθμός clusters (ομάδων).

Ο υπολογιστής εκκινεί την διαδικασία σειριακής ανάλυσης των εικόνων που ανιχνεύει μέσα στον φάκελο **framez** με τον δείκτη **image** που εξισώνεται σε κάθε επανάληψη (με κάθε νέα εικόνα που αντλαμβάνεται ο υπολογιστής μέσα στον φάκελο) με την εικόνα που πρόκειται να αναλυθεί. Προγραμματιστικά αυτό συμβαίνει όταν ο υπολογιστής καλεί την συνάρτηση **cv2.imread(imgName)** και της προσάπτει τον δείκτη **image**, η οποία μετατρέπεται στην συνέχεια από bgr (προκαθορισμένος χρωματικός χώρος στο `opencv`) σε `rgb`. Οκτώ ξεχωριστές υπολειτουργίες του κώδικα μελετάνε τεμάχιο στην αντίστοιχη θέση σε κάθε καρτέ για να μελετηθεί η κίνηση μεταξύ των διαδοχικών καρτέ στα σχετικά τεμάχια. Αυτό πραγματοποιείται με την βοήθεια των συναρτήσεων **gray0 = cv2.cvtColor(tile0, cv2.COLOR_BGR2GRAY)** και **gray0 = cv2.GaussianBlur(gray0, (21, 21), 0)** που αποθηκεύουν προσωρινά στην μεταβλητή **gray0** το αντίστοιχο greyscale της εικόνας που μελετάται, αφού πρώτα έχουν ομαλοποιηθεί οι σκληρές ακμές και τα άκρα της εικόνας (`gaussian blur`). Αν πρόκειται να μελετηθεί το πρώτο καρτέ, τότε το περιεχόμενο της μεταβλητής **gray0** μεταφέρεται στην μεταβλητή **firstframe0** που πρόκειται για δείκτης παρόμοιας φύσης με εκείνη του **image**. Ο υπολογιστής συγκρίνει το περιεχόμενο του **firstframe0** με εκείνο του `gray0`, το οποίο μετά το πρώτο καρτέ, διατηρεί το περιεχόμενο του προηγούμενου καρτέ. Έτσι, καλείται η συνάρτηση **frameDelta = cv2.absdiff(firstFrame0, gray0)** που υπολογίζει την 'απόλυτη τιμή της διαφοράς' κατά `pixel`, αποθηκεύοντας την στην μεταβλητή **frameDelta**. Κατόπιν καλείται η συνάρτηση **thresh = cv2.threshold(frameDelta, 25, 255, cv2.THRESH_BINARY)[1]** που μετατρέπει το περιεχόμενο της **frameDelta** σε `binary` εικόνα (δυαδικό χρώμα, 1-λευκό, 0-μαύρο) για να εξεταστούν οι περιοχές του χώρου της εικόνας που διαφέρουν από τον ίδιο χώρο του προηγούμενου καρτέ. Με την συνάρτηση **(_, cnts, _) = cv2.findContours(thresh.copy(), cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)**, ο υπολογιστής αντλαμβάνεται της περιοχές αυτές και δημιουργεί ένα περίγραμμα γύρω από αυτές (`contour tracing`).

```
# Foreach image file in /Art directory do the analysis
for imgName in sorted(glob.glob('framez/*.jpg')+glob.glob('framez/*.png'),
key=lambda f: int(filter(str.isdigit, f)) ):

    # Close Previous plots
    plt.close('all')

    # Create graphs figure analogies
    gs = gridspec.GridSpec(3,2,height_ratios=[8,1,1],width_ratios=[2,1])
    fig = plt.figure(1,figsize=(17,7),dpi=100)

    # Load image
    image = cv2.imread(imgName)
    image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)

    if count%8 == 0:
```

```

tile0 = image
gray0 = cv2.cvtColor(tile0, cv2.COLOR_BGR2GRAY)
gray0 = cv2.GaussianBlur(gray0, (21, 21), 0)
if firstFrame0 is None:
    firstFrame0 = gray0
frameDelta = cv2.absdiff(firstFrame0, gray0)
firstFrame0 = gray0

if count%8 == 1:
    tile1 = image
    gray1 = cv2.cvtColor(tile1, cv2.COLOR_BGR2GRAY)
    gray1 = cv2.GaussianBlur(gray1, (21, 21), 0)
    if firstFrame1 is None:
        firstFrame1 = gray1
    frameDelta = cv2.absdiff(firstFrame1, gray1)
    firstFrame1 = gray1

(...)
if count%8 == 7:
    tile7 = image
    gray7 = cv2.cvtColor(tile7, cv2.COLOR_BGR2GRAY)
    gray7 = cv2.GaussianBlur(gray7, (21, 21), 0)
    if firstFrame7 is None:
        firstFrame7 = gray7
    frameDelta = cv2.absdiff(firstFrame7, gray7)
    firstFrame7 = gray7

thresh = cv2.threshold(frameDelta, 25, 255, cv2.THRESH_BINARY)[1]

thresh = cv2.dilate(thresh, None, iterations=2)
(_, cnts, _) = cv2.findContours(thresh.copy(), cv2.RETR_EXTERNAL,
cv2.CHAIN_APPROX_SIMPLE)

```

Στην συνέχεια, ο κώδικας ακολουθεί μια διαδικασία σύγκρισης των χωρίων που εντοπίστηκαν με μία ελάχιστη τιμή των pixel που έχει προκαθορίσει ο χρήστης (εδώ 1500). Η τιμή αυτή πρόκειται για το κατώτερο όριο χωρίου που μπορεί να αποτελέσει η διαφορά των αντίστοιχων τεμαχίων των διαδοχικών καρέ, δηλαδή υποδηλώνει την σχέση για να εντοπιστεί η ελάχιστη κίνηση σε συγκεκριμένο σημείο της εικόνας. Συνεπώς αν κάποιο χωρίο ξεπεράσει το όριο αυτό, 'πυροδοτεί' την νότα του αντίστοιχου τεμαχίου και ο υπολογιστής επιτρέπει την προσπέλαση στην επόμενη εντολή. Αν αποδειχθεί αληθής η σύγκριση, τότε προσθέτει ένα δεκατοέκτο στην αξία της νότας. Η διαδικασία επαναλαμβάνεται ως που να αποδειχθεί ψευδής η σχέση (δεν υπάρχει η ελάχιστη κίνηση στο σημείο), ή να αλλάξει ο τόνος της νότας (αλλαγή χρώματος), ή με την αλλαγή του μέτρου.

```

trigger = False
for c in cnts:
    # if the contour is too small, ignore it
    if cv2.contourArea(c) > 1500:
        trigger = True

if trigger == True:
    beats +=1
    beat_end = False
else:
    beat_end = True

```

Επιπλέον, ο κώδικας εφαρμόζει K-means ομαδοποίηση με βάση το χρώμα των pixels καλώντας το αντικείμενο **clt = KMeans(clusters)**, μετατρέπει τον χρωματικό χώρο από rgb σε hsv, αποθηκεύει προσωρινά τα κυρίαρχα χρώματα (τόνος νότας) στον πίνακα **dominantColor** και μετατρέπει τις τιμές των saturation και value σε μία κλίμακα αντίστοιχη για τον καθορισμό οκτάβας και δυναμικών της νότας. Επιλέγεται από τον υπολογιστή το χρώμα που εκφράζεται περισσότερο ποσοστιαία στο τεμάχιο και το αποθηκεύει στην μεταβλητή **ind**.

```

# Find Dominant colors
image = image.reshape((image.shape[0] * image.shape[1], 3))
clt = KMeans(clusters)
clt.fit(image)

# Find the most Dominant color and create the second dominant color bar with s=1
and v=255 (brightest possible)
hsv = []
saturated = []
velocity = []
sat2oct = []

dominantColor = maxColorPercent = 0
for i,color in enumerate(clt.cluster_centers_):
    newColor = colorsys.rgb_to_hsv(color[0],color[1],color[2])
    hsv.append(newColor)
    velocity.append(int(round(newColor[2]*127/255)))

    sat2oct.append(int(round(newColor[1]*8)))
    saturated.append(colorsys.hsv_to_rgb(newColor[0],1,255))

```

```
if hist[i] > maxColorPercent:
    maxColorPercent = hist[i]
    dominantColor = hsv[i]
    ind = i
```

Ο κώδικας συμπληρώνει τα στοιχεία που άντλησε προηγουμένως στις μεταβλητές καθορισμού τόνου, οκτάβας και δυναμικής της κλάσης αρμονίας (class Instrument, εδώ ως play). Οι εντολές που ακολουθούν καθορίζουν την αλλαγή τόνου στο έκαστο τεμάχιο στην περίπτωση που αλλάξει το κυρίαρχο χρώμα.

```
play.root = ind
play.oct0 = sat2oct[ind]
play.dynamics0 = velocity[ind]

if count%16 == 0:
    previous_root = ind
    buff = 0

if previous_root != ind:
    beat_end = True
    play.root = previous_root
    play.oct0 = sat2oct[previous_root]
    play.dynamics0 = velocity[previous_root]
    previous_root = ind
```

Με την καθολική μεταβλητή **beat_end** ορίζεται εάν και πότε πρέπει ο υπολογιστής να αποτυπώσει στο μέτρο την νότα που παρήχθη. Αυτό συμβαίνει όταν αλλάξει ο τόνος (δηλαδή το χρώμα) ή όταν πρόκειται να υπάρξει κίνηση (όταν δεν ξεπεραστεί το ελάχιστο όριο χωρίου κίνησης) και έπειτα μηδενίζεται πάλι. Στην συνέχεια, ο υπολογιστής προσθέτει την αντίστοιχη αξία στο μέτρο με αυτή του αριθμού δεκατοέκτων που έχουν καθοριστεί από την μεταβλητή **beats**. Παρακάτω φαίνεται ο καθορισμός της αξίας του φθόγγου στο υποπρόγραμμα αρμονίας για **beats = 4** (4 δεκαταέκτα/1 τέταρτο) και **beats = 8** (8 δεκαταέκτα/1 μισό).

```
if beat_end == True:

    if beats == 4:
        play.quarter = True
        play.eighth = False
        play.half = False
        play.sixteenth = False
```

```
play.whole = False

elif beats == 8:
    play.half = True
    play.sixteenth = False
    play.quater = False
    play.eighth = False
    play.whole = False
```

Το κάθε μέτρο συμπληρώνεται στον κώδικα μέσω του κλάσης αρμονίας με την συνάρτηση **opbar = play.bar(b)** και στην συνέχεια ο υπολογιστής επαναφέρει τις μεταβλητές **beats** και **beat_end**. Εάν γίνει προσπέλαση για 16 τεμάχια (δύο καρέ) τότε συμπληρώνεται το μέτρο σε μία στοίβα για να προστεθεί καινούργιο μέτρο στο κομμάτι αργότερα και υπολογιστής εκκινεί την διαδικασία για την κατασκευή νέου μέτρου.

```
opbar = play.bar(b)
beats = 0
beat_end == False

if count%16 == 15:
    t.add_bar(b)
```

Τέλος, ο υπολογιστής κατασκευάζει το midi αρχείο που προκύπτει από τις παραπάνω λειτουργίες.

```
midi_file_out.write_Track('iceland-melody.MIDI', t)
```

Η διαδικασία επαναλαμβάνεται για το πολυφωνικό μέρος, με την διαφορά ότι η ανάλυση γίνεται σε όλο το καρέ με κάθε επανάληψη και συνεπώς προστίθεται σε κάθε νέο μέτρο μία συγχορδία πέντε νοτών (πέντε clusters-πέντε κυρίαρχα χρώματα).

Το κύριο πρόγραμμα επιτελεί και άλλες λειτουργίες ανάλυσης της εικόνας οι οποίες όμως, δεν αξιοποιήθηκαν στα πλαίσια τις εργασίας αλλά θα αναλυθούν στο επόμενο κεφάλαιο.

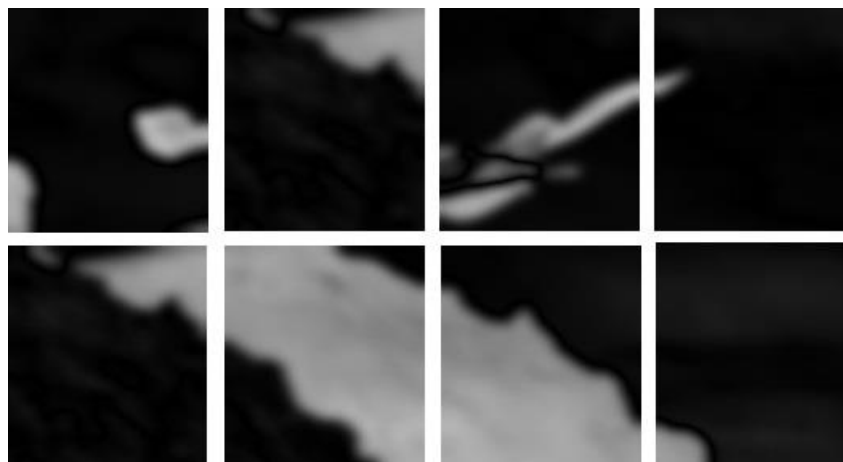
5.3.4 Επεξήγησης αλγορίθμου χρωματικής και κινητικής παραμετροποίησης

Η επιλογή της εικόνας από τον υπολογιστή (είτε τεμάχιο, είτε ολόκληρο καρέ) γίνεται με σειριακό τρόπο με αριθμητική σειρά. Η αρίθμηση των εικόνων αυτών έχει προκύψει από τον αλγόριθμο δειγματοληψίας, οι οποίες στην συνέχεια αποθηκεύονται με την ίδια σειρά στους φακέλους.

Οι διαφορές από καρέ σε καρέ, που συμβάλλουν στον εντοπισμό της κίνησης προκύπτουν με τέτοιο τρόπο ώστε να μειωθεί ή αφαιρεθεί αρχικά τυχόν θόρυβος με την συνάρτηση ομαλοποίησης (gaussian blur) σε κάθε τεμάχιο, για να μην προκύψουν στίγματα και χωρία που μπορεί ο υπολογιστής να τα αντιληφθεί ως κίνηση, χωρίς όμως να είναι, καθώς και να εντοπίζει με μεγαλύτερη ευκολία τα περιγράμματα των χωρίων. Ο υπολογιστής αντιλαμβάνεται την κάθε εικόνα που αναλύει ως έναν πίνακα (array) με τις τιμές των pixels σε κάθε σημείο ως τα στοιχεία αυτού του πίνακα. Συνεπώς, όταν υπολογίζει την απόλυτη διαφορά των εικόνων, ουσιαστικά αφαιρεί τις τιμές των pixels της μίας εικόνας από την άλλη. Αυτό που προκύπτει οπτικά είναι η αποτύπωση του 'αρνητικού' των εικόνων, δηλαδή τα σημεία στα οποία διαφέρουν οι εικόνες είναι φωτεινά, ενώ τα σημεία που είναι όμοια παραμένουν σκοτεινά (όπως φαίνεται στην εικόνα 29). (Howse, Joshi, & Beyeler, 2016) Οι εικόνες, αφού έχουν μετατραπεί σε αντίστοιχες greyscale εικόνες, έχουν περάσει από το φίλτρο της ομαλοποίησης και υπολογίστηκε η απόλυτη διαφορά τους, μετατρέπονται σε binary. Δηλαδή, τα φωτεινά σημεία της εικόνας έχουν μοναδική τιμή και απόχρωση, εκείνη του λευκού χρώματος, ενώ τα σκοτεινά αντίστοιχα αποτυπώνονται αποκλειστικά με μαύρο χρώμα. Ο υπολογιστής λαμβάνει ως είσοδο μια εικόνα πηγής (*src*) και μια τιμή κατωφλίου (*thresh*) και παράγει μια εικόνα εξόδου (*dst*) συγκρίνοντας την τιμή του pixel πηγής (x, y) με το κατώφλι. Εάν $src(x, y) > thresh$, τότε για την εικόνα εξόδου $dst(x, y)$ ορίζει την τιμή του αντίστοιχου pixel της στο *maxValue*, διαφορετικά το θέτει στο μηδέν. Εδώ το *maxValue* είναι 255 (μέγιστη τιμή φωτεινότητας-λευκό). Όταν παραχθεί η επιθυμητή δυαδική εικόνα, διαστέλλεται για να μεγαλώσει η ευκρίνεια των χωρίων κίνησης και στην συνέχεια ανιχνεύονται οι περιμετροί των χωρίων από τον υπολογιστή.

Εφαρμόζεται μία λογική συνθήκη στο σημείο αυτό του κώδικα, που καθορίζει ποιες νότες θα παιχτούν και τι διάρκεια θα έχουν. Εάν ο αριθμός των pixels στην οριοθετημένη περιοχή (που περικλείεται από το περίγραμμα το οποίο βρέθηκε προηγουμένως) ξεπερνάει τα 1500 (π.χ. 42x60, 50x50, κλπ.) τότε επιτρέπει στην προσθήκη αξίας στον φθόγγο που θα παιχτεί. Έτσι, εάν δεν υπάρχει ιδιαίτερη κίνηση σε κάποια σημεία του καρέ και εφόσον δεν υπάρχει προσθήκη αξίας προηγουμένως, δηλαδή ο οριοθετημένος χώρος *contour* να είναι λιγότερος των 1500 pixels ($contour < 1500$), δεν θα παραχθεί νέος φθόγγος και θα συλληφθεί ως παύση στο μουσικό κομμάτι. Ειδάλλως, εφόσον έχει υπάρξει προσθήκη αξίας από το προηγούμενο τεμάχιο και είναι $contour > 1500$, τότε θα προσαυξηθεί η αξία της νότας κατά ένα δέκατο-έκτο και θα περιμένει στην σε μία στοίβα μέχρι να

υπάρξει παύση όπου και γίνεται η προσθήκη αυτή και της νότας στο μέτρο. Συνεπώς για κάθε εικόνα που ισχύει η σχέση $contour > 1500$ προσαυξάνεται ένα δέκατο-έκτο στην αξία μιας νότας που πρόκειται να παιχτεί, εάν η νότα αυτή δεν έχει καμία αξία τότε αυτή θα γίνει αξίας ενός δέκατου-έκτου.



Εικόνα 29: Η απόλυτη διαφορά των καρέ (και ξεχωριστών τεμαχίων) των καρέ 55 και 56. Αντιπροσωπεύει την κίνηση μεταξύ αυτών των καρέ.

Συνολικά, η εφαρμογή του K-means αποδίδει K ξεχωριστές ομάδες (clusters) των αρχικών δεδομένων σημείων n. Τα σημεία αυτά, μέσα σε μία συγκεκριμένη ομάδα θεωρούνται ότι είναι "περισσότερο πανομοιότυπα" το ένα με το άλλο σε σχέση με τα σημεία που ανήκουν σε άλλη ομάδα. Εδώ έχουν επιλεχθεί πέντε ομάδες λόγω του ότι μετά από διάφορες δοκιμές, φάνηκε να είναι η βέλτιστη τιμή για την ποσοτικοποίηση των χρωμάτων. Αυτό διότι άνω των 5 ομάδων συνήθως εμφανιζόταν χρώμα που ενώ υπήρχε στην εικόνα ή στο τεμάχιο σε κάποια αναλογία, ήταν πολύ μικρή. Κατά συνέπεια, ο υπολογιστής έχοντας το δεδομένο χρώμα και τις αντίστοιχες εντολές, πρόσθετε στην 'μελωδία' την νότα που αντιστοιχούσε στο ασθενές χρώμα και ως επίπτωση, δεν υπήρχε σωστή αναλογία ήχου και χρώματος διότι οι δυναμικές της νότας ακολουθούν την παράμετρο της φωτεινότητας⁶ και όχι εκείνη της χρωματικής αναλογίας⁷ στην εικόνα.

⁶ Για λόγους που έχουν εξηγηθεί στην προηγούμενη ενότητα.

⁷ Η χρωματική αναλογία εκφράζεται στην συνέχεια μέσω των τεμαχίων της εικόνας, στις οποίες επίσης γίνεται K-means ποσοτικοποίηση για να βρεθεί το κυρίαρχο χρώμα και συνεπώς την νότα που το εκπροσωπεί.

5.3.5 Αλγόριθμος αρμονίας

Κατά την διάρκεια της εκτέλεσης του κύριου προγράμματος, καλείται συχνά ο αλγόριθμος που δημιουργήθηκε για να κατασκευάζει την αρμονία της μουσικής καθώς και την μουσική την ίδια. Ο αλγόριθμος αυτός πρόκειται για μία κλάση εν ονόματι Instrument. Κάθε φορά που δέχεται ορίσματα που αφορούν το τονικό ύψος, την οκτάβα, τις δυναμικές, το μέτρο, κλπ., κατασκευάζει ένα αντικείμενο με τις παραμέτρους αυτές. Το αντικείμενο αυτό πρόκειται αρχικά για την προσθήκη νότας στο μέτρο και στην συνέχεια το κάθε μέτρο που θα προστεθεί στο τελικό κομμάτι και αντιστοιχεί σε δύο καρτέ. Αφενός, οι λειτουργίες που αφορούν την κατασκευή του κομματιού σε πρακτικό επίπεδο βρίσκονται σε αυτό το κομμάτι του συνολικού κώδικα και αφετέρου, το τελικό προϊόν (midi αρχείο) είναι αποτέλεσμα των δεδομένων που παρέχονται από το κομμάτι αυτό του κώδικα.

Αρχικά, ο υπολογιστής καλεί όλες τις βιβλιοθήκες που χρησιμοποιεί στην συνέχεια του κώδικα.

```
from mingus import *
from mingus.midi.fluidsynth import FluidSynthSequencer
from mingus.midi import fluidsynth
from mingus.containers import *
import time
import mingus.core.scales as scales
import random
import mingus.core.intervals as intervals
import mingus.core.chords as chords
from mingus.midi import midi_file_out
```

Στην συνέχεια, ορίζονται οι λειτουργίες της κλάσης για να μπορέσει να κατασκευαστεί το αντικείμενο, στην προκειμένη το κάθε μέτρο.

```
class Instrument:

    def __init__(self):

    def settempo(self, tempo):
        self.tempo = tempo

        print('tempo set at:', tempo)

    def settimbre(self, timbre):
        self.timbre = timbre

        print('timbre set at:', timbre)
```

```

def setrootnote(self, root):
    self.root = root
    Raw = ['G', 'D', 'A', 'E', 'B', 'Gb', 'Db', 'Ab', 'Eb', 'Bb', 'F', 'C']
    print('Root note set at:', Raw[self.root])

def veloc(self, dynamics0, dynamics1, dynamics2, dynamics3, dynamics4,
dynamics5, dynamics6):
    self.dynamics0 = dynamics0
    self.dynamics1 = dynamics1
    self.dynamics2 = dynamics2
    self.dynamics3 = dynamics3
    self.dynamics4 = dynamics4
    self.dynamics5 = dynamics5
    self.dynamics6 = dynamics6

def octave(self, oct0, oct1, oct2, oct3, oct4, oct5, oct6):
    self.oct0 = oct0
    self.oct1 = oct1
    self.oct2 = oct2
    self.oct3 = oct3
    self.oct4 = oct4
    self.oct5 = oct5
    self.oct6 = oct6

def note_value(self, rest, half, quarter, eighth, sixteenth):
    self.half = half
    self.quarter = quarter
    self.eighth = eighth
    self.sixteenth = sixteenth
    self.rest = rest
def newbar(self, new, done):
    self.new = new

    self.done = done
    return d

```

Όλα τα παραπάνω ορίσματα καθορίζουν το πως θα ακουστεί ο φθόγγος (δυναμικές, οκτάβα, τόνος), την αξία που θα έχει, καθώς και την τοποθέτηση του στο μέτρο, την ολοκλήρωση και προσθήκη του μέτρου στο κομμάτι και την δημιουργία νέου, με βάση τα δεδομένα από το κύριο πρόγραμμα. Στο επόμενο μέρος του κώδικα όμως χρησιμοποιεί τις παραπάνω δομές δεδομένων για να υλοποιηθεί στην πράξη το ζητούμενο αποτέλεσμα. Χρησιμοποιούνται οι συναρτήσεις κατασκευής και τοποθέτησης φθόγγου στο μέτρο της βιβλιοθήκης `mingus`. Το πρώτο μέρος αφορά την μονοφωνική μελωδία που παράγεται με βάση το κυρίαρχο χρώμα του κάθε τεμαχίου:

```

def bar(self, b):
    self.b = b

    Raw = ["G", "D", "A", "E", "B", "Gb", "Db", "Ab", "Eb", "Bb", "F", "C"]

    fluidsynth.set_instrument(1, self.timbre)

    n0 = Note(Raw[self.root])
    n0.octave = self.oct0
    n0.velocity = self.dynamics0

    if self.sixteenth == True:

        b.place_notes(n0, 16)
    elif self.eighth == True:
        b.place_notes(n0, 8)
    elif self.quater == True:
        b.place_notes(n0, 4)
    elif self.half == True:
        b.place_notes(n0, 2)
    elif self.sixteenth == True and self.eighth == True:
        b.place_notes(n0, 8)
        b.place_notes(n0, 16)
    elif self.sixteenth == True and self.quater == True:
        b.place_notes(n0, 4)
        b.place_notes(n0, 16)
    elif self.eighth == True and self.quater == True:
        b.place_notes(n0, 4)
        b.place_notes(n0, 8)
    elif self.eighth == True and self.quater == True and self.sixteenth:
        b.place_notes(n0, 4)
        b.place_notes(n0, 8)
        b.place_notes(n0, 16)
    elif self.whole == True:
        b.place_notes(n0, 1)

    if self.rest == True and self.whole == False:
        b.place_rest(16)

    if self.done == True:
        return b
    self.done = False

```

Το δεύτερο κομμάτι αφορά το πολυφωνικό μέρος κατασκευής συγχορδιών, το λεγόμενο 'χαλί' για την μελωδία, που προκύπτει από τα πέντε κυρίαρχα χρώματα του καρέ:

```
def chordz(self, second, third, fourth, fifth):

    self.second = second
    self.third = third
    self.fourth = fourth
    self.fifth = fifth

    Raw = ["G", "D", "A", "E", "B", "Gb", "Db", "Ab", "Eb", "Bb", "F", "C"]

    n0 = Note(Raw[self.root])
    n0.octave = self.oct0
    n0.velocity = self.dynamics0

    n1 = Note(Raw[self.second])
    n1.octave = self.oct0
    n1.velocity = self.dynamics0

    n2 = Note(Raw[self.third])
    n2.octave = self.oct0
    n2.velocity = self.dynamics0

    n3 = Note(Raw[self.fourth])
    n3.octave = self.oct0
    n3.velocity = self.dynamics0

    n4 = Note(Raw[self.fifth])
    n4.octave = self.oct0
    n4.velocity = self.dynamics0

    n = NoteContainer()
    n.add_note(n0)
    n.add_note(n1)
    n.add_note(n2)
    n.add_note(n3)
    n.add_note(n4)

    br = Bar()

    br.place_notes(n, 2)

    return br
```


6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό θα μελετηθούν τα αποτελέσματα εφαρμογής των αλγορίθμων παραγωγής μουσικής από βίντεο, τα συμπεράσματα που μπορούν να αντληθούν από τα αποτελέσματα αυτά καθώς και κάποιες προτάσεις για την επέκταση, την τροποποίηση και βελτίωση των αλγορίθμων.

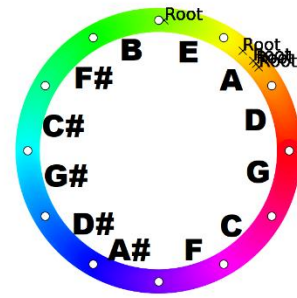
6.1 ΑΠΟΤΕΛΕΣΜΑΤΑ ΥΛΟΠΟΙΗΣΗΣ

Με την εφαρμογή της επεξεργασίας και ανάλυσης στο βίντεο προέκυψε επιτυχώς το κομμάτι μουσικής που του αντιστοιχεί. Λόγω της χρωματικής δομής του βίντεο, παρατηρήθηκαν πολλές συμπληρωματικές αποχρώσεις του κόκκινου (έδαφος, χώμα, πετρώματα, κλπ.), του μπλε (ουρανός, θάλασσα) και του πράσινου (δέντρα, χορτάρι, κλπ.) μιας και το βίντεο δείχνει διάφορα τοπία της Ισλανδικής φύσης. Αντίστοιχα, οι αξίες που διατυπώθηκαν υπήρξαν σχετικά μεγάλες και μακριές, όπως και αραιές εντός του μέτρου, μιας και τα πλάνα στο συγκεκριμένο βίντεο ήταν αρκετά στατικά, χωρίς πολλή κίνηση. Αντίστοιχα, λόγω του φωτισμού των πλάνων (ασθενής φωτισμός, χαμηλός φυσικός φωτισμός) και του χαμηλού contrast των χρωμάτων της φύσης της Ισλανδίας, πολλά από τα καρέ ήταν χαμηλού κορεσμού (saturation) και συνεπώς λόγω του mapping που έχει γίνει στον αλγόριθμο, υπήρχαν αρκετές χαμηλής οκτάβας νότες που όμως εναλλάσσονται σε κάποια πιο κορεσμένα πλάνα με μεσαίες ή ψηλές νότες.

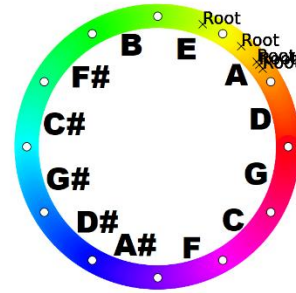
Παρακάτω φαίνεται η κυματομορφή όπως προκύπτει με τα μονοφωνικά και πολυφωνικά μέρη μαζί (εικόνα 30). Παρατηρείται γύρω από το 01:15 λεπτό μία απότομη αύξηση στην ένταση του ήχου, ενώ μετά το 01:30 λεπτό ξαναπέφτει η ένταση του ήχου. Αυτό συμβαίνει διότι, στο 01:09 λεπτό μεταξύ των καρέ του βίντεο αυξάνεται επίσης απότομα η φωτεινότητα των χρωμάτων (εικόνα 31), ή μειώνεται αντίστοιχα στο 01:25.



Εικόνα 30: Η κυματομορφή ήχου της μουσικής που προκύπτει συνολικά μετά την εφαρμογή του αλγορίθμου.



1 2 3 4 5

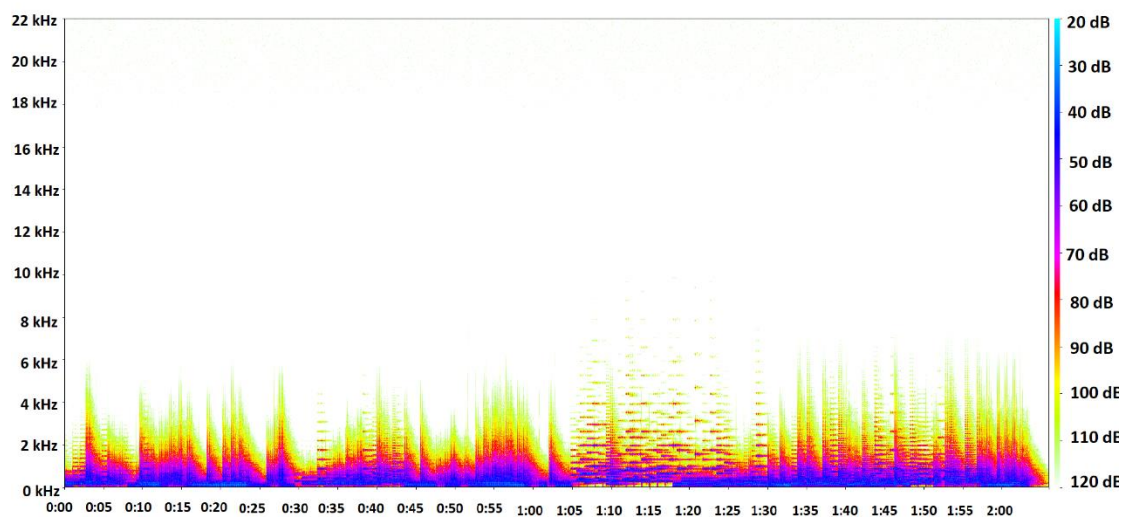


1 2 3 4 5



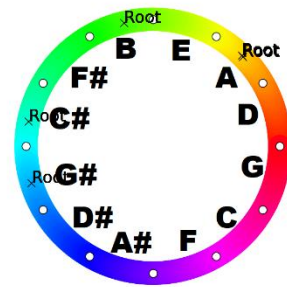
Εικόνα 31: Η ανάλυση των καρέ 77 και 78 που αντιστοιχούν περίπου στο 01:09 λεπτό του βίντεο. Είναι εμφανής η αλλαγή της φωτεινότητας από το ένα καρέ στο άλλο.

Επιπρόσθετα, με το φασματογράφημα της εικόνας 32 μπορεί να διακρίνει κανείς πως μεταξύ των 1:05 και 1:17 λεπτών περίπου, οι χαμηλές συχνότητες εξαφανίζονται και οι μεσαίες έχουν αρκετή ένταση. Αυτό συμβαίνει διότι τα χρώματα των καρέ στο διάστημα αυτό είναι επίσης αρκετά κορεσμένα (saturated), όπως και φαίνεται στα παραπάνω καρέ. Επίσης, είναι εμφανές πως το κομμάτι έχει κυρίως νότες χαμηλού συχνотικού επιπέδου, διότι όπως προαναφέρθηκε, τα χρώματα του βίντεο δεν είναι ιδιαίτερα saturated και ως επί το πλείστον προκύπτουν οι χαμηλές νότες.

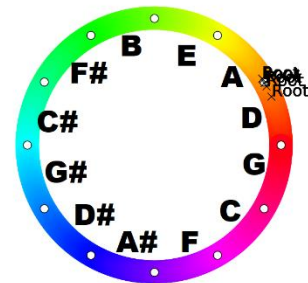


Εικόνα 32: Φασματογράφημα του κομματιού που κατασκεύασε ο υπολογιστής. Ο αριστερός κατακόρυφος άξονας εκφράζει την συχνότητα (kHz) συναρτήσει του χρόνου (οριζόντιος άξονας) και δεξιά μια χρωματική κλίμακα για την ένταση του ήχου (σε dB) συναρτήσει του χρόνου.

Αρμονικά, τα αποτελέσματα ποικίλουν ώστε ορισμένες φορές να παρατηρείται κάποιο είδος κλασικής αρμονίας ή κάποιος άλλος συνηθισμένος μουσικός ήχος, χωρίς όμως να περιορίζονται οι ήχοι σε συγκεκριμένες μουσικές φόρμες. Για παράδειγμα, στην εικόνα 33, γίνεται εύκολα αντιληπτό πως στο πρώτο καρέ η ανάλυση δείχνει μία συγχορδία Λα μείζονα με την προσθήκη της έβδομης και ένατης βαθμίδας (Σολ δίεση και Σι αντίστοιχα). Ενώ στο αμέσως επόμενο καρέ όλες οι νότες επικεντρώθηκαν στην Λα, με αποτέλεσμα να ακούγεται μόνο εκείνη η νότα.



1 2 3 4 5



1 2 3 4 5



Εικόνα 33: Η ανάλυση δύο τυχαίων διαδοχικών καρέ και η αντίστοιχη αρμονία που προκύπτει.

6.2 ΠΡΟΣΘΗΚΕΣ ΑΛΓΟΡΙΘΜΟΥ & ΕΠΙΠΡΟΣΘΕΤΕΣ ΛΕΙΤΟΥΡΓΙΕΣ

Σχεδιάζοντας τον συνολικότερο αλγόριθμο και καθορίζοντας τις βασικές λειτουργίες που θα επιτελούσε, υπήρχαν διάφορες προσεγγίσεις και προτάσεις οι οποίες ενδεχομένως να οδηγούσαν μία μορφή ηχητικοποίησης του βίντεο εισόδου. Προφανώς ο καθορισμός των δεδομένων, η αντιστοίχιση και η παραμετροποίηση τους, οδήγησε σε ένα θεμιτό αποτέλεσμα στο οποίο γίνεται αντιληπτό εύκολα το χρωματικό περιεχόμενο αλλά και το κινησιολογικό περιεχόμενο του βίντεο, ακόμη και αν δεν υπάρχει εκείνη η στιγμή της ακρόασης το αντίστοιχο οπτικό μέρος του γενικότερου προϊόντος της εφαρμογής. Όμως κάποιες από τις προτάσεις αυτές, ενώ δοκιμάστηκαν στον αλγόριθμο, δεν αξιοποιήθηκαν ιδιαίτερα. Παρακάτω γίνεται μία αναφορά σε αυτές.

6.2.1 Χρωματική αρμονία

Έχει προταθεί ότι «τα χρώματα που βλέπουμε από κοινού για να παραχθεί μια ευχάριστη συναισθηματική απόκριση λέγεται ότι είναι σε αρμονία» (Burchett, 2002). Ωστόσο, η αρμονία χρώματος είναι μια σύνθετη αντίληψη επειδή οι ανθρώπινες αποκρίσεις στο χρώμα είναι τόσο συναισθηματικές όσο και γνωστικές, με συναισθηματική απόκριση και κρίση. Ως εκ τούτου, η ανθρώπινη αντίδραση στο ερέθισμα του χρώματος και η αντίληψη της χρωματικής αρμονίας είναι ανοικτές στην επίδραση διαφόρων παραγόντων. Αυτοί οι παράγοντες περιλαμβάνουν μεμονωμένες διαφορές (όπως ηλικία, φύλο, προσωπική προτίμηση, συναισθηματική κατάσταση κλπ.), καθώς και πολιτισμικοί παράγοντες και κοινωνικές διαφορές, οι οποίες δημιουργούν κλινικές και μνημένες αποκρίσεις σχετικά με το χρώμα. Επιπλέον, το πλαίσιο μέσα στο οποίο γίνεται αντιληπτό το χρώμα έχει πάντα επιρροή στις αποκρίσεις σχετικά με το χρώμα και την έννοια της χρωματικής αρμονίας και αυτή η έννοια επηρεάζεται επίσης από χρονικούς παράγοντες (όπως μεταβαλλόμενες τάσεις) και αντιληπτικούς παράγοντες (όπως ταυτόχρονη αντίθεση). Το ακόλουθο εννοιολογικό μοντέλο απεικονίζει αυτήν την προσέγγιση του 21ου αιώνα για την αρμονία των χρωμάτων:

$$Color\ harmony = f(Col\ 1,2,3, \dots, n) \times (AD + PE + EP + A + T)$$

Όπου η χρωματική αρμονία είναι μια συνάρτηση (f) της αλληλεπίδρασης μεταξύ των χρωμάτων ($Col\ 1, 2, 3, \dots, n$) και των παραγόντων που επηρεάζουν τη θετική αισθητική απόκριση στο χρώμα: ατομικές διαφορές (AD) όπως η ηλικία, προσωπικότητα και συναισθηματική κατάσταση, πολιτισμικές εμπειρίες (PE), το επικρατούμενο πλαίσιο (EP) που περιλαμβάνει φωτισμό περιβάλλοντος και γενικότερου φωτισμού, οι παρεμβατικές επιδράσεις της αντίληψης (A) και τις συνέπειες του χρόνου (T) σε σχέση με τις επικρατούσες κοινωνικές τάσεις. (O'Connor, 2010)

Κάποια συνηθισμένα αρμονικά σχήματα χρωμάτων είναι τα εξής:

- Συμπληρωματικά χρώματα - Τα χρώματα που είναι απέναντι στον χρωματικό κύκλο θεωρούνται συμπληρωματικά χρώματα. Δημιουργούν την μεγαλύτερη αντίθεση και ως εκ τούτου την μεγαλύτερη οπτική ένταση λόγω του πόσο διαφορετικά είναι τα χρώματα μεταξύ τους. (π.χ. κόκκινο και πράσινο).
- Ανάλογα χρώματα - Η απλούστερη και πιο σταθερή αρμονία είναι αυτή των ανάλογων χρωμάτων. Αποτελείται από ένα χρώμα ρίζα και δύο ή περισσότερα κοντινά χρώματα. Αποτελεί τη βάση ενός χρωματικού σχήματος και στην πράξη πολλά χρωματικά σχήματα είναι συνδυασμοί

ανάλογων και συμπληρωματικών αρμονιών προκειμένου να επιτευχθεί τόσο το οπτικό ενδιαφέρον μέσω της ποικιλίας ή της χρωματικής σταθερότητας και της χρωματικής έντασης μέσω της αντίθεσης των χρωμάτων.

- Μερικώς συμπληρωματικά χρώματα – Τα μερικώς συμπληρωματικά χρώματα είναι όμοια με τα συμπληρωματικά με την διαφορά του ότι ένα από τα δύο συμπληρώματα χωρίζεται στα δύο κοντινά ανάλογα χρώματα του χρωματικού κύκλου. Αυτό διατηρεί την ένταση εκείνη των συμπληρωματικών χρωμάτων ενώ εισάγει ταυτόχρονα περισσότερο οπτικό ενδιαφέρον με περισσότερη ποικιλία.
- Τριάδες χρωμάτων - Ομοίως με τα μερικώς συμπληρωματικά χρώματα που αναφέρθηκαν παραπάνω, οι χρωματικές τριάδες περιλαμβάνουν τρία χρώματα σε μια γεωμετρική σχέση. Αντίθετα όμως με τα μερικώς συμπληρωματικά χρώματα, και τα τρία χρώματα ισαπέχουν το ένα από το άλλο στον χρωματικό κύκλο με αποτέλεσμα να δημιουργούν ένα ισόπλευρο τρίγωνο. Η πιο κοινή τριάδα χρώματος είναι τα πρωταρχικά χρώματα (κόκκινο, κίτρινο, μπλε).

Ενώ στον αλγόριθμο το παραπάνω μοντέλο χρησιμοποιήθηκε αμιγώς για το τελικό αποτέλεσμα, δεν αξιοποιήθηκε πλήρως. Το μέρος του κώδικα που ανιχνεύει την χρωματική αρμονία είναι το παρακάτω:

```
for i in range(len(note_index)):
    for j in range(4,i,-1):
        for k in range(0,12):
            if note_index[i] == k:

                if note_index[j]== (k + 6)%12:
                    print('complementary')
                    Comp = True

                if note_index[j] == (k + 1)%12:
                    print('analogous counter clockwise')
                    AnalogCC = True

                if note_index[j] == abs(k + 12 - 1)%12:
                    print('analogous clockwise')
                    AnalogC = True

                if note_index[j] == (k + 4)%12 or note_index[j] == (k +
3)%12:
                    for l in range(0,j):
                        if note_index[l]== (k+8)%12 or
note_index[l]== (k+7)%12:
```

```

print('triadic')
Triad = True

if note_index[j] == (k + 2)%12:
    for h in range(0,j):
        if note_index[h]== (k+5)%12 or
note_index[h]== (k+6)%12:
            print('split complementary counter
clockwise')

            SplitcomCC = True
if note_index[j] == (k - 2 + 12)%12:
    for p in range(0,j):
        if note_index[p]== (k - 5 + 12)%12 or
note_index[p]== (k- 6 +12)%12:
            print('split complementary
clockwise')

            Splitcom = True

if note_index[j] == (k + 2)%12 or note_index[j] == (k +
3)%12:
    for u in range(0,j):
        if note_index[u]== (k+4)%12 or
note_index[u]== (k+5)%12 or note_index[u]== (k+6)%12:
            for t in range(u,0):
                if note_index[t]==
(k+7)%12 or note_index[t]== (k+8)%12 or note_index[t]== (k+9)%12:
                    print('tetradic')

Tetrad = True

```

Το παραπάνω κομμάτι κώδικα ουσιαστικά εντοπίζει τα χρώματα στον χρωματικό κύκλο και καθορίζει την χρωματική αναλογία με βάση την διάταξή τους σε αυτόν αλλά και την θέση του μέγιστου ποσοστιαία χρώματος.

6.2.2 Θερμά & ψυχρά χρώματα

Η αντίθεση μεταξύ των θερμών και ψυχρών χρωμάτων, φαίνεται να σχετίζεται με την παρατηρούμενη αντίθεση του φωτός σε κάποιο τοπίο, μεταξύ των "θερμών" χρωμάτων που σχετίζονται με το φως της ημέρας ή το ηλιοβασίλεμα και των "ψυχρών" χρωμάτων του χειμώνα ή των νεφώσεων. Τα θερμά χρώματα συχνά λέγεται ότι είναι αποχρώσεις χρωμάτων από το κόκκινο έως το κίτρινο, το καφέ και των σκουρόχρωμων χρωμάτων που γίνονται αντιληπτά ως αποχρώσεις του καφέ. Τα ψυχρά χρώματα συχνά λέγεται ότι είναι οι αποχρώσεις χρωμάτων από το μπλε-πράσινο μέχρι το μπλε-ιώδες, συμπεριλαμβανομένων των περισσότερων γκριζών

συναφών χρωμάτων. Υπάρχει ιστορική διαφωνία σχετικά με τα χρώματα που αγκυροβολούν την πολικότητα αυτή, αλλά οι πηγές του 19ου αιώνα καθόρισαν την κορυφαία αντίθεση του είδους αυτή μεταξύ κόκκινου πορτοκαλί και πρασινωπού μπλε.

Το κομμάτι του κώδικα που εκπληρώνει αυτόν τον διαχωρισμό είναι το παρακάτω:

```
warm = 0
cool = 0
major = False
minor = False

for i in range(len(note_index)):

    if note_index[i] in range(0,5) or note_index[i]==11:
        #print('warm')
        warm += 1
    elif note_index[i] in range(5,11):
        #print('cool')
        cool += 1

if warm > cool:
    major = True
else:
    minor = True
```

Ο απλός αυτός κώδικας χωρίζει τον χρωματικό κύκλο σε δύο ίσα μέρη των έξι χρωμάτων. Τα πρώτα έξι χρώματα αντιστοιχούν στο 'θερμό' μέρος και τα υπόλοιπα έξι στο 'ψυχρό' μέρος. Εντοπίζει σε πιο μέρος βρίσκεται η πλειοψηφία των χρωμάτων και αναλόγως πραγματοποιεί κάποια αρμονική αντιστοιχία (εδώ έχει γίνει αντιστοιχία πλειοψηφίας θερμών χρωμάτων σε μείζονα κλίμακα και πλειοψηφίας ψυχρών χρωμάτων σε ελάσσονα κλίμακα).

6.3 ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΕΠΕΚΤΑΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ

Με μία σύντομη περιγραφή, ο παραπάνω αλγόριθμος πρόκειται για μία συγκεκριμένη υλοποίηση της ηχητικοποίησης βιντεοσκοπημένου υλικού, βασισμένη στην ανάλυση, παραμετροποίηση και χαρτογράφηση του χρώματος και της κίνησης. Οι τεχνικές που χρησιμοποιήθηκαν αποδείχθηκαν αρκετά αποτελεσματικές σε βαθμό τέτοιο που επιτρέπει σε οποιονδήποτε χρήστη να 'αξιοποιήσει' τις δυνατότητες του σύγχρονου υπολογιστή και να δημιουργήσει μία καλλιτεχνική ή εναλλακτική προσέγγιση χρήσης και αξιοποίησης οπτικού υλικού. Τα μουσικό προϊόν του αλγορίθμου μπορεί να χρησιμοποιηθεί είτε ανεξάρτητα ως μουσικό κομμάτι για κάποιον που επιθυμεί να συνθέσει μουσική κατά αυτόν τον τρόπο, είτε ως επένδυση του οπτικού υλικού ενός ψηφιακού βίντεο με σκοπό την συσχέτιση του οπτικού και ακουστικού περιεχομένου. Ωστόσο, έχοντας ήδη αρκετές δυνατότητες, προκύπτει και η ανάγκη επέκτασης των δυνατοτήτων αυτών και της δημιουργίας ακόμα περισσότερων.

Αναφορικά, οι δύο υπολειτουργίες που σημειώθηκαν παραπάνω, του καθορισμού της χρωματικής αρμονίας και των θερμών χρωμάτων έναντι των ψυχρών, μπορούν να αξιοποιηθούν για μεγαλύτερο εύρος αποτελεσμάτων ή για μία διαφορετική προσέγγιση. Στην προκειμένη, η χρωματική αρμονία μπορεί να συνδεθεί με την εναρμόνιση του μουσικού αποτελέσματος και να την 'χτίσει' με βάση την αντίστοιχη διάταξη της χρωματικής αρμονίας. Για παράδειγμα, δύο συμπληρωματικά χρώματα που βρίσκονται απέναντι στον χρωματικό κύκλο, θα μπορούσαν να εκφράζουν μία σχέση πέμπτης μεταξύ δύο νοτών, μιας και η πέμπτη βαθμίδα σε σχέση με το τονικό κέντρο είναι σύμφωνα και πρόκειται για ένα είδος συμπληρωματικών νοτών. Αντίστοιχα, εφόσον η ανάλυση δείξει πως υπάρχει μια σχέση ανάλογων χρωμάτων, η αρμονία που θα χτιζόταν θα αφορούσε κάποιο τρίχορδο που αντιστοιχεί στην σχέση αυτή (π.χ. μείζονα). Επιπλέον, με την χρήση της σχέσης μεταξύ των ψυχρών και θερμών χρωμάτων, μία αξιόλογη προσέγγιση θα ήταν ο διαχωρισμός της αρμονίας σε ελάσσονα και μείζονα αντίστοιχα μιας και τα θερμά χρώματα ορισμένες φορές εκφράζουν την χαρά ή την ζωντάνια σε μία εικόνα, ενώ τα ψυχρά την λύπη ή την νοσταλγία αντίστοιχα.

Σε σχέση με την κίνηση που εκφράζεται σε ένα βίντεο, σημαντικό παράγοντα θα μπορούσε να αποτελεί και το είδος της κίνησης. Αυτό μπορεί να επιτευχθεί ενδεχομένως με την αναγνώριση προτύπων, μέθοδος που έχει εφαρμοσθεί ευρέως με την χρήση της υπολογιστικής όρασης. Για παράδειγμα, εάν ο υπολογιστής εκπαιδευτεί (deep learning-machine learning) να αναγνωρίζει τα διάφορα είδη παραδοσιακών χορών και τους βηματισμούς τους μέσα από την συνεχή τροφοδότηση του με βιντεοσκοπημένο υλικό τέτοιων χορών, τότε εφόσον αναγνωρίσει κάτι τέτοιο, να παίζει την 'αντίστοιχη' μουσική βασισμένη μεν στα χρώματα της εικόνας. Αντίστοιχα, μία προσθήκη θα ήταν η αναγνώριση εκφράσεων του προσώπου με την χρήση του facial recognition, όπου εάν ο υπολογιστής αναγνωρίσει τις εκφράσεις ενός συνηθισμένου χαρούμενου προσώπου ή

στεναχωρημένου προσώπου, να παίζει και την αντίστοιχη μουσική, με βάση κάποια ψυχοακουστικά πρότυπα.

Σε επίπεδο χρηστικότητας, μία σημαντική προσθήκη για τον αλγόριθμο της εργασίας θα ήταν η επεξεργασία του βίντεο σε πραγματικό χρόνο (real-time video processing). Με την βελτίωση της πολυπλοκότητας του αλγορίθμου και ενδεχομένως την προσθήκη λειτουργιών για την λήψη από κάμερα, θα αποκτούσε μία νέα πρακτική διάσταση ο αλγόριθμος. Αντίστοιχα, στα πλαίσια του video streaming μέσω του διαδικτύου, θα γινόταν εφικτή η ζωντανή μετάδοση του ακουστικού αποτελέσματος ανάλογη των χρωμάτων και της κίνησης της λήψης.

Τέλος, η σημαντικότερη προσθήκη θα ήταν η μείωση του ανθρώπινου παράγοντα όπως για παράδειγμα ο καθορισμός ορίων και χωρίων με βάση κάποιο μέγιστο επιθυμητό από τον χρήστη. Αυτό μπορεί να επιτευχθεί με συνεχείς δοκιμασίες που έχουν σκοπό την αύξηση της αποδοτικότητας του αλγορίθμου και την εξάρτηση του αποκλειστικά στα τεχνολογικά μέσα που διατίθενται όπως και την παραμετροποίηση των δεδομένων της εικόνας.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Erich Moritz von Hornbostel, Curt Sachs. (1961). Hornbostel–Sachs. *Galpin Society Journal*.
- Albright, D. (2004). *Modernism and Music: An Anthology of Sources*. Chicago: University Of Chicago Press.
- Alpern, A. (1995). *Techniques for algorithmic composition of music*. Hampshire College.
- Ballard, D. H., & Brown, C. M. (1982). *Computer Vision*. Prentice Hall.
- Boden, M. A. (2006). *Mind as Machine: A History of Cognitive Science*. Clarendon Press.
- Briggs, D. (2013, April). *PART 7. THE DIMENSION OF HUE*. Retrieved from The Dimensions of Color: http://www.huevaluechroma.com/071.php#Newton_s_hue_system
- Burchett, K. E. (2002). Color Harmony. *Color Research & Application*, 28-31.
- Erickson, R. (1975). *Sound and Structure in Music*. University of California Press.
- Fairchild, M. (2004). Color Appearance Models: CIECAM02 and Beyond. *IS&T/SID 12th Color Imaging Conference* (p. 68). RIT Munsell Color Science Laboratory.
- Forsyth, D., & Ponce, J. (2003). *Computer Vision: A Modern Approach*. Prentice Hall.
- GÖNCZ, Z. (2011). *The Sacred Codes of the Six-Part Ricercar*. the Hungarian Bach Society.
- Howse, J., Joshi, P., & Beyeler, M. (2016). *OpenCV: Computer Vision Projects with Python*. Packt Publishing Ltd.
- Hunter, J. (2007). Matplotlib: A 2D graphics environment. *Computing In Science \& Engineering*, 90-95.

- Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active Contour Models. *International Journal of Computer Vision*, 321-331.
- Kaufman, A. (1993). *Rendering, Visualization and Rasterization Hardware*. Springer Science & Business Media.
- Klette, R. (2014). *Concise Computer Vision*. London: Springer-Verlag London.
- Kramer, G. (1994). *Auditory Display: Sonification, Audification, And Auditory Interfaces*. Avalon Publishing.
- McConnell, S. (2009). Code Complete. In S. McConnell, *Code Complete* (p. 100). Microsoft Press.
- Neill, S., & Delehanty, C. (Directors). (2012). *Yosemite HD* [Motion Picture].
- Newton, I. (1704). *Opticks: or, A Treatise of the Reflexions, Refractions, Inflexions and Colours of Light*. London: the Royal Society.
- Nick Collins, Julio d'Escrivan. (2007). *The Cambridge Companion to Electronic Music*. Cambridge: Cambridge University Press.
- NumPy Developers. (2017). *About Numpy*. Retrieved from Numpy: <http://www.numpy.org/>
- O'Connor, Z. (2010). Colour Harmony Revisited. *Color Research and Application*, 267-273.
- O'Grady, S. (2013, February 28). *The RedMonk Programming Language Rankings: January 2013*. Retrieved from RedMonk: <http://redmonk.com/sogrady/2013/02/28/language-rankings-1-13/>
- Olson, H. F. (1972, February). The Measurement of Loudness. *Audio*, pp. 18-22.
- Oxford University. (2010). *Oxford Dictionary of English (3 ed.)*. Oxford: Oxford University Press.

Peng, D., Merriman, B., Osher, S., Zhao, H., & Kang, M. (1999). A PDE-Based Fast Local Level Set Method. *Journal of Computational Physics*, 410-438.

Philip, J. A. (1966). *Pythagoras and early Pythagoreanism*. Toronto: University of Toronto Press.

Pisarevsky, V. (2017, August 2). *OpenCV Wiki*. Retrieved from Github: <https://github.com/opencv/opencv/wiki>

Read, P., & Meyer, M.-P. (2000). *Restoration of Motion Picture Film*. Butterworth-Heinemann.

Roads, C. (2004). *Microsound*. MIT Press.

Rosebrock, A. (2017, May 1). *Imutils*. Retrieved from Github: <https://github.com/jrosebr1/imutils>

Sacks, O. (2008). *Musicophilia*. New York: Random House, Inc.

Schopenhauer, A. (1966). *The World as Will and Representation, Vol. 1*. Dover Publications.

scikit-learn developers (BSD License). (2017). *scikit-learn documentation*. Retrieved from scikit-learn: <http://scikit-learn.org/stable/documentation.html>

Seashore, C. E. (1967). *The Psychology of Music*. Dover Publications.

Shapiro, L. G., & Stockman, G. C. (2001). *Computer Vision*. Prentice Hall.

Sles, M., & Nick, F. (Directors). (2015). *EYLENDI | Iceland 4k* [Motion Picture].

Spaans, B. (2015). *mingus*. Retrieved from Github: <https://bspaans.github.io/python-mingus/#>

Stone, H. S. (1971). *Introduction to computer organization and data structures*. the University of Michigan: McGraw-Hill.

Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer Science & Business Media.

Xenakis, I. (1955). the Crisis of Serial Music. *the Gravesaner Blätter*.