

**ΕΛΛΗΝΙΚΟ ΜΕΣΟΓΕΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΤΜΗΜΑ ΜΟΥΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΑΚΟΥΣΤΙΚΗΣ  
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ:  
ΤΕΧΝΟΛΟΓΙΕΣ ΗΧΟΥ ΚΑΙ ΜΟΥΣΙΚΗΣ**

**ΟΠΤΙΚΟΠΟΙΗΣΗ ΕΚΦΡΑΣΤΙΚΩΝ  
ΑΠΟΚΛΙΣΕΩΝ ΚΑΤΑ ΤΗ ΜΟΥΣΙΚΗ  
ΕΚΤΕΛΕΣΗ**

**ΤΡΙΑΝΤΑΦΥΛΛΟΥ ΙΩΑΝΝΗΣ  
ΡΕΘΥΜΝΟ, ΙΑΝΟΥΑΡΙΟΣ 2021**

**ΕΠΙΒΛΕΠΟΥΣΑ ΚΑΘΗΓΗΤΡΙΑ: ΔΡ. ΑΛΕΞΑΝΔΡΑΚΗ ΧΡΥΣΟΥΛΑ**



## Περίληψη

Η μουσική ερμηνεία αφορά στην εκ προθέσεως απόκλιση της μουσικής εκτέλεσης από το σχεδιασμό του συνθέτη όπως αυτός καταγράφεται και διανέμεται μέσω σημειογραφίας, για παράδειγμα μέσω της παρτιτούρας ενός μουσικού έργου. Για την περίπτωση της μουσικής που είναι προσημειωμένη και δεν βασίζεται στον αυτοσχεδιασμό, η μουσική έκφραση εκδηλώνεται με τη χρήση ανεπαίσθητων αποκλίσεων στο tempo, τη δυναμική και την άρθρωση κατά τη διάρκεια της μουσικής εκτέλεσης. Η παρούσα εργασία παρουσιάζει μία εφαρμογή που ονομάζεται FlameCens και η οποία, δοθέντων δύο ηχογραφήσεων του ίδιου μουσικού έργου, κατά κανόνα εκτελεσμένων από διαφορετικούς εκτελεστές, οπτικοποιεί τις αποκλίσεις στο tempo και τη δυναμική κατά τη διάρκεια της αναπαραγωγής του έργου. Η εφαρμογή μπορεί επίσης να συγκρίνει μία συγκεκριμένη ερμηνεία με την αντίστοιχη παρτιτούρα (δοσμένη σε μορφή MIDI). Καθώς η παρτιτούρα θεωρείται ως η εκφραστικά ουδέτερη αναπαράσταση ενός μουσικού έργου, η εφαρμογή σε αυτή την περίπτωση απεικονίζει τις εκφραστικές αποκλίσεις που υιοθετούνται από συγκεκριμένους ερμηνευτές. Η εφαρμογή FlameCens χρησιμοποιεί τον αλγόριθμο DTW (Dynamic Time Warping – Αλγόριθμος Δυναμικής Χρονικής Στρέβλωσης) για να συγκρίνει δύο ηχητικές ακολουθίες, με χρήση των ηχητικών χαρακτηριστικών που στη βιβλιογραφία αναφέρονται ως CENS features (Chroma Energy distribution Normalized Statistics). Οι αποκλίσεις αυτές αναπαρίστανται γραφικά με τη μορφή μίας κινούμενης φλόγας, η οποία διαμορφώνεται με χρήση ενός πλήθους σωματιδίων (particles). Το μέγεθος της φλόγας απεικονίζει την απόκλιση στη δυναμική, ενώ η κλίση της φλόγας απεικονίζει την απόκλιση στο tempo έτσι ώστε γρηγορότερο tempo να προκαλεί αλλαγή στην κλίση προς τα δεξιά και αργότερο tempo να προκαλεί αλλαγή στην κλίση προς τα αριστερά. Σταθερή κλίση της φλόγας υποδηλώνει μηδενική απόκλιση στο tempo. Οι αποκλίσεις στο tempo και στις δυναμικές που ανιχνεύονται κατά την αναπαραγωγή, καταγράφονται επιπρόσθετα σε ένα αρχείο κειμένου, επιτρέποντας έτσι την περαιτέρω μουσικολογική διερεύνηση. Επιπλέον, στην περίπτωση της μονοφωνικής μουσικής, ο τόνος των νοτών οπτικοποιείται μέσω του χρώματος των σωματιδίων. Η εφαρμογή FlameCens υλοποιήθηκε στη γλώσσα προγραμματισμού Python και διατίθεται ως λογισμικό ανοιχτού κώδικα στο αποθετήριο GitHub. Η εφαρμογή αξιολογήθηκε πειραματικά σε διάφορα μουσικά είδη, όπως στην κλασική μουσική, στη σύγχρονη, στη jazz και στην pop. Τα αποτελέσματα των πειραμάτων καταδεικνύουν ότι, καθώς η εφαρμογή FlameCens επιτρέπει τη διερεύνηση των εκφραστικών τακτικών που υιοθετούνται από διαφορετικούς εκτελεστές, μπορεί να αποτελέσει ένα χρήσιμο εργαλείο, τόσο για εξειδικευμένο κοινό (μουσικούς και μουσικολόγους), όσο και για τον απλό ακροατή ενισχύοντας την εμπειρία της μουσικής ακρόασης με οπτικά μέσα.

**Λέξεις – κλειδιά:** Συγχρονισμός ηχητικού σήματος, υπολογιστική ανάλυση της μουσικής, μουσική έκφραση, οπτικοποίηση πληροφορίας.

## Abstract

Music interpretation accounts to the way musicians shape their performance by deliberately deviating from composers' intentions, which are commonly communicated via some form of music transcription, such as a music score. For transcribed and non-improvised music, music expression is manifested by introducing subtle deviations in tempo, dynamics and articulation during the evolution of performance. This paper presents an application, named FlameCens, which, given two recordings of the same piece of music, presumably performed by different musicians, allow visualising deviations in tempo and dynamics during playback. The application may also compare a certain performance to the music score of that piece (i.e. MIDI file), which may be thought of as an expression-neutral representation of that piece, hence depicting the expressive queues employed by certain performers. FlameCens uses the Dynamic Time Warping algorithm to compare two audio sequences, based on CENS (Chroma Energy distribution Normalized Statistics) audio features. Expressive deviations are illustrated in a moving flame, which is generated by an animation of particles. The length of the flame is mapped to deviations in dynamics, while the slope of the flame is mapped to tempo deviations so that faster tempo changes the slope to the right and slower tempo changes the slope to the left. Constant slope signifies no tempo deviation. The detected deviations in tempo and dynamics can be additionally recorded in a text file, which allows for offline investigation. Moreover, in the case of monophonic music, the color of particles is used to convey the pitch of the notes during performance. FlameCens has been implemented in Python and it is openly available via GitHub. The application has been experimentally validated for different music genres including classical, contemporary, jazz and popular music. These experiments revealed that FlameCens can be a valuable tool for music specialists (i.e. musicians or musicologists) to investigate the expressive performance strategies employed by different musicians, as well as for music audience to enhance their listening experience.

**Keywords:** Audio synchronization, computational music analysis, expressive music performance, information visualization.

## Περιεχόμενα

1. Εισαγωγή.....	6
1.1 Αναπαραστάσεις της μουσικής.....	7
1.1.2 Τονικότητα και σημειογραφία της Δυτικής μουσικής.....	7
1.1.3 Πρωτόκολλο MIDI.....	9
2. Μουσικός Συγχρονισμός.....	11
2.1 Ηχητικές Αναπαραστάσεις (Audio).....	12
2.1.1 Συχνότητα και τονικότητα.....	13
2.1.2 Χροιά.....	14
2.2 Χαρακτηριστικά του Ήχου.....	15
2.2.1 Φασματογράφημα λογαριθμικών συχνοτήτων.....	15
2.2.2 Chroma Features.....	18
2.2.3 Λογαριθμική Συμπύεση.....	19
2.2.4 Κανονικοποίηση.....	20
2.2.5 Αλλαγή Τονικότητας.....	21
2.2.6 CENS Features.....	22
2.3 Ο Αλγόριθμος DTW.....	24
2.3.1 Βασική Προσέγγιση για Μουσικό Συγχρονισμό.....	24
2.3.2 Warping Path.....	26
2.3.3 Βέλτιστο Warping Path και απόσταση DTW.....	27
2.3.4 Αλγόριθμος Δυναμικού Προγραμματισμού.....	28
2.3.5 Παράδειγμα Εφαρμογής του Αλγορίθμου DTW.....	31
2.3.6 Παραλλαγές του DTW.....	33
2.3.6.1 Συνθήκη Μεγέθους βήματος.....	33
2.3.6.2 Τοπικά βάρη.....	33
2.3.6.3 Καθολικοί Περιορισμοί.....	34
2.3.6.4 Πολύ-κλιμακούμενος DTW.....	35
3. Υπολογιστική μοντελοποίηση της μουσικής έκφρασης.....	37
3.1 Το μοντέλο KTH.....	38
3.1.1 Βασική Λειτουργία του μοντέλου.....	38
3.1.2 Εμπειρική αξιολόγηση.....	39
3.2 Το μοντέλο Todd.....	40
3.2.2 Το βασικό μοντέλο.....	40
3.2.3 Εμπειρική αξιολόγηση.....	41
3.3 Το μοντέλο Mazzola.....	41
3.3.1 Το βασικό μοντέλο.....	42

3.4 Το μοντέλο μηχανικής μάθησης .....	42
3.4.1 Το μοντέλο σε επίπεδο νότας .....	42
3.4.2 Το πολύ-επίπεδο μοντέλο .....	43
3.4.3 Εμπειρική αξιολόγηση.....	44
4. Οπτικοποίηση της μουσικής έκφρασης .....	46
4.1 Ερευνητικές Προσεγγίσεις .....	46
4.2 Animus .....	48
4.3 Οπτικοποίηση με σύστημα σωματιδίων .....	49
4.4 Comp-i .....	51
4.5 Οπτικοποίηση της αρμονίας .....	51
4.6 Οπτικοποίηση με χρήση MAX/MSP και Flash .....	52
4.7 Performance Worm .....	53
5. Υλοποίηση.....	55
5.1 Η εφαρμογή FlameCens .....	55
5.1.1 Η παραλλαγή FlameCens-Pitch.....	57
5.2 Βιβλιοθήκες της Python .....	58
5.2.1 Η βιβλιοθήκη Librosa .....	58
5.2.2 Η βιβλιοθήκη PyGame.....	59
5.2.3 Η βιβλιοθήκη Numba .....	59
5.2.4 Άλλες βιβλιοθήκες.....	59
5.3 Εξαγωγή Ηχητικών Χαρακτηριστικών .....	59
5.3.1 Υλοποίηση με chroma features .....	60
5.3.2 Υλοποίηση με χρήση CENS features .....	62
5.4 Υλοποίηση του Dynamic Time Warping .....	63
5.4.1 Εμπειρική επαλήθευση του DTW .....	64
5.5 Οπτικοποίηση της μουσικής έκφρασης .....	65
5.5.1 Αντιστοίχιση χρονικής απόκλισης και δυναμικών .....	66
5.5.2 Αντιστοίχιση τονικότητας με χρώματα .....	67
5.6 Καταγραφή Εκφραστικών Αποκλίσεων .....	68
6. Αξιολόγηση .....	69
6.1 Αξιολόγηση της μουσικής έκφρασης.....	69
6.2 Πειραματική Αξιολόγηση .....	70
6.2.1 Αντιστοίχιση ηχητικού αρχείου με τον εαυτό του .....	70
6.2.2 Σύγκριση ερμηνείας με παρτιτούρα.....	71
6.2.3 Σύγκριση ανάμεσα σε διαφορετικές ερμηνείες του ίδιου έργου .....	74
6.2.4 Οπτικοποίηση για διαφορετικά μουσικά είδη .....	76

<b>7. Συμπεράσματα .....</b>	<b>80</b>
<b>7.1 Επισκόπηση .....</b>	<b>80</b>
<b>7.2 Περιορισμοί και μελλοντικές προεκτάσεις .....</b>	<b>81</b>
<b>8. Αναφορές .....</b>	<b>83</b>
<b>Παράρτημα 1 – Υλοποίηση αλγορίθμου DTW σε Python .....</b>	<b>88</b>
<b>Παράρτημα 2 – Υλοποίηση οπτικοποίησης σε Python .....</b>	<b>89</b>

# 1. Εισαγωγή

Η μουσική αποτελεί μία σύνθετη και πολυδιάστατη δραστηριότητα της ανθρώπινης νόησης και δημιουργικότητας και ταυτόχρονα έναν παράγοντα συναισθηματικής έκφρασης που συνοδεύει τον άνθρωπο από τα πρώτα στάδια του πολιτισμού. Ο βαθμός που ο καθένας εμπλέκεται στη μουσική διαδικασία είναι φυσικά διαφορετικός, με τον εκπαιδευμένο μουσικό να αντιλαμβάνεται τη δομή και τη λειτουργία της μουσικής σε πολύ μεγαλύτερο βαθμό από τον απλό ακροατή. Παρόλα αυτά, ο κάθε ακροατής μπορεί να αντιληφθεί βασικές μουσικές έννοιες (π.χ. το tempo ή διαφορές στην ένταση) ακόμη και αν δεν μπορεί να τις εκφράσει με αυστηρά μουσικούς όρους.

Η εκτέλεση επομένως ενός μουσικού έργου, ως καλλιτεχνική και συναισθηματική δραστηριότητα, δεν θα μπορούσε να είναι μια μηχανική αναπαραγωγή νοτών ή οδηγιών που αναγράφονται σε κάποιο μουσικό κείμενο. Διαβάζοντας τη μουσική σημειογραφία (παρτιτούρα) ο μουσικός ακολουθεί τις οδηγίες που αυτή κωδικοποιεί, όπως τις έχει ορίσει ο συνθέτης του εκάστοτε έργου. Πάνω στις οδηγίες που λαμβάνει από την παρτιτούρα προσθέτει τις δικές του ερμηνευτικές παραλλαγές, ανάλογα με την αισθητική – καλλιτεχνική του αντίληψη και το μουσικό στόχο που θέλει να μεταδώσει στον ακροατή. Αυτές ακριβώς οι εσκεμμένες διαμορφώσεις μουσικών παραμέτρων όπως το tempo, οι δυναμικές και η μουσική άρθρωση ορίζουν την έννοια της μουσικής έκφρασης και αποτελούν και το στόχο της παρούσης εργασίας.

Στο δεύτερο κεφάλαιο της εργασίας γίνεται μία ανασκόπηση της συσχετιζόμενης έρευνας πάνω στα θέματα που απασχολούν την εργασία. Συγκεκριμένα γίνεται ανασκόπηση της διαδικασίας χρονικής αντιστοίχισης μεταξύ δύο ηχητικών ακολουθιών, καθώς η χρονική απόκλιση αποτελεί βασικό εκφραστικό μέσο για τον ερμηνευτή που μπορεί να αυξάνει ή να μειώνει το tempo σε διάφορα σημεία του κομματιού (μέρη, μουσικές φράσεις ή και συγκεκριμένες νότες). Ερευνάται η εξαγωγή των κατάλληλων ηχητικών χαρακτηριστικών (features) που μπορούν να αξιοποιηθούν για το σκοπό αυτό και η λειτουργία του αλγορίθμου χρονικής αντιστοίχισης Dynamic Time Warping (DTW). Καθώς η παραπάνω διαδικασία μπορεί να γίνει αρκετά χρονοβόρα σε μεγάλα μουσικά έργα, η απόδοση του DTW αποτελεί επίσης αντικείμενο έρευνας. Το κεφάλαιο 3 ασχολείται με την υπολογιστική μοντελοποίηση της μουσικής έκφρασης. Γίνεται αναφορά στα πιο σημαντικά υπολογιστικά μοντέλα που έχουν αναπτυχθεί τα τελευταία με στόχο να ορίσουν με ακρίβεια τις φυσικές παραμέτρους που καθορίζουν μία ερμηνεία και τις ποσοτικές σχέσεις μεταξύ των παραμέτρων αυτών. Τέλος στο κεφάλαιο 4 παρουσιάζονται διάφορες ερευνητικές προσεγγίσεις πάνω στην οπτικοποίηση (visualization) της μουσικής. Αναλύονται πιο εκτεταμένα κάποιες υλοποιήσεις που αποτέλεσαν πηγή έμπνευσης για τη συγκεκριμένη εργασία, αν και οι περισσότερες από αυτές αφορούν την οπτικοποίηση της μουσικής έκφρασης σε μεμονωμένα μουσικά έργα και όχι συγκριτικά με την παρτιτούρα ή μία διαφορετική ερμηνεία.

Η εργασία συνοδεύεται από εξειδικευμένη εφαρμογή λογισμικού που υλοποιήθηκε στη γλώσσα προγραμματισμού Python. Στο κεφάλαιο 5 γίνεται αναλυτική παρουσίαση της υλοποίησής της και των προγραμματιστικών μεθόδων που χρησιμοποιήθηκαν. Η εφαρμογή αρχικά πραγματοποιεί σύγκριση υπαρχόντων ηχογραφήσεων μουσικών έργων από διαφορετικούς ερμηνευτές με την αντίστοιχη σημειογραφία (παρτιτούρα), προκειμένου να υπάρξει χρονική αντιστοίχιση και να διαπιστωθούν οι αποκλίσεις από την προσημειωμένη αυτή μορφή. Η παρτιτούρα κωδικοποιείται σε μορφή MIDI ενώ από τις ηχογραφήσεις εξάγονται τα καταλληλά ηχητικά χαρακτηριστικά (Chroma και CENS features) που επιτρέπουν



τη σύγκριση μέσω του αλγορίθμου DTW. Στη συνέχεια πραγματοποιείται οπτικοποίηση των ερμηνευτικών χαρακτηριστικών με χρήση της τεχνικής γραφικών των σωματιδίων (particles). Η υλοποίηση εμφανίζει μία γραφική απεικόνιση η κλίση της οποίας ως προς ένα σταθερό άξονα αλλάζει δυναμικά, ανάλογα με τη χρονική απόκλιση από την παρτιτούρα. Η διαφορά των δυναμικών των δύο ερμηνευτών επίσης οπτικοποιείται, αυξομειώνοντας την ταχύτητα των σωματιδίων (και κατά συνέπεια και το μήκος της συνολικής απεικόνισης). Τέλος υπάρχει δυνατότητα και για χρωματική απεικόνιση της κάθε νότας για περιπτώσεις μονοφωνικής μουσικής. Στόχος της συγκεκριμένης υλοποίησης δεν ήταν τόσο η δημιουργία εντυπωσιακών γραφικών όσο η δημιουργία μια οπτικοποίησης όπου οι μεταβολές της μουσικής έκφρασης θα είναι σαφείς και κατανοητές ακόμη και από ακροατές που δεν έχουν υψηλό επίπεδο μουσικής εκπαίδευσης. Στο κεφάλαιο 6 πραγματοποιούνται μια σειρά από πειραματικές δοκιμές της εφαρμογής πάνω σε διαφορετικά μουσικά έργα προκειμένου να αξιολογηθεί η λειτουργία της.

### 1.1 Αναπαραστάσεις της μουσικής

Η μουσική μπορεί να αναπαρασταθεί με διάφορους τρόπους. Για παράδειγμα ένας συνθέτης μπορεί να γράφει μουσική με τη μορφή σημειογραφίας όπου μουσικά σύμβολα χρησιμοποιούνται για την οπτική κωδικοποίηση των νοτών και το πώς αυτές θα αναπαραχθούν από τους μουσικούς. Η τυπωμένη μουσική σημειογραφία αναφέρεται συχνά και ως παρτιτούρα. Αν και το αρχικό μέσο παρουσίασης της παρτιτούρας ήταν το χαρτί σήμερα μπορεί να υπάρχει και σε ψηφιακή μορφή. Για τα ηλεκτρονικά μουσικά όργανα και τους υπολογιστές η μουσική μπορεί να αναπαρασταθεί με τη χρήση συγκεκριμένων πρωτοκόλλων όπως το ευρέως διαδεδομένο πρωτόκολλο MIDI (Musical Instruments Digital Interface – Ψηφιακή Διεπαφή Μουσικών Οργάνων), όπου μηνύματα γεγονότων νοτών χρησιμοποιούνται για να καθορίσουν τον τόνο, τη διάρκεια και άλλες παραμέτρους που απαιτούνται για την αναπαραγωγή ενός μουσικού έργου. Και οι δύο παραπάνω μορφές αποτελούν ένα συμβολικό τρόπο αναπαραστάσης της μουσικής. Σε αντίθεση με τις συμβολικές αναπαραστάσεις, οι ηχητικές (audio) αναπαραστάσεις, όπως τα αρχεία WAV και MP3 κωδικοποιούν σήματα ήχου, που δημιουργούνται από μία ηχητική πηγή και διαδίδονται ως ακουστικό κύμα, δηλαδή ως πυκνώματα και αραιώματα του αέρα.

Το πρόβλημα του μουσικού συγχρονισμού προφανώς απαιτεί την εύρεση αντιστοιχίσεων ανάμεσα σε διαφορετικές μορφές αναπαράστασης της μουσικής. Για το λόγο αυτό ακολουθεί μία αναφορά των διάφορων μορφών αναπαράστασης και των βασικών χαρακτηριστικών τους.

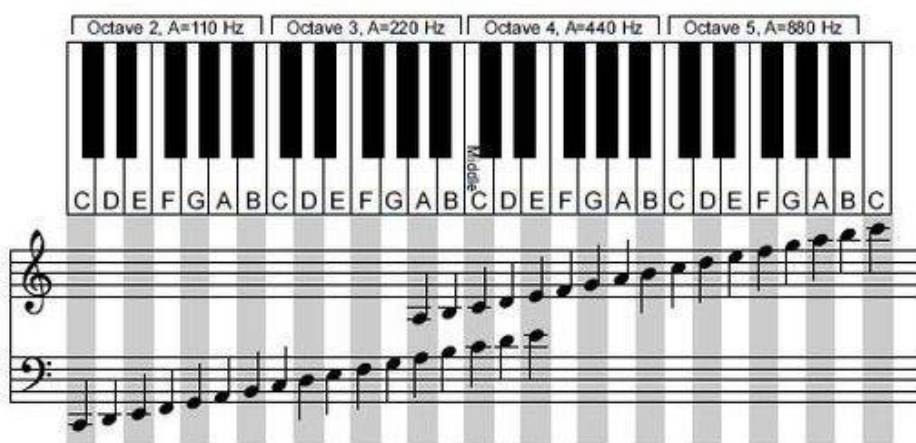
#### 1.1.2 Τονικότητα και σημειογραφία της Δυτικής μουσικής

Μια παρτιτούρα περιγράφει ένα μουσικό έργο χρησιμοποιώντας μία γλώσσα μουσικών συμβόλων. Διαβάζοντας μια παρτιτούρα ο μουσικός μπορεί να αναπαράγει ένα μουσικό έργο ακολουθώντας τις οδηγίες που αυτή κωδικοποιεί. Η απόδοση της μουσικής όμως δεν απαιτεί απλά συγκεκριμένες γνώσεις, όπως η ικανότητα κάποιου να καταλαβαίνει τη μουσική σημειογραφία, αλλά αποτελεί μία δημιουργική διεργασία. Ένα μουσικό κείμενο σπάνια αναπαράγεται μηχανικά. Ο μουσικός διαμορφώνει τη μουσική με παραλλαγές στο ρυθμό, τις δυναμικές και την άρθρωση, δημιουργώντας μία προσωπική ερμηνεία του κομματιού. Η παρτιτούρα λειτουργεί ως οδηγός που αφήνει χώρο για προσωπική ερμηνεία.

Στη μουσική ο όρος νότα χρησιμοποιείται με ένα σχετικά αόριστο τρόπο, που μπορεί να αναφέρεται είτε σε ένα αντίστοιχο μουσικό σύμβολο είτε στην τονικότητα ενός ήχου. Στη σημειογραφία της δυτικής μουσικής κάθε νότα έχει διάφορες ιδιότητες που καθορίζουν τη σχετική διάρκεια και τον τόνο του ήχου που πρέπει να εκτελεστεί από τον μουσικό. Η έννοια του τόνου είναι μία αντιληπτική έννοια που επιτρέπει στον ακροατή να ταξινομήσει τον ήχο σε μία κλίμακα συχνοτήτων. Η αναπαραγωγή μίας νότας σε κάποιο μουσικό όργανο οδηγεί σε ένα περιοδικό ήχο με κάποια θεμελιώδη συχνότητα. Η θεμελιώδης συχνότητα αυτή σχετίζεται άμεσα με αυτό που αποκαλούμε τονικότητα της νότας. Η τονικότητα αποτελεί το αισθητικό κριτήριο με βάση το οποίο ταξινομούμε τους ήχους σε «χαμηλούς» και «υψηλούς» σε αντιστοιχία με τα πλήκτρα του πιάνου από αριστερά προς τα δεξιά. Δύο νότες με λόγο θεμελιωδών συχνοτήτων ίσο με δύναμη του δύο, γίνονται αντιληπτές ως παρόμοιες. Νότες που συνδέονται με μία τέτοιου είδους σχέση χαρακτηρίζονται της ίδιας κλάσης (pitch class). Στην παρατήρηση αυτή βασίζεται και η βασική έννοια της οκτάβας που ορίζεται ως το διάστημα μεταξύ μιας νότας και μιας άλλης με τη μισή ή τη διπλάσια θεμελιώδη συχνότητα. Αντίστοιχα μία κλάση νοτών ορίζεται ως το σύνολο των τόνων που απέχουν μεταξύ τους ακέραιο αριθμό οκτάβων (Randel 2003)

Για να περιγράψουμε τη μουσική απαιτείται η χρήση κάποιας μουσικής κλίμακας, ενός πεπερασμένου αριθμού συμβόλων που θα αναπαριστά τις διάφορες τονικότητες. Πολλές μουσικές κλίμακες έχουν υπάρξει σε διαφορετικές ιστορικές περιόδους της μουσικής και σε διαφορετικά πολιτιστικά περιβάλλοντα (π.χ. Ινδική, Βυζαντινή κ.τ.λ.). Εμείς θα επικεντρωθούμε στην κλασική Δυτική κλίμακα, στην οποία κάθε οκτάβα χωρίζεται σε 12 ίσα διαστήματα. Η διαφορά ανάμεσα στις θεμελιώδεις συχνότητες δύο διαδοχικών βημάτων της κλίμακας ονομάζεται ημιτόνιο κι αποτελεί το μικρότερο διάστημα της Δυτικής κλίμακας.

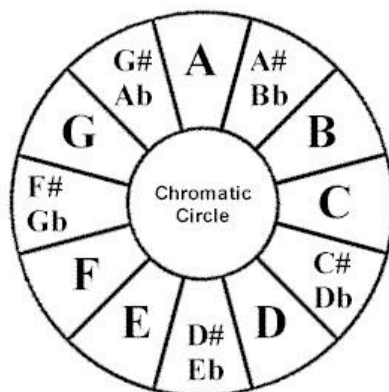
Στη δυτική σημειογραφία, οι δώδεκα τονικές κλάσεις αναπαρίστανται με γράμματα του λατινικού αλφαβήτου και τις αλλοιώσεις (δίεση και ύφεση). Μια δίεση (#) αυξάνει τη νότα κατά ένα ημιτόνιο και μια ύφεση (♭) μειώνει τη νότα κατά ένα ημιτόνιο. Μια κλίμακα επομένως αποτελείται από τις νότες C,D,E,F,G,A,B και τις C<sup>#</sup>,D<sup>#</sup>,F<sup>#</sup>,G<sup>#</sup>,A<sup>#</sup> ή D<sup>♭</sup>,E<sup>♭</sup>,G<sup>♭</sup>,A<sup>♭</sup>,B<sup>♭</sup> (Οι νότες C<sup>#</sup> και D<sup>♭</sup> για παράδειγμα αντιπροσωπεύουν την ίδια τονική κλάση) **(Εικόνα 1-1)**



**Εικόνα 1-1**

*Τμήμα του κλαβιέ ενός πιάνου με έκταση πλήκτρων από C2 ως C5 και τις αντίστοιχες νότες σε Δυτική σημειογραφία*

Κάθε νότα ακολουθείται από τον αριθμό της οκτάβας της, με τον αριθμό της οκτάβας να αλλάζει κάθε φορά στη νότα C. Η νότα A4 έχει οριστεί να έχει θεμελιώδη συχνότητα 440Hz και λειτουργεί ως σημείο αναφοράς. Ταξινομώντας τις νότες ανάλογα με την τονικότητά τους, παίρνουμε τη χρωματική κλίμακα. Η έννοια της χρωματικότητας έχει άμεση σχέση με τις τονικές κλάσεις, για παράδειγμα οι νότες D2 και D5 έχουν την ίδια χρωματική αξία D (η νότα ρε σε διαφορετικές οκτάβες) και γίνονται αντιληπτές ως παρόμοιες νότες. Η κυκλική φύση των χρωματικών αξιών φαίνεται στο λεγόμενο χρωματικό κύκλο (**Εικόνα 1-2**). Γενικά ο χρωματικός παράγοντας της τονικότητας μπορεί να χρησιμοποιηθεί για τη δημιουργία ενδιάμεσων αναπαραστάσεων που χρησιμοποιούνται σε πολλές εφαρμογές ανάλυσης και ανάκτησης μουσικής πληροφορίας.



Εικόνα 1-2  
Ο χρωματικός κύκλος

### 1.1.3 Πρωτόκολλο MIDI

Το πρωτόκολλο MIDI έχει περιγραφεί ανά περίπτωση ως μια διασύνδεση μεταξύ υπολογιστών και οργάνων, ως ένα σύνολο οδηγιών για τη μεταφορά δεδομένων από το ένα όργανο στο άλλο και ως μια γλώσσα για μετάδοση μουσικής σημειογραφίας μεταξύ υπολογιστών και synthesizers. Όλοι οι παραπάνω ορισμοί καλύπτουν κάποια από τις όψεις του πρωτόκολλου MIDI (Roads 1996). Αν και το MIDI δεν σχεδιάστηκε αρχικά ως μια συμβολική μουσική σημειογραφία, η σπουδαιότητά του οφείλεται στην τεράστια εξάπλωση που έχει γνωρίσει τις τρεις τελευταίες δεκαετίες.

Το MIDI σχεδιάστηκε για τον έλεγχο ψηφιακών μουσικών συσκευών σε πραγματικό χρόνο. Το πρωτόκολλο καθορίζει έναν τρόπο διασύνδεσης υλικού και μια μέθοδο για επικοινωνία δεδομένων. Ορίζει επίσης μία γραμματική για την κωδικοποίηση πληροφοριών μουσικής εκτέλεσης. Οι πληροφορίες αυτές ομαδοποιούνται σε σύντομα μηνύματα που αποστέλλονται από μία συσκευή σε μια άλλη. Για παράδειγμα ένα μήνυμα μπορεί να ορίζει την αρχή (noteon message) και ένα δεύτερο το τέλος (noteoff message) μιας νότας. Στα μηνύματα αυτά πέρα από τη χρονική στιγμή εμφάνισης ενός μουσικού γεγονότος προσδιορίζεται η τονικότητά, η ένταση και η χροιά της νότας. Άλλου τύπου μηνύματα αφορούν την εφαρμογή ηχητικών εφέ (modulation, pitch bend, κ.λπ.), τον συγχρονισμό και τη μίξη (panning, stereo balance, volume) μουσικών οργάνων, κ.ο.κ. Τα MIDI μηνύματα επομένως αναπαριστούν εντολές μουσικής εκτέλεσης και όχι κυματομορφές ήχου. Τα πακέτα MIDI δεν καθορίζουν αυστηρά το ηχόχρωμα του τελικού ήχου, αφήνοντας το ως επιλογή στην συσκευή – παραλήπτη (γεννήτρια ήχου). Αυτό σημαίνει ότι το ίδιο MIDI μήνυμα

μπορεί να αντιστοιχεί σε δύο διαφορετικές χροιές οι οποίες ορίζονται από τον αλγόριθμο σύνθεσης ήχου που χρησιμοποιεί η κάθε συσκευή.

Τα πιο σημαντικά μηνύματα MIDI είναι ίσως τα note-on και note-off που σηματοδοτούν την αρχή και το τέλος μιας νότας αντίστοιχα. Κάθε ένα από αυτά τα μηνύματα περιλαμβάνει ένα αριθμό MIDI νότας, μία τιμή velocity (ταχύτητα του πλήκτρου – καθορίζει συνήθως και το αρχικό πλάτος), τον προσδιορισμό του καναλιού και το χρονικό σημείο (timestamp). Ο αριθμός MIDI νότας είναι ένας ακέραιος από 0 ως 127 που κωδικοποιεί την τονικότητα των νοτών σε αντιστοιχία με την κλίμακα ίσων αποστάσεων που αναφέρθηκε στην προηγούμενη παράγραφο. Κωδικοποιούνται οι νότες με τονικότητα από C0 ως G#9 με την νότα A4 για παράδειγμα να αντιστοιχεί στον αριθμό 69 (Πίνακας 1). Η τιμή του velocity κυμαίνεται επίσης από 0 ως 127 και καθορίζει την ένταση της νότας σε περίπτωση note-on και το «σβήσιμο» (decay) της νότας σε περίπτωση note-off, αν και η ερμηνεία του εξαρτάται σε μεγάλο βαθμό από το είδος του synthesizer ή ψηφιακού οργάνου. Το MIDI κανάλι είναι ένας ακέραιος από 0 ως 15, με κάθε αριθμό να αντιστοιχεί σε ένα όργανο που έχει ανατεθεί σε κάθε κανάλι. Τέλος το timestamp είναι ένας ακέραιος αριθμός που υποδηλώνει πόσους παλμούς ρολογιού πρέπει να περιμένει το συγκεκριμένο note-on ή note-off μήνυμα για να εκτελεστεί.

MIDI Τιμές												
Οκτάβα	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
0	0	1	2	3	4	5	6	7	8	9	10	11
1	12	13	14	15	16	17	18	19	20	21	22	23
2	24	25	26	27	28	29	30	31	32	33	34	35
3	36	37	38	39	40	41	42	43	44	45	46	47
4	48	49	50	51	52	53	54	55	56	57	58	59
5	60	61	62	63	64	65	66	67	68	69	70	71
6	72	73	74	75	76	77	78	79	80	81	82	83
7	84	85	86	87	88	89	90	91	92	93	94	95
8	96	97	98	99	100	101	102	103	104	105	106	107
9	108	109	110	111	112	113	114	115	116	117	118	119
10	120	121	122	123	124	125	126	127				

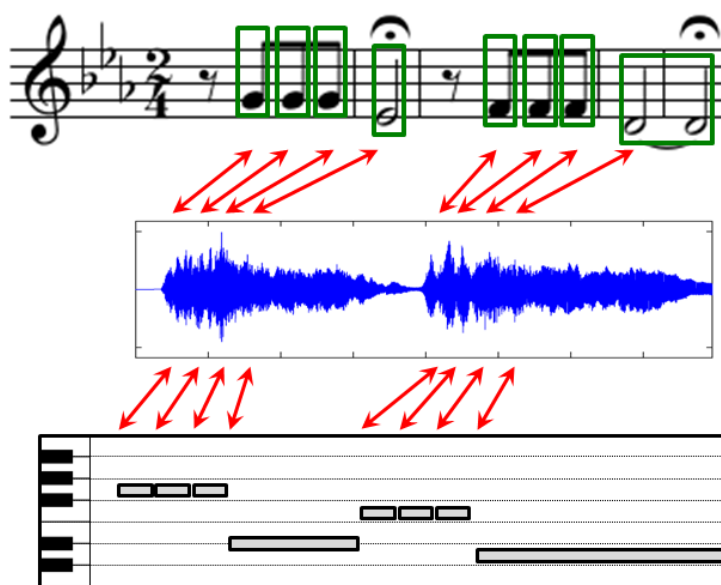
Πίνακας 1  
Αντιστοιχία νοτών με MIDI τιμές

Ένα σημαντικό χαρακτηριστικό του MIDI είναι ότι, όπως και στη κλασική μουσική σημειογραφία, μπορεί να εκφράσει πληροφορίες χρονισμού με όρους μουσικών γεγονότων και όχι με τη χρήση απόλυτων χρονικών μονάδων όπως π.χ. millisecond. Για το σκοπό αυτό χρησιμοποιεί ως χρονική μονάδα κάποια υποδιαίρεση του τέταρτου της νότας, τους παλμούς ρολογιού (clock pulses). Ο αριθμός των παλμών ρολογιού ανά τέταρτο της νότας (PPQN) αναφέρεται στην αρχή κάθε αρχείου MIDI (header των Standard Midi Files - .SMF) και ισχύουν για όλα τα μηνύματα που ακολουθούν. Ο αριθμός των μονάδων χρόνου που αναφέρεται στο timestamp του κάθε μηνύματος ισχύει σε σχέση με το προηγούμενο μήνυμα (σχετικός και όχι απόλυτος χρονισμός). Επίσης το πρωτόκολλο MIDI επιτρέπει την αποθήκευση πληροφορίας απόλυτου χρονισμού, μέσω μηνυμάτων tempo που ορίζουν αριθμό microsecond ανά τέταρτο νότας. Ενώ ο αριθμός των παλμών ανά τέταρτο νότας παραμένει σταθερός σε όλη τη διάρκεια του αρχείου, η απόλυτη χρονική πληροφορία μπορεί να μεταβάλλεται παρεμβάλλοντας μηνύματα tempo μεταξύ των υπόλοιπων μηνυμάτων. Με τον τρόπο αυτό μπορούμε να ορίσουμε και τοπικές εναλλαγές στη χρονική πληροφορία για λόγους μουσικής εκφραστικότητας (π.χ. ritardandi δηλαδή τοπικά σημεία επιβράδυνσης)

## 2. Μουσικός Συγχρονισμός

Η μουσική μπορεί να περιγραφεί και να αναπαρασταθεί με διάφορους τρόπους, όπως παρτιτούρα, συμβολική αναπαράσταση και ηχογραφήσεις. Για κάθε μία από αυτές τις αναπαραστάσεις, μπορεί να υπάρχουν διαφορετικές εκδοχές που αντιστοιχούν στο ίδιο μουσικό έργο. Για παράδειγμα η 5<sup>η</sup> συμφωνία του Beethoven μπορεί να βρεθεί σε μεγάλο αριθμό ηχογραφήσεων, εκτελεσμένη από διάφορες ορχήστρες και ερμηνευτές. Ο γενικός στόχος του μουσικού συγχρονισμού είναι να συνδέει αυτόματα τις διαφορετικές μορφές δεδομένων και κατ' επέκταση τα διαφορετικά σύνολα πληροφοριών που αφορούν ένα μουσικό έργο. Πιο συγκεκριμένα, συγχρονισμός είναι η διαδικασία κατά την οποία, για ένα χρονικό σημείο μίας μορφής αναπαράστασης, καθορίζεται το αντίστοιχο χρονικό σημείο σε μια άλλη αναπαράσταση.

Για παράδειγμα η **Εικόνα 2-1** δείχνει την αναπαράσταση ενός μουσικού θέματος σε μορφή παρτιτούρας, κυματομορφής και MIDI. Η παρτιτούρα δίνει μία αυστηρή περιγραφή των νωτών όσο αφορά την τονικότητα και τη διάρκειά τους, αφήνει όμως περιθώριο σε εκφραστικούς παράγοντες όπως ο χρονισμός, οι δυναμικές και η άρθρωση (articulation), γεγονός που οδηγεί σε διαφορετικές μουσικές ερμηνείες του ίδιου κομματιού. Αυτές οι ερμηνείες μπορεί να αποθηκευτούν σε μορφή Audio ή MIDI. Σε κάθε περίπτωση θα υπάρχει μία χρονική ακολουθία η οποία μπορεί να εκφράζεται σε φυσικές μονάδες (π.χ. δευτερόλεπτα) ή μουσικές μονάδες (μέτρα, beats κλπ.). Συγχρονισμός τέτοιου είδους χρονικά εξαρτώμενων ροών δεδομένων είναι η διαδικασία της αντιστοίχισης κάθε χρονικού σημείου της μίας μορφής με το μουσικά αντίστοιχο χρονικό σημείο της άλλης μορφής (Müller 2015). Στην **Εικόνα 2-1** (Wikipedia.org 2020a) κάθε τέτοια αντιστοίχιση σημειώνεται με ένα κόκκινο αμφίδρομο βέλος



**Εικόνα 2-1**

*Το αρχικό θέμα της Συμφωνίας Νο.5 του Ludwig van Beethoven σε παρτιτούρα, κυματομορφή και piano roll αναπαράσταση. Τα κόκκινα αμφίδρομα βέλη υποδεικνύουν την αντιστοίχιση των χρονικών σημείων των σχετικών νωτών στις διαφορετικές αναπαραστάσεις.*

Ο μουσικός συγχρονισμός αυτού του είδους αποτέλεσε τη βάση για τη δημιουργία καινοτόμων εφαρμογών που επιτρέπουν στους χρήστες την πρόσβαση, αναζήτηση και

περιήγηση μουσικού περιεχόμενου με πρακτικό τρόπο, όπως για παράδειγμα τη χρονικά σύγχρονη εμφάνιση της παρτιτούρας κατά την αναπαραγωγή της ηχογράφησης του αντίστοιχου κομματιού ή την ταυτόχρονη συγχρονισμένη πρόσβαση σε πολλαπλές ηχογραφήσεις του ίδιου κομματιού

Μια τυπική διαδικασία συγχρονισμού για δύο διαφορετικές μουσικές αναπαραστάσεις περιλαμβάνει συνήθως δύο βήματα (Müller 2015): Στο πρώτο βήμα οι δύο αναπαραστάσεις μετατρέπονται σε δύο ακολουθίες από κατάλληλα ηχητικά χαρακτηριστικά (audio features). Τα χαρακτηριστικά αυτά θα πρέπει να ικανοποιούν ένα συμβιβασμό μεταξύ δύο αντικρουόμενων στόχων. Θα πρέπει από τη μια να παραμένουν όσο γίνεται ανεπηρέαστα από μεταβολές που δεν αφορούν την αντιστοίχιση στο χρόνο, από την άλλη όμως θα πρέπει να καταγράφουν αρκετή πληροφορία ώστε να πετύχουμε τον στόχο μας. Επιθυμούμε δηλαδή, κάποια features που να καταγράφουν τα αρμονικά και μελωδικά χαρακτηριστικά της μουσικής αλλά να είναι όσο γίνεται ανεξάρτητα από τη χροιά και το είδος των μουσικών οργάνων. Στο δεύτερο βήμα οι ακολουθίες από features που προέκυψαν από το πρώτο βήμα, πρέπει να συγκριθούν και να αντιστοιχισθούν χρονικά με τη χρήση κάποιας κατάλληλης τεχνικής αντιστοίχισης (alignment) και με χρήση κάποιου μέτρου απόστασης. Για να υλοποιηθεί μια τέτοιου είδους διαδικασία πρέπει να λάβουμε υπόψιν κάποια θεμελιώδη ερωτήματα, όπως το είδος της μουσικής που θα χρησιμοποιηθεί (στην περίπτωσή μας μιλάμε για μουσική που ακολουθεί τη Δυτική τονική κλίμακα) και πώς η έννοια της ομοιότητας μπορεί να ενσωματωθεί στα features και στο μέτρο απόστασης που θα χρησιμοποιήσουμε. Για μεγάλα σύνολα δεδομένων, το ερώτημα της απόδοσης του αλγορίθμου που θα χρησιμοποιηθεί αποτελεί επίσης θεμελιώδες στοιχείο.

## 2.1 Ηχητικές Αναπαραστάσεις (Audio)

Η μουσική είναι πολλά περισσότερα από συμβολικές αναπαραστάσεις νοτών και οδηγιών για το πώς οι νότες πρέπει να αναπαραχθούν. Είναι μια δημιουργική και καλλιτεχνική διαδικασία που περιλαμβάνει αισθητικές, καλλιτεχνικές και συναισθηματικές διαδικασίες με τελικό σκοπό τη δημιουργία και διαμόρφωση ήχου. Οι μουσικοί δεν αναπαράγουν μηχανικά τη μουσική, αλλά εμπλέκονται συναισθηματικά με αυτή παρεμβαίνοντας στο ρυθμό και τις δυναμικές, αυξομειώνοντας την ένταση και τονίζοντας κάποιες συγκεκριμένες νότες, δημιουργώντας τελικά τη δική τους προσωπική ερμηνεία για κάθε μουσικό κείμενο.

Από φυσικής άποψης, η εκτέλεση μουσικής οδηγεί στη δημιουργία ακουστικών κυμάτων (ήχου). Ένα ηχητικό σήμα (audio signal) αποτελεί την αναπαράσταση ενός ήχου (κυματομορφή – διάγραμμα μεταβολής της ακουστικής πίεσης ως προς το χρόνο). Σε αντίθεση με την παρτιτούρα, ένα σήμα audio περιλαμβάνει όλες τις πληροφορίες για την ακουστικά αντιληπτή αναπαραγωγή της μουσικής, όπως τις ρυθμικές, δυναμικές και τονικές μικρό-μεταβολές που διαμορφώνουν την ερμηνεία του κάθε μουσικού. Από την άλλη, ένα ηχητικό σήμα δεν υποδηλώνει επακριβώς πληροφορίες όπως τους χρόνους έναρξης των νοτών (Onsets), την τονικότητα ή τη διάρκεια των νοτών. Για το λόγο αυτό η σύγκριση μουσικών σημάτων αποτελεί μια πολύπλοκη διαδικασία, ειδικά σε περιπτώσεις πολυφωνικής μουσικής. Επίσης, η αντίληψη του ήχου πέρα από τα αντικειμενικά φυσικά χαρακτηριστικά του, περιλαμβάνει και υποκειμενικές διεργασίες που έχουν να κάνουν με τον τρόπο που ο ανθρώπινος εγκέφαλος αντιλαμβάνεται τα ακουστικά ερεθίσματα. Η μελέτη της υποκειμενικής αντιληπτικής ικανότητας των ήχων από τον άνθρωπο αποτελεί το αντικείμενο μελέτης του πεδίου της ψυχοακουστικής.

### 2.1.1 Συχνότητα και τονικότητα

Η απλούστερη μορφή κυματομορφής που μπορεί να υπάρξει είναι η ημιτονοειδής η οποία και θεωρείται το πρωτότυπο για την ακουστική αντίληψη μίας νότας. Για το λόγο αυτό ο ήχος που προκύπτει από μία ημιτονοειδή κυματομορφή αναφέρεται και ως αρμονικός ήχος ή καθαρός τόνος. Η έννοια της συχνότητας είναι άμεσα συνδεδεμένη με την έννοια της τονικότητας, για παράδειγμα μία ημιτονοειδής κυματομορφή 440Hz αντιστοιχεί στην τονικότητα A4. Καθώς μία μικρή αλλαγή στη συχνότητα δεν οδηγεί σε αντιληπτή αλλαγή της τονικότητας, ένα ολόκληρο εύρος συχνοτήτων αντιστοιχείται με μία συγκεκριμένη τονικότητα.

Όπως αναφέρθηκε και σε προηγούμενη παράγραφο, δύο νότες γίνονται αντιληπτές ως όμοιες όταν διαφέρουν κατά μία οκτάβα (ο λόγος τους ισούται με δύναμη του 2). Για παράδειγμα οι νότες A3 (220Hz), A4 (440Hz) και A5 (880Hz) ακούγονται όμοιες. Με άλλα λόγια, η ανθρώπινη αντίληψη της τονικότητας είναι λογαριθμικής φύσης. Χρησιμοποιώντας τον Πίνακα 1 με τις MIDI τιμές, μπορούμε να αποδώσουμε σε κάθε τονικότητα  $p \in [0,127]$  μία κεντρική συχνότητα  $F_{pitch}(p)$  (σε Hz):

$$F_{pitch}(p) = 2^{(p-69)/12} \cdot 440 \quad (2.1)$$

Η συγκεκριμένη σχέση δίνει την τιμή  $F_{pitch}(p) = 440$  για τη νότα αναφοράς  $p=69$  (A4). Αυξάνοντας την τονικότητα κατά 12 (μία οκτάβα) έχουμε διπλασιασμό της συχνότητας:

$$F_{pitch}(p+12) = 2 \cdot F_{pitch}(p) \quad (2.2)$$

επίσης προκύπτει ότι ο λόγος δύο διαδοχικών τόνων  $p+1$  και  $p$  είναι:

$$F_{pitch}(p+1) / F_{pitch}(p) = 2^{1/12} \approx 1.059463 \quad (2.3)$$

που σημαίνει ότι πολλαπλασιάζοντας μία οποιαδήποτε τονικότητα με αυτόν τον αριθμό, αυξάνεται κατά ένα ημιτόνιο.

Οι πραγματικοί ήχοι απέχουν πολύ από τους απλούς ημιτονοειδείς. Μια απλή νότα ενός οργάνου αποτελεί ένα σύνθετο ήχο που αποτελείται από τη σύνθεση πολλών συχνοτήτων που αλλάζουν με τον χρόνο. Μια τέτοια μουσική νότα μπορεί να περιγραφεί από την υπέρθεση απλών ημιτονοειδών κυμάτων, κάθε ένα από τα οποία έχει τη δική του συχνότητα. Η μικρότερη από αυτές τις συχνότητες ονομάζεται θεμέλια συχνότητα και είναι αυτή που χαρακτηρίζει την τονικότητα της νότας. Η θεμέλια συχνότητα είναι αυτή που προκαλείται από την ταλάντωση ολόκληρου του μήκους της χορδής ή του αέρα στο σωλήνα ενός οργάνου. Αντίστοιχα, μία αρμονική είναι η συχνότητα ενός συμβαλλόμενου κύματος που είναι ακέραιο πολλαπλάσιο της θεμέλιας. Οι αρμονικές επομένως μίας συχνότητας  $\omega$  είναι μια αριθμητική σειρά  $\omega, 2\omega, 3\omega, 4\omega, \dots$ , όπου η απόσταση ανάμεσα σε δύο αρμονικές είναι σταθερή και ίση με τη θεμέλια. Από την άλλη οι οκτάβες είναι μία γεωμετρική πρόοδος  $\omega, 2\omega, 4\omega, 8\omega, \dots$  και αντιλαμβανόμαστε τις αποστάσεις τους ως «ίσες» από άποψη μουσικών διαστημάτων. Καθώς λοιπόν η αντίληψή μας περί τονικότητας είναι λογαριθμική ως προς τη συχνότητα, αντιλαμβανόμαστε τις υψηλότερες αρμονικές ως πιο «κοντινές» μεταξύ τους σε σχέση με τις χαμηλότερες.

### 2.1.2 Χροιά

Η χροιά αποτελεί ένα βασικό χαρακτηριστικό του ήχου και είναι αυτό που επιτρέπει στον ακροατή να ξεχωρίζει για παράδειγμα τα όργανα μεταξύ τους. Όπως και η τονικότητα, η χροιά είναι μία αντιληπτική ιδιότητα του ήχου. Η χροιά είναι δύσκολο να περιγραφεί και λόγω της ασάφειάς της περιγράφεται συχνά με πλάγιο τρόπο: χροιά είναι η ιδιότητα με την οποία ο ακροατής κρίνει δύο ήχους ως διαφορετικούς, χρησιμοποιώντας οποιοδήποτε κριτήριο πέρα από την τονικότητα, την ένταση ή τη διάρκεια του ήχου. Για παράδειγμα, η χροιά μας επιτρέπει να ξεχωρίσουμε τον ήχο ενός όμποε από αυτόν ενός βιολιού, ακόμη και όταν η τονικότητα και η έντασή τους είναι ταυτόσημες (Müller και Klaruri 2013). Σε ερευνητικό επίπεδο γίνονται προσπάθειες η χροιά να προσεγγιστεί με τη χρήση πιο αντικειμενικών χαρακτηριστικών του ήχου, όπως η χρονική και φασματική του εξέλιξη, η παρουσία ή απουσία τονικών παραγόντων και παραγόντων με μορφή θορύβου ή η κατανομή της ενέργειας στις αρμονικές του τόνου.

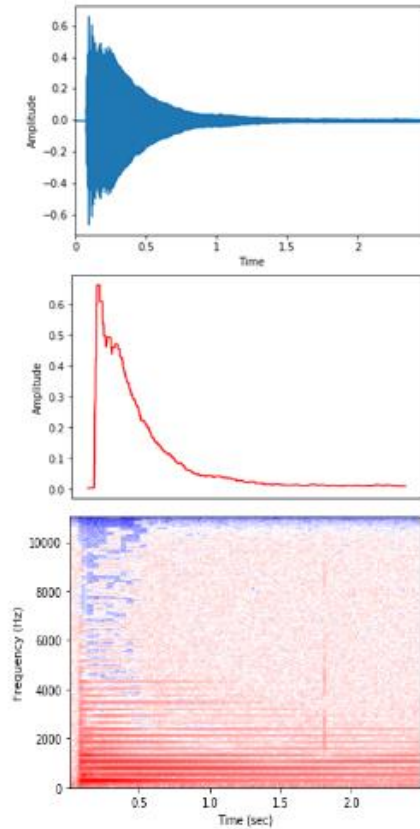
Το παίξιμο μίας νότας, για παράδειγμα ενός πλήκτρου του πιάνου, δημιουργεί μία πολύπλοκη ηχητική μίξη τα χαρακτηριστικά της οποίας αλλάζουν με το χρόνο. Στην αρχή του μουσικού τόνου συνήθως παρατηρείται μία απότομη αύξηση της ενέργειας και έχουμε τη φάση κατά την οποία ο ήχος δημιουργείται (attack). Στη φάση αυτή παρατηρείται μεγάλος βαθμός μη περιοδικών παραγόντων που απλώνονται σε όλο το εύρος των συχνοτήτων, χαρακτηριστικό που παρατηρείται και στο θόρυβο. Στην επόμενη φάση (decay) ο ήχος σταθεροποιείται και περνάει σε μία σταθερή περιοδική κατάσταση. Η τρίτη φάση (sustain) καταλαμβάνει τη μεγαλύτερη διάρκεια του τόνου και κατά τη διάρκειά της η ενέργεια παραμένει σταθερή ή μειώνεται ελαφρώς. Στη τέταρτη και τελευταία φάση (release) ο μουσικός τόνος σβήνει και στο παράδειγμα του πιάνου η φάση αυτή ξεκινάει όταν αφηθεί το πλήκτρο.

Η περιβάλλουσα πλάτους (amplitude envelope) μιας κυματομορφής αποτελεί μία καμπύλη που περιγράφει τη μεταβολή του συνολικού πλάτους της (**Εικόνα 2-2**) και σχετίζεται με την ένταση και την ενέργεια του παραγόμενου ήχου. Οι τέσσερις φάσεις που αναφέρθηκαν καθορίζονται από τον τρόπο μεταβολής της περιβάλλουσας ενός μουσικού τόνου. Στην ηχητική σύνθεση, η περιβάλλουσα ενός σήματος αναφέρεται συνήθως με τα αρχικά ADSR, από τα αρχικά των τεσσάρων φάσεων που προαναφέρθηκαν (Attack, Decay, Sustain, Release). Η διάρκεια και τα πλάτη των φάσεων επηρεάζουν καθοριστικά το πώς θα ακουστεί ο ήχος.

Ένας από τους πιο σημαντικούς παράγοντες που χαρακτηρίζουν τη χροιά είναι η παρουσία συγκεκριμένων αρμονικών και η σχετική τους ένταση (Muller και Klaruri 2013). Στην **Εικόνα 2-2** φαίνεται η περιβάλλουσα πλάτους και το φασματογράφημα μιας κυματομορφής (νότα C4 ενός πιάνου). Τόσο η θεμέλια συχνότητα (261,6 Hz) όσο και οι αρμονικές της (πολλαπλάσια του 261,6) διακρίνονται με ξεκάθαρες κόκκινες γραμμές. Στο φασματογράφημα φαίνεται ανάλογα και η φθορά (decay) των αρμονικών. Το μεγαλύτερο μέρος της ενέργειας του τόνου είναι συγκεντρωμένου στις χαμηλότερες δυναμικές.

Γενικά οι αρμονικές δεν εμφανίζονται με την ίδια ένταση σε όλες τις περιπτώσεις. Έγχορδα όργανα όπως το βιολί τείνουν να έχουν πιο πλούσιο φάσμα αρμονικών με μεγάλο μέρος της ενέργειας να συγκεντρώνεται και στις υψηλές αρμονικές, ενώ σε πνευστά κλειστού σωλήνα οι αρμονικές που είναι άρτια πολλαπλάσια της θεμέλιας δεν εμφανίζονται καθόλου. Σε ένα φαγκότο, η θεμέλια συχνότητα συχνά περιλαμβάνει πολύ λιγότερη ενέργεια από τις ψηλότερες αρμονικές, ενώ σε ένα διαπασών το μεγαλύτερο μέρος της ενέργειας είναι συγκεντρωμένο στη θεμέλια συχνότητα.





**Εικόνα 2-2**

*Η κυματομορφή, η περιβάλλουσα πλάτους και το φασματογράφημα μιας νότας C στο πιάνο.*

## 2.2 Χαρακτηριστικά του Ήχου

Προκειμένου να καταστήσουμε τα διάφορα μουσικά έργα συγκρίσιμα και προσβάσιμα από αλγορίθμους, το πρώτο βήμα στις περισσότερες διαδικασίες επεξεργασίας ήχου και μουσικής είναι η εξαγωγή κάποιων χαρακτηριστικών (features) του ήχου (Ο όρος feature θα χρησιμοποιείται από εδώ και πέρα στην παρούσα εργασία). Από τα διάφορα features που μπορούν να εξαχθούν, το καθένα αναπαριστά κάποιο επιμέρους αντιληπτικό χαρακτηριστικό του ήχου όπως είναι π.χ. το τονικό ύψος, η χροιά, η ενέργεια κ.α. Προφανώς η αντίληψη της ομοιότητας αποτελεί κρίσιμο παράγοντα στη σχεδίαση και εξαγωγή μουσικών features. Όπως προαναφέρθηκε, σε μία εφαρμογή μουσικού συγχρονισμού θα επιθυμούσαμε να χαρακτηρίσουμε τον ήχο ανεξάρτητα από λεπτομέρειες που αφορούν την ερμηνευτική χροιά ή το είδος των οργάνων. Αντίθετα, άλλες εφαρμογές μουσικής έκφρασης μπορεί να απαιτούν τη μέτρηση τέτοιων λεπτομερειών που συνδέονται με την προσωπική μουσική άρθρωση και ερμηνευτική έκφραση του κάθε μουσικού.

### 2.2.1 Φασματογράφημα λογαριθμικών συχνοτήτων

Ο μετασχηματισμός Fourier και συγκεκριμένα ο Short Term Fourier Transform (STFT) αποτελεί το πρώτο υπολογιστικό βήμα για την εξαγωγή ενός μεγάλου αριθμού από features. Αν  $x$  είναι ένα διακριτό σήμα (πραγματικών τιμών) με ρυθμό δειγματοληψίας  $F_s$ ,  $w: [0: N-1] \rightarrow \mathbb{R}$  μία πραγματική διακριτού χρόνου παραθυρική συνάρτηση μεγέθους  $N \in \mathbb{N}$ , και  $H \in \mathbb{N}$  το μέγεθος του hop, τότε ο STFT  $X$  του  $x$  θα είναι:

$$X(n, k) := \sum_{l=0}^{N-1} x(l+nH)w(l) \exp(-2\pi ikl / N) \quad (2.4)$$

Με  $n \in \mathbb{Z}$  και  $k \in [0: K]$ , όπου  $K = N / 2$  ο δείκτης των συχνοτήτων σε αντιστοιχία με την συχνότητα Nyquist. Ο μιγαδικός αριθμός  $X(n, k)$  υποδηλώνει τον  $k$ -οστό συντελεστή Fourier για το  $n$ -οστό χρονικό πλαίσιο. Κάθε συντελεστής Fourier  $X(n, k)$  σχετίζεται με το φυσικό χρονικό σημείο  $T_{coef}(n) = nH / F_s$  σε δευτερόλεπτα και με τη φυσική συχνότητα  $F_{coef}(k) = kF_s / N$  σε Herz. Η χρήση της παραθυρικής συνάρτησης δίνει στον STFT το πλεονέκτημα ότι δεν δείχνει απλά ποιες συχνότητες περιέχονται στο σήμα αλλά και τα χρονικά διαστήματα στα οποία οι συχνότητες αυτές εμφανίζονται. Από τον STFT μπορεί να προκύψει το φασματογράφημα, το οποίο αποδίδει οπτικά το χρονικό – συχνοτικό περιεχόμενο του σήματος. Το φασματογράφημα είναι μια δισδιάστατη απεικόνιση του τετραγώνου του μέτρου του STFT:

$$Y(n, k) := |X(n, k)|^2 \quad (2.5)$$

Όπως είδαμε σε προηγούμενη παράγραφο, κάθε τονικότητα  $p \in [0, 127]$  αντιστοιχεί σε μία κεντρική συχνότητα  $F_{pitch}(p)$  (σε Hz):

$$F_{pitch}(p) = 2^{(p-69)/12} \cdot 440 \quad (2.6)$$

Η λογαριθμική φύση της αντίληψής μας για τη συχνότητα οδήγησε στη χρήση μιας αναπαράστασης της σχέσης χρόνου – συχνοτήτων όπου ο άξονας των συχνοτήτων εκφράζεται σε λογαριθμική κλίμακα, σε αντιστοιχία με τις τονικότητες του πρωτοκόλλου MIDI. Για να προκύψει μία τέτοια αναπαράσταση από ένα φασματογράφημα, η βασική ιδέα είναι να αντιστοιχίσουμε κάθε παράγοντα  $X(n, k)$  του φάσματος στην τονικότητα της οποίας η κεντρική συχνότητα είναι πιο κοντά στη συχνότητα  $F_{coef}(k)$ . Συγκεκριμένα, για κάθε τονικότητα  $p \in [0: 127]$  ορίζουμε το σύνολο:

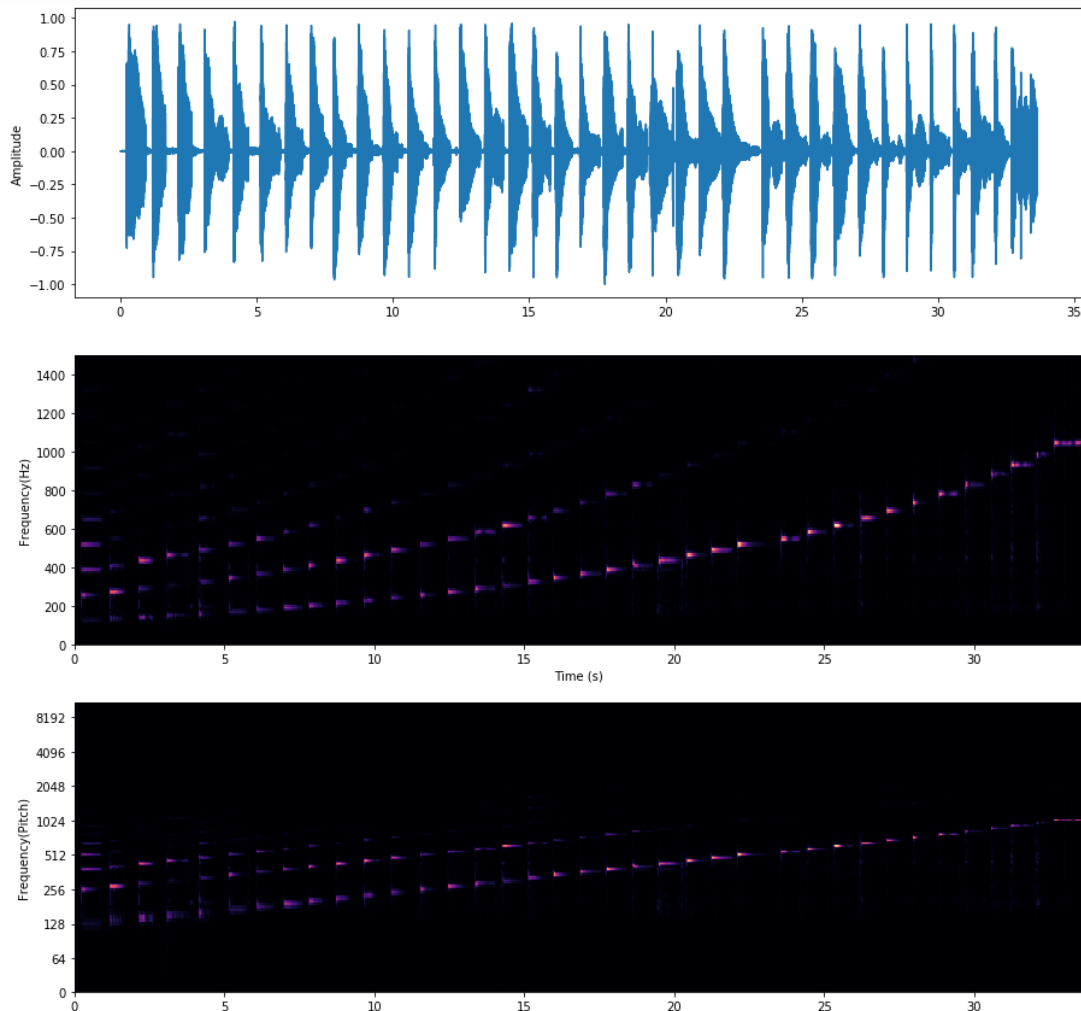
$$P(p) := \{k : F_{pitch}(p-0,5) \leq F_{coef}(k) < F_{pitch}(p+0,5)\} \quad (2.7)$$

προκύπτει έτσι ένα φασματογράφημα λογαριθμικών συχνοτήτων (Log-frequency spectrogram)  $Y_{LF} : \mathbb{Z} \times [0: 127]$  που ορίζεται ως:

$$Y_{LF}(n, p) := \sum_{k \in P(p)} |X(n, k)|^2 \quad (2.8)$$

Με την παραπάνω διαδικασία ο άξονας των συχνοτήτων χωρίζεται λογαριθμικά και αριθμείται γραμμικά (Εικόνα 2-3). Το αποτέλεσμα είναι ένα φασματογράφημα λογαριθμικών συχνοτήτων, στο οποίο οι νότες της χρωματικής κλίμακας έχουν μία γραμμικά αυξανόμενη απεικόνιση. Από την εικόνα αυτή επίσης φαίνεται ότι το παίξιμο μίας νότας σε ένα πραγματικό όργανο δημιουργεί ένα πολύπλοκο μίγμα διαφορετικών συχνοτήτων. Οι υψηλότερες συχνότητες φαίνεται να έχουν ένα πιο «καθαρό» φάσμα από άποψη αρμονικών. Στις χαμηλές συχνότητες το μεγαλύτερο μέρος της ενέργειας κατανέμεται συχνά σε

ψηλότερες αρμονικές, παρόλα αυτά ο ακροατής έχει την αντίληψη ενός ήχου χαμηλής τονικότητας. Το συχνοτικό περιεχόμενο των χαμηλών συχνοτήτων μπορεί να επηρεάζεται και από την απόκριση των μικροφώνων τα οποία συχνά καταγράφουν συχνότητες πάνω από ένα συγκεκριμένο όριο.



**Εικόνα 2-3**

α) Η κυματομορφή μιας χρωματικής κλίμακας στο πιάνο από C3 ως C6, β) Το φασματογράφημά της και γ) Το φασματογράφημα σε λογαριθμική κλίμακα

Ένας άλλος λόγος για την ελλiptή αναπαράσταση των χαμηλών τόνων έχει να κάνει με το γεγονός ότι ο διακριτός STFT πραγματοποιεί γραμμική δειγματοληψία του άξονα των συχνοτήτων ενώ το εύρος συχνοτήτων που αντιστοιχεί σε κάθε τονικότητα  $p$  υπολογίζεται λογαριθμικά. Συγκεκριμένα σε κάθε τονικότητα  $p$  αντιστοιχεί ένα εύρος  $BW(p)$  τέτοιο ώστε:

$$BW(p) := F_{pitch}(p+0,5) - F_{pitch}(p-0,5) \quad (2.9)$$

το οποίο μειώνεται όσο μικραίνει η τονικότητα. Για παράδειγμα για τη MIDI τονικότητα  $p=66$  έχουμε ένα εύρος 21,4Hz ενώ για  $p=54$  το εύρος πέφτει κάτω από τα 10,7 Hz. Αν για παράδειγμα έχουμε ρυθμό δειγματοληψίας 44100 Hz και μήκος παραθύρου  $N=4096$ , από τη

σχέση  $F_{coef}(k) = kF_s / N$  προκύπτει ότι ο STFT θα έχει μία διακριτική ικανότητα συχνοτήτων της τάξης των 10,8 Hz περίπου. Στην περίπτωση αυτή, η ανάλυση του φασματογραφήματος δεν είναι αρκετή για να ξεχωρίσει τονικότητα με MIDI τιμές μικρότερες του  $p=54$ , γεγονός που αποτελεί από τους βασικούς λόγους για την ελλιπή αναπαράσταση των χαμηλών συχνοτήτων σε λογαριθμική κλίμακα. Η συχνοτική ανάλυση του διακριτού STFT μπορεί να αυξηθεί αν μεγαλώσουμε το μέγεθος του παραθύρου  $N$ , αυτό όμως θα οδηγήσει σε ταυτόχρονη μείωση της χρονικής ανάλυσης και μπορεί να χάσουμε σημαντικές πληροφορίες όπως onset των νοτών. Μία εναλλακτική είναι να χρησιμοποιήσουμε πολλά φασματογραφήματα με διαφορετικούς ρυθμούς δειγματοληψίας και διαφορετικό μέγεθος παραθύρων. Καθώς τα απότομα γεγονότα όπως τα onset των νοτών γίνονται πιο εμφανή στις υψηλές συχνότητες, μπορούμε να χρησιμοποιήσουμε για αυτές μεγάλο ρυθμό δειγματοληψίας και μικρό παράθυρο προκειμένου να πετύχουμε υψηλή χρονική ανάλυση. Αυτό βέβαια θα έχει επίπτωση στην ανάλυση των συχνοτήτων γεγονός που είναι αποδεκτό λόγω της εκθετικής αύξησης των θεμέλιων σε αυτές τις περιοχές συχνοτήτων. Αντίστροφα για τις χαμηλές συχνότητες όπου η ανάλυση των συχνοτήτων γίνεται πιο σημαντική, μπορούμε να χρησιμοποιήσουμε χαμηλότερη δειγματοληψία και μεγαλύτερο παράθυρο με όποιο κόστος αυτό συνεπάγεται στη χρονική ανάλυση. (Müller 2007).

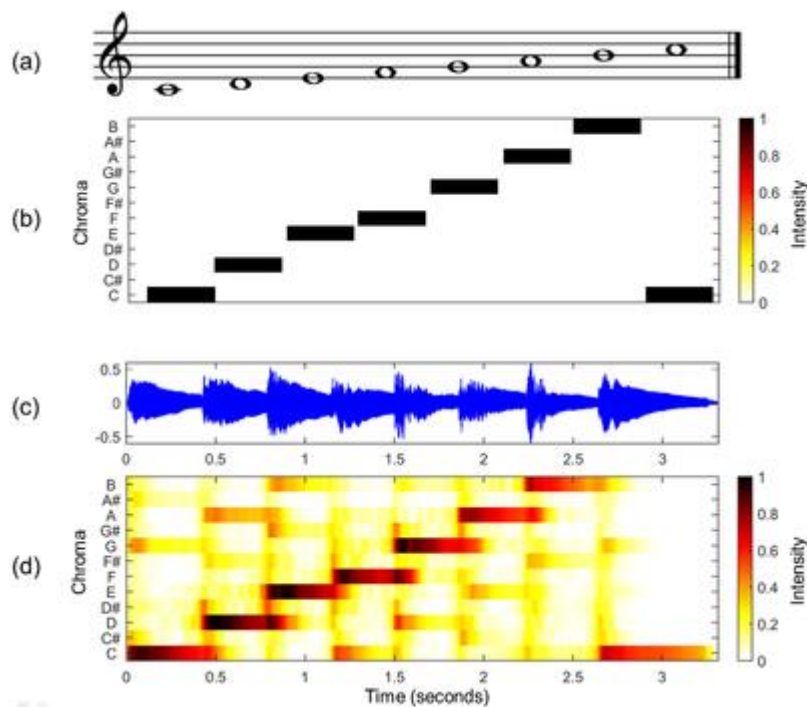
## 2.2.2 Chroma Features

Έχει ήδη αναφερθεί ότι η ανθρώπινη αντίληψη της τονικότητας είναι περιοδική με την έννοια ότι ο άνθρωπος αντιλαμβάνεται δύο τόνους ως χρωματικά ίδιους όταν αυτοί διαφέρουν κατά μία οκτάβα. Με βάση αυτή την παρατήρηση ένας τόνος μπορεί να χωριστεί σε δύο συνθετικά, το τονικό ύψος και το χρώμα. Το τονικό ύψος αναφέρεται στον αριθμό της οκτάβας. Το χρώμα αντιστοιχεί στο σύνολο  $\{C, C^\#, D, \dots, B\}$  που αποτελείται από τους δώδεκα τονικούς συμβολισμούς της Δυτικής μουσικής σημειογραφίας. Αντίστοιχα ένα chroma feature αντιπροσωπεύεται από το διάνυσμα  $x = (x(1), x(2), \dots, x(12))^T$ , όπου το  $x(1)$  αντιστοιχεί στο χρώμα C, το  $x(2)$  στο  $C^\#$  κ.ο.κ. (Muller and Ewert 2011). Μια τονική κλάση ορίζεται ως το σύνολο των τόνων που είναι ίδιοι χρωματικά.

Η βασική ιδέα των χρωματικών features είναι να συγκεντρώσουμε όλες τις φασματικές πληροφορίες που συνδέονται με μία συγκεκριμένη τονική κλάση σε ένα χρωματικό παράγοντα. Έχοντας ένα φασματογράφημα με λογαριθμική κλίμακα συχνότητας  $Y_{LF} : \mathbb{Z} \times [0:127] \rightarrow \mathbb{R}_{\geq 0}$  όπως το ορίσαμε στη παράγραφο 3.1 τότε μία χρωματική αναπαράσταση ή χρωματογράφημα (chromagram)  $\mathbb{Z} \times [0:11] \rightarrow \mathbb{R}_{\geq 0}$  προκύπτει αθροίζοντας όλους του τονικούς παράγοντες που ανήκουν στο ίδιο χρώμα:

$$C(n, c) := \sum_{\{p \in [0:127]; p \bmod 12 = c\}} Y_{LF}(n, p) \quad (2.10)$$

Στην **Εικόνα 2-4** (Wikipedia.org 2020b) φαίνεται το χρωματογράφημα της κλίμακας C ματζόρε όπως προκύπτει από την παρτιτούρα (2.6b) και από την ηχογράφιση της κλίμακας σε πιάνο (2.6d). Η επαναληπτική φύση των χρωματικών χαρακτηριστικών φαίνεται στην νότα C. Από το χρωματογράφημα της ηχογράφισης φαίνεται ότι λόγω της ύπαρξης ανώτερων αρμονικών μέρος της ενέργειας του σήματος κατανέμεται σε διαφορετικές χρωματικές ζώνες, αν και στο πιάνο αναπαράγεται μία νότα κάθε φορά



Εικόνα 2-4

a) Παρτιτούρα μία κλίμακας C ματζόρε b) Χρωματογράφημα όπως προκύπτει από τη σημειογραφία c) Ηχογράφηση της κλίμακας από ένα πιάνο d) Χρωματογράφημα της ηχογράφησης

Κατά την παραπάνω διαδικασία οι 128 τονικές ζώνες του φασματογραφήματος λογαριθμικών συχνοτήτων συμπύσσονται στις 12 χρωματικές ζώνες του χρωματογραφήματος. Αυτό έχει σαν αποτέλεσμα τη μη αναστρέψιμη απώλεια κάποιων πληροφοριών. Για παράδειγμα στο χρωματογράφημα μία νότα A4 δεν διακρίνεται από μια νότα A5. Αυτή η απώλεια πληροφοριών είναι επιθυμητή σε κάποια είδη εφαρμογών καθώς προσδίδει σε μεγάλο βαθμό ανεξαρτησία από τις διακυμάνσεις της χροιάς. Όπως αναφέρθηκε σε προηγούμενη παράγραφο, η χροιά ενός οργάνου εξαρτάται σε μεγάλο βαθμό από την κατανομή της ενέργειας στις διάφορες αρμονικές. Στην περίπτωση του χρωματογραφήματος, η 1<sup>η</sup>, 2<sup>η</sup>, 4<sup>η</sup>, 8<sup>η</sup> κ.ο.κ. αρμονικές απέχουν μεταξύ τους μία οκτάβα και έτσι η ενέργειά τους συγκεντρώνεται στην ίδια χρωματική ζώνη. Αντίστοιχα η 3<sup>η</sup>, 6<sup>η</sup>, 12<sup>η</sup> κ.ο.κ. αρμονικές μοιράζονται το ίδιο ηχόχρωμα. Έτσι πολλές πληροφορίες που σχετίζονται με τη διάκριση της χροιάς διαφορετικών οργάνων προσπίπτουν στις ίδιες χρωματικές ζώνες.

### 2.2.3 Λογαριθμική Συμπύεση

Η λογαριθμική συμπύεση  $\log(1+x)$  αποτελεί μια μέθοδο γραφικής απεικόνισης σχέσεων όπου η εξαρτώμενη μεταβλητή έχει μεγάλο δυναμικό εύρος. Χρησιμοποιείται προκειμένου η μικρές τιμές να μην κυριαρχούνται από τις μεγαλύτερες όπως θα συνέβαινε σε μία γραμμική αναπαράσταση, αλλά και να μην έχουμε αρνητικές τιμές όπως σε μία λογαριθμική αναπαράσταση. Αντίστοιχα σε ένα φασματογράφημα ή χρωματογράφημα, μικρές αλλά χρήσιμες τιμές δεδομένων μπορεί να επισκιάζονται από τις μεγαλύτερες και να «χάνονται» μέσα στην απεικόνιση. Σκοπός της λογαριθμικής συμπύεσης είναι η εξισορρόπηση αυτής της ανακολουθίας ενισχύοντας τις μικρές τιμές και μειώνοντας έτσι τη διαφορά στην απεικόνιση

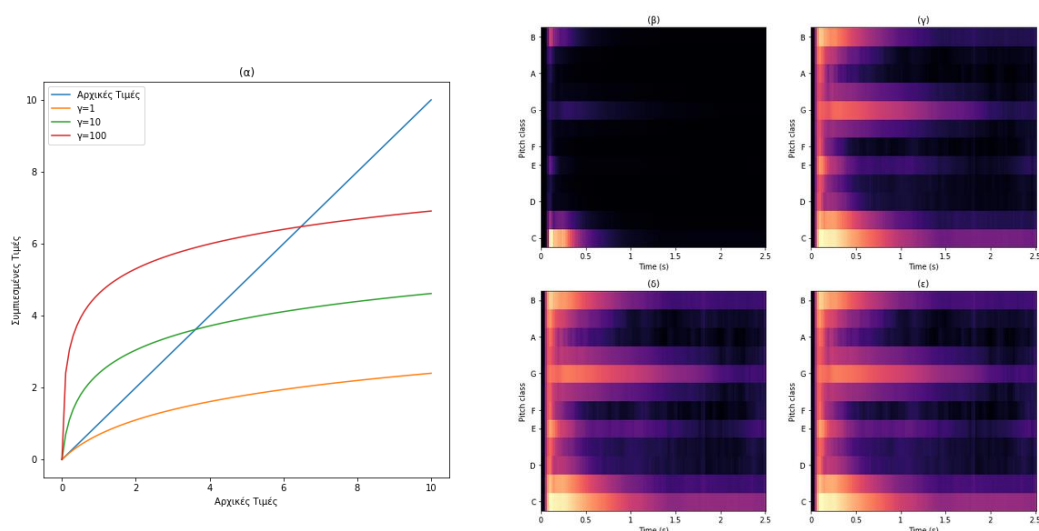
από τις μεγαλύτερες. Για το σκοπό αυτό ορίζουμε μία θετική σταθερά  $\gamma \in \mathbb{R}_{>0}$  και την συνάρτηση  $\Gamma_\gamma : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$  τέτοια ώστε:

$$\Gamma_\gamma(v) := \log(1 + \gamma \cdot v) \quad (2.11)$$

για θετικές τιμές  $v \in \mathbb{R}_{>0}$ . Η συνάρτηση θα αποδίδει θετικές τιμές για κάθε  $v \in \mathbb{R}_{>0}$ . Έτσι για παράδειγμα, για ένα χρωματογράφημα  $C$ , η συμπιεσμένη έκδοσή του ορίζεται ως :

$$(\Gamma_\gamma \circ C)(n, c) := \log(1 + \gamma \cdot C(n, c)) \quad (2.12)$$

Η σταθερά  $\gamma$  καθορίζει τον βαθμό της συμπίεσης και η κατάλληλη επιλογή της εξαρτάται από τα χαρακτηριστικά των δεδομένων και το είδος της εφαρμογής. Για παράδειγμα όταν υπάρχει παρουσία θορύβου πρέπει να βρεθεί μία ισορροπία ώστε να ενισχυθούν τα επιθυμητά αδύναμα στοιχεία του σήματος χωρίς την ενίσχυση του ανεπιθύμητου θορύβου (Müller 2015). Στην **Εικόνα 2-5** φαίνεται η εφαρμογή της συνάρτησης  $\Gamma_\gamma$  στο χρωματογράφημα μιας νότας  $C$  στο πιάνο για διάφορες τιμές της σταθεράς  $\gamma$ .



**Εικόνα 2-5**

(α) Η συνάρτηση συμπίεσης για διαφορετικές τιμές της σταθεράς  $\gamma$  (β) Το αρχικό χρωματογράφημα της νότας  $C$  και τα αντίστοιχα χρωματογραφήματα για (γ)  $\gamma=1$  (δ)  $\gamma=10$  και (ε)  $\gamma=100$

## 2.2.4 Κανονικοποίηση

Όπως είδαμε τα χρωματικά χαρακτηριστικά παραμένουν (τουλάχιστον σε μεγάλο βαθμό) αμετάβλητα στις μεταβολές της χροιάς. Για να γίνουν αμετάβλητα και στις αλλαγές των δυναμικών μπορούν να κανονικοποιηθούν. Κάθε ένα από τα 12-διάστατα χρωματικά διανύσματα  $x = C(n, \cdot) \in \mathbb{R}^{12}$  για ένα δείκτη  $n \in \mathbb{Z}$  (n-ιοστή στήλη του  $C$ ) αντικαθίσταται από το διάνυσμα  $x / \|x\|$  (Müller 2015), όπου  $\|x\|$  η Ευκλείδεια νόρμα του διανύσματος  $x$ . Από τη γραμμική άλγεβρα γνωρίζουμε ότι η Ευκλείδεια ή  $L_2$  νόρμα ενός διανύσματος  $x$ ,  $N$  διαστάσεων είναι:

$$\|x\| = \sqrt{\sum_{i=1}^N |x_i|^2} \quad (2.13)$$

Επομένως για ένα χρωματικό διάνυσμα  $x = (x(0), x(1), \dots, x(11))^T$  θα έχουμε:

$$\|x\| := \sqrt{\sum_{i=0}^{11} |x(i)|^2} \quad (2.14)$$

Η διαδικασία της κανονικοποίησης είναι δυνατή μόνο αν  $\|x\| > 0$ . Επίσης για πολύ μικρές τιμές του  $\|x\|$  που μπορεί να προκύψουν σε διαστήματα σιγής (π.χ. στην αρχή της ηχογράφησης ή σε μεγάλες παύσεις) η κανονικοποίηση θα οδηγήσει σε χρωματικές τιμές χωρίς νόημα. Για αυτό τον λόγο, όταν το  $\|x\|$  είναι κάτω από ένα συγκεκριμένο όριο, το διάνυσμα  $x$  μπορεί να αντικατασταθεί από ένα ομοιόμορφο διάνυσμα αντί να διαιρεθεί με το  $\|x\|$ . Έτσι δοθέντος ενός ορίου  $\varepsilon > 0$ , ορίζουμε την συνάρτηση  $\pi^\varepsilon(x)$  ως:

$$\pi^\varepsilon(x) := \begin{cases} x/\|x\| & \alpha \nu \quad \|x\| > \varepsilon \\ (1, 1, \dots, 1)^T / \sqrt{12} & \alpha \nu \quad \|x\| \leq \varepsilon \end{cases} \quad (2.15)$$

Το όριο  $\varepsilon$  αποτελεί ένα παράγοντα που πρέπει να οριστεί προσεκτικά και ανάλογα με τις απαιτήσεις της εκάστοτε εφαρμογής. Καθώς το ανθρώπινο αυτί είναι ιδιαίτερα ευαίσθητο ακόμη και ηχητικοί παράγοντες πολύ μικρής ενεργείας μπορεί να έχουν σημαντικό ρόλο. Για αυτό το όριο  $\varepsilon$  συνήθως ορίζεται αρκετές τάξεις μεγέθους (διαιρέτης από  $10^4$  ως  $10^6$ ) κάτω από το μέσο επίπεδο του ήχου (Müller 2015).

### 2.2.5 Αλλαγή Τονικότητας

Μια ιδιαίτερα χρήσιμη διαδικασία στις χρωματογραφικές αναπαραστάσεις είναι η αντιμετάθεση (transposition) των χρωματικών χαρακτηριστικών σε αντιστοιχία με τη μουσική διαδικασία της αλλαγής τονικότητας (για μια μελωδία ή για ολόκληρο μουσικό έργο). Μια τέτοια διαδικασία είναι αρκετά συνηθισμένη στη μουσική προκειμένου να προσαρμοστεί ένα μουσικό κομμάτι στο τονικό εύρος ενός άλλου οργάνου ή ενός άλλου τραγουδιστή. Για παράδειγμα ένα κομμάτι γραμμένο για μία σοπράνο (η υψηλότερη γυναικεία φωνή) μπορεί να αντιμετωπιστεί επτά ημιτόνια κάτω προκειμένου να προσαρμοστεί για μία κοντράλτο (η χαμηλότερη γυναικεία φωνή). Από τεχνικής άποψης η διαδικασία αυτή περιλαμβάνει τη μετακίνηση μίας ακολουθίας από νότες χαμηλότερα ή ψηλότερα κατά ένα συγκεκριμένο αριθμό διαστημάτων (ημιτόνιων). Σε χρωματικό επίπεδο, για ένα διάνυσμα  $x = (x(0), x(1), \dots, x(10), x(11))^T \in \mathbb{R}^{12}$  και καθώς οι χρωματικές αξίες ακολουθούν μία κυκλική διάταξη (**Εικόνα 1-2**), μπορούμε να ορίσουμε ένα τελεστή κυκλικής ολίσθησης  $\rho: \mathbb{R}^{12} \rightarrow \mathbb{R}^{12}$  τέτοιο ώστε:

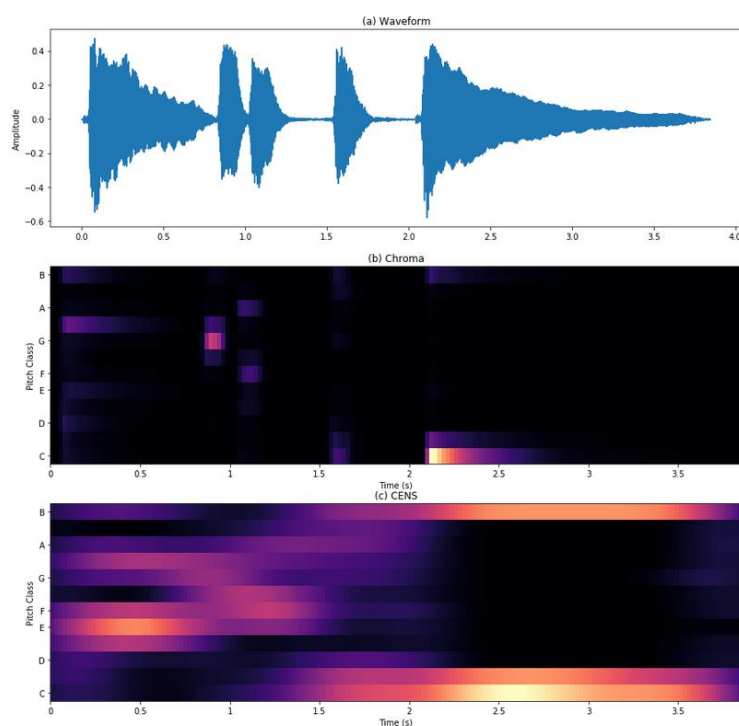
$$\rho(x) = (x(11), x(0), x(1), \dots, x(10))^T \quad (2.16)$$

Με άλλα λόγια, η νότα C στο διάνυσμα  $x$  γίνεται η C<sup>#</sup> στο  $\rho(x)$ , η νότα C<sup>#</sup> γίνεται D και αντίστοιχα μέχρι την τελική νότα B που γίνεται C. Ο τελεστής κυκλικής ολίσθησης μπορεί να

εφαρμοστεί διαδοχικά δίνοντας  $\rho^i = \rho \circ \rho^{i-1}$  για  $i \in \mathbb{N}$ , που σημαίνει ολίσθηση κατά  $i$  ημιτόνια ψηλότερα. Προφανώς ισχύει  $\rho^{12}(x) = x$ , που σημαίνει ότι αν ένα διάνυσμα ολισθήσει κατά μία οκτάβα, προκύπτει πάλι το αρχικό διάνυσμα. Εφαρμόζοντας την κυκλική ολίσθηση σε κάθε ένα από τα frames ενός χρωματογραφήματος οδηγούμαστε στην κυκλική ολίσθηση ολόκληρου του χρωματογραφήματος σε κάθετη κατεύθυνση.

## 2.2.6 CENS Features

Τα Chroma Energy Normalized Statistics (CENS) αποτελούν παραλλαγή των chroma features. Η βασική τους ιδέα είναι ότι η εξαγωγή στατιστικών με χρήση παραθύρου μεγάλου μεγέθους εξομαλύνει τις τοπικές αποκλίσεις στο tempo που μπορεί να οφείλονται σε λόγους μουσικής άρθρωσης ή σε διάφορες μουσικές τεχνικές όπως για παράδειγμα τα αρπίσματα (arpeggio chords). Τα CENS features επιδεικνύουν χαμηλή επιρροή σε μεταβολές της χροιάς, των δυναμικών και της μουσικής άρθρωσης και για αυτό χρησιμοποιούνται ευρέως σε τεχνικές μουσικού συγχρονισμού, αντιστοίχισης και ανάκτησης μουσικής πληροφορίας. Στην **Εικόνα 2-6** φαίνεται η κυματομορφή, τα chroma features και τα CENS features για μία σύντομη ηχογράφηση σε πιάνο.



**Εικόνα 2-6**

a) Κυματομορφή b) Chroma Features και c) CENS features μιας σύντομης ηχογράφησης πιάνου

Για να δημιουργήσουμε μια αναπαράσταση CENS features ξεκινώντας από ένα χρωματογράφημα ακολουθούμε τα παρακάτω βήματα (Müller, Kurth and Clausen 2005b):

1. Κάθε χρωματικό διάνυσμα κανονικοποιείται με χρήση της νόρμας Manhattan, έτσι ώστε οι δώδεκα χρωματικές αξίες να αθροίζονται στη μονάδα. Προκύπτει έτσι η ακολουθία  $X = (x_1, x_2, \dots, x_N)$  των κανονικοποιημένων χρωματικών διανυσμάτων  $x_n \in [0, 1]^{12}$ ,  $n \in [1 : N]$ . Κάθε ένα από αυτά τα διανύσματα έχει μόνο θετικές τιμές



μεταξύ 0 και 1. Στη συνέχεια ορίζουμε μία συνάρτηση κβαντισμού  $Q: [0, 1] \rightarrow \{0, 1, 2, 3, 4\}$  τέτοια ώστε:

$$Q(a) := \begin{cases} 0 & \alpha \nu \quad 0 \leq \alpha < 0.05, \\ 1 & \alpha \nu \quad 0.05 \leq \alpha < 0.1, \\ 2 & \alpha \nu \quad 0.1 \leq \alpha < 0.2, \\ 3 & \alpha \nu \quad 0.2 \leq \alpha < 0.4, \\ 4 & \alpha \nu \quad 0.4 \leq \alpha \leq 1. \end{cases} \quad (2.17)$$

Εφαρμόζοντας τη συνάρτηση  $Q$  σε κάθε χρωματικό διάνυσμα  $x_n$  έχουμε:

$$Q(x_n) = (Q(x_n(0)), \dots, Q(x_n(11)))^T \quad (2.18)$$

Η συνάρτηση  $Q$  αποδίδει την τιμή 4 σε κάποιο χρωματικό παράγοντα αν η αντίστοιχη χρωματική κλάση περιέχει περισσότερο από το 40% της συνολικής ενέργειας του σήματος και αντίστοιχα για τις υπόλοιπες τιμές. Οι χρωματικοί παράγοντες που βρίσκονται κάτω από το όριο του 5% μηδενίζονται, γεγονός που καθιστά τα CENS features ανθεκτικά στο θόρυβο. Τα όρια έχουν οριστεί σε λογαριθμική μορφή σε αναλογία με τη λογαριθμική αντίληψη που έχουμε για την ένταση του ήχου.

2. Στο δεύτερο βήμα, η ακολουθία  $Q(x_n)$  που προέκυψε από τη διαδικασία του κβαντισμού εξομαλύνεται περεταίρω. Αρχικά ορίζουμε έναν αριθμό  $\ell \in \mathbb{N}$  που καθορίζει το μήκος ενός παράθυρου εξομάλυνσης (π.χ. παράθυρο Hann) και στη συνέχεια υπολογίζουμε τοπικές μέσες τιμές, με χρήση της παραθυρικής συνάρτησης, για κάθε έναν από τους 12 παράγοντες της ακολουθίας  $Q(x_n)$ . Αυτό οδηγεί και πάλι σε μία ακολουθία διανυσμάτων με θετικές τιμές που αντιπροσωπεύουν ένα είδος «ζυγισμένων» στατιστικών της κατανομής της ενέργειας μέσα σε ένα παράθυρο  $\ell$  συνεχόμενων χρωματικών διανυσμάτων. Στο τελικό βήμα, ο ρυθμός δειγματοληψίας της ακολουθίας μειώνεται κατά ένα παράγοντα  $d$  (downsample) και τα διανύσματα που προκύπτουν κανονικοποιούνται με χρήση της Ευκλείδειας νόρμας.

Το αποτέλεσμα της παραπάνω διαδικασίας είναι ένα νέο feature που ονομάζεται  $CENS_d^\ell$ , μια παραλλαγή των chroma features που βασίζεται στους παράγοντες  $\ell$  (μέγεθος παραθύρου) και  $d$  (μειωμένη δειγματοληψία). Τα CENS αποτελούν ένα feature ευέλικτο και υπολογιστικά συμφέρον. Η κανονικοποίηση στην αρχή του βήματος 1 καθιστά τα CENS ανεπηρέαστα από μεταβολές των δυναμικών ενώ ο ορισμός ενός ορίου (threshold) ενέργειας τα καθιστά ανεπηρέαστα και από παράγοντες θορύβου που εμφανίζονται στην αρχή (attack) των νοτών. Επίσης μπορούμε να μεταβάλλουμε παράγοντες όπως το μήκος του παραθύρου ή τα όρια του βήματος 1, προκειμένου να ενισχύσουμε ή να αποκρύψουμε συγκεκριμένα χαρακτηριστικά, χωρίς να χρειάζεται να επαναλάβουμε το υπολογιστικά περίπλοκο βήμα του υπολογισμού των χρωματογραφικών χαρακτηριστικών (Müller, Kurth and Clausen 2005b).

Τα CENS features μπορούν να υπολογιστούν ξεκινώντας από οποιαδήποτε μορφή αναπαράστασης ενός χρωματογραφήματος. Για παράδειγμα, αντί να ξεκινήσουμε από ένα χρωματογράφημα βασισμένο στον STFT θα μπορούσαμε να ξεκινήσουμε από ένα βασισμένο σε multirate φίλτρα που ήταν και η προσέγγιση της πρωτότυπης δημοσίευσης (Müller, Kurth and Clausen 2005a).

## 2.3 Ο Αλγόριθμος DTW

Ο αλγόριθμος Δυναμικής Χρονικής Στρέβλωσης (Dynamic Time Warping - DTW) είναι ένας αλγόριθμος για τη μέτρηση ομοιότητας ανάμεσα σε δύο χρονικές ακολουθίες που μπορεί να διαφέρουν σε ταχύτητα. Για παράδειγμα ομοιότητες στο περπάτημα δύο ατόμων θα μπορούσαν να ανιχνευθούν με τη χρήση του DTW ακόμη και αν το ένα άτομο περπατάει γρηγορότερα από το άλλο ή αν υπήρχαν επιταχύνσεις και επιβραδύνσεις κατά τη διάρκεια της παρατήρησης. Οι δύο ακολουθίες υφίστανται μια μη γραμμική «παραμόρφωση» (warping) στο πεδίο του χρόνου προκειμένου να αντιστοιχηθούν μεταξύ τους. Ο DTW βρίσκει εφαρμογές σε δεδομένα βίντεο, ήχου και γραφικών – γενικά οποιοδήποτε είδος δεδομένων μπορεί να μετατραπεί σε μία γραμμική ακολουθία, μπορεί να αναλυθεί με τον DTW. Από τις πρώτες και πιο γνωστές εφαρμογές του είναι στην αυτόματη αναγνώριση ομιλίας, προκειμένου να αντιμετωπιστούν οι διαφορετικές ταχύτητες των ομιλητών (Wikipedia.org 2020c).

Γενικά, Ο DTW αποτελεί μία μέθοδο που υπολογίζει τη βέλτιστη αντιστοίχιση μεταξύ δύο ακολουθιών με συγκεκριμένους περιορισμούς και κανόνες:

- Κάθε δείκτης της πρώτης ακολουθίας πρέπει να αντιστοιχηθεί με ένα ή περισσότερους δείκτες της δεύτερης ακολουθίας και αντίστροφα.
- Ο πρώτος δείκτης της πρώτης ακολουθίας πρέπει να αντιστοιχηθεί με τον πρώτο δείκτη της άλλης ακολουθίας (χωρίς απαραίτητα να είναι η μόνη του αντιστοίχιση)
- Ο τελευταίος δείκτης της πρώτης ακολουθίας πρέπει να αντιστοιχηθεί με τον τελευταίο δείκτη της άλλης ακολουθίας (χωρίς απαραίτητα να είναι η μόνη του αντιστοίχιση)
- Η αντιστοίχιση των δεικτών της πρώτης ακολουθίας με δείκτες της δεύτερης ακολουθίας πρέπει να αυξάνεται μονοτονικά και αντίστροφα. Αν για παράδειγμα  $j > i$  είναι δείκτες της πρώτης ακολουθίας, δεν πρέπει να υπάρχουν δύο δείκτες  $l > k$  στην δεύτερη ακολουθία τέτοιοι ώστε ο δείκτης  $i$  να αντιστοιχίζεται με τον δείκτη  $l$  και ο δείκτης  $j$  να αντιστοιχίζεται με τον δείκτη  $k$ .

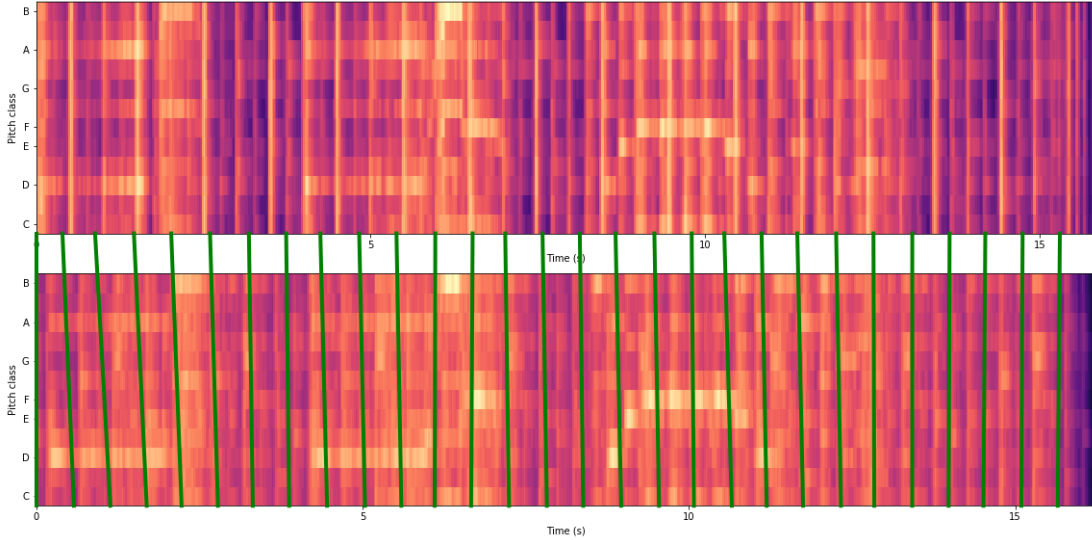
Ως βέλτιστη θεωρείται η αντιστοίχιση εκείνη που ικανοποιεί όλους τους περιορισμούς και κανόνες και έχει το ελάχιστο κόστος. Το κόστος υπολογίζεται ως το άθροισμα από τα τοπικά κόστη που προκύπτουν για κάθε ζεύγος δεικτών που αντιστοιχίζονται μεταξύ τους. Το τοπικό κόστος για ένα ζεύγος δεικτών υπολογίζεται ως η απόσταση μεταξύ των τιμών που τους αντιστοιχούν (π.χ Ευκλείδεια Απόσταση ή απόσταση συνημίτονου). Με βάση τις αποστάσεις μεταξύ όλων των ζευγών δεικτών κατασκευάζεται ένας πίνακας κόστους (cost matrix). Σκοπός είναι να βρεθεί ένα «μονοπάτι»  $P$  μέσα στον πίνακα κόστους για το οποίο το άθροισμα των τιμών τοπικού κόστους ελαχιστοποιείται (musicinformationretrieval.com 2020):

$$\min \sum_{(i,j) \in P} d(x[i], y[j]) \quad (2.19)$$

### 2.3.1 Βασική Προσέγγιση για Μουσικό Συγχρονισμό

Σε μία εφαρμογή μουσικού συγχρονισμού οι χρονικές ακολουθίες τις οποίες θα πρέπει να αντιστοιχήσει ο DTW θα μπορούσαν να είναι, για παράδειγμα, δύο χρωματογραφήματα από δύο διαφορετικές εκτελέσεις του ίδιου μουσικού κομματιού. Έστω ότι η πρώτη ακολουθία είναι η  $X = (x_1, x_2, \dots, x_N)$  με  $N \in \mathbb{N}$  να είναι το μήκος της και η δεύτερη ακολουθία είναι η

$Y = (y_1, y_2, \dots, y_M)$  με  $M \in \mathbb{N}$  να είναι το μήκος της. Τα στοιχεία  $x_n$  και  $y_m$  των δύο ακολουθιών είναι χρωματογραφικά διανύσματα και οι δύο ακολουθίες αντιστοιχούν στο ίδιο μουσικό περιεχόμενο με διαφορετικό tempo. Καθώς οι δύο ακολουθίες δεν έχουν το ίδιο μήκος, προκειμένου να αντιστοιχηθούν μεταξύ τους είναι πιθανό είτε κάποια στοιχεία τους να παραλειφθούν είτε να χρησιμοποιηθούν παραπάνω από μία φορά (**Εικόνα 2-7**).



**Εικόνα 2-7**  
Χρονική αντιστοίχιση δύο σύντομων χρονικών ακολουθιών (chromagrams)

Οι δείκτες των δύο ακολουθιών αντιστοιχούν σε διαδοχικά χρονικά σημεία που απέχουν ίσα διαστήματα μεταξύ τους (σε περίπτωση που αυτό δεν συμβαίνει μπορεί να αντιμετωπιστεί με resampling). Αν οι δύο ακολουθίες παίρνουν τιμές από ένα πεδίο τιμών  $\phi$ , τότε προκειμένου να συγκριθούν οι δύο ακολουθίες  $X, Y \in \phi$  χρειαζόμαστε ένα μέτρο τοπικού κόστους (συνά αναφέρεται και ως μέτρο τοπικής απόστασης) το οποίο ορίζεται ως μια συνάρτηση:

$$c : \phi \times \phi \in \mathbb{R} \quad (2.20)$$

Η συνάρτηση  $c$  θα έχει μικρές τιμές (χαμηλό κόστος) όταν οι δύο ακολουθίες είναι παρόμοιες και μεγάλες τιμές (υψηλό κόστος) όταν είναι αρκετά διαφορετικές. Ο στόχος επομένως της βέλτιστης αντιστοίχισης του DTW έγκειται στην τακτοποίηση όλων των σημείων της ακολουθίας με ταυτόχρονη ελαχιστοποίηση αυτής της συνάρτησης κόστους (Senin 2008). Υπολογίζοντας το μέτρο του τοπικού κόστους για κάθε ζεύγος στοιχείων των ακολουθιών  $X$  και  $Y$  προκύπτει ο πίνακας κόστους (cost matrix)  $C \in \mathbb{R}^{N \times M}$  ο οποίος ορίζεται ως:

$$C(n, m) = c(x_n, y_m) \quad (2.20)$$

για  $n \in [1 : N]$  και  $m \in [1 : M]$ . Το ζεύγος  $(n, m)$  αντιπροσωπεύει ένα κελί του πίνακα.

Υπάρχουν διάφοροι τρόποι για να υπολογιστεί η απόσταση μεταξύ δύο στοιχείων  $x, y \in \phi$ . Από τους πιο συχνά χρησιμοποιούμενους είναι η Ευκλείδεια απόσταση  $\|x - y\|$  και η απόσταση Manhattan. Ένας άλλος τρόπος είναι η απόσταση συνημίτονου, η οποία και είναι:

$$c(x, y) = 1 - \frac{\langle x | y \rangle}{\|x\| \cdot \|y\|} \quad (2.21)$$

επειδή ισχύει ότι:

$$\cos(\varphi) = \frac{\langle x | y \rangle}{\|x\| \cdot \|y\|} \quad (2.22)$$

όπου  $\varphi$  η γωνία μεταξύ των διανυσμάτων  $x$  και  $y$ , θα ισχύει ότι  $c(x, y) \in [0, 1]$ ,  $c(x, y) = 0$  όταν τα  $x$  και  $y$  έχουν την ίδια κατεύθυνση και  $c(x, y) = 1$  όταν τα  $x$  και  $y$  είναι κάθετα μεταξύ τους. Σε αντίθεση με την Ευκλείδεια απόσταση, η απόσταση συνημίτονου δεν εξαρτάται από το πραγματικό μήκος των διανυσμάτων. Έτσι, όταν συγκρίνουμε χρωματογραφικά διανύσματα, η μέτρηση λαμβάνει υπόψιν μόνο την κατανομή της ενέργειας στις δώδεκα χρωματικές ζώνες και αγνοεί την πραγματική τοπική ενέργεια. Σε μία εφαρμογή μουσικού συγχρονισμού η παράμετρος αυτή είναι ιδιαίτερη χρήσιμη όταν συγκρίνουμε δύο εκτελέσεις με σημαντικές διαφορές στις δυναμικές τους (π.χ. η εκτέλεση ενός κομματιού από ορχήστρα και από πιάνο) (Müller 2015).

### 2.3.2 Warping Path

Όπως αναφέρθηκε παραπάνω, για δύο ακολουθίες  $X = (x_1, x_2, \dots, x_N)$  και  $Y = (y_1, y_2, \dots, y_M)$  η αντιστοιχία μεταξύ δύο στοιχείων  $x_n$  και  $y_m$  μπορεί να μοντελοποιηθεί με το ζεύγος δεικτών ή κελί  $(n, m)$ . Για να μοντελοποιηθεί μία γενική αντιστοίχιση μεταξύ των ακολουθιών  $X$  και  $Y$ , η βασική ιδέα είναι να θεωρήσουμε μία ακολουθία από ζεύγη δεικτών που ικανοποιεί συγκεκριμένους περιορισμούς. Έτσι ένα warping path (μονοπάτι «παραμόρφωσης») είναι μία ακολουθία μήκους  $L \in \mathbb{N}$ :

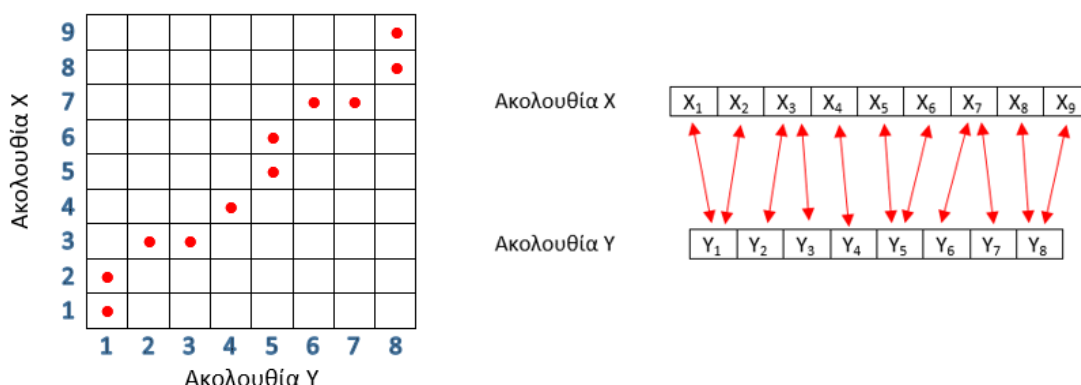
$$P = (p_1, p_2, \dots, p_L) \quad (2.23)$$

με το  $p_l = (n_l, m_l) \in [1 : N] \times [1 : M]$  για κάθε  $l \in [1 : L]$  να ικανοποιεί τις παρακάτω τρεις συνθήκες:

- Οριακή Συνθήκη:  $p_1 = [1, 1]$  και  $p_L = [N, M]$ . Το πρώτο και το τελευταίο σημείο του warping path πρέπει να αντιστοιχεί στο πρώτο και στο τελευταίο σημείο των δύο ακολουθιών.
- Συνθήκη Μονοτονίας:  $n_1 \leq n_2 \leq \dots \leq n_L$  και  $m_1 \leq m_2 \leq \dots \leq m_L$ . Η συνθήκη αυτή διατηρεί τη χρονική τάξη των σημείων.
- Συνθήκη μεγέθους βήματος:  $p_{l+1} - p_l \in \{(1, 1), (1, 0), (0, 1)\}$  για κάθε  $l \in [1 : L - 1]$ . Η συνθήκη αυτή περιορίζει το warping path από το να πραγματοποιεί μεγάλα χρονικά άλματα και εξασφαλίζει ότι κάθε στοιχείο της ακολουθίας  $X$  θα αντιστοιχισθεί με ένα τουλάχιστον στοιχείο της ακολουθίας  $Y$  και αντίστροφα. (Senin 2008)

Το warping path αντιστοιχίζει το στοιχείο  $x_{n_l}$  της ακολουθίας  $X$  στο στοιχείο  $y_{m_l}$  της ακολουθίας  $Y$ . Στην **Εικόνα 2-8** φαίνεται το παράδειγμα ενός warping path που πληροί τις

παραπάνω συνθήκες και η ερμηνεία του στην αντιστοίχιση δύο ακολουθιών X με μήκος N=9 και Y με μήκος M=8.



Εικόνα 2-8  
Παράδειγμα ενός warping path για δύο ακολουθίες X και Y

### 2.3.3 Βέλτιστο Warping Path και απόσταση DTW

Στην προηγούμενη παράγραφο περιγράφηκε η γενική μορφή ενός warping path χωρίς να λαμβάνεται καθόλου υπόψιν η ποιότητά του και η σχέση του με τον πίνακα κόστους (cost matrix). Το συνολικό κόστος  $C_p(X, Y)$  ενός warping path P μεταξύ δύο ακολουθιών X και Y σε σχέση με το τοπικό κόστος c ορίζεται ως:

$$C_p(X, Y) = \sum_{l=1}^L c(x_{n_l}, y_{m_l}) = \sum_{l=1}^L C(n_l, m_l) \quad (2.24)$$

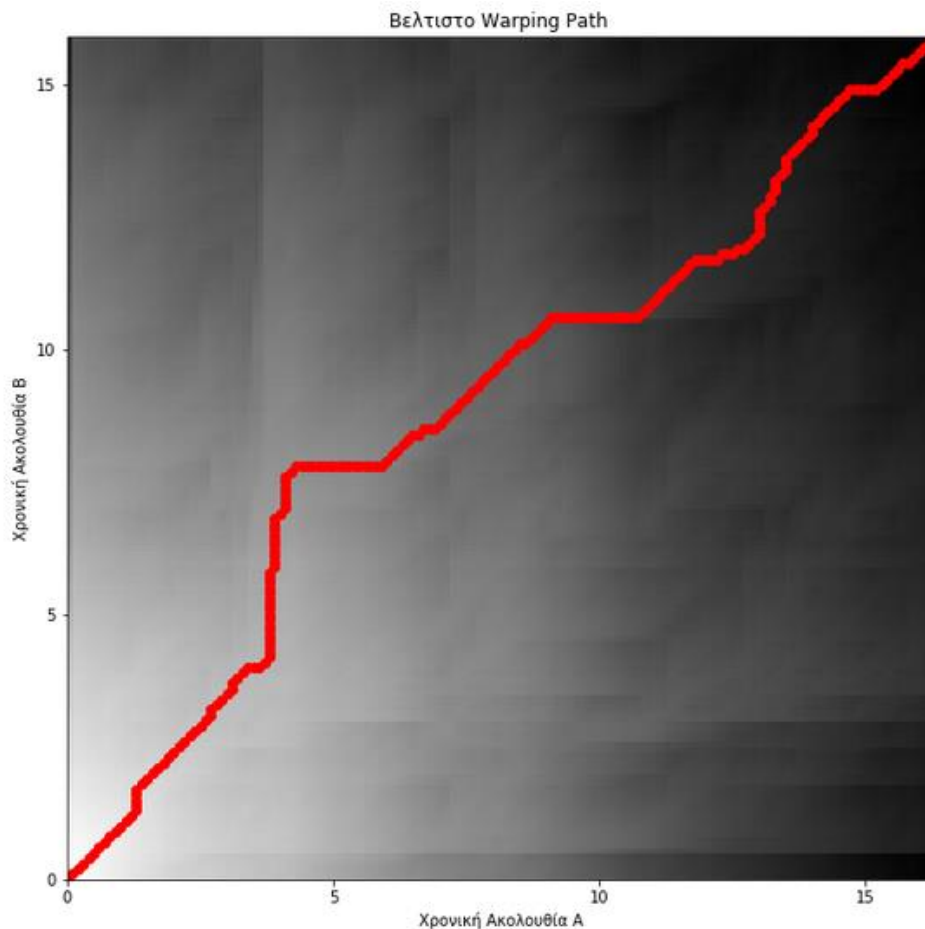
Από την παραπάνω σχέση φαίνεται ότι ένα warping path συσσωρεύει το κόστος όλων των κελιών από τα οποία περνάει. Ένα warping path είναι καλό όταν το συνολικό κόστος είναι χαμηλό και κακό όταν το συνολικό κόστος είναι μεγάλο. Στόχος είναι να βρεθεί το βέλτιστο warping path μεταξύ των ακολουθιών X και Y, το οποίο θα είναι το warping path  $P^*$  το οποίο θα έχει το μικρότερο συνολικό κόστος από όλα τα πιθανά warping paths.

Από τον παραπάνω ορισμό του βέλτιστου warping path φαίνεται επίσης ότι προκειμένου αυτό να εντοπιστεί, πρέπει να υπολογιστούν όλα τα πιθανά warping paths που μπορεί να υπάρξουν ανάμεσα στις ακολουθίες X και Y. Αυτό μπορεί να γίνει υπολογιστικά επίπονο, καθώς ο αριθμός των πιθανών warping paths αυξάνεται εκθετικά όσο το μέγεθος των X και Y αυξάνεται γραμμικά. Για να αντιμετωπιστεί το παραπάνω πρόβλημα έχει αναπτυχθεί ο αλγόριθμος DTW δυναμικού προγραμματισμού με τον οποίο θα ασχοληθούμε στην επόμενη παράγραφο. Πρώτα όμως θα προχωρήσουμε στον ορισμό της απόστασης DTW (DTW distance) ως το συνολικό κόστος του βέλτιστου warping path:

$$DTW(X, Y) = c_{p^*}(X, Y) = \min \{ c_p(X, Y), p \in P^{N \times M} \} \quad (2.25)$$

όπου  $P^{N \times M}$  είναι το σύνολο όλων των πιθανών warping paths (Senin 2008). Είναι πιθανό να υπάρχουν περισσότερα από ένα βέλτιστα warping path. Αν για παράδειγμα ο πίνακας

κόστους  $C$  είναι μηδενικός, τότε κάθε πιθανό μονοπάτι είναι βέλτιστο warping path. Παρόλα αυτά η απόσταση DTW είναι ξεκάθαρα ορισμένη καθώς όλα τα βέλτιστα warping paths θα έχουν το ίδιο συνολικό κόστος. Στην **Εικόνα 2-9** φαίνεται ένα παράδειγμα βέλτιστου warping Path όπως προκύπτει για δύο χρονικές ακολουθίες – ηχητικά σήματα.



**Εικόνα 2-9**  
Βέλτιστο Warping Path ανάμεσα σε δύο ηχητικά σήματα

### 2.3.4 Αλγόριθμος Δυναμικού Προγραμματισμού

Όπως αναφέρθηκε η εύρεση του βέλτιστου warping path προϋποθέτει τον υπολογισμό του κόστους για όλα τα πιθανά warping paths, γεγονός που έχει μεγάλο υπολογιστικό κόστος. Το πρόβλημα αυτό μπορεί να αντιμετωπιστεί με τη χρήση δυναμικού προγραμματισμού. Η βασική ιδέα του δυναμικού προγραμματισμού είναι η διαίρεση ενός προβλήματος σε μικρότερα υπό-προβλήματα και στη συνέχεια ο αναδρομικός συνδυασμός των λύσεων των υπό-προβλημάτων για την εύρεση της συνολικής λύσης. Στην περίπτωση του DTW η ιδέα είναι να υπολογιστεί ένα βέλτιστο warping path για τις αρχικές ακολουθίες από τα βέλτιστα warping paths περικομμένων τμημάτων των ακολουθιών. Ορίζουμε επομένως τα τμήματα των ακολουθιών  $X(1:n) = (x_1, \dots, x_n)$  για  $n \in [1:N]$  και  $Y(1:m) = (y_1, \dots, y_m)$  για  $m \in [1:M]$  και θέτουμε:

$$D(n, m) = DTW(X(1:n), Y(1:m)) \quad (2.26)$$

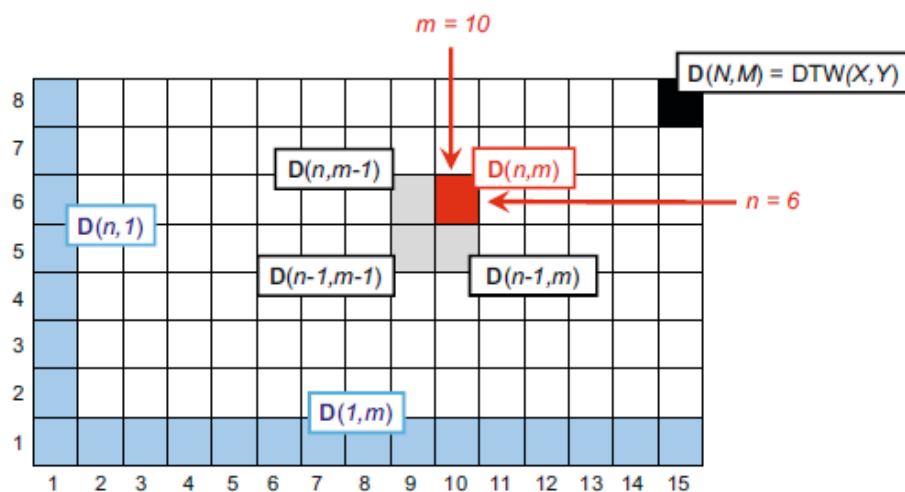
Οι τιμές  $D(n,m)$  ορίζουν ένα  $(N \times M)$  πίνακα  $D$  που αναφέρεται ως πίνακας συγκεντρωτικού κόστους (accumulated cost matrix). Κάθε τιμή  $D(n,m)$  ορίζει το συνολικό (συγκεντρωτικό) κόστος ενός βέλτιστου warping path που ξεκινάει από το κελί  $(1,1)$  και τελειώνει στο κελί  $(n,m)$ . Οι παρακάτω εξισώσεις δείχνουν πώς ο συγκεντρωτικός πίνακας κόστους  $D$  μπορεί να υπολογιστεί αναδρομικά και ο Meinard Müller απεικονίζει τα βήματα της αναδρομικής διαδικασίας στην **Εικόνα 2-10** (2015):

$$D(n,1) = \sum_{k=1}^n C(k,1) \quad \text{για } n \in [1:N] \quad (2.27)$$

$$D(1,m) = \sum_{k=1}^m C(1,k) \quad \text{για } m \in [1:M] \quad (2.28)$$

$$D(n,m) = C(n,m) + \min \begin{cases} D(n-1,m-1) \\ D(n-1,m) \\ D(n,m-1) \end{cases} \quad (2.29)$$

για  $n \in [2:N]$  και  $m \in [2:M]$ . Οι δύο πρώτες σχέσεις αρχικοποιούν την αναδρομική διαδικασία υπολογίζοντας τις τιμές για την πρώτη στήλη και την πρώτη γραμμή αντίστοιχα. Οι υπόλοιπες τιμές  $D(n,m)$  για  $n > 1$  και  $m > 1$  υπολογίζονται από την σχέση (2.29). Ο υπολογισμός των τιμών του  $D$  πρέπει να γίνει αυξάνοντας διαδοχικά τις τιμές των  $n$  και  $m$  ξεκινώντας από την κάτω αριστερή γωνία  $(1,1)$  και τελειώνοντας στην πάνω δεξιά γωνία  $(N,M)$ . Το τελικό κελί  $(N,M)$  μας δίνει και την απόσταση του DTW. Οι εγγραφές  $D(n,m)$  μπορούν να υπολογιστούν με διαφορετική σειρά (π.χ. ανά γραμμή ή ανά στήλη) αρκεί να τηρείται η μονοτονία στην αύξηση των τιμών των  $m$  και  $n$ . Σε κάθε περίπτωση το αναδρομικό βήμα του αλγορίθμου (2.29) θα κληθεί  $(n-1) \cdot (m-1)$  φορές. Η πολυπλοκότητα του αλγορίθμου για τον υπολογισμό του  $D$  από ένα γνωστό πίνακα κόστους  $C$  είναι της τάξης  $O(NM)$ .



**Εικόνα 2-10**

Αναδρομικός υπολογισμός της απόστασης DTW. Τα γαλάζια κελία υποδηλώνουν τις εγγραφές  $D(n,1)$  και  $D(1,m)$ , το μαύρο κελί την τελική εγγραφή  $D(N,M)$  και το κόκκινο κελί την τρέχουσα εγγραφή  $D(n,m)$

Μέχρι στιγμής έχουμε υπολογίσει την απόσταση DTW αλλά το βέλτιστο warping path δεν έχει ακόμη οπτικοποιηθεί. Εφαρμόζοντας μία ανάστροφη (backtracking) διαδικασία το βέλτιστο μονοπάτι μπορεί να κατασκευαστεί ακολουθώντας αντίθετη φορά, ξεκινώντας δηλαδή από το κελί  $q_1 = (N, M)$  και υπολογίζοντας κάθε φορά το  $q_l = (n, m)$ . Στην περίπτωση που  $(n, m) = (1, 1)$  έχουμε τελειώσει, ορίζουμε  $L = l$  και το μονοπάτι  $P^* = (q_L, q_{L-1}, \dots, q_1)$  ορίζει ένα βέλτιστο warping path:

$$q_{l+1} = (1, m-1) \quad \text{αν } n = 1 \quad (2.30)$$

$$q_{l+1} = (n-1, m) \quad \text{αν } m = 1 \quad (2.31)$$

$$q_{l+1} = \operatorname{argmin} \begin{cases} D(n-1, m-1), \\ D(n-1, m), \\ D(n, m-1) \end{cases} \quad \text{αν } n \in [2: N] \text{ και } m \in [2: M] \quad (2.32)$$

όπου ο όρος  $\operatorname{argmin}$  αποδίδει το κελί με τη μικρότερη από τις τρεις τιμές. Το  $\operatorname{argmin}$  δεν είναι απαραίτητα μοναδικό, δίνοντας έτσι την πιθανότητα να υπάρχουν περισσότερα από ένα βέλτιστα warping paths. Στην περίπτωση αυτή μπορεί να επιλεγεί το μικρότερο αλφαβητικά κελί, εφόσον επιθυμούμε τον ορισμό ενός μοναδικού warping path.

Παρακάτω παρουσιάζεται ο αλγόριθμος υπολογισμού της απόστασης DTW και του βέλτιστου warping path (σε ψευδοκώδικα), ενώ στο **Παράρτημα 1 – Υλοποίηση αλγορίθμου DTW σε Python** παρουσιάζεται κώδικας υλοποίησης του αλγορίθμου DTW σε Python.

#### Αλγόριθμος: DTW

**Είσοδος:** Πίνακας κόστους  $C$  μεγέθους  $N \times M$

**Έξοδος :** Πίνακας Συγκεντρωτικού Κόστους  $D$   
Βέλτιστο Warping Path  $P^*$

**Διαδικασία:** Αρχικοποίησε τον  $(N \times M)$  πίνακα  $D$  με  $D(n, 1) = \sum_{k=1}^n C(k, 1)$  για  $n \in [1: N]$  και

$D(1, m) = \sum_{k=1}^m C(1, k)$  για  $m \in [1: M]$ . Μετά υπολόγισε σε μία εμφωλευμένη επανάληψη για  $n=2, \dots, N$  και  $m=2, \dots, M$ :

$$D(n, m) = C(n, m) + \min \{ D(n-1, m-1), D(n-1, m), D(n, m-1) \}$$

Όρισε  $l=1$  και  $q_l=(N, M)$ . Επανάλαβε μέχρις ότου  $q_l=(1, 1)$ :

Αύξησε το  $l$  κατά 1 και όρισε  $(n, m)=q_{l-1}$

Αν  $n=1$  τότε  $q_l=(1, m-1)$

αλλιώς αν  $m=1$  τότε  $q_l=(n-1, m)$

αλλιώς  $q_l=\operatorname{argmin}\{D(n-1, m-1), D(n-1, m), D(n, m-1)\}$

(Αν το  $\operatorname{argmin}$  δεν είναι μοναδικό επέλεξε το αλφαβητικά μικρότερο κελί)

Όρισε  $L=l$ . Επέστρεψε  $P^*=(q_L, q_{L-1}, \dots, q_1)$ . Επέστρεψε  $D$



### 2.3.5 Παράδειγμα Εφαρμογής του Αλγορίθμου DTW

Έστω δύο ακολουθίες  $X = [0, 4, 4, 0, -4, -4, 0]$  και  $Y = [1, 3, 4, 3, 1, -1, -2, -1, 0]$ . Υπολογίζοντας το τοπικό κόστος ως  $c(x, y) = |x - y|$ ,  $x, y \in \mathbb{R}$  προκύπτει ο πίνακας κόστους (cost matrix)  $C$  όπως φαίνεται στον **Πίνακας 2**:

0	1	3	4	3	1	1	2	1	0
-4	5	7	8	7	5	3	2	3	4
-4	5	7	8	7	5	3	2	3	4
0	1	3	4	3	1	1	2	1	0
4	3	1	0	1	3	5	6	5	4
4	3	1	0	1	3	5	6	5	4
0	1	3	4	3	1	1	2	1	0

1	3	4	3	1	-1	-2	-1	0
---	---	---	---	---	----	----	----	---

**Πίνακας 2**

Πίνακας κόστους (cost matrix) για τις ακολουθίες  $X$  και  $Y$

Στον πίνακα κόστους  $C$  θεωρούμε τους δείκτες των γραμμών από τη βάση προς την κορυφή και των στηλών από αριστερά προς τα δεξιά (σε αντίθεση με τη συνηθισμένη σύμβαση για τους δείκτες των πινάκων, καθώς οι δύο διαστάσεις αποτελούν δύο χρονικούς άξονες). Από τον παραπάνω πίνακα προκύπτει ο πίνακας συγκεντρωτικού κόστους (accumulated cost matrix)  $D$  όπως φαίνεται στον **Πίνακας 3**:

0	19	21	24	22	15	11	11	10	10
-4	18	20	21	19	14	10	9	10	14
-4	13	13	14	12	9	7	7	10	12
0	8	6	6	5	4	5	7	8	8
4	7	3	2	3	6	11	17	22	24
4	4	2	2	3	6	11	17	20	20
0	1	4	8	11	12	13	15	16	16

1	3	4	3	1	-1	-2	-1	0
---	---	---	---	---	----	----	----	---

**Πίνακας 3**

Πίνακας συγκεντρωτικού κόστους (accumulated cost matrix) για τις ακολουθίες  $X$  και  $Y$

Στη συνέχεια, υπολογίζεται το βέλτιστο warping path το οποίο, όπως φαίνεται στον **Πίνακας 4**, είναι το  $P^* = ((1,1), (2,2), (2,3), (3,4), (4,5), (4,6), (5,7), (6,8), (7,9))$

0	19	21	24	22	15	11	11	10	10
-4	18	20	21	19	14	10	9	10	14
-4	13	13	14	12	9	7	7	10	12
0	8	6	6	5	4	5	7	8	8
4	7	3	2	3	6	11	17	22	24
4	4	2	2	3	6	11	17	20	20
0	1	4	8	11	12	13	15	16	16

1	3	4	3	1	-1	-2	-1	0
---	---	---	---	---	----	----	----	---

**Πίνακας 4**

Βέλτιστο warping path

Το βέλτιστο warping path υπολογίζεται ξεκινώντας από το κελί  $q_1 = (N, M) = (7, 9)$ . Εφαρμόζοντας την ανάστροφη (backtracking) μέθοδο, ελέγχουμε τα κελιά από τα οποία θα μπορούσαμε να φτάσουμε στο  $q_1$  μέσω ενός έγκυρου βήματος. Αυτά είναι τα κελιά  $(N-1, M)$ ,  $(N, M-1)$  και  $(N-1, M-1)$ . Από αυτά επιλέγεται εκείνο που έχει το μικρότερο συγκεντρωτικό κόστος. Όπως φαίνεται στον **Πίνακας 5**, υπάρχουν δύο κελιά που ικανοποιούν αυτές τις συνθήκες, τα κελιά D(6,8) και D(7,8) με κόστος 10, γεγονός που σημαίνει ότι υπάρχουν περισσότερα από ένα βέλτιστα warping path. Στην περίπτωση αυτή επιλέγεται το κελί με την μικρότερη αλφαριθμητική – αλφαριθμητική αξία, δηλαδή το D(6,8). Η διαδικασία επαναλαμβάνεται μέχρι να φτάσουμε το κελί  $q_L = (1,1)$ , όπου και ολοκληρώνεται επιστρέφοντας ένα βέλτιστο Warping Path 9 βημάτων.

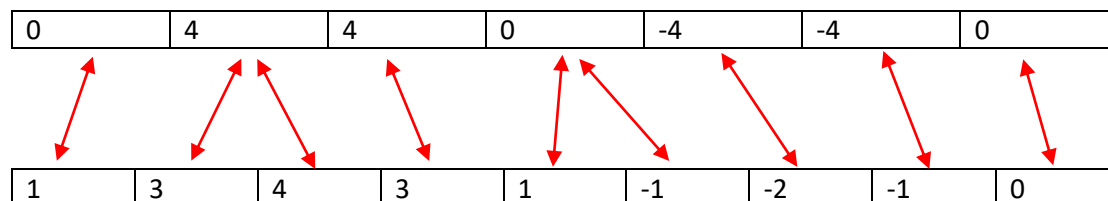
0	19	21	24	22	15	11	11	10	10
-4	18	20	21	19	14	10	9	10	14
-4	13	13	14	12	9	7	7	10	12
0	8	6	6	5	4	5	7	8	8
4	7	3	2	3	6	11	17	22	24
4	4	2	2	3	6	11	17	20	20
0	1	4	8	11	12	13	15	16	16

1	3	4	3	1	-1	-2	-1	0
---	---	---	---	---	----	----	----	---

**Πίνακας 5**  
Warping path χωρίς τη χρήση backtracking

Στο παράδειγμα φαίνεται επίσης ότι η χρήση της backtracking τεχνικής είναι απαραίτητη προκειμένου να υπολογίσουμε το βέλτιστο warping path. Όπως προκύπτει από τον **Πίνακας 5**, αν ξεκινούσαμε από το κελί (1,1) και χρησιμοποιούσαμε μία προς τα εμπρός μέθοδο (greedy method) θα προέκυπτε ένα Warping Path 10 βημάτων και συγκεκριμένα το μονοπάτι  $P = ((1,1), (2,2), (3,3), (3,4), (4,5), (4,6), (5,7), (6,7), (7,8), (7,9))$ . Το Warping Path αυτό δεν είναι βέλτιστο καθώς περιλαμβάνει περισσότερα βήματα από το αντίστοιχο του **Πίνακας 4**. Η αντιστοίχιση που προκύπτει τελικά μεταξύ των δύο ακολουθιών, με χρήση του βέλτιστου Warping Path, φαίνεται στην **Εικόνα 2-11**:



**Εικόνα 2-11**  
Αντιστοίχιση μεταξύ των δύο ακολουθιών σύμφωνα με το βέλτιστο warping path

### 2.3.6 Παραλλαγές του DTW

Διάφορες παραλλαγές του αλγορίθμου DTW έχουν προταθεί προκειμένου να αυξηθεί η ταχύτητα των υπολογισμών καθώς και για τον καλύτερο έλεγχο της διαδικασίας του warping path. Παρακάτω παρουσιάζονται μερικές από αυτές τις παραλλαγές

#### 2.3.6.1 Συνθήκη Μεγέθους βήματος

Όπως αναφέρθηκε στην ενότητα **2.3.2 Warping Path** η συνθήκη  $p_{l+1} - p_l \in \{(1,1), (1,0), (0,1)\}$  για κάθε  $l \in [1: L-1]$  εξασφαλίζει ότι το warping path θα αντιστοιχίσει κάθε σημείο της ακολουθίας  $X = (x_1, x_2, \dots, x_N)$  σε ένα τουλάχιστον σημείο της ακολουθίας  $Y = (y_1, y_2, \dots, y_M)$  και αντίστροφα. Το μειονέκτημα αυτής της συνθήκης είναι ότι ένα σημείο της μιας ακολουθίας μπορεί να αντιστοιχηθεί σε περισσότερα από ένα σημεία της άλλης ακολουθίας. Στην περίπτωση αυτή δημιουργούνται οριζόντια ή κάθετα ευθύγραμμα τμήματα στο warping path που σημαίνουν ότι αυτό προχωράει χρονικά στη μία ακολουθία ενώ παραμένει χρονικά σταθερό στην άλλη. Από άποψη φυσικού χρόνου, μια τέτοια κατάσταση δημιουργεί μία σημαντική χρονική παραμόρφωση κατά την αντιστοίχιση των δύο χρονικών ακολουθιών. Για να αποφευχθεί αυτό, μπορούμε να τροποποιήσουμε τη συνθήκη μεγέθους βήματος περιορίζοντας την κλίση των αποδεκτών warping path. Για παράδειγμα, μπορούμε να ορίσουμε:

$$p_{l+1} - p_l \in \{(2,1), (1,2), (1,1)\} \quad (2.33)$$

Η σχέση αυτή οδηγεί σε warping paths που έχουν τοπική κλίση ανάμεσα στα όρια  $\frac{1}{2}$  και 2. Ο πίνακας συγκεντρωτικού κόστους μπορεί τότε να υπολογιστεί από τη σχέση:

$$D(n, m) = C(n, m) + \min \begin{cases} D(n-1, m-1) \\ D(n-2, m-1) \\ D(n-1, m-2) \end{cases} \quad (2.34)$$

για  $n \in [1: N]$  και  $m \in [1: M]$  με  $(n, m) \neq (1, 1)$ . Για την αρχικοποίηση χρησιμοποιούμε την τεχνική της επέκτασης του πίνακα D με 2 επιπλέον γραμμές και στήλες (με δείκτες -1 και 0) και ορίζουμε  $D(1, 1) = C(1, 1)$ ,  $D(n, -1) = D(n, 0) = \infty$  για  $n \in [-1: N]$  και  $D(-1, m) = D(0, m) = \infty$  για  $m \in [-1: M]$ . Η παραπάνω τροποποίηση δεν εξασφαλίζει ότι όλα τα στοιχεία της ακολουθίας X θα αντιστοιχηθούν με κάποιο στοιχείο της ακολουθίας Y και αντίστροφα. Κάποια στοιχεία μπορεί να παρακαμφθούν και επομένως να μη προκαλέσουν κανένα κόστος στον αλγόριθμο.

#### 2.3.6.2 Τοπικά βάρη

Προκειμένου να ωθήσουμε προς την οριζόντια, την κάθετη ή τη διαγώνια κατεύθυνση του warping path μπορούμε να εισάγουμε τοπικά βάρη  $w_h, w_v, w_d \in \mathbb{R}$  αντίστοιχα. Η αρχικοποίηση και η αναδρομική διαδικασία για τον πίνακα συγκεντρωτικού κόστους θα προκύπτει ως εξής:

$$D(1,1) = C(1,1) \quad (2.35)$$

$$D(n,1) = \sum_{k=1}^n w_h \cdot C(k,1) \quad \text{για } n \in [2:N] \quad (2.36)$$

$$D(1,m) = \sum_{k=1}^m w_v \cdot C(1,k) \quad \text{για } m \in [2:M] \quad (2.37)$$

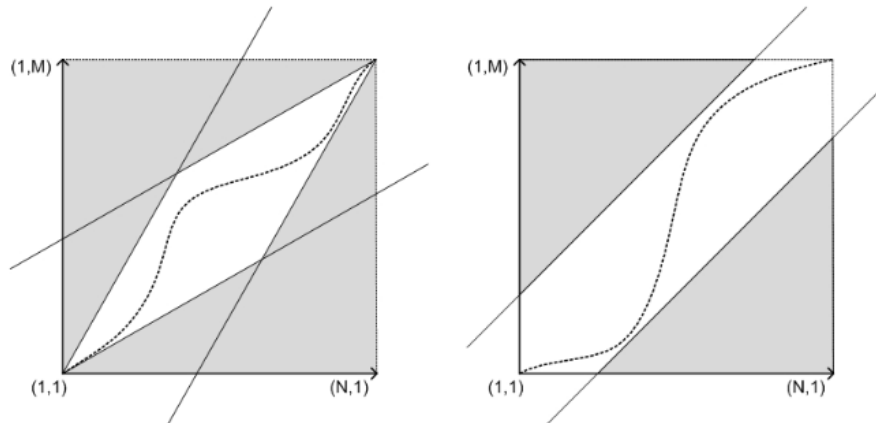
$$D(n,m) = \min \begin{cases} D(n-1,m-1) + w_d \cdot C(n,m) \\ D(n-1,m) + w_v \cdot C(n,m) \\ D(n,m-1) + w_h \cdot C(n,m) \end{cases} \quad \text{για } n \in [2:N] \text{ και } m \in [2:M] \quad (2.38)$$

Για την περίπτωση που  $w_d = w_h = w_v = 1$  τότε έχουμε τον κλασικό DTW όπου υπάρχει η προτίμηση των διαγώνιων βημάτων, καθώς ένα διαγώνιο βήμα (κόστος ενός κελιού) αντιστοιχεί στο συνδυασμό ενός οριζοντίου και ενός κάθετου (κόστος 2 κελιών).

### 2.3.6.3 Καθολικοί Περιορισμοί

Μια συνηθισμένη τεχνική είναι ο καθορισμός καθολικών περιορισμών προκειμένου να αυξηθεί η ταχύτητα του DTW και να αντιμετωπιστούν παθογένειες στην αντιστοίχιση περιορίζοντας τη διαδρομή των αποδεκτών warping path. Πιο συγκεκριμένα αν  $R \subseteq [1:N] \times [1:M]$  είναι ένα υποσύνολο που ορίζεται ως η περιοχή καθολικού περιορισμού, τότε ένα warping path σχετικό με το R είναι ένα warping path που το σύνολό του βρίσκεται εντός του R. Το βέλτιστο warping path σχετικό με το R, που ορίζεται ως  $P_R^*$  είναι το warping path με το μικρότερο κόστος απ' όλα τα warping paths σχετικά με το R.

Δύο πολύ γνωστές μέθοδοι καθολικού περιορισμού είναι η ζώνη Sakoe-Chiba (1978) και το παραλληλόγραμμο Itakura (1975), οι οποίες πρωτοπαρουσιάστηκαν στο πεδίο της αναγνώρισης ομιλίας. Όπως φαίνεται στην **Εικόνα 2-12** (Senin 2008), η ζώνη Sakoe-Chiba βρίσκεται κατά μήκος της κύριας διαγώνιου και έχει σταθερό πλάτος, ενώ το παραλληλόγραμμο Itakura ορίζει μία περιοχή που περιορίζει την κλίση του warping path.



**Εικόνα 2-12**

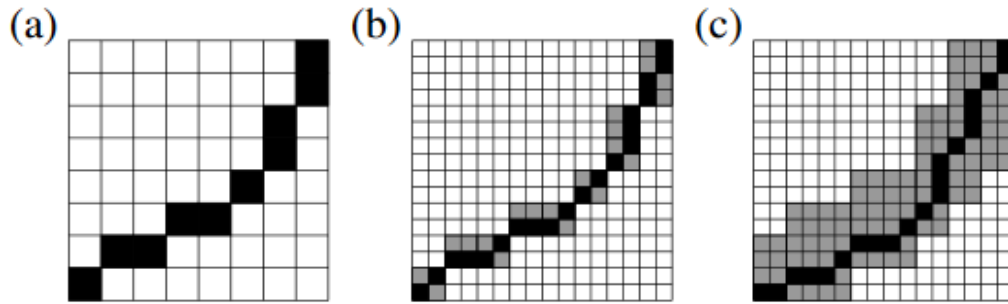
Παραλληλόγραμμο Itakura και ζώνη Sakoe-Chiba – Pavel Senin, 2008

Για μία περιορισμένη περιοχή, το βέλτιστο warping path  $P_R^*$  μπορεί να υπολογιστεί με παρόμοιο τρόπο με τον γενικό υπολογισμό του warping path, ορίζοντας  $C(n, m) = \infty$  για κάθε  $(n, m) \in [1: N] \times [1: M] / R$ . Έτσι για τον υπολογισμό του  $P_R^*$  μόνο τα κελιά που βρίσκονται εντός της περιοχής  $R$  χρειάζεται να εκτιμηθούν, γεγονός που επιταχύνει σημαντικά τον αλγόριθμο DTW. Για παράδειγμα, στην περίπτωση μια ζώνης Sakoe-Chiba με πλάτος  $\Delta$ , θα πρέπει να εκτελεστούν  $O(\Delta \cdot \max(N, M))$  υπολογισμοί σε αντίθεση με τους  $O(NM)$  του κλασικού DTW. Το μειονέκτημα της χρήσης καθολικών περιορισμών είναι ότι το βέλτιστο warping path  $P^*$  πιθανόν να περιλαμβάνει κελιά έξω από την περιορισμένη περιοχή. Στην περίπτωση αυτή τα  $P^*$  και  $P_R^*$  δεν ταυτίζονται. Έτσι, η χρήση μίας πολύ «αυστηρής» περιορισμένης περιοχής μπορεί να οδηγήσει σε μη επιθυμητά ή και τελείως άχρηστα αποτελέσματα αντιστοίχισης.

#### 2.3.6.4 Πολύ-κλιμακούμενος DTW

Όπως είδαμε στην προηγούμενη παράγραφο, στη μέθοδο των καθολικών περιορισμών είναι πιθανό τα warping path  $P^*$  και  $P_R^*$  να μην ταυτίζονται. Καθώς το βέλτιστο warping path  $P^*$  δεν είναι γνωστό εκ των προτέρων, είναι συχνά πολύ δύσκολο να βρεθεί ένα ικανοποιητικό μέγεθος της περιοχής περιορισμού, ώστε αυτή να είναι όσο το δυνατό μικρότερη ώστε να επιταχυνθεί ο αλγόριθμος αλλά ταυτόχρονα και αρκετά μεγάλη ώστε να περιλαμβάνει το επιθυμητό warping path. Μια άλλη στρατηγική είναι να μειωθεί ο ρυθμός δειγματοληψίας (συχνά αναφέρεται και ως μείωση διαστάσεων ή αφαίρεση δεδομένων) που οδηγεί σε μείωση του μήκους  $N$  και  $M$  των ακολουθιών που αντιστοιχίζονται. Όμως, το βέλτιστο warping path που προκύπτει, γίνεται αρκετά ανακριβές ή και τελείως άχρηστο όσο η ανάλυση μειώνεται. Ο πολύ-κλιμακούμενος (multiscale) DTW (MsDTW) συνδυάζει τις δύο παραπάνω στρατηγικές σε μία επαναληπτική μορφή για να δημιουργήσει περιοχές περιορισμού εξαρτώμενες από τα δεδομένα (Müller, Mattes and Kurth 2006).

Έστω  $X_1 = X$  και  $Y_1 = Y$  οι δύο ακολουθίες που πρέπει να συγχρονιστούν με μήκος  $N_1 = N$  και  $M_1 = M$  αντίστοιχα. Οι δύο αυτές ακολουθίες αναπαριστούν τα δεδομένα στο υψηλότερο επίπεδο ανάλυσης, που θα αναφέρεται ως Επίπεδο 1. Μειώνοντας τον ρυθμό δειγματοληψίας κατά ένα παράγοντα  $f_2 \in \mathbb{N}$  προκύπτουν οι ακολουθίες  $X_2$  με μήκος  $N_2 = N_1 / f_2$  και  $Y_2$  με μήκος  $M_2 = M_1 / f_2$  (υποθέτοντας ότι το  $f_2$  διαιρεί ακριβώς τα  $N_1$  και  $M_1$ ). Στη συνέχεια υπολογίζουμε το βέλτιστο warping path  $P_2^*$  μεταξύ των  $X_2$  και  $Y_2$  στο επίπεδο ανάλυσης που έχει προκύψει (Επίπεδο 2) όπως φαίνεται στην **Εικόνα 2-13** (a) (Müller, Mattes and Kurth 2006). Αυτό το warping path προβάλλεται πάνω στο Επίπεδο 1 και εκεί ορίζει μία περιορισμένη περιοχή  $R$  (**Εικόνα 2-13** (b)). Το  $R$  αποτελείται από  $L_2 \times f_2^2$  κελιά, όπου  $L_2$  το μήκος του  $P_2^*$ . Τέλος υπολογίζεται ένα βέλτιστο warping path  $P_R^*$  σχετικό με το  $R$ . Ο συνολικός αριθμός κελιών που υπολογίζεται στη διαδικασία είναι  $N_2 M_2 + L_2 \cdot f_2^2$ , αρκετά μικρότερος από τον συνολικό αριθμό  $N_1 M_1$  κελιών του Επιπέδου 1. Η διαδικασία μπορεί να επαναλαμβάνεται αναδρομικά εισάγοντας περισσότερα επίπεδα όλο και μικρότερης ανάλυσης.



**Εικόνα 2-13**

(a) Βέλτιστο *Warping Path* στο επίπεδο 2 (b) Προβολή στο Επίπεδο 1 και υπολογισμός του βέλτιστου *Warping Path* στην περιορισμένη περιοχή  $R$  (c) Βέλτιστο *warping path* με χρήση της αυξημένης περιοχής  $R^\delta$  – Από τους Müller, Mattes and Kurth (2006)

Η διαδικασία θεωρείται επιτυχής όταν  $P^* = P_R^*$ . Σε περίπτωση που αυτό δεν συμβαίνει μπορούμε να επεκτείνουμε την περιορισμένη περιοχή  $R$  - με κόστος στην απόδοση του αλγορίθμου – προσθέτοντας  $\mathcal{D}$  κελιά αριστερά, δεξιά, πάνω και κάτω από κάθε κελί του  $R$ , για κάποια παράμετρο  $\delta \in \mathbb{N}$ , δημιουργώντας έτσι την περιοχή  $R^\delta$  (Εικόνα 2-13 (c)).

### 3. Υπολογιστική μοντελοποίηση της μουσικής έκφρασης

Η Μουσική απόδοση, ως μια πράξη ερμηνείας, δόμησης και φυσικής αντίληψης ενός μουσικού κειμένου, αποτελεί μία πολυσύνθετη ανθρώπινη δραστηριότητα με φυσικές, ακουστικές, ψυχολογικές, κοινωνικές και καλλιτεχνικές προεκτάσεις (Gabriellsson 1999). Στόχος του συγκεκριμένου κεφαλαίου είναι η μελέτη του παράγοντα της έκφρασης της μουσικής απόδοσης (expressive performance), της εσκεμμένης δηλαδή διαμόρφωσης από τον εκτελεστή – ερμηνευτή διάφορων μουσικών παραμέτρων όπως ο ρυθμός, το tempo, ο χρονισμός, οι δυναμικές των νοτών και η άρθρωση (articulation). Στη συνέχεια παρουσιάζονται σύγχρονες ερευνητικές προσεγγίσεις στην ποσοτική και υπολογιστική μοντελοποίηση της μουσικής έκφρασης. Με τον όρο υπολογιστική μοντελοποίηση εννοούμε τον σχηματισμό υποθέσεων σχετικά με τη μουσική έκφραση με βάση ακριβή και μετρήσιμα δεδομένα της ερμηνείας. Τα περισσότερα υπολογιστικά μοντέλα προσπαθούν να καταγράψουν κοινές ερμηνευτικές αρχές και επικεντρώνονται στις ομοιότητες ανάμεσα σε διαφορετικές ερμηνείες και διαφορετικούς ερμηνευτές.

Η εγκυκλοπαίδεια Britannica ορίζει ένα επιστημονικό μοντέλο γενικά ως «μία φυσική, εννοιολογική ή μαθηματική αναπαράσταση ενός πραγματικού φαινομένου που είναι δύσκολο να παρατηρηθεί απ' ευθείας» (Britannica.com 2012). Ένα μαθηματικό μοντέλο προβλέπει ποσοτικές σχέσεις, ντετερμινιστικές ή στατιστικές, μεταξύ κάποιων μεταβλητών που χρησιμοποιούνται για να περιγράψουν ένα σύστημα ή ένα φαινόμενο, το οποίο είναι πιθανά εξαρτώμενο από τις τιμές κάποιων παραμέτρων του μοντέλου. Ένα υπολογιστικό μοντέλο αποτελεί την υλοποίηση ενός μαθηματικού μοντέλου σε λογισμικό υπολογιστή, το οποίο μπορεί άμεσα να εφαρμοστεί σε κάποια δεδομένα για έλεγχο ή πρόβλεψη τιμών. Τα υπολογιστικά μοντέλα μουσικής έκφρασης στοχεύουν στον ακριβή καθορισμό των φυσικών παραμέτρων που χαρακτηρίζουν μία ερμηνεία (χρονισμός των Onset, διαστήματα ανάμεσα στα onset, επίπεδα έντασης, διάρκεια νότας κτλ.) και τον προσδιορισμό ποσοτικών σχέσεων ανάμεσα σε αυτές τις παραμέτρους.

Τις τελευταίες δεκαετίες έχουν πραγματοποιηθεί πολλές ποσοτικές έρευνες σε λεπτομερείς παράγοντες της μουσικής έκφρασης, βασισμένες σε μετρήσεις του χρονισμού, των δυναμικών ( Shaffer et al. 1985, Clarke 1985, Palmer 1996, Repp 1998 & 1999, Goebel 2001 κ.α.). Κάποιες από αυτές επικεντρώνονται περισσότερο στην στατιστική ανάλυση απ' ό τι στην πρόβλεψη της ερμηνείας, μετρώντας ερμηνευτικές λεπτομέρειες και ορίζοντας κλάσεις κοινών μοτίβων . Άλλες έρευνες οδήγησαν σε ποσοτικά και υπολογιστικά μοντέλα πολύ συγκεκριμένων χαρακτηριστικών της μουσικής έκφρασης, όπως για παράδειγμα το τελικό ritardando (ritardando – μουσικός όρος για τη σταδιακή επιβράδυνση του tempo που είθισται να χρησιμοποιείται στο τέλος ενός μουσικού έργου) και η εμφανής σχέση του με την ανθρώπινη κίνηση (Todd 1995, Sundberg 2000, Friberg et al. 2000, Honing 2003), ο χρονισμός των νοτών διάνθισης (fioritura - (Timmers et al. 2002)) ή η διαφοροποίηση μουσικής άρθρωσης με την παρουσία vibrato (Schoonderwaldt and Friberg 2001), legato (Bresin and Battel 2000) ή staccato (Bresin and Widmer, 2000, Bresin 2001) στις μουσικές φράσεις.

Ένα βήμα παραπέρα προς τη δημιουργία ενός γενικού μοντέλου για την εκφραστικότητα του χρονισμού και των δυναμικών βρίσκουμε στη δουλειά του Clarke (1988). Ο Clarke πρότεινε εννέα κανόνες που εκφράζουν τη συνολική δόμηση της μουσικής μέσω των αλλαγών στο ρυθμό και τις δυναμικές. Συγκεκριμένα, για το πεδίο του χρονισμού, ο Clarke υπέδειξε ότι:

1. Οι διαβαθμισμένες αλλαγές στο χρονισμό μπορούν να χρησιμοποιηθούν για να υποδείξουν την ομαδοποιημένη δομή της μουσικής, με τα όρια των ομάδων να χαρακτηρίζονται από τοπικά μέγιστα στο πεδίο του χρονισμού.

2. Μία νότα μπορεί να επιμηκυνθεί χρονικά προκαλώντας καθυστέρηση στην επόμενη νότα, με σκοπό να αυξηθεί η επιρροή της καθυστερούμενης σημαντικής νότας. Ο κανόνας αυτός σχετίζεται και με τον προηγούμενο, καθώς η νότα στην οποία προκαλείται καθυστέρηση αποτελεί συχνά την αρχή μια νέας ομάδας νοτών.
3. Μια νότα που είναι δομικά σημαντική μπορεί να τονιστεί αυξάνοντας το μεσοδιάστημα πριν από την έναρξή της (*agogic accent*).

Για το πεδίο των δυναμικών, ο Clarke ορίζει τους παρακάτω τρεις κανόνες:

1. Διαβαθμισμένες αλλαγές στις δυναμικές μπορούν να χρησιμοποιηθούν για να υποδείξουν ένα μοτίβο από «ασθενή» και «ισχυρά» μέτρα και κατ' επέκταση μία μετάβαση μέσα στην ομαδοποιημένη δομή της μουσικής. Ο κανόνας αυτός αποτελεί μία απεικόνιση του πρώτου κανόνα χρονισμού στο πεδίο των δυναμικών.
2. Μία αντίθεση μεταξύ των δυναμικών μπορεί να χρησιμοποιηθεί για να αρθρώσει τα όρια μιας ομάδας νοτών, δημιουργώντας μία συσχέτιση μεταξύ μιας ασυνέχειας στις δυναμικές με μία υποκείμενη δομική ασυνέχεια.
3. Σημαντικά δομικά γεγονότα (π.χ. αρμονικές αλλαγές ή μελωδικές κορυφώσεις) μπορεί να ενταθούν δυναμικά, ένας κανόνας αντίστοιχος με τον τρίτο κανόνα χρονισμού.

Τέλος, για το πεδίο της μουσικής άρθρωσης ορίζονται τρεις επιπλέον κανόνες, αντίστοιχοι με αυτούς των δυναμικών:

1. Διαβαθμισμένες αλλαγές στην άρθρωση μπορούν να χρησιμοποιηθούν για να υποδείξουν ένα μοτίβο από «ασθενή» και «ισχυρά» μέτρα και κατ' επέκταση μία μετάβαση μέσα στην ομαδοποιημένη δομή της μουσικής. Ο κανόνας αυτός μπορεί να μην είναι τόσο διακριτός, όσο ο αντίστοιχος κανόνας των δυναμικών.
2. Τα όρια των δομικών ομάδων μπορεί να υποδηλώνονται χρησιμοποιώντας μία ασυνέχεια στην άρθρωση. Αυτό θα μπορούσε να έχει τη μορφή μίας αντίθεσης διάρκειας στην άρθρωση που θα ορίζει διαφορετικές ομάδες νοτών ή τη χρήση μιας μεμονωμένης *staccato* νότας στα όρια ως σημείο διαχωρισμού. Αντίστροφα, μία *legato* άρθρωση μπορεί να χρησιμοποιηθεί ως μέσο συνοχής μεταξύ νοτών της ίδιας ομάδας.
3. Μεμονωμένα μουσικά γεγονότα μπορούν να τονιστούν με χρήση της άρθρωσης.

### 3.1 Το μοντέλο ΚΤΗ

Πρόκειται για ένα σύστημα κανόνων που αναπτύχθηκε στο Βασιλικό Ινστιτούτο Τεχνολογίας της Στοκχόλμης (ΚΤΗ)<sup>1</sup> με έρευνα που διήρκησε πάνω από 20 χρόνια ξεκινώντας με τον Sundberg (1983). Στην πορεία επεκτάθηκε και αναλύθηκε ποικιλοτρόπως και εφαρμόστηκε και στην σύγχρονη μουσική.

#### 3.1.1 Βασική Λειτουργία του μοντέλου

Το μοντέλο αποτελείται από ένα σύνολο κανόνων που προβλέπουν ή περιγράφουν παράγοντες του χρονισμού, των δυναμικών και της μουσικής άρθρωσης (*articulation*), βασιζόμενοι στο τοπικό μουσικό περιεχόμενο. Οι κανόνες αναφέρονται σε ένα περιορισμένο εύρος μουσικών καταστάσεων και θεωρητικών εννοιών (λόγος διάρκειας μεταξύ δύο διαδοχικών νοτών, μελωδικές και αρμονικές μεταβάσεις, δομή μουσικών φράσεων κτλ). Οι περισσότεροι κανόνες λειτουργούν σε πολύ χαμηλό επίπεδο, αναζητώντας σε περιορισμένο μουσικό εύρος και επηρεάζοντας μεμονωμένες νότες. Υπάρχουν πάντως και κανόνες που

<sup>1</sup> ΚΤΗ Royal Institute of Technology, <https://www.kth.se/>



επηρεάζουν ολόκληρες φράσεις. Όλοι οι κανόνες καθορίζονται από κάποιον αριθμό παραμέτρων.

Το μοντέλο αναπτύχθηκε χρησιμοποιώντας την προσέγγιση «*ανάλυση μέσω σύνθεσης*» (Sundberg et al. 1983, Sundberg et al. 1991a) που περιλαμβάνει έναν επαγγελματία μουσικό ο οποίος αξιολογεί απευθείας κάθε υποψήφιο κανόνα που προτείνει ο ερευνητής. Ο μουσικός και ο ερευνητής βρίσκονται σε ένα κύκλο συνεχούς ανατροφοδότησης για την καλύτερη διαμόρφωση και παραμετροποίηση του κάθε κανόνα. Η μέθοδος έχει το πλεονέκτημα ότι μοντελοποιεί μία διάδραση ακροατή – ερμηνευτή. Μειονεκτεί στο ότι θέτει υψηλές απαιτήσεις αξιολόγησης στο μουσικό και ότι έχει μικρό αριθμό αξιολογήσεων. Η μέθοδος εμπνεύστηκε και χρησιμοποιήθηκε σε μεθόδους σύνθεσης ομιλίας (Friberg 1995).

Ως παράδειγμα παρατίθεται ο κανόνας που αναφέρεται ως Duration Contrast (Friberg 1995). Ο κανόνας αυτός τροποποιεί τον λόγο της διάρκειας μεταξύ διαδοχικών νोटών προκειμένου να ενισχύσει και επομένως να κάνει περισσότερο αντιληπτές τις διαφορές στη διάρκεια τους. Αλλάζει τη διάρκεια μίας μεγάλης και μίας σύντομης νότας ώστε η πρώτη να γίνει ακόμη μεγαλύτερη και η δεύτερη ακόμη μικρότερη. Όπως σε όλους του κανόνες, υπάρχει ένας ποσοτικός παράγοντας ( $k$ ) που ορίζεται από τον ερευνητή. Ο παράγοντας λειτουργεί ώστε να δίνει καλά αποτελέσματα σε τιμές κοντά στο 1. Για τιμή 0 ο συγκεκριμένος κανόνας απενεργοποιείται ενώ για αρνητικές τιμές ο κανόνας αναστρέφεται, γεγονός που μπορεί να οδηγήσει σε περίεργα, μη μουσικά αποτελέσματα. Ένας σημαντικός παράγοντας του μοντέλου είναι ότι οι κανόνες λειτουργούν αθροιστικά (πολλοί κανόνες μπορούν για παράδειγμα να τροποποιήσουν τη διάρκεια της νότας) γεγονός που δημιουργεί πρόβλημα στη ρύθμιση των παραμέτρων σε πραγματικές ηχογραφήσεις.

### 3.1.2 Εμπειρική αξιολόγηση

Το μοντέλο ΚΤΗ περιλαμβάνει ένα μεγάλο αριθμό παραμέτρων οι οποίες πρέπει να οριστούν σε σταθερές τιμές προκειμένου να λειτουργήσει ως προγνωστικό μοντέλο. Ήδη από τα πρώτα βήματα του μοντέλου, η ομάδα του Sundberg πραγματοποίησε διάφορες αντιληπτικές αξιολογήσεις προκειμένου να αποδείξει την εγκυρότητα και αξιοπιστία του μοντέλου. Σε δύο ακουστικά πειράματα, αξιολογήθηκε η αντιληπτική απόκριση μουσικών και μη μουσικών σε διαφοροποιήσεις της μουσικής έκφρασης (Sundberg et al. 1991). Στο πρώτο πείραμα καθορίστηκε το αντιληπτικό όριο για τις ποσοτικές τιμές του κανόνα ( $k$  values), για τις οποίες οι ακροατές μπορούσαν να αντιληφθούν διαφορές ανάμεσα σε δύο παραδείγματα (οι μουσικοί φάνηκαν πιο ευαίσθητοι σε διαφορές από τους μη μουσικούς). Τα όρια αυτά εξαρτώνται άμεσα από τον τύπο του κανόνα. Στο δεύτερο πείραμα, εξειδικευμένοι μουσικοί προσάρμοσαν τις τιμές του  $k$  για έξι κανόνες. Για ένα κανόνα (Duration Contrast) προσάρμοσαν την τιμή στο μηδέν, ενώ για τους υπόλοιπους κοντά στο όριο της αντιληπτικότητας.

Σε μία άλλη μελέτη, η ομάδα του Sundberg μελέτησε κατά πόσο το μοντέλο μπορεί να προσαρμοστεί σε μία συγκεκριμένη ερμηνεία μια αργής σονάτας του Mozart με την προσέγγιση της *δοκιμής και του λάθους* (trial and error). Επικεντρώθηκαν αποκλειστικά στον εκφραστικό χρονισμό της μελωδίας (με την έννοια της παρέκκλισης από την σημειογραφία) και καθόρισαν την ομοιότητα ανάμεσα στην ερμηνεία και στο μοντέλο με χρήση του συντελεστή συσχέτισης. Δοκιμάστηκαν τιμές του  $k$  για κάθε φράση ξεχωριστά. Ο κανόνας PHRASE ARCH εμφάνισε τη μεγαλύτερη συσχέτιση, ακολουθούμενος από τους κανόνες HARMONIC CHARGE και DURATION CONTRAST.

Οι Zanon και De Poli αποπειράθηκαν να προσαρμόσουν το μοντέλο σε πραγματικά δεδομένα, αρχικά με σταθερές και στη συνέχεια με χρονικά μεταβαλλόμενες παραμέτρους (2003). Στην

πρώτη τους απόπειρα, χρησιμοποίησαν διάφορες πραγματικές εκτελέσεις από τα πρώτα 24 μέτρα της σονατίνας σε σολ ματζόρε του Beethoven, τα οποία περιλαμβάνουν διάφορους συναισθηματικούς στόχους (θυμό, φόβο, χαρά κτλ.). Παρατηρήθηκε ότι οι περισσότεροι κανόνες του μοντέλου ήταν σχεδόν κάθετοι μεταξύ τους, με ελάχιστες εξαιρέσεις. Οι συναισθηματικοί στόχοι διαφοροποιούνταν από συγκεκριμένους κανόνες, για παράδειγμα ο κανόνας PUNCTUATION ξεχώριζε τις φοβικές ή λυπητερές ερμηνείες από τις χαρούμενες ή ουδέτερες.

Σε μια δεύτερη απόπειρα οι Zanon και De Poli προσπάθησαν να ξεπεράσουν το περιορισμό των σταθερών τιμών των παραμέτρων  $k$  κατά τη διάρκεια του κομματιού. Για τον σκοπό αυτό χρησιμοποίησαν χρονικά παράθυρα μεγέθους 1-2 μέτρων, που ολισθαίνουν σε αντιστοιχία με τις νότες της σημειογραφίας. Καθώς οι βαθμοί ελευθερίας γίνονταν πάρα πολλοί, περιόρισαν τον αριθμό των κανόνων σε πρώτη φάση. Σε δεύτερη φάση το μέγεθος του παραθύρου και ο αριθμός των κανόνων αυξήθηκε. Αν και συγκεκριμένοι κανόνες έλαβαν τυπικές τιμές για τις παραμέτρους  $k$  σε συγκεκριμένες φράσεις, το συνολικό αποτέλεσμα που προέκυψε ήταν αρκετά πολύπλοκο. Η εκτίμηση με χρονικά μεταβαλλόμενες παραμέτρους  $k$  ήταν πολύ ευαίσθητη σε εκφραστικούς παράγοντες όπως οι νότες φιοριτούρας.

Συνοψίζοντας, υπάρχουν ενδείξεις ότι το μοντέλο ΚΤΗ αποτελεί μία υπολογίσιμη γλώσσα αναπαράστασης της μουσικής έκφρασης. Το κατά πόσο μπορεί να εφαρμοστεί σε μεγάλες συλλογές σύνθετης μουσικής παραμένει ανοικτό θέμα. (Widmer and Goebel 2004)

## 3.2 Το μοντέλο Todd

Σε αντίθεση με το μοντέλο ΚΤΗ που χρησιμοποιεί την προσέγγιση «ανάλυση μέσω σύνθεσης», το μοντέλο Todd συνοψίζεται με την έννοια «ανάλυση μέσω μέτρησης» καθώς αντλεί την απόδειξή του απ' ευθείας από μετρήσεις ανθρώπινων εκφραστικών ερμηνειών. Αναπτύχθηκε από τον Neil Todd στα τέλη των 80's και τις αρχές των 90's

### 3.2.2 Το βασικό μοντέλο

Η ουσία του μοντέλου βρίσκεται στη θεώρηση ότι υπάρχει μία άμεση σύνδεση ανάμεσα σε κάποιους παράγοντες της μουσικής δομής (π.χ grouping structure – *ομαδοποίηση μουσικών δομών*) και την ερμηνεία, καθώς και ότι η σχέση αυτή μπορεί να μοντελοποιηθεί με τη χρήση ενός μόνο απλού κανόνα. Ο όρος grouping structure (Lerdahl and Jackendoff 1983) αναφέρεται στο γεγονός πως όταν ακούμε ένα μουσικό κομμάτι αυτόματα χωρίζουμε τον ήχο που ακούμε σε μονάδες ή τμήματα το μέγεθος των οποίων ποικίλει, όπως μοτίβα, θέματα, φράσεις, ομάδες θεμάτων, μέρη κτλ. Ο Todd θεώρησε ότι οι ερμηνευτές τείνουν να ελαττώνουν το tempo στα όρια τέτοιου είδους τμημάτων προκειμένου να επιτρέψουν στον ακροατή να αντιληφθεί την ιεραρχική δόμηση της μουσικής (αντίστοιχα έχει παρατηρηθεί ότι οι τραγουδιστές και οι μουσικοί πνευστών συνήθως παίρνουν ανάσα ανάμεσα σε αυτά τα τμήματα και όχι κατά τη διάρκειά τους). Αυτό βρίσκεται σε συμφωνία με μετρήσεις που έχουν πραγματοποιηθεί σχετικά με τη μεταβολή του tempo στα όρια των φράσεων και κατά τη διάρκειά τους.

Ο Todd αρχικά μοντελοποίησε αποκλειστικά το χρονισμό και στη συνέχεια περιέλαβε τη συσχέτιση μεταξύ χρονισμού και δυναμικών. Η σχέση μεταξύ χρονισμού και δυναμικών μοντελοποιήθηκε από την απλή σχέση «όσο πιο γρήγορο, τόσο πιο δυνατό», όπου η ένταση είναι ανάλογη με το τετράγωνο του tempo (Todd 1992). Κατά συνέπεια, το ιεραρχικό grouping structure της μουσικής καθορίζει άμεσα τόσο το στιγμιαίο τοπικό tempo όσο και τις δυναμικές της μουσικής έκφρασης. Σε κάθε ιεραρχικό επίπεδο, το μοντέλο αυξάνει tempo και δυναμικές προς το μέσο της φράσης και αντίστροφα.

### 3.2.3 Εμπειρική αξιολόγηση

Η απλοϊκή φύση του μοντέλου του Todd έχει το πλεονέκτημα ότι οι θεωρητικές του υποθέσεις μπορούν να δοκιμαστούν σχετικά εύκολα. Στις έρευνες του ο Todd συνέκρινε τα αποτελέσματα του μοντέλου με τις καμπύλες του tempo και των δυναμικών από μία ή δύο εκτελέσεις επιλεγμένων κομματιών των Haydn, Mozart και Chopin. Οι πραγματικές και οι αλγοριθμικές καμπύλες ήταν αρκετά παρόμοιες αν και δεν πραγματοποιήθηκαν ποσοτικές εκτιμήσεις.

Σε πιο πρόσφατη έρευνα (Clarke and Windsor 2000) μια ομάδα ακροατών αξιολόγησε τόσο ανθρώπινες όσο και αλγοριθμικές εκτελέσεις που πραγματοποιήθηκαν με το μοντέλο Todd. Δύο διαφορετικές παραλλαγές μουσικών φράσεων από την σονάτα K331 του Mozart υλοποιήθηκαν τόσο από το μοντέλο Todd όσο και από δύο επαγγελματίες πιανίστες. Στις ανθρώπινες εκτελέσεις οι ακροατές αντιλήφθηκαν τις δύο διαφορετικές εκδοχές των μουσικών φράσεων με βάση τις αλλαγές στις δυναμικές και στο χρονισμό. Η τεχνητή εκτέλεση δεν κατάφερε να ερμηνεύσει τις δυο διαφορετικές παραλλαγές, ενώ σε κάποιες περιπτώσεις έκανε το αντίθετο από τους πιανίστες: ενώ οι πιανίστες έδιναν έμφαση στη μη αναμενόμενη όγδοη νότα ως ένα δυναμικό πέρασμα σε νέα φράση, το μοντέλο την έπαιζε πιο απαλά σαν ξεκίνημα καινούργιας φράσης. Το μοντέλο απέδωσε καλύτερα στην ερμηνεία δύο φράσεων μίας απλής μελωδίας οχτώ μέτρων σε μουσικά εκπαιδευμένο κοινό. Ένα γενικό συμπέρασμα των παραπάνω πειραμάτων είναι ότι ο εκφραστικός χρονισμός και οι δυναμικές δεν συνδέονται μεταξύ τους με τον απλό τρόπο που προτείνει το μοντέλο Todd (Widmer and Goebel 2004).

Σε μία άλλη εμπειρική μελέτη οι Clarke και Windsor δημιούργησαν διάφορες τεχνητές εκτελέσεις του ίδιου κομματιού χρησιμοποιώντας διαφορετικά επίπεδα βαρών στο μοντέλο του Todd. Τις αξιολόγησαν με τη στατιστική τεχνική της παλινδρόμησης ως προς δύο επαναλαμβανόμενες εκτελέσεις ενός επαγγελματία πιανίστα. Η καλύτερη αλγοριθμική εκτέλεση βρισκόταν κάπου ανάμεσα στις δύο εκτελέσεις του πιανίστα, χρησιμοποιώντας διαφορετικά επίπεδα βαρών για το χρονισμό και τις δυναμικές (η λεγόμενη «υβριδική ερμηνεία» - Windsor and Clarke 1997). Ο χρονισμός απαιτεί περισσότερη έμφαση σε δομικά στοιχεία χαμηλότερου επιπέδου ενώ οι δυναμικές σε υψηλότερου επιπέδου. Παρατηρώντας περαιτέρω τις διαφορές ανάμεσα στις ανθρώπινες εκτελέσεις και αυτές του μοντέλου, οι Windsor και Clarke εισήγαγαν την έννοια των “residuals”, με τον ισχυρισμό ότι τα πιο ενδιαφέροντα είναι όσα δεν εξηγούνται από το μοντέλο. Με τον τρόπο αυτό το μοντέλο του Todd χρησιμοποιήθηκε ως ένα εργαλείο ανάλυσης που ερμηνεύει την ιδιοσυγκρασία των ανθρώπινων μουσικών εκτελέσεων

### 3.3 Το μοντέλο Mazzola

Ένα διαφορετικό μοντέλο βασισμένο σε καθαρά μαθηματική προσέγγιση είναι το μοντέλο Mazzola. Ο Ελβετός μαθηματικός και πιανίστας της jazz Guerino Mazzola, ανέπτυξε τη δική του θεωρία μαθηματικής μουσικής και το αντίστοιχο μοντέλο ερμηνείας τη δεκαετία του '80, με την υπολογιστική εφαρμογή του να συνεχίζει να αναπτύσσεται ως και σήμερα. Η υπολογιστική εφαρμογή του μοντέλου του διατίθεται ελεύθερα μέσω της ιστοσελίδας του περιβάλλοντος λογισμικού RUBATO<sup>2</sup> (GNU public license) .

---

<sup>2</sup> <http://www.rubato.org>

### 3.3.1 Το βασικό μοντέλο

Το μοντέλο Mazzola βασίζεται σε ένα τεράστιο θεωρητικό υπόβαθρο, τη «μαθηματική μουσική θεωρία» (mathematical music theory – Mazzola 1990), η οποία καλύπτει διάφορους παράγοντες της μουσικής θεωρίας και ανάλυσης μέσω πολύπλοκων μαθηματικών προσεγγίσεων αλλά ταυτόχρονα εμπεριέχει και φιλοσοφικές, σημειολογικές και αισθητικές προσεγγίσεις. Κάθε βήμα της θεωρίας εξηγείται με συγκεκριμένους μαθηματικούς όρους και ορολογία που σε γενικές γραμμές διαφέρει από αυτή που συνήθως χρησιμοποιείται στην έρευνα της μουσικής εκτέλεσης.

Το μοντέλο αποτελείται από το μέρος της ανάλυσης και το μέρος της ερμηνείας. Το μέρος της ανάλυσης περιλαμβάνει υπολογιστικά εργαλεία ανάλυσης διαφόρων παραγόντων της μουσικής δομής όπως π.χ. μέτρα, μελωδία και αρμονία. Καθένα από αυτά υλοποιείται μέσω ενός plugin (τα επονομαζόμενα RUBETTES) που αναθέτει συγκεκριμένα «βάρη» σε κάθε νότα. Το κομμάτι της ερμηνείας μετασχηματίζει τα δομικά χαρακτηριστικά σε μια τεχνητή ερμηνεία βασιζόμενο σε δύο θεωρίες, τις επονομαζόμενες «Stemma Theory» και «Operator Theory» (μία μορφή συσσωρευτικής και βασισμένης σε κανόνες αντιστοίχισης δομών προς ερμηνεία). Τροποποιεί επαναληπτικά τα πεδία του «διανύσματος ερμηνείας (performance vector), καθένα από τα οποία ελέγχει μία παράμετρο έκφρασης στην τεχνητή ερμηνεία (Widmer and Goebel 2004).

### 3.4 Το μοντέλο μηχανικής μάθησης

Μια εναλλακτική μέθοδος για τη δημιουργία μοντέλων μουσικής έκφρασης είναι να συγκεντρώσουμε ένα μεγάλο όγκο εμπειρικών δεδομένων (επακριβώς μετρημένες ερμηνείες ικανών μουσικών) και να αναθέσουμε στον υπολογιστή να ανακαλύψει κανονικότητες και νόρμες στα δεδομένα με τη χρήση τεχνικών επαγωγικής μηχανικής μάθησης και εξόρυξης δεδομένων (data mining). Κάποιοι από αυτούς τους αλγορίθμους παράγουν γενικούς κανόνες ερμηνείας που μπορούν να χρησιμοποιηθούν απευθείας ως προγνωστικά υπολογιστικά μοντέλα (Widmer 2000 & 2002, Widmer et al. 2003). Παρακάτω αναλύεται πως η μηχανική μάθηση μπορεί να χρησιμοποιηθεί για να προβλέψει μοντέλα, τόσο για εκφραστικές αποκλίσεις σε επίπεδο νότας, όσο και για υψηλότερου επιπέδου μοτίβα σε επίπεδα μουσικών φράσεων, καθώς και πως οι δύο αυτοί τύποι μοντέλων μπορούν να συνδυαστούν για τη δημιουργία ενός πολυεπίπεδου ολοκληρωμένου μοντέλου εκφραστικού χρονισμού και δυναμικών.

#### 3.4.1 Το μοντέλο σε επίπεδο νότας

Σε μια πρώτη προσπάθεια δημιουργίας γενικών κανόνων ερμηνείας, ο Widmer (2003) ανέπτυξε ένα νέο αλγόριθμο επαγωγικής μηχανικής μάθησης και τον εφάρμοσε πάνω στη διαδικασία της μάθησης κανόνων για τον χρονισμό, τις δυναμικές και την άρθρωση σε επίπεδο νότας. Με την έκφραση «σε επίπεδο νότας» εννοούμε το πώς ένας πιανίστας θα παίξει μία συγκεκριμένη νότα σε ένα κομμάτι – πιο αργά ή πιο γρήγορα από τη σημειογραφία, πιο δυνατά ή πιο σιγά από την προηγούμενη, πιο staccato ή πιο legato. Τα δεδομένα αυτά αντιπαραβάλλονται με υψηλότερου επιπέδου τεχνικές έκφρασης, όπως ο σχηματισμός μια ολόκληρης μουσικής φράσης. Η εκπαίδευση του μοντέλου βασίστηκε σε δεκατρείς (13) σονάτες του Mozart οι οποίες εκτελέστηκαν από επαγγελματία πιανίστα σε πιάνο ελεγχόμενο από υπολογιστή (Bösendorfer 290SE) μέσω του οποίου καταγράφονταν κάθε λεπτομέρεια χρονισμού, δυναμικών και άρθρωσης. Το σύνολο των δεδομένων

εκπαίδευσης αποτελούνταν από πάνω από 106.000 νότες που αντιστοιχούσε σε περίπου 4 ώρες μουσικής.

Τα πειράματα πραγματοποιήθηκαν μόνο πάνω στη μελωδία, κάτι που δίνει ένα ενεργό σύνολο δεδομένων εκπαίδευσης 41116 νότες. Κάθε νότα περιγράφονταν από 29 χαρακτηριστικά που αντιπροσωπεύαν τόσο ενδογενή χαρακτηριστικά της (όπως κλίμακα, διάρκεια, μετρική θέση) όσο και παράγοντες τοπικών συνθηκών (π.χ. το μελωδικό περίγραμμα γύρω από τη νότα). Από αυτό ο υπολογιστής «έμαθε» ένα σύνολο 17 σχετικά απλών κανόνων ταξινόμησης που προβλέπουν ένα μεγάλο μέρος των επιλογών του πιανίστα σε επίπεδο νότας. Συγκεκριμένα βρέθηκαν 4 κανόνες οι οποίοι συνδυαζόμενοι μεταξύ τους προέβλεψαν σωστά το 23% των περιπτώσεων όπου ο πιανίστας μεγάλωσε τη διάρκεια μιας νότας σε σχέση με τη σημειογραφία. Για παράδειγμα, ο παρακάτω απλός και γενικός κανόνας προβλέφθηκε από τον υπολογιστή:

**RULE TL2:**

```
abstract_duration_context = equal-longer  
& metr_strength ≤ 1  
⇒ Lengthen
```

*«αν έχουμε δύο νότες ίδιας διάρκειας ακολουθούμενες από μία νότα μεγαλύτερης διάρκειας, επιμήκυνε την νότα που προηγείται της τελικής μεγαλύτερης νότας, αν αυτή η νότα βρίσκεται σε μετρικά ασθενή θέση ( $metr\_strength \leq 1$ )»*

Η χρήση μόνο του κανόνα TL2 προέβλεψε με επιτυχία το 14,12% από όλες τις περιπτώσεις επιμήκυνσης κάποιας νότας των δεδομένων εκπαίδευσης, με το υπόλοιπο των περιπτώσεων να εμπίπτει σε άλλους κανόνες του μοντέλου. Το ποσοστό των λαθεμένων προβλέψεων του ήταν μόλις 2.86%.

Το σύνολο των κανόνων περιγράφεται αναλυτικά στην έρευνα του Widmer (2002), όπου επίσης η γενικότητα και η ισχύς του κάθε κανόνα ποσοτικοποιείται με βάση εκτενή πειράματα σε πραγματικά δεδομένα. Ενδιαφέρον είναι το γεγονός ότι πολλοί από τους κανόνες που ανακαλύφθηκαν μέσω της μηχανικής μάθησης παρουσιάζουν μεγάλη ομοιότητα με τους κανόνες του μοντέλου ΚΤΗ. Έτσι η προσέγγιση της μηχανικής μάθησης παρέχει έμμεσα περαιτέρω ενδείξεις για την εγκυρότητα του μοντέλου ΚΤΗ.

### 3.4.2 Το πολύ-επίπεδο μοντέλο

Το μοντέλο επιπέδου νότας αναφέρεται σε ένα μικρό μέρος των εκφραστικών τεχνικών που παρατηρούνται σε μία πραγματική μουσική ερμηνεία. Οι μουσικοί αντιλαμβάνονται τη μουσική ως μία πολυπλοκότητα που αποτελείται από πιο αφηρημένες δομές (μοτίβα, μέρη, φράσεις κτλ.) και χρησιμοποιούν το ρυθμό, τις δυναμικές και την άρθρωση για τη δημιουργία αυτών των δομών. Η μουσική ερμηνεία αποτελεί ένα πολύ-επίπεδο φαινόμενο, με μουσικές δομές και ερμηνευτικά μοτίβα που σε πολλά σημεία ενσωματώνονται το ένα μέσα στο άλλο.

Προκύπτει επομένως το θέμα της δημιουργίας μοντέλων πολύ-επίπεδης μάθησης εκφραστικών χρονισμών και δυναμικών. Στόχος είναι ο υπολογιστής να μπορεί να προβλέψει τι είδους βασικό ρυθμό και δυναμικές μορφές θα χρησιμοποιήσει ο ερμηνευτής σε μία μουσική φράση, σε ένα συγκεκριμένο τμήμα του κομματιού και σε ένα συγκεκριμένο επίπεδο ιεραρχίας φράσεων. Θεμελιώδεις για το μοντέλο αυτό είναι ένας αριθμός απλοϊκών υποθέσεων. Πρώτον, ότι οι τεχνικές εκφραστικού χρονισμού και δυναμικών που

εφαρμόζονται σε μία φράση μπορούν να προσεγγισθούν από μια οικογένεια καμπύλων τετραγωνικών πολυώνυμων . Δεύτερο, ότι η συνολική πολύ-επίπεδη ερμηνεία μπορεί να αναπαρασταθεί λογικά ως γραμμικός συνδυασμός τέτοιων εκφραστικών σχημάτων σε διαφορετικό ιεραρχικό επίπεδο. Και τρίτο, παρόμοιες φράσεις τείνουν να παίζονται με παρόμοιο τρόπο από τον πιανίστα (Widmer and Goebel 2004).

Καμία από τις παραπάνω υποθέσεις δεν αναμένεται να ισχύει απόλυτα, αλλά παρέχουν τη βάση για ένα λειτουργικό πολύ-επίπεδο μοντέλο εκφραστικής απόδοσης μουσικής φράσης (phrasing) που ενσωματώνεται σε ένα σύστημα μηχανικής μάθησης (Widmer and Tobudic 2003). Το σύστημα παίρνει ως είσοδο ένα σύνολο εκτελέσεων από μουσικούς που αναπαρίσταται από τη μουσική σημειογραφία, μια ιεραρχική ανάλυση των φράσεων της μουσικής και επιπλέον, τις καμπύλες ρυθμού και δυναμικών που αναπαριστούν τους αντίστοιχους παράγοντες ως μέρη της ερμηνευτικής έκφρασης. Αποσυνθέτει τις καμπύλες μιας ερμηνείας εφαρμόζοντας προσεγγιστικές τετραγωνικές συναρτήσεις στα σημεία των καμπυλών που σχετίζονται με συγκεκριμένες φράσεις, συσχετίζοντας ένα στοιχειώδες εκφραστικό σχήμα με κάθε φράση σε κάθε επίπεδο. Στη συνέχεια προβλέπει στοιχειώδη εκφραστικά σχήματα για φράσεις σε νέα κομμάτια, βασιζόμενο στην ομοιότητα με γνωστές φράσεις, και συνδυάζει τα ανεξάρτητα σχήματα σε σύνθετες καμπύλες έκφρασης (ρυθμού και δυναμικών). Η μάθηση έτσι πραγματοποιείται με έναν ανά περίπτωση τρόπο, βασισμένη σε αναλογίες μεταξύ κομματιών. Δεν υπάρχει ένα ξεκάθαρο προγνωστικό μοντέλο, αλλά η ουσία του μοντέλου ενσωματώνεται στις τρεις υποθέσεις που αναφέρθηκαν παραπάνω.

Υπάρχει τρόπος να συνδυαστεί το μοντέλο πρόβλεψης σε επίπεδο φράσεων με την μάθηση βασισμένη σε κανόνες που περιεγράφηκε προηγουμένως. Αφού εφαρμόσουμε τετραγωνικά προσεγγιστικά πολυώνυμα σε μια καμπύλη χρονισμού ή δυναμικών, αφαιρούμε το αποτέλεσμα της προσέγγισης από την αρχική καμπύλη. Το αποτέλεσμα είναι αυτό που οι Windsor & Clarke ονόμασαν «residuals», χαμηλού επιπέδου τοπικές αποκλίσεις στο χρονισμό και στις δυναμικές που δεν μπορούν να εξηγηθούν με αναφορά σε εκτεταμένες δομικές ενότητες όπως οι φράσεις. Ο αλγόριθμος μάθησης κανόνων που περιεγράφηκε προηγουμένως μπορεί να χρησιμοποιηθεί για να εκπαιδευθεί ένα μοντέλο πρόβλεψης επί αυτών των τοπικών φαινομένων. Συνδυάζοντας τα δύο μοντέλα προκύπτει ένα προγνωστικό υπολογιστικό μοντέλο που λαμβάνει υπόψιν του τόσο την ιεραρχική μουσική δομή όσο και το τοπικό μουσικό περιεχόμενο (Widmer and Tobudic 2003).

### 3.4.3 Εμπειρική αξιολόγηση

Τόσο το μοντέλο σε επίπεδο νότας όσο και το πολύ-επίπεδο μοντέλο έχουν δοκιμαστεί εκτενώς σε πραγματικά δεδομένα. Στο πείραμά του Widmer (2002) η ακρίβεια για κάθε νότα των δεδομένων εκπαίδευσης (41116 νότες) καταγράφεται λεπτομερώς για κάθε κανόνα που ανακαλύφθηκε, δίνοντας έτσι μια πλήρη εικόνα για τη γενικότητα και την αξιοπιστία των κανόνων. Επίσης παρουσιάζονται ποσοτικά πειράματα σε μεγάλο αριθμό νέων ερμηνειών που χρησιμοποιήθηκαν ως δεδομένα ελέγχου, δείχνοντας ότι οι κανόνες ισχύουν και για άλλους ερμηνευτές, ακόμη και για μουσική διαφορετικού στυλ, χωρίς σημαντική απώλεια στην ακρίβειά τους. Για παράδειγμα οι κανόνες δοκιμάστηκαν σε μουσική ελαφρώς διαφορετική (Chopin) και κάποιιοι από αυτούς επέδειξαν ακόμη και μεγαλύτερη ακρίβεια απ' ότι στα δεδομένα με τα οποία εκπαιδεύτηκαν (Mozart). Φαίνεται ότι η μηχανική μάθηση ανακάλυψε κάποιες θεμελιώδεις, αν και επί το πλείστο απλές, αρχές της ερμηνείας.

Η απόδοση το πολύ-επίπεδου μοντέλου ποσοτικοποιήθηκε από τους Tobudic και Widmer (2003a) μετρώντας πόσο καλά μπορεί να προβλέψει τις λεπτομέρειες στην ερμηνεία ενός πιανίστα. Συγκεκριμένα μετρήθηκε κατά πόσο ταιριάζουν οι καμπύλες ρυθμού και δυναμικών που προέβλεψε το μοντέλο με τις αντίστοιχες καμπύλες πραγματικών ανθρώπινων ερμηνειών. Πειράματα με ένα σημαντικό αριθμό εκτελέσεων από ένα επαγγελματία πιανίστα έδειξαν ότι οι καμπύλες ταιριάζουν ικανοποιητικά. Τα αποτελέσματα ήταν καλύτερα όσο αφορά τις δυναμικές, παρά τον χρονισμό και τον ρυθμό. Μετά από πιο λεπτομερή διερεύνηση φάνηκε ότι η φτωχή απόδοση του μοντέλου στο επίπεδο του ρυθμού οφείλεται σε προβλήματα προσέγγισης των πραγματικών καμπύλων ρυθμού μέσω ιεραρχικών τετραγωνικών συναρτήσεων. Τα αποτελέσματα βελτιώθηκαν σε επόμενη έρευνα των Tobudic and Widmer (2003b) με βελτιστοποίηση του αλγορίθμου μάθησης.

## 4. Οπτικοποίηση της μουσικής έκφρασης

Όπως αναφέρθηκε και σε προηγούμενες παραγράφους, κάθε φορά που ένα μουσικό κομμάτι ερμηνεύεται από ένα διαφορετικό εκτελεστή, η παρτιτούρα μπορεί να είναι κοινή, ο ερμηνευτής όμως προσθέτει τις δικές του ερμηνευτικές διαφοροποιήσεις οι οποίες αφορούν σε διακυμάνσεις στο ρυθμό ή/και το tempo, σε δυναμικές φράσεων ή/και επιμέρους νοτών, καθώς και στη μουσική άρθρωση, δηλαδή τον τρόπο εκτέλεσης αναφορικά με την περιβάλλουσα πλάτους κάθε νότας. Το γεγονός αυτό καθιστά την κάθε ερμηνεία μία μοναδική μουσική εμπειρία, αντιληπτή από τον μέσο ακροατή έστω και αν δεν μπορεί να την εξηγήσει με μουσικούς όρους. Η μουσική σημειογραφία (παρτιτούρα) μπορεί να θεωρηθεί ως μια πρώτη μορφή απεικόνισης της μουσικής, απαιτεί όμως ανάλογη μουσική παιδεία για να γίνει κατανοητή και δεν αναπαριστά πλήρως τις ερμηνευτικές αποκλίσεις των διάφορων ερμηνευτών (οι διαφορές στη μουσική έκφραση του κάθε εκτελεστή αντικατοπτρίζουν το πώς αυτός ερμηνεύει το ίδιο μουσικό κείμενο).

Η οπτική απεικόνιση επομένως της μουσικής ερμηνείας αποτελεί μια πρόκληση τόσο σε ερευνητικό όσο και σε καλλιτεχνικό επίπεδο. Όσον αφορά συγκεκριμένα τη μουσική έκφραση, μία αποτελεσματική απεικόνιση της μουσικής μπορεί να βελτιώσει τη μουσική εμπειρία και να βοηθήσει τη μουσική κατανόηση του ακροατή. Ένα οπτικό «στιγμιότυπο» της μουσικής έκφρασης βοηθάει στη βαθύτερη κατανόηση της μουσικής ερμηνείας. Για το σκοπό αυτό πρέπει να καθοριστεί το ποια στοιχεία θα οπτικοποιηθούν, πως αυτά θα υπολογιστούν και ποια θα είναι η σχέση ανάμεσα στο καθένα από τα στοιχεία αυτά και την παραγόμενη εικόνα (Hiraga και Matsuda 2004). Στόχος είναι να βρεθεί μία κατάλληλη συσχέτιση (mapping) μεταξύ των χαρακτηριστικών της μουσικής και κάποιων οπτικών απεικονίσεων που θα καθιστούν τα χαρακτηριστικά αυτά αντιληπτά και σε οπτικό επίπεδο.

Η οπτικοποίηση της μουσικής εμπίπτει στον γενικότερο ερευνητικό τομέα της *οπτικοποίησης πληροφορίας* (Information Visualisation) που ορίζεται ως η οπτική αναπαράσταση αφηρημένων δεδομένων με σκοπό να ενισχυθεί η ανθρώπινη αντίληψη και νόηση. Τα αφηρημένα δεδομένα μπορεί να είναι είτε αριθμητικά είτε μη αριθμητικά. Η χρήση της αυξάνεται συνεχώς σε τομείς επιστημονικής έρευνας που άπτονται στο χώρο των μεγάλων δεδομένων (big data) και έχουν εφαρμογές όπως οι ψηφιακές βιβλιοθήκες, η μελέτη οικονομικών δεδομένων, η μετεωρολογία, η φαρμακευτική έρευνα, κ.ο.κ. Η σπουδαιότητα της οπτικοποίησης δεδομένων οφείλεται στην άμεση και εκτεταμένη σύνδεση ανάμεσα στην όραση και τη νόηση που επιτρέπει στον άνθρωπο να βλέπει, να εξερευνά και να κατανοεί μεγάλο όγκο πληροφοριών άμεσα και με ενστικτώδη τρόπο (Wikipedia.org 2020d).

### 4.1 Ερευνητικές Προσεγγίσεις

Διάφορες έρευνες έχουν πραγματοποιηθεί τα τελευταία χρόνια πάνω στην οπτικοποίηση της μουσικής με διαφορετικούς στόχους και προσεγγίσεις η καθεμία. Ενδεικτικά, ο Schoonderwaldt και οι συνεργάτες του ερεύνησαν τη χρηστικότητα μεθόδων οπτικοποίησης σε ένα διαδραστικό σύστημα εκπαίδευσης με στόχο τη βελτίωση της ικανότητας μουσικής έκφρασης των ερμηνευτών (Schoonderwaldt et al., 2002). Οι Hirada και Matsuda (2004) πρότειναν ένα διαδραστικό σύστημα που επιτρέπει στους χρήστες να εξερευνούν στοιχεία της ερμηνείας σε τρισδιάστατη απεικόνιση. Οι Dixon, Goebel και Widmer (2002) χρησιμοποίησαν μία οπτικοποίηση της μουσικής έκφρασης σε ένα πεδίο tempo / δυναμικών όπου δύναται να εντοπιστούν ομοιότητες και διαφορές μεταξύ διαφορετικών ερμηνειών. Ο Gasser (2005) ερευνά τη μουσική έκφραση στο πιάνο αντιστοιχίζοντας πραγματικές



ηχογραφήσεις με την παρτιτούρα η οποία αναπαρίσταται σε μορφή MIDI. Οι Fonteles, Rodrigues και Basso (2013) χρησιμοποιούν την τεχνική γραφικών των σωματιδίων (particles)<sup>3</sup> για να δημιουργήσουν τρισδιάστατες απεικονίσεις σε πραγματικό χρόνο και παρατηρούν οπτικά μοτίβα που αντιστοιχούν στα μουσικά θέματα. Στην έρευνα των Gang, Lundberg κ.α. (2010) εξάγονται τρία βασικά ηχητικά χαρακτηριστικά (η τονικότητα, το επίπεδο των δυναμικών και η χροιά ως κατανομή ενέργειας στο φάσμα) και απεικονίζονται ως προς το χρόνο. Άλλα χαρακτηριστικά του ήχου, όπως η μορφή και η διάρκεια της ενεργοποίησης (attack) και απενεργοποίησης μια νότας (release) μπορούν επίσης να υπολογισθούν ως ηχητικά χαρακτηριστικά και να παρουσιαστούν στον χρήστη.

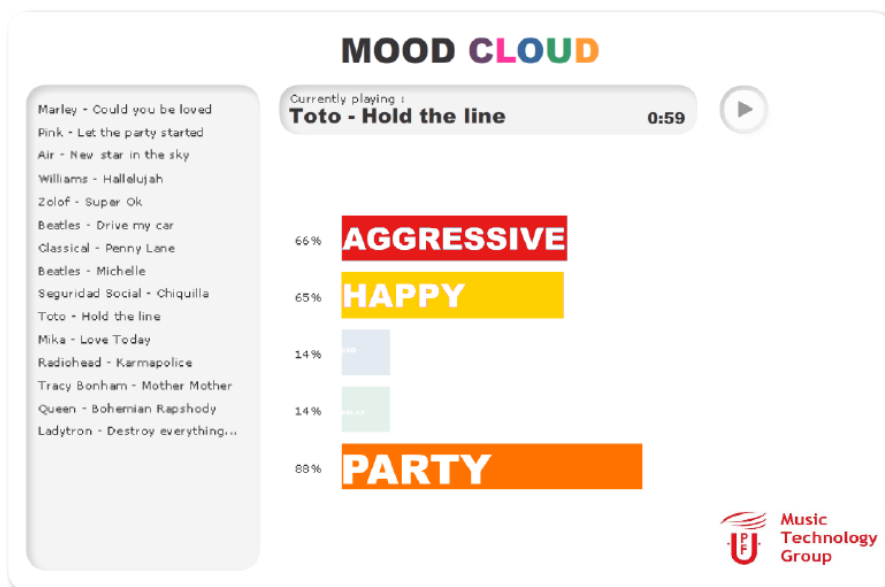
Σύμφωνα με άρθρο των Hiraga κ.α. (2002), η οπτικοποίηση της ερμηνείας μπορεί να φανεί πρακτικά χρήσιμη σε περιπτώσεις όπως:

1. *Διαμοιρασμός στην αντίληψη της ερμηνείας ενός μουσικού συνόλου μεταξύ των μουσικών.* Η δημιουργία μουσικής μαζί με άλλους ανθρώπους είναι μια πολύ ευχάριστη εμπειρία, στην οποία όμως δεν είναι πάντα εύκολη η ακριβής επικοινωνία όσο αφορά την άποψη του καθενός σχετικά με την ερμηνεία. Οπτικοποιώντας μια εκτέλεση που αποτελείται από πολλά όργανα, οι μουσικοί μπορούν διαμοιραστούν την ερμηνευτική τους άποψη.
2. *Σύγκριση μουσικών εκφράσεων σε διαφορετικές ερμηνείες.* Δεν είναι εύκολο να εντοπίσουμε τα σημεία στη σημειογραφία, καθώς και το βαθμό συμφωνίας ή διαφοροποίησης μεταξύ δύο ερμηνειών του ίδιου κομματιού μόνο με την ακοή. Η μετατροπή σε εικόνα της διαφοροποίησης ως προς το χρόνο μιας ερμηνείας, βοηθάει στον εντοπισμό των σημείων αυτών.
3. *Κατανόηση του μουσικού «στόχου» του ερμηνευτή.* Η πλήρης δομή της μουσικής δεν απεικονίζεται απαραίτητα με λεπτομέρεια στη σημειογραφία. Παρόλα αυτά, είναι πολύ σημαντικό για τον ερμηνευτή να αποδώσει το νόημα της ερμηνείας του και για τον ακροατή να καταλάβει τον μουσικό στόχο του ερμηνευτή. Με την οπτικοποίηση μπορούμε να εμβαθύνουμε την κατανόησή μας όσο αφορά τη μουσική ερμηνεία.
4. *Εξόρυξη δεδομένων και κατηγοριοποίηση μουσικών βάσεων δεδομένων σύμφωνα με τη «διάθεση» (mood) του κομματιού.* Με την οπτικοποίηση της μουσικής έκφρασης, μπορούμε να πετύχουμε την ανάκτηση μουσικής πληροφορίας σύμφωνα με τη διάθεση του κομματιού, χωρίς τη χρήση ετικετών (tags) και μετά-δεδομένων τα οποία συχνά περιέχουν υποκειμενικούς χαρακτηρισμούς. Μπορούμε έτσι να έχουμε πρόσβαση σε μία μουσική βάση δεδομένων με κριτήριο την επιθυμητή ατμόσφαιρα.

Ενδεικτικά, για την περίπτωση της κατηγοριοποίησης σύμφωνα με την διάθεση (mood), οι Feng, Zhuang και Pan (2003) στην έρευνά τους κατηγοριοποίησαν την μουσική με χρήση νευρωνικών δικτύων και κριτήρια το tempo και την άρθρωση, ενώ οι Laurier και Herrera (2008) δημιούργησαν την εφαρμογή Mood Cloud που κατηγοριοποιεί τη μουσική χρησιμοποιώντας πέντε κατηγορίες διάθεσης (happy, sad, aggressive, relax και party – **Εικόνα 4-1**). Για την κατηγοριοποίηση χρησιμοποίησαν πολλαπλά ηχητικά χαρακτηριστικά (π.χ. MFCC, Spectral Centroid, Onset rate κ.α.) των οποίων ο μέσος όρος και η τυπική απόκλιση χρησιμοποιήθηκαν ως είσοδο για την εκπαίδευση ενός ταξινομητή που βασίζονταν στον αλγόριθμο Support Vector Machine (SVM).

---

<sup>3</sup> Τεχνική 3D γραφικών που χρησιμοποιείται για να αποδώσει ασαφή φαινόμενα που δεν είναι εύκολο να δημιουργηθούν με τις κλασσικές rendering τεχνικές (π.χ. σύννεφα, βροχή, φλόγες κ.τ.λ).



**Εικόνα 4-1**  
*Η εφαρμογή Mood Cloud (Laurier and Herrera 2008)*

Από τις περιπτώσεις που αναφέρει ο Hiraga (2002) προκύπτει ότι η οπτικοποίηση της ερμηνείας μπορεί να εισάγει νέους τρόπους πρόσβασης και κατανόησης των δεδομένων της ερμηνείας. Η οπτικοποίηση μας επιτρέπει να εξηγήσουμε στον ακροατή τη μουσική έκφραση της ερμηνείας, όχι με υποκειμενικά επίθετα ή με αόριστες περιγραφές, αλλά με μία αντικειμενική μέθοδο εφαρμόσιμη στα δεδομένα της. Η οπτική παρουσίαση των αποτελεσμάτων εμπεριέχει μια πιο καλλιτεχνική αισθητική καθώς δεν αποτελεί μία απλή παράθεση τιμών και μετρήσεων. Παράλληλα, η οπτικοποίηση μπορεί να συμβάλει στην εξαγωγή άγνωστων μουσικών δομών και ποιοτικών χαρακτηριστικών εφαρμόζοντας μεθόδους ανάλυσης που έχουν καθιερωθεί από την γενικότερη έρευνα στην οπτικοποίηση πληροφορίας.

Παρακάτω παρουσιάζονται αναλυτικά ορισμένες εφαρμοσμένες έρευνες πάνω στην οπτικοποίηση της μουσικής.

#### 4.2 Animus

Η συγκεκριμένη εφαρμογή επιτρέπει τη διάδραση μεταξύ ενός μουσικού και ενός εικονικού χαρακτήρα σε πραγματικό χρόνο. Ο εικονικός χαρακτήρας ανταποκρίνεται ανάλογα με τη μουσική που «ακούει» (**Εικόνα 4-2**) και οπτικοποιεί την αντίληψη της μουσικής με ανάλογη ανταπόκριση στη συμπεριφορά του. Η εφαρμογή έχει βασιστεί στην έρευνα του Cooke (1963) που αναφέρεται στα συναισθήματα που βρίσκονται ενσωματωμένα στη μουσική. Τα συναισθήματα είναι εγγενή στη μουσική χωρίς να είναι ένα τεχνικό μέρος της αλλά κάτι πιο αόριστο και απροσδιόριστο. Ο Cooke χρησιμοποιώντας τη δυτική σημειογραφία ως μέσο αναφοράς, πραγματοποίησε υποθέσεις για τη σχέση ανάμεσα στις μουσικές δομές και τον συναισθηματικό στόχο του συνθέτη



**Εικόνα 4-2**

*Χρήστης του Animus αλληλοεπιδρά με έναν εικονικό χαρακτήρα (Taylor, Boulanger and Torres 2006)*

Βασιζόμενοι στην έρευνα του Cooke, οι Taylor, Boulanger και Torres (2006) δημιούργησαν μία οπτική απόκριση στα μουσικά συναισθήματα με χρήση ενός πραγματικού μεγέθους εικονικού χαρακτήρα τον οποίο ονόμασαν ANIMUS. Ο χαρακτήρας δέχεται τη μουσική ως είσοδο και σε πραγματικό χρόνο την κωδικοποιεί σύμφωνα με προκαθορισμένα μοντέλα συσχετίζοντας μουσικά χαρακτηριστικά (features) με συναισθηματικές καταστάσεις, σε συμφωνία με την έρευνα του Cooke. Ο εικονικός χαρακτήρας μπορεί για παράδειγμα να αποδώσει χαρούμενα συναισθήματα σε κάποια μελωδία που θεωρεί «ευχάριστη» ή συναισθήματα φόβου σε μία ακολουθία συγχορδιών που τη θεωρεί «απειλητική». Ο χαρακτήρας θα πρέπει να είναι αρκετά ευέλικτος ώστε να παρουσιάζει συναισθηματική απόκριση σε μουσικά χαρακτηριστικά εντός μιας μελωδίας. Για παράδειγμα, αν μια μελωδία περιέχει πολλά διαστήματα 3<sup>ης</sup> μικρής που παραπέμπουν σε ελάσσονα (minore) κλίμακα, λαμβάνεται ως «δραματική» ενώ αν περιέχει πολλά διαστήματα 3<sup>ης</sup> μεγάλης (majore) λαμβάνεται ως «χαρούμενη» (Taylor, Boulanger and Torres 2006)

### 4.3 Οπτικοποίηση με σύστημα σωματιδίων

Τα τελευταία χρόνια αρκετές έρευνες έχουν ασχοληθεί με τη χρήση του τρισδιάστατου συστήματος γραφικών των σωματιδίων (particle systems). Σε ένα τέτοιο σύστημα, ένα φαινόμενο κατασκευάζεται από ένα σύνολο σωματιδίων απλού σχήματος που κάθε ένα έχει συγκεκριμένο κύκλο ζωής. Κάθε σωματίδιο έχει ένα σύνολο από αρχικές ιδιότητες που καθορίζουν και την κίνησή του, όπως η θέση, το σχήμα, το μέγεθος, την ταχύτητα, την κατεύθυνση, το χρώμα και τη διάρκεια ζωής. Πρόκειται επομένως για ένα δυναμικό σύστημα όπου νέα σωματίδια δημιουργούνται συνεχώς, ενώ παλαιότερα που έκλεισαν τον κύκλο ζωής τους εξαφανίζονται. Για τον υπολογισμό του κάθε καρέ (frame) σε μία ακολουθία σωματιδίων, εκτελούνται τα παρακάτω βήματα (Reeves 1983):

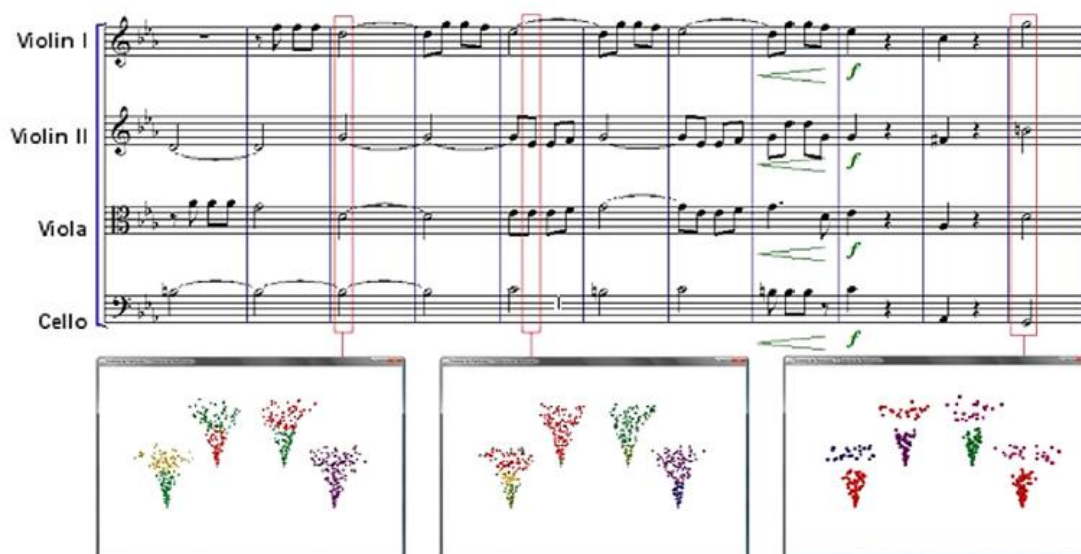
1. Νέα σωματίδια δημιουργούνται στο σύστημα.
2. Σε κάθε νέο σωματίδιο αποδίδονται τα μοναδικά του χαρακτηριστικά.
3. Όσα σωματίδια ολοκλήρωσαν τον κύκλο ζωής τους εξαφανίζονται.

4. Τα σωματίδια που παραμένουν στο σύστημα κινούνται και μετασχηματίζονται σύμφωνα με τα δυναμικά χαρακτηριστικά τους.
5. Ένα στιγμιότυπο των «ζωντανών» σωματιδίων αποθηκεύεται ως ένα καρτέ.

Οι Fonteles, Rodrigues και Basso (2013) χρησιμοποίησαν την τεχνική των particles για να δημιουργήσουν μία οπτικοποίηση της μουσικής, σε περιβάλλον OpenGL 3D, που έχει τη μορφή ενός «μουσικού σιντριβανιού». Στην εφαρμογή τους, το αρχικό χρώμα των σωματιδίων καθορίζεται ανάλογα με τη νότα στην οποία αντιστοιχούν, οδηγώντας επομένως σε δώδεκα διαφορετικούς χρωματισμούς. Το τρέχον χρώμα του σωματιδίου (σε κάθε καρτέ) υπολογίζεται με μία γραμμική προσέγγιση του αρχικού χρώματος και του χρώματος στο τέλος της ζωής του. Γενικά, η «ηλικία» και η διάρκεια ζωής του κάθε σωματιδίου χρησιμοποιούνται από το σύστημα σε διάφορες παραμέτρους (μέγεθος, χρώμα, θέση και ταχύτητα). Κατά τη διάρκεια του animation, τα σωματίδια κινούνται ανοδικά και απλώνονται με κάποια κλίση, δημιουργώντας μια εικόνα με μορφή σιντριβανιού. Ένας παράγοντας βαρύτητας έχει οριστεί και επηρεάζει το κάθε σωματίδιο ανάλογα με το μέγεθός του.

Οι Fonteles, Rodrigues και Basso χρησιμοποίησαν MIDI αρχεία ως είσοδο στην εφαρμογή, από τα οποία εξάγουν τρία χαρακτηριστικά για κάθε νότα: την τονικότητα, την ένταση και τη χροιά. Η τονικότητα αντιστοιχίζεται στο χρώμα του κάθε σωματιδίου (με τα σκούρα χρώματα να αντιστοιχούν στις πιο χαμηλές νότες και τα ανοιχτά στις πιο υψηλές), η ένταση καθορίζει το μέγεθος των σωματιδίων ενώ η χροιά δημιουργεί μία γεννήτρια σωματιδίων για κάθε όργανο που συμμετέχει. Τα note-on και note-off μηνύματα του MIDI καθορίζουν την αρχή και το τέλος λειτουργίας μία γεννήτριας σωματιδίων.

Η εφαρμογή δοκιμάστηκε σε μουσική για κουαρτέτο εγχόρδων, με το κάθε όργανο να αντιστοιχεί σε μία «γεννήτρια σωματιδίων» (**Εικόνα 4-3**). Δοκιμάστηκε σε τρεις ομάδες ακροατών διαβαθμισμένης μουσικής παιδείας. Η πλειοψηφία των ακροατών κατανόησε μέσω της απεικόνισης τις έννοιες της χροιάς, της έντασης και της τονικότητας. Επίσης οι ακροατές μπόρεσαν να αντιληφθούν, μέσω των γραφικών, επαναλαμβανόμενα μελωδικά μοτίβα (μουσικά θέματα) κατά τη διάρκεια του μουσικού έργου.

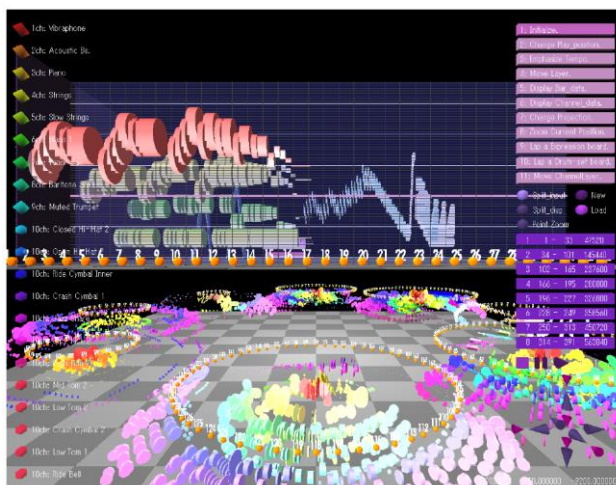


**Εικόνα 4-3**

*Σύγκριση τριών σημείων της παρτιτούρας με τα αντίστοιχα frames που δημιουργήθηκαν με την τεχνική των particles (Fonteles, Rodrigues and Basso 2013).*

#### 4.4 Comp-i

Βασικό στοιχείο στην οπτικοποίηση της μουσικής είναι να συμπεριλάβει τα στοιχεία εκείνα που έχουν νόημα για τη μουσική αντίληψη του ακροατή, όπως η άρθρωση, το tempo και οι δυναμικές. Για το σκοπό αυτό οι Hiraga, Mizaki και Fujishiro (2002) δημιούργησαν ένα τρισδιάστατο εικονικό χώρο τον οποίο ονόμασαν comp-i (**Εικόνα 4-4**), όπου η αναπαράσταση της μουσικής πληροφορίας γίνεται με χρήση κυλίνδρων, καθένας από τους οποίους αντιστοιχεί σε μία νότα.



**Εικόνα 4-4**  
Το περιβάλλον comp-i

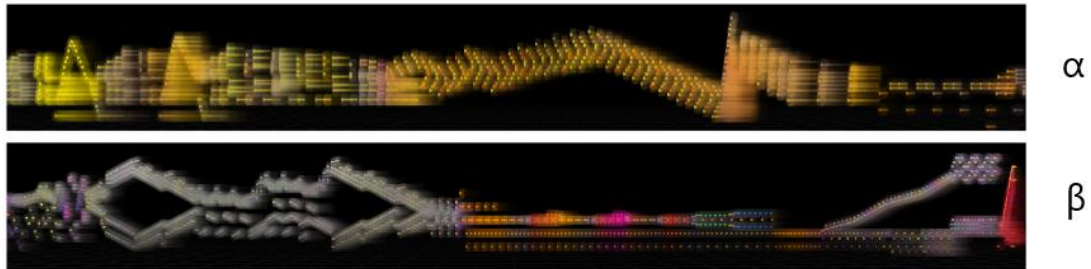
Το περιβάλλον δέχεται ως είσοδο MIDI αρχεία, με τα διαφορετικά κανάλια του MIDI να αναπαρίστανται στον άξονα z. Η τονικότητα αναπαρίσταται στον άξονα y, η ένταση από τη διάμετρο του κυλίνδρου και το tempo από την απόχρωσή του. Καθώς η μουσική αναπαράγεται, ένας δρομέας κάθετος στο άξονα x μετακινείται από αριστερά προς τα δεξιά δείχνοντας στο ακροατή την αντιστοιχία των απεικονίσεων με τη μουσική που ακούει. Ο ακροατής έχει τη δυνατότητα να αλλάξει την οπτική γωνία (πρόσοψη, κάτοψη κτλ.) ή να ζουμάρει σε συγκεκριμένα σημεία.

#### 4.5 Οπτικοποίηση της αρμονίας

Η έννοια της αρμονίας αποτελεί έναν από τους πιο σημαντικούς παράγοντες της μουσικής. Νότες που παίζονται ταυτόχρονα δεν τις αντιλαμβανόμαστε ως ξεχωριστές οντότητες αλλά ως ένα σύνολο. Επίσης η ανθρώπινη αντίληψη της αρμονίας δεν επικεντρώνεται σε συγκεκριμένη χρονική στιγμή ή νότα. Λαμβάνοντας υπόψιν τους τα παραπάνω, οι Klemenc κ.α. (2011) δημιούργησαν μία οπτικοποίηση στην οποία αναπαριστούν αρμονικούς παράγοντες με χρώματα (**Εικόνα 4-5**). Για το σκοπό αυτό, το μουσικό έργο χωρίζεται σε μουσικά τμήματα, και σε κάθε τμήμα αποδίδεται μία απόχρωση.

Η εφαρμογή, για την απόδοση του χρώματος, συνδυάζει ένα κύκλο χρωματικών αποχρώσεων με δύο μουσικούς χρωματικούς κύκλους (**Εικόνα 1-2**) που έχουν περιστραφεί μεταξύ τους έτσι ώστε το επόμενο σημείο για κάθε τονικότητα να αναπαριστά διάστημα 3<sup>ης</sup> μεγάλης (η απεικόνιση αυτή συχνά αποκαλείται και circle of thirds). Οι δώδεκα τονικότητες της χρωματικής κλίμακας αντιστοιχούν σε δώδεκα διανύσματα που ξεκινούν από το κέντρο του κύκλου προς την αντίστοιχη τονική κλάση. Για ένα συνδυασμό νοτών, τα διανύσματα προστίθενται και η κατεύθυνση του διανύσματος που προκύπτει δείχνει την απόχρωση (hue)

ενώ το μήκος του τον κορεσμό (saturation). Καθώς οι ματζόρε κλίμακες συνήθως γίνονται αντιληπτές ως πιο χαρούμενες ενώ οι μινόρε ως πιο μελαγχολικές, έχουν αντιστοιχηθεί σε πιο έντονα και πιο σκοτεινά χρώματα αντίστοιχα (Klemenc et.al 2011)



**Εικόνα 4-5**

Οπτικοποίηση για τμήματα από τα μουσικά έργα α) *Waltz of the Flowers* του *Tchaikovsky* και β) *Toccata in D minor, Op. 11* του *Prokofiev* (Klemenc et.al 2011).

#### 4.6 Οπτικοποίηση με χρήση MAX/MSP και Flash

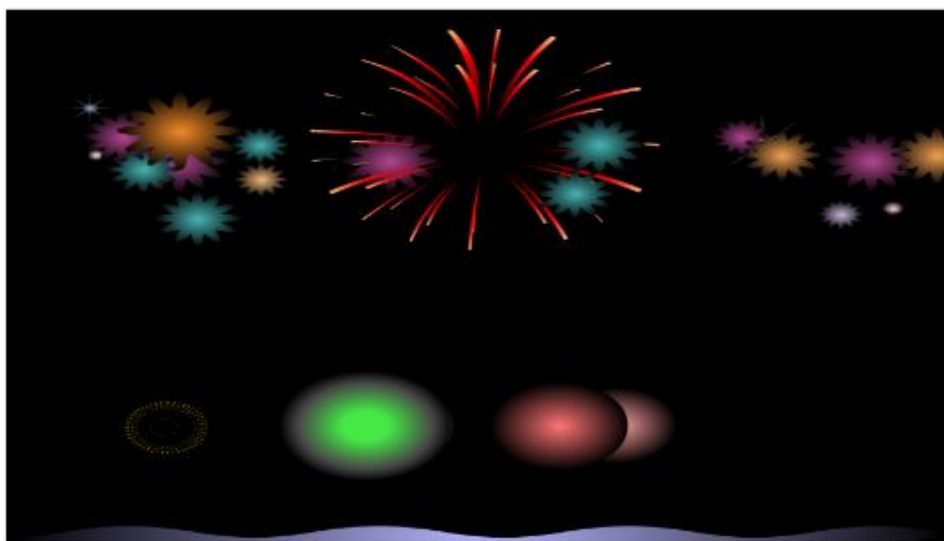
Οι Nanayakkara κ.α. (2007) πρότειναν ένα σύστημα οπτικής αναπαράστασης της μουσικής που βασίζεται στη γλώσσα μουσικού προγραμματισμού MAX/MSP και στο Flash της Adobe. Το σύστημα δέχεται δεδομένα σε μορφή MIDI, είτε από αρχείο είτε σε πραγματικό χρόνο μέσω MIDI keyboard και δημιουργεί απεικονίσεις που αναπαριστούν την τονικότητα, τη διάρκεια της νότας, την ένταση (velocity) του πλήκτρου και το είδος του οργάνου. Η αντιστοίχιση ηχητικών και οπτικών αντικειμένων έγινε με βάση κάποιες γενικές αντιληπτικές αρχές. Για παράδειγμα, καθώς τα μικρότερα όργανα συνήθως παράγουν πιο οξείς ήχους και τα μεγαλύτερα όργανα πιο μπάσους ήχους, οι υψηλές τονικότητες αντιστοιχίζονται σε μικρότερα αντικείμενα και αντίστροφα. Επίσης το πλάτος συσχετίστηκε με την φωτεινότητα καθώς και τα δύο αποτελούν μέτρο της έντασης στο ηχητικό και οπτικό πεδίο αντίστοιχα. Τέλος η αλλαγή μουσικής κλίμακας κατά τη διάρκεια του κομματιού συνοδεύεται από αλλαγή του χρώματος παρασκηνίου. Στην **Εικόνα 4-6** φαίνεται ένα παράδειγμα αντιστοίχισης μεταξύ διαφόρων οργάνων και ήχων με κάποια σχηματική μορφή, θέση και εφέ κίνησης.

Instrument	Visual Effect	Instrument	Visual Effect
<b>Bass Drum</b> Effect: Wave (Pulsating up and down) Colour: Blue Screen position: Bottom edge of the screen		<b>Closed Hi-Hat</b> Effect: Bursting effect Colour: Yellow Screen position: Lower-Left of the screen	
<b>Snare Drum</b> Effect: Sphere fading away Colour: Red Screen position: Lower-Middle of the screen		<b>Ride Cymbal</b> Effect: Star falling down Colour: Silver Screen position: Top-Middle of the screen	
<b>Hi Tom</b> Effect: Sphere fading away Colour: Gold Screen position: Lower-Right of the screen		<b>Crash Cymbal</b> Effect: Firework effect Colour: Red Screen position: High-Middle of the screen	
<b>Hi Bongo</b> Effect: Sphere fading away Colour: Green Screen position: Lower-left of the screen		<b>Piano note</b> Effect: Star-like object fading away Colour and screen position: depends on the note class (C, C# ect)	

**Εικόνα 4-6**

Αντιστοίχιση μεταξύ διάφορων οργάνων και μορφών απεικόνισης (Nanayakkara et.al 2007)

Η εφαρμογή αποτελείται από τρία τμήματα, της επεξεργασίας, του server και της εμφάνισης. Στο πρώτο τμήμα η γλώσσα MAX/MSP χρησιμοποιείται για να επεξεργαστεί το MIDI σήμα και να εξάγει την τονικότητα, το velocity, το είδος του οργάνου και πιθανές αλλαγές κλίμακας. Οι πληροφορίες που εξάγονται περνάνε στο Adobe Flash Player μέσω του αντικειμένου *flashserver* της MAX/MSP. Η σύνδεση γίνεται με χρήση TCP/IP πρωτοκόλλου, γεγονός που επιτρέπει τα δεδομένα να μεταφέρονται μέσω τοπικού δικτύου ή του Internet. Κάθε πακέτο δεδομένων που λαμβάνεται από τον flash player δημιουργεί ένα τμήμα ταινίας – animation μέσα στο συνολικό animation που προβάλλεται. Το σχήμα, το χρώμα και η θέση του τμήματος αυτού μέσα στο σύνολο εξαρτάται από τη μουσική πληροφορία που λαμβάνεται. Στην **Εικόνα 4-7** φαίνεται ένα στιγμιότυπο της ολοκληρωμένης οπτικοποίησης.



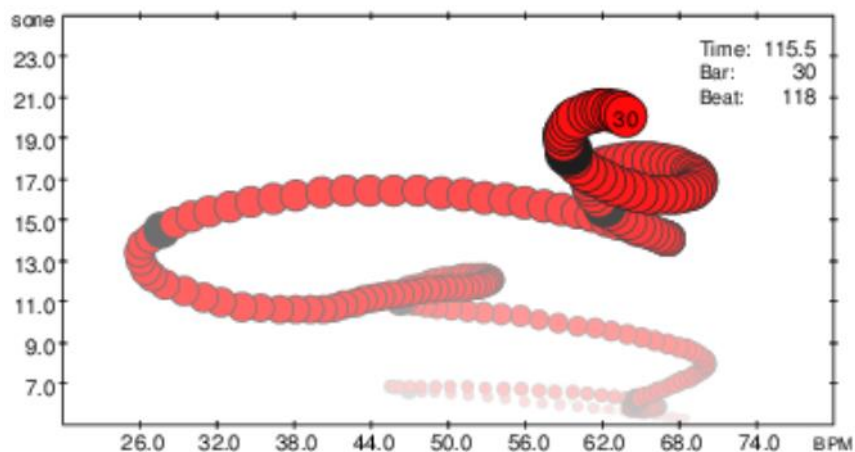
*Εικόνα 4-7*

*Στιγμιότυπο της μουσικής οπτικοποίησης (Nanayakkara et.al 2007)*

#### 4.7 Performance Worm

Οι παράγοντες του tempo και των δυναμικών αποτελούν τα βασικά στοιχεία μουσικής έκφρασης τα οποία ένας μουσικός τροποποιεί προκειμένου να ερμηνεύσει τα δομικά και συναισθηματικά χαρακτηριστικά της μουσικής. Η εφαρμογή Performance Worm δέχεται ως είσοδο ήχο σε μορφή PCM, ιχνηλατεί τις αλλαγές στο tempo και τις δυναμικές σε πραγματικό χρόνο και απεικονίζει γραφικά την εξέλιξη αυτών των εκφραστικών παραγόντων στο χρόνο. Με τον τρόπο αυτό παρέχει στο χρήστη μία κατανοητή και καλαίσθητη εικόνα των εκφραστικών μοτίβων που χρησιμοποιεί ο ερμηνευτής σε όλη την εξέλιξη του μουσικού έργου (Dixon, Goebel και Widmer 2002).

Η εφαρμογή εισάγει ήχο από αρχείο ή απευθείας από την κάρτα ήχου (π.χ. από μικρόφωνο) και οπτικοποιεί δυναμικά τόσο τις τρέχουσες τιμές όσο και το ιστορικό των δυναμικών και του tempo. Η απεικόνιση φθίνει εκθετικά όσο προχωράμε προς πιο παρελθοντικές στιγμές (**Εικόνα 4-8**). Η εφαρμογή λειτουργεί διαδραστικά καθώς ο χρήστης μπορεί να τροποποιεί παραμέτρους της απεικόνισης (π.χ. να αλλάζει την κλίμακα των αξόνων) ή να επιλέγει μεταξύ των διάφορων υποθέσεων για το tempo, εξαναγκάζοντας τον αλγόριθμο ανίχνευσης να επαναξιολογήσει την συσταδοποίησή του.



Εικόνα 4-8

*Στιγμιότυπο της εφαρμογής Performance Worm. Η θέση της κεφαλής του «σκουληκιού» δίνει το tempo (οριζόντιος άξονας) και τις δυναμικές (κάθετος άξονας) για την τρέχουσα χρονική στιγμή με την «ουρά» του να απεικονίζουν παρελθοντικές στιγμές. (Dixon, S., Goebel, W & Widmer, G. 2005)*

Για τη μέτρηση και ανίχνευση των δυναμικών η εφαρμογή βασίστηκε σε γνωστές και δοκιμασμένες τεχνικές επεξεργασίας σήματος. Για την ανίχνευση του tempo, το εισερχόμενο σήμα αρχικά υπόκειται σε μία προ-επεξεργασία προκειμένου να ανιχνευτούν τα σημεία έναρξης των νοτών (Onsets). Με βάση τα onsets, ένας αλγόριθμος συσταδοποίησης (clustering) ομαδοποιεί τα ενδιάμεσα χρονικά διαστήματα σε συστάδες που αντιστοιχούν σε μουσικές μονάδες χρόνου (τέταρτα, όγδοα, κτλ.). Δημιουργείται έτσι ένα σύνολο υποθέσεων σχετικά με το τοπικό tempo, οι οποίες όμως ανανεώνονται δυναμικά όσο καταφθάνουν νέα δεδομένα. Η υπόθεση που θα αξιολογηθεί από τον αλγόριθμο ως πιο πιθανή δίνεται ως έξοδος. Η παραπάνω διαδικασία υπόκειται και σε μια διαδικασία εξομάλυνσης (smoothing) προκειμένου ο αλγόριθμος να μην επηρεάζεται από τοπικές χρονικές παρεκκλίσεις όπως για παράδειγμα η καθυστέρηση μίας και μόνο νότας.



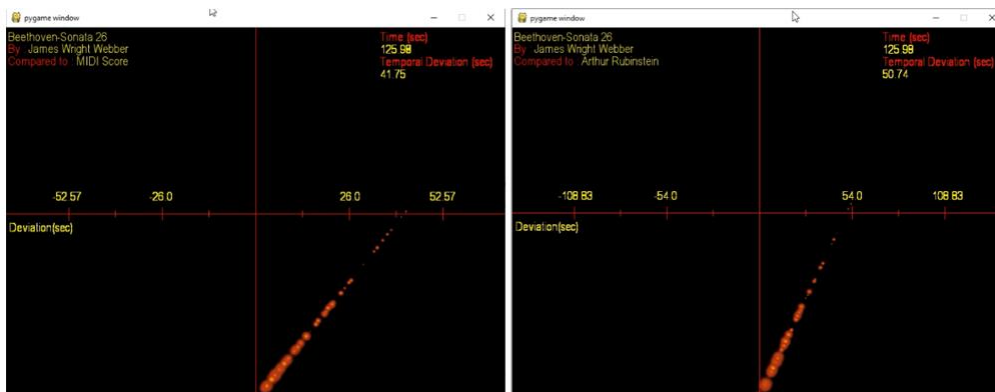
## 5. Υλοποίηση

Στο πλαίσιο της παρούσης εργασίας υλοποιήθηκε η εφαρμογή μουσικού συγχρονισμού και οπτικοποίησης της μουσικής έκφρασης που ονομάστηκε FlameCens, σύμφωνα με όσα παρουσιάστηκαν στο κεφάλαιο 2 και σε αναλογία με αντίστοιχες ερευνητικές εφαρμογές όπως παρουσιάστηκαν στο κεφάλαιο 4. Η εφαρμογή υλοποιήθηκε σε γλώσσα Python, τόσο σε περιβάλλον Jupyter Notebook όσο και σε ανεξάρτητο πρόγραμμα Python. Το περιβάλλον Jupyter Notebook επιλέχθηκε καθώς παρέχει τη δυνατότητα τμηματικής εκτέλεσης και εμφάνισης αποτελεσμάτων του κώδικα σε πραγματικό χρόνο. Το γεγονός αυτό δίνει ιδιαίτερη δυναμική, ευελιξία και δυνατότητα πειραματισμών, τόσο κατά τη φάση της ανάπτυξης όσο και κατά τη φάση της εκσφαλμάτωσης (debugging) της εφαρμογής. Ταυτόχρονα καθιστά τον κώδικα ευκολότερο στη δημοσίευση και κατανόηση από τρίτους.

Η εφαρμογή διατίθεται ως ελεύθερο / ανοικτού κώδικα λογισμικό στο αποθετήριο GitHub στη ηλεκτρονική διεύθυνση <https://github.com/triantafyllou/FlameCens>, τόσο σε έκδοση για περιβάλλον Jupyter Notebook όσο και σε ανεξάρτητο πρόγραμμα Python.

### 5.1 Η εφαρμογή FlameCens

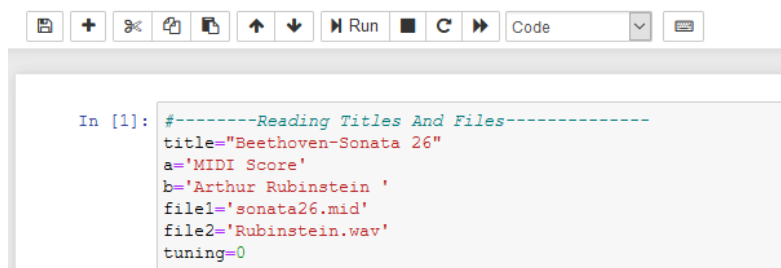
Σκοπός της εφαρμογής είναι η χρονική σύγκριση και αντιστοίχιση μεταξύ μιας πραγματικής εκτέλεσης ενός μουσικού έργου με την αντίστοιχη σημειογραφία (ή με κάποια άλλη εκτέλεση) και ο εντοπισμός έτσι του τρόπου με τον οποίο διαφορετικοί ερμηνευτές χρησιμοποιούν παραλλαγές στο tempo και στο χρονισμό των νωτών για λόγους μουσικής έκφρασης και απόδοσης. Η εφαρμογή αναπαριστά τη μουσική ως μια δυναμικά μεταβαλλόμενη απεικόνιση με μορφή φλόγας. Το μέγεθος της φλόγας σε κάθε χρονική στιγμή αναπαριστά τις δυναμικές του ερμηνευτή και τη διαφορά τους από τις δυναμικές της ερμηνείας με την οποία συγκρίνεται. Η κλίση της φλόγας αναπαριστά τη χρονική απόκλιση από τη σημειογραφία. Όταν το τοπικό tempo της ερμηνείας είναι πιο αργό η πιο γρήγορο σε σχέση με τη παρτιτούρα, η κλίση της «φλόγας» αλλάζει προς τα αριστερά ή προς τα δεξιά αντίστοιχα. Όταν το τοπικό tempo παραμένει ίδιο με το αντίστοιχο της παρτιτούρας, η κλίση της «φλόγας» παραμένει σταθερή. Ο κάθετος άξονας αντιστοιχεί στον χρονισμό της παρτιτούρας (απόκλιση 0) ενώ ο οριζόντιος δείχνει τη χρονική απόκλιση που αντιστοιχεί στην εκάστοτε κλίση της φλόγας (**Εικόνα 5-1**). Η εφαρμογή μπορεί να πραγματοποιήσει σύγκριση είτε μεταξύ μιας εκτέλεσης και της αντίστοιχης παρτιτούρας (**Εικόνα 5-1** αριστερά) είτε μεταξύ δύο εκτελέσεων διαφορετικών ερμηνευτών (**Εικόνα 5-1** δεξιά).



**Εικόνα 5-1**

Η εκτέλεση της σονάτας No.26 του Beethoven από τον πιανίστα James Wright Webber, συγκρινόμενη με: (α) την παρτιτούρα και (β) την εκτέλεση του ίδιου έργου από τον Arthur Rubinstein

Για την εκτέλεση της εφαρμογής σε περιβάλλον Jupyter Notebook (αρχείο FlameCens.ipynb) ο χρήστης καλείται να συμπληρώσει τα πεδία του πρώτου κελιού, όπως φαίνεται στην **Εικόνα 5-2**. Τα πεδία file1 και file2 αντιστοιχούν στα αρχεία που περιέχουν τη σημειογραφία και την προς απεικόνιση εκτέλεση αντίστοιχα. Τα πεδία title, a, b, αποτελούν προαιρετικές πληροφορίες που θα εμφανιστούν στην τελική οπτικοποίηση (Τίτλος μουσικού έργου, ερμηνευτής αρχείου σύγκρισης και ερμηνευτής του προς απεικόνιση αρχείου) και μπορούν να αφεθούν κενά. Το πεδίο tuning χρησιμοποιείται στην περίπτωση που ο χρήστης θέλει να συγκρίνει δύο εκτελέσεις του ίδιου μουσικού έργου σε διαφορετική τονικότητα (διαφορά ημιτονίων).



```
In [1]: #-----Reading Titles And Files-----
title="Beethoven-Sonata 26"
a='MIDI Score'
b='Arthur Rubinstein '
file1='sonata26.mid'
file2='Rubinstein.wav'
tuning=0
```

**Εικόνα 5-2**

*Εισαγωγή δεδομένων για την εφαρμογή FlameCens σε περιβάλλον Jupyter Notebook*

Η εφαρμογή FlamCens αναπτύχθηκε και ως αυτόνομο πρόγραμμα Python3 (αρχείο FlameCens.py). Για την εκτέλεσή του σε γραμμή εντολών, ο χρήστης πρέπει να δώσει την εντολή.

```
python FlameCens.py file1 file2 artist1 artist2 title tuning
```

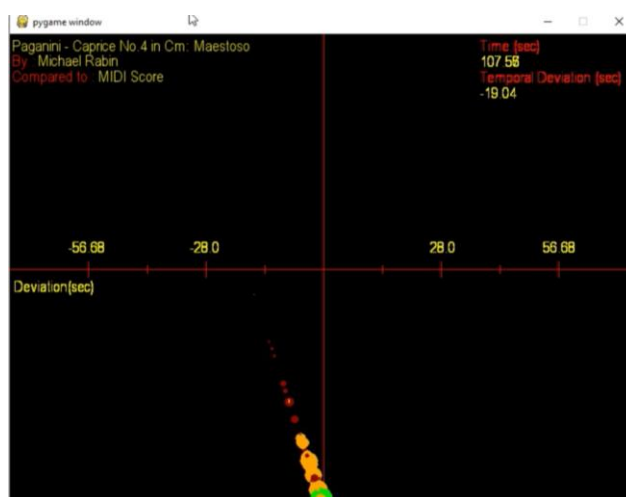
Τα ορίσματα file1 και file2 είναι υποχρεωτικά (αρχείο ήχου με το οποίο θα γίνει η σύγκριση και αρχείο προς οπτικοποίηση αντίστοιχα) και σε περίπτωση που δεν δοθούν το πρόγραμμα εμφανίζει μήνυμα σφάλματος και τερματίζεται. Τα υπόλοιπα ορίσματα είναι προαιρετικά και σε περίπτωση που δεν δοθούν λαμβάνουν προεπιλεγμένες (default) τιμές. Τα προαιρετικά ορίσματα αντιστοιχούν:

- artist1            Καλλιτέχνης του αρχείου file1 (Default τιμή: Unknown)
- artist2            Καλλιτέχνης του αρχείου file2 (Default τιμή: Unknown)
- title                Τίτλος του μουσικού έργου (Default τιμή: Unknown)
- tuning              Διαφορά τονικότητας σε ημιτόνια (Default τιμή: 0)

Η παρτιτούρα εισάγεται κωδικοποιημένη σε μορφή MIDI και μετατρέπεται από το πρόγραμμα σε μορφή .WAV προκειμένου να εξαχθούν τα κατάλληλα χαρακτηριστικά του ήχου (**5.3 Εξαγωγή Ηχητικών Χαρακτηριστικών**). Το πρόγραμμα ελέγχει κατά πόσο έχει δοθεί ως όρισμα κάποιο αρχείο MIDI (καθώς μπορεί να εισάγονται εξ αρχής δύο διαφορετικές εκτελέσεις σε μορφή .WAV) και ανάλογα προχωρά ή όχι στη διαδικασία μετατροπής. Η μετατροπή γίνεται σε πραγματικό χρόνο – το αρχείο MIDI αναπαράγεται ολόκληρο και παράλληλα ηχογραφείται. Η διαδικασία αυτή είναι μεν χρονοβόρα, καθιστά όμως την εφαρμογή ανεξάρτητη από τρίτες εφαρμογές, πλατφόρμες synthesizer και εκδόσεις του λειτουργικού συστήματος. Από τη στιγμή που η μετατροπή ολοκληρωθεί, αποθηκεύεται το αρχείο .WAV που αντιστοιχεί στη συγκεκριμένη παρτιτούρα και ο χρήστης μπορεί να το χρησιμοποιήσει σε μελλοντικές συγκρίσεις άλλων εκτελέσεων του ίδιου μουσικού έργου.

### 5.1.1 Η παραλλαγή FlameCens-Pitch

Παράλληλα με την εφαρμογή FlamceCens αναπτύχθηκε και η παραλλαγή της FlameCens-Pitch (αρχεία FlameCens-Pitch.ipynb και FlameCens-Pitch.py). Στην παραλλαγή αυτή πραγματοποιείται ανίχνευση τονικότητας και η «φλόγα» αλλάζει χρώμα δυναμικά, ανάλογα με τη νότα που εκτελείται τη συγκεκριμένη χρονική στιγμή (**Εικόνα 5-3**). Η παραλλαγή αυτή μπορεί να λειτουργήσει ικανοποιητικά μόνο σε μονοφωνικά μουσικά έργα. Για τη περίπτωση της πολυφωνικής μουσικής η εφαρμογή θα απαιτούσε κάποια τεχνική αναγνώρισης συγχορδιών, γεγονός που δεν αποτέλεσε στόχο της παρούσης εργασίας.



Εικόνα 5-3

Οπτικοποίηση του έργου *Caprice No.4 σε Cm* του *Niccolo Paganini* σε εκτέλεση του βιολιστή *Michael Rabin*

Η ιδέα της αντιστοίχισης των χρωμάτων με νότες είναι πολύ παλιά. Πρώτος ο Isaac Newton τον 18<sup>ο</sup> αιώνα είχε την ιδέα να αντιστοιχίσει τα επτά χρώματα που έβλεπε μέσω ενός πρίσματος με τις επτά νότες της χρωματικής κλίμακας. Η διαδικασία επαναλήφθηκε πολλές φορές στο πέρασμα των αιώνων με καλλιτέχνες όπως ο Castel, ο Bainbridge και ο Rimington να δημιουργούν τις δικές τους αντιστοιχίσεις μεταξύ νοτών και χρωμάτων (Collory 2009). Η αντιστοίχιση που χρησιμοποιήθηκε στη συγκεκριμένη εφαρμογή βασίζεται στην ιδέα του Louis Bertrand Castel (1734) ο οποίος, σε αντίθεση με τον Newton, έλαβε υπόψιν του και τις αλλοιώσεις (υφέσεις και διέσεις) δημιουργώντας μια χρωματική αντιστοίχιση για κάθε μία από τις δώδεκα τονικές κλάσεις. Συγκεκριμένα, για κάθε τονική κλάση, έγινε μία αντιστοίχιση σε τιμή RGB, όπως φαίνεται στον **Πίνακας 6**.

Pitch Class	Color	RGB Value
C	Blue	(0,0,255)
C#	Cyan	(0,255,255)
D	Lime	(0,255,0)
D#	Green	(0,128,0)
E	Yellow	(255,255,0)
F	Pink	(255,105,180)
F#	Gold	(255,215,0)
G	Orange	(255,165,0)
G#	Maroon	(128,0,0)
A	Magenta	(255,0,255)
A#	Indigo	(75,0,130)
B	Purple	(128,0,128)

Πίνακας 6

Αντιστοίχιση των νοτών σε χρώματα

Στις επόμενες παραγράφους ακολουθεί αναλυτική περιγραφή των βημάτων και τεχνικών που ακολουθήθηκαν για την ολοκλήρωση της εφαρμογής οπτικοποίησης καθώς και των βιβλιοθηκών της Python που χρησιμοποιήθηκαν.

## 5.2 Βιβλιοθήκες της Python

Για τις ανάγκες υλοποίησης της εφαρμογής χρησιμοποιήθηκαν οι παρακάτω βιβλιοθήκες (modules) της γλώσσας Python

### 5.2.1 Η βιβλιοθήκη Librosa

Για τις ανάγκες της υλοποίησης έγινε χρήση της βιβλιοθήκης librosa της Python. Πρόκειται για μια βιβλιοθήκη επεξεργασίας ηχητικού και μουσικού σήματος. Η βιβλιοθήκη παρέχει σε υψηλό επίπεδο μία σειρά από υλοποιήσεις συνηθισμένων διαδικασιών που χρησιμοποιούνται στο πεδίο της Ανάκτησης Μουσικής Πληροφορίας (Music Information Retrieval-MIR), ένα πεδίο που συνδυάζει ζητήματα μουσικολογίας, ψηφιακής επεξεργασίας σήματος, μηχανικής μάθησης και ανάκτησης πληροφοριών (McFee et al. 2015). Συγκεκριμένα, από τη βιβλιοθήκη librosa χρησιμοποιήθηκαν οι συναρτήσεις:

- `librosa.load` : φορτώνει ένα αρχείο ήχου ως μια χρονική ακολουθία πραγματικών αριθμών (μονοδιάστατος πίνακας)
- `librosa.stft`: Υπολογίζει το διακριτό μετασχηματισμό Fourier με χρήση παραθυρικής συνάρτησης με μικρό μέγεθος παραθύρου. Επιστρέφει ένα δισδιάστατο μιγαδικό πίνακα  $D$  τέτοιο ώστε:
  - $np.abs(D[f,t])$  είναι το μέγεθος της συχνοτικής ζώνης  $f$  στο χρονικό παράθυρο  $t$  και
  - $np.angle(D[f,t])$  είναι η φάση της συχνοτικής ζώνης  $f$  στο χρονικό παράθυρο  $t$ .
- `librosa.feature.chroma_stft`: Υπολογίζει ένα χρωματογράφημα (chromagram) από μία κυματομορφή ή ένα φασματογράφημα. Επιστρέφει ένα δισδιάστατο πίνακα  $C[n,t]$ , όπου  $n$  ο αριθμός των χρωματικών ζωνών (default τιμή 12) και  $t$  ο αριθμός των χρονικών παραθύρων. Ο πίνακας περιέχει την ενέργεια της κάθε χρωματικής ζώνης για κάθε χρονικό παράθυρο.
- `librosa.feature.chroma_cens`: Υπολογίζει ένα CENS χρωματογράφημα (CENS chromagram) από μία κυματομορφή ή ένα φασματογράφημα. Επιστρέφει ένα πίνακα με τα CENS χρωματικά διανύσματα ως προς το χρόνο.
- `librosa.util.normalize`: Κανονικοποιεί ένα πίνακα στον επιλεγμένο άξονα (γραμμές ή στήλες) σύμφωνα με μία δοθείσα νόρμα. Η συνάρτηση στηρίζει τον ορισμό κάποιου κατωφλίου (thresholding) ώστε γραμμές ή στήλες με νόρμα κάτω από το συγκεκριμένο κατώφλι να αφήνονται μη κανονικοποιημένες, να μηδενίζονται ή να συμπληρώνονται με μη μηδενικές τιμές που κανονικοποιούνται στο 1.
- `librosa.amplitude_to_db`: Μετατρέπει ένα φασματογράφημα πλάτους σε κλίμακα db σύμφωνα με μία δοθείσα ένταση αναφοράς.
- `librosa.time_to_frames`: Μετατρέπει χρονικά σημεία (ή διανύσματα χρόνου) σε αντίστοιχα frames του STFT. Αν δεν δοθεί ρυθμός δειγματοληψίας και μέγεθος hop χρησιμοποιεί της εξ ορισμού τιμές (22050 και 512 αντίστοιχα).

### 5.2.2 Η βιβλιοθήκη PyGame

Για τη δημιουργία της οπτικοποίησης χρησιμοποιήθηκε το cross-platform λογισμικό PyGame, ένα σύνολο αρθρωμάτων (modules) της Python που έχει σχεδιαστεί για τη δημιουργία παιχνιδιών. Περιλαμβάνει βιβλιοθήκες γραφικών και ήχου (Wikipedia.org 2020e). Η Pygame διαθέτει μία πολύ ενεργή κοινότητα χρηστών<sup>4</sup>. Αξιοποιήθηκαν οι δυνατότητες της να αναπαράγει απευθείας αρχεία ήχου MIDI (και άλλων μορφών ήχου) και να παρέχει επιφάνεια γραφικής σχεδίασης η οποία μπορεί να ανανεώνεται σε επιλεγμένο ρυθμό fps. Η υλοποίηση της οπτικοποίησης με την τεχνική των σωματιδίων πραγματοποιήθηκε εξ ολοκλήρου με χρήση των δυνατοτήτων γραφικών της PyGame.

### 5.2.3 Η βιβλιοθήκη Numba

Πέραν από τα πλεονεκτήματα του και τη συνεχώς αυξανόμενη χρήση του στην ανάπτυξη και δημοσίευση επιστημονικών εφαρμογών, το περιβάλλον Jupyter έχει το μειονέκτημα ότι δεν στηρίζει πολυεπεξεργασία και επομένως δεν εκμεταλλεύεται τους πολλαπλούς πυρήνες των επεξεργαστών σύγχρονης αρχιτεκτονικής. Αυτό μπορεί να οδηγήσει την εκτέλεση της εφαρμογής σε σημαντική επιβράδυνση, ειδικά σε αλγορίθμους με μεγάλο υπολογιστικό κόστος, όπως είναι ο DTW. Για το σκοπό αυτό, οι συναρτήσεις υλοποίησης του DTW έχουν οριστεί να εκτελούνται με χρήση του μεταφραστή (compiler) Numba. Ο μεταφραστής Numba εισάγεται σε οποιοδήποτε τμήμα κώδικα Python όπως μία οποιαδήποτε βιβλιοθήκη και μεταφράζει συναρτήσεις της Python σε βελτιστοποιημένο κώδικα μηχανής, κάνοντας χρήση της δομής LLVM. Τμήματα κώδικα που εκτελούνται με χρήση του μεταφραστή Numba προσεγγίζουν ταχύτητες εκτέλεσης αντίστοιχες με αυτές της γλώσσας C. Το τμήμα κώδικα που εκτελείται μέσω Numba θα πρέπει να περιέχει μόνο εντολές Python και numpy (να μη καλεί τρίτες βιβλιοθήκες).

### 5.2.4 Άλλες βιβλιοθήκες

Πέρα από τις παραπάνω βιβλιοθήκες, για την υλοποίηση της εφαρμογής χρησιμοποιήθηκαν επίσης η βιβλιοθήκη scipy για το υπολογισμό των αποστάσεων του πίνακα κόστους, οι βιβλιοθήκες wave και pyaudio για τις ανάγκες της μετατροπής των αρχείων MIDI σε μορφή .WAV και η βιβλιοθήκη OS για λόγους διαχείρισης αρχείων.

## 5.3 Εξαγωγή Ηχητικών Χαρακτηριστικών

Προκειμένου να εξαχθούν τα χαρακτηριστικά του ήχου με βάση τα οποία θα γίνει η χρονική αντιστοίχιση, τα αρχεία της παρτιτούρας μετατρέπονται από μορφή MIDI σε αρχεία ήχου (.wav). Προκειμένου η εφαρμογή να είναι ανεξάρτητη από τρίτες εφαρμογές, δεν χρησιμοποιήθηκε κάποια από τις βιβλιοθήκες Python που πραγματοποιούν απευθείας μετατροπή καθώς αυτές προϋποθέτουν την εγκατάσταση κάποιας εφαρμογής ηχητικής σύνθεσης (synthesizer) όπως το Fluidsynth ή το TiMidity. Η εφαρμογή ελέγχει κατά πόσο το αρχείο σύγκρισης είναι σε μορφή MIDI και, αν αυτό ισχύει, το αρχείο αναπαράγεται μέσω της βιβλιοθήκης PyGame και ηχογραφείται σε μορφή .wav σε πραγματικό χρόνο. Η ηχογράφηση πραγματοποιείται με χρήση ενός αντικειμένου της βιβλιοθήκης PyAudio (Audio Stream) .

---

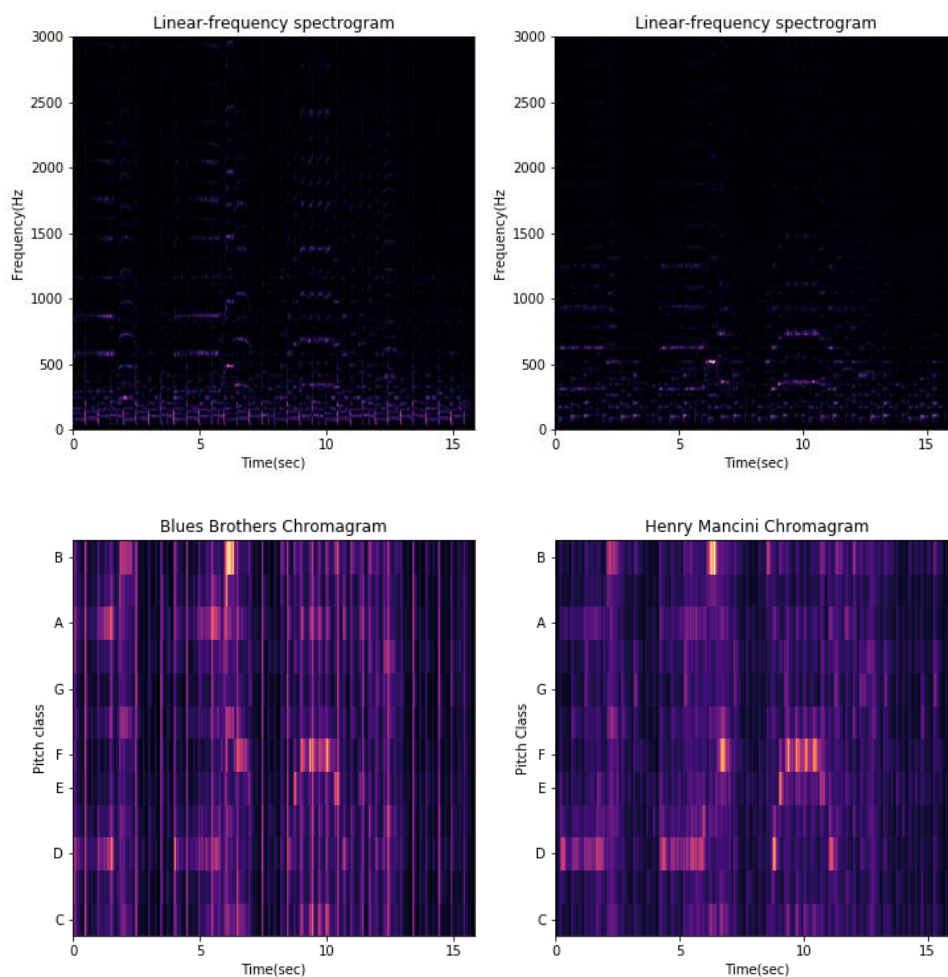
<sup>4</sup> <https://www.pygame.org>

Στη συνέχεια τα αρχεία ήχου φορτώνονται ως μονοδιάστατοι πίνακες πραγματικών αριθμών με χρήση της συνάρτησης *librosa.load*. Η συνάρτηση εξ ορισμού φορτώνει όλα τα αρχεία ήχου ως μονοφωνικά (ως μέσο όρο των δύο στερεοφωνικών καναλιών) και πραγματοποιεί επαναδειγματοληψία σε ρυθμό 22050Hz, γεγονός που εξασφαλίζει ομοιομορφία των δεδομένων σε περίπτωση που τα αρχεία ήχου έχουν διαφορετικά χαρακτηριστικά ποιότητας. Στα αρχεία ήχου που φορτώθηκαν πραγματοποιήθηκε διακριτός μετασχηματισμός Fourier μικρού χρόνου (Short Time Fourier Transform – STFT) με χρήση της συνάρτησης *librosa.stft*. Χρησιμοποιήθηκε η παραθυρική συνάρτηση Hann με μήκος παραθύρου 2048 δείγματα και hop length 512 δείγματα. Στη συνέχεια υλοποιήθηκαν δύο παραλλαγές ηχητικών χαρακτηριστικών, μία με χρήση chroma features τα οποία στη συνέχεια κανονικοποιήθηκαν και συμπιέστηκαν λογαριθμικά, και μία με χρήση CENS features.

Στην πορεία της ανάπτυξης της εφαρμογής, και σε δοκιμές που έγιναν πριν ακόμη φτάσουμε στη φάση της οπτικοποίησης (εμπειρική επαλήθευση του DTW, **5.4.1 Εμπειρική επαλήθευση του DTW**) παρατηρήθηκε ότι η χρήση των Chroma features οδηγούσε σε μία λιγότερο αξιόπιστη αντιστοίχιση από τον αλγόριθμο DTW, ενώ ταυτόχρονα πρόσθετε σημαντική χρονική καθυστέρηση στην εκτέλεση του. Το γεγονός αυτό βρίσκεται σε συμφωνία και με τα συμπεράσματα της υπάρχουσας βιβλιογραφίας και έρευνας όπως αυτή παρουσιάστηκε στο Κεφάλαιο 2, καθώς τα CENS features επηρεάζονται λιγότερο από αλλαγές στη χροιά και στις δυναμικές, ενώ γενικά είναι ένα ηχητικό χαρακτηριστικό υπολογιστικά συμφέρων (βλέπε Ενότητα **2.2.6 CENS Features**). Ταυτόχρονα, τα CENS features ενσωματώνουν και όλες τις υπόλοιπες δυνατότητες των χρωματογραφικών χαρακτηριστικών, όπως η αλλαγή τονικότητας ή η κανονικοποίηση κατά στήλη. Για τους λόγους αυτούε, η υλοποίηση με χρήση των CENS features προκρίθηκε για την τελική μορφή της εφαρμογής. Παρόλα αυτά, για λόγους ερευνητικής τεκμηρίωσης, στην επόμενη ενότητα παρουσιάζεται και ο τρόπος που πραγματοποιήθηκε η εξαγωγή και επεξεργασία των χρωματογραφικών χαρακτηριστικών.

### 5.3.1 Υλοποίηση με chroma features

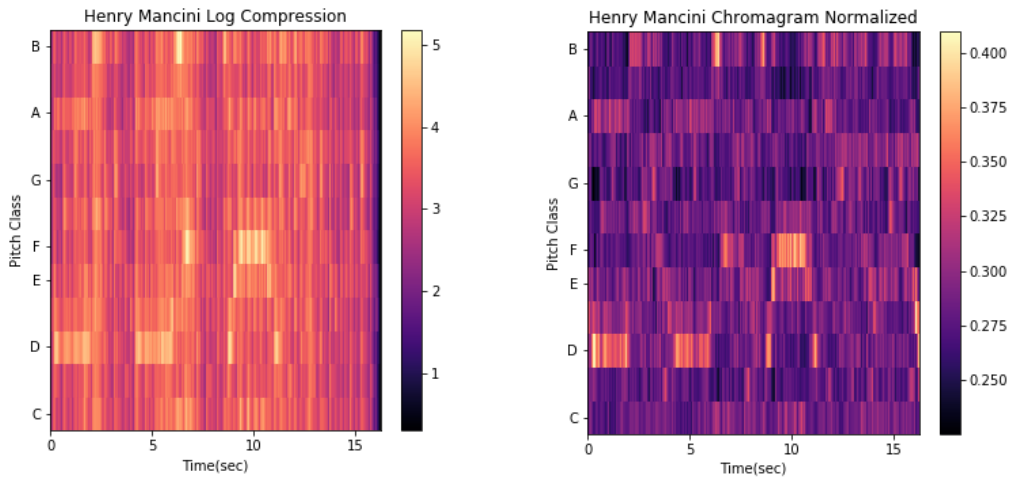
Το χαρακτηριστικό του ήχου που χρησιμοποιήθηκε στην αρχική προσέγγιση είναι το χρωματογράφημα (**Εικόνα 5-4**) καθώς, όπως παρουσιάστηκε στη ενότητα **2.2.2 Chroma Features**, αποδίδει την τονική κλάση του ήχου, ανεξάρτητα από την οκτάβα της και χωρίς να επηρεάζεται σημαντικά από τη χροιά των διαφορετικών μουσικών οργάνων. Για το σκοπό αυτό χρησιμοποιήθηκε η συνάρτηση *librosa.feature.chroma\_stft*, με χρήση της παραθυρικής συνάρτησης Hann με μήκος παραθύρου 2048 δείγματα και hop size 512 δείγματα (default τιμές). Η συνάρτηση στηρίζει επίσης και την αλλαγή τονικότητας (βλέπε ενότητα **2.2.5 Αλλαγή Τονικότητας**) μέσω της παραμέτρου tuning, για την περίπτωση που κάποια ερμηνεία του μουσικού έργου εκτελείται σε διαφορετικό τονικό ύψος. Η διαδικασία της αλλαγής τονικότητας είναι συνηθισμένη κατά την διασκευή – ερμηνεία ενός μουσικού έργου ώστε αυτό να προσαρμοστεί στις απαιτήσεις του εκάστοτε μουσικού (π.χ. προσαρμογή στις φωνητικές δυνατότητες διαφορετικών τραγουδιστών.) Η συνάρτηση επίσης στηρίζει την κανονικοποίηση κατά στήλη των χρωματικών χαρακτηριστικών, η παράμετρος αυτή όμως δεν χρησιμοποιήθηκε, καθώς η κανονικοποίηση έγινε ξεχωριστά στο επόμενο βήμα ακολουθώντας την μεθοδολογία που περιεγράφηκε στις ενότητες **2.2.3 Λογαριθμική Συμπίεση** (Λογαριθμική Συμπίεση) και **2.2.4 Κανονικοποίηση** (Κανονικοποίηση) (Müller 2015).



**Εικόνα 5-4**

Φασματογράφημα και χρωματογράφημα για τα πρώτα 8 μέτρα του *Peter Gunn Theme*, σε εκτελέσεις των *Blues Brothers* και *Henry Mancini* αντίστοιχα. Στο χρωματογράφημα της εκτέλεσης των *Blues Brothers* έχει εφαρμοστεί αλλαγή τονικότητας κατά 1 ημιτόνιο ψηλότερα.

Στο επόμενο βήμα τα χαρακτηριστικά χρώματος περάσαν από διαδικασία λογαριθμικής συμπίεσης και κανονικοποίησης όπως περιγράφονται στις παραγράφους **2.2.3 Λογαριθμική Συμπίεση** και **2.2.4 Κανονικοποίηση**. Η λογαριθμική συμπίεση, με τιμή της σταθεράς  $\gamma=0,5$ , έδωσε ένα πολύ πιο «πλούσιο» χρωματογράφημα (**Εικόνα 5-5**), καθώς πρόκειται για πολυφωνική μουσική όπου η παρουσία πολλών διαφορετικών οργάνων σημαίνει και παρουσία πολλών διαφορετικών αρμονικών. Στη συνέχεια το χρωματογράφημα κανονικοποιήθηκε κατά στήλη ώστε να επιτευχθεί ανεξαρτησία των τιμών των χαρακτηριστικών χρώματος από την ένταση (π.χ. οι χαμηλές δυναμικές της μπασσογραμμής σε σχέση με τις πολύ υψηλότερες δυναμικές των πνευστών). Για την κανονικοποίηση χρησιμοποιήθηκε η συνάρτηση *librosa.util.normalize* όπου στις παραμέτρους της ορίστηκε η χρήση της Ευκλείδειας (ή L2) νόρμας. Επίσης στις παραμέτρους της συνάρτησης ορίστηκε ένα κατώφλι (threshold) της τάξης  $10^{-4}$  ώστε να μη προκύψουν χρωματικές τιμές χωρίς νόημα π.χ. σε περιόδους σιγής. Οι τιμές κάτω από το όριο ορίστηκε να μηδενίζονται κατά την κανονικοποίηση.

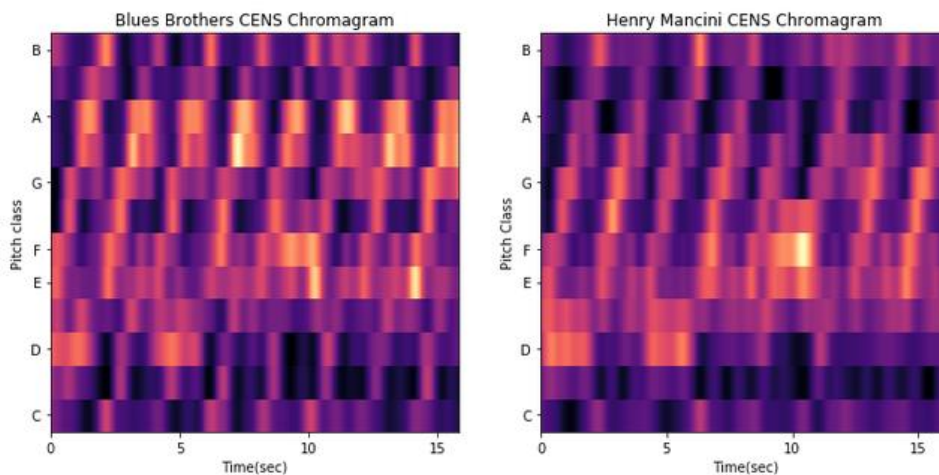


**Εικόνα 5-5**

Χρωματογράφημα για τα 8 πρώτα μέτρα του *Peter Gunn Theme* του Henry Mancini ύστερα από λογαριθμική συμπίεση και κανονικοποίηση αντίστοιχα.

### 5.3.2 Υλοποίηση με χρήση CENS features

Στη δεύτερη υλοποίηση χρησιμοποιήθηκαν τα χαρακτηριστικά CENS (Chroma Energy Normalized Statistics ) αντί των χρωματογραφικών χαρακτηριστικών. Για το σκοπό αυτό χρησιμοποιήθηκε η συνάρτηση `librosa.feature.chroma_cens` με χρήση παραθύρου Hann και μήκος παραθύρου  $\ell=41$ , σε αντιστοιχία με την πρωτότυπη δημοσίευση των CENS (Müller, Kurth and Clausen – 2005a). Επίσης πραγματοποιήθηκε κανονικοποίηση κατά στήλη με χρήση της νόρμας Manhattan. Ο παράγοντας  $\ell$  μπορεί εύκολα να οριστεί σε διαφορετικές τιμές μήκους παραθύρου προκειμένου να γίνουν δοκιμές και να αναδειχθούν ή να αποκρυφτούν παράγοντες που μπορεί να επηρεάσουν τη διαδικασία του μουσικού συγχρονισμού.



**Εικόνα 5-6**

CENS chromagram για τα πρώτα 8 μέτρα του *Peter Gunn Theme*, σε εκτελέσεις των *Blues Brothers* και *Henry Mancini* αντίστοιχα.

Στην **Εικόνα 5-6** παρουσιάζονται τα CENS χρωματογραφήματα που προέκυψαν για τα ίδια δείγματα ήχου που χρησιμοποιήθηκαν τόσο για την **Εικόνα 5-4** όσο και για την **Εικόνα 5-5**.



Συγκρίνοντας οπτικά τις εικόνες, η εξομάλυνση των χρωματικών χαρακτηριστικών είναι εμφανής.

#### 5.4 Υλοποίηση του Dynamic Time Warping

Ο αλγόριθμος που υλοποιήθηκε για τη χρονική αντιστοιχία δύο εκτελέσεων βάσει των ηχητικών τους χαρακτηριστικών είναι εκείνος της Δυναμικής Στρέβλωσης Χρόνου (Dynamic Time Warping).

Ο αλγόριθμος Dynamic Time Warping (DTW) είναι ένας κλασσικός αλγόριθμος σύγκρισης χρονικών ακολουθιών, όπως παρουσιάζεται στην ενότητα **2.3 Ο Αλγόριθμος DTW**. Για την υλοποίηση του DTW αναπτύχθηκαν δύο συναρτήσεις σε γλώσσα Python.

Αρχικά υπολογίζεται ο πίνακας κόστους μεταξύ των διανυσμάτων CENS των δύο εκτελέσεων. Για το σκοπό αυτό γίνεται αντιμετάθεση γραμμών και στηλών (transpose) των δύο πινάκων που προέκυψαν από τα προηγούμενα βήματα. Ο υπολογισμός του τοπικού κόστους γίνεται με χρήση της απόστασης συνημίτονου καθώς αυτή είναι πιο ευέλικτη όταν συγκρίνονται διανύσματα με έντονες διαφορές στις δυναμικές τους και συνήθως προτιμάται στις εφαρμογές μουσικού συγχρονισμού (Ενότητα **2.3.1 Βασική Προσέγγιση για Μουσικό Συγχρονισμό**).

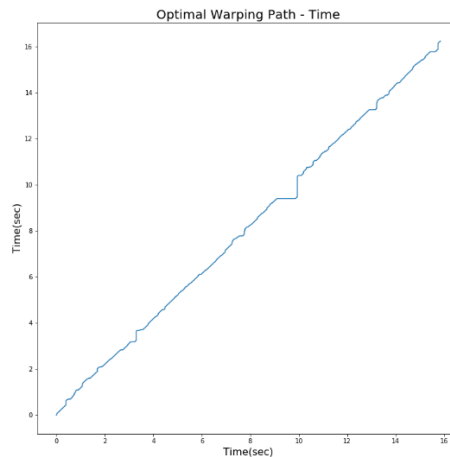
Στη συνέχεια η συνάρτηση `compute_accumulated_cost_matrix` υπολογίζει το συγκεντρωτικό πίνακα κόστους του DTW. Δέχεται ως είσοδο τον πίνακα κόστους που προέκυψε από το προηγούμενο βήμα και επιστρέφει το συγκεντρωτικό πίνακα κόστους. Ο συγκεντρωτικός πίνακας κόστους θα έχει μέγεθος  $N \times M$  αντίστοιχο με αυτό του πίνακα κόστους. Οι τιμές της πρώτης στήλης (index 0) ορίζονται ως το συγκεντρωτικό άθροισμα των επόμενων στηλών και η πρώτη γραμμή αντίστοιχα ορίζεται ως το συγκεντρωτικό άθροισμα των επόμενων γραμμών, σε αντιστοιχία με όσα αναφέρθηκαν στη ενότητα **2.3.4 Αλγόριθμος Δυναμικού Προγραμματισμού**. Τα υπόλοιπα κελιά του πίνακα υπολογίζονται από τη σχέση

$$D(n, m) = C(n, m) + \min \begin{cases} D(n-1, m-1) \\ D(n-1, m) \\ D(n, m-1) \end{cases}$$

για τιμές  $n=1$  ως  $N$  και  $m=1$  ως  $M$ . Προφανώς οι τιμές της πρώτης γραμμής και της πρώτης στήλης ( $n=0$  και  $m=0$ ) που ορίστηκαν εκ των προτέρων, δεν θα μπορούσαν να υπολογιστούν από την παραπάνω σχέση, καθώς αυτό θα οδηγούσε σε μη έγκυρους δείκτες για τον πίνακα  $D$  (index=-1)

Στη συνέχεια η συνάρτηση `compute_optimal_warping_path` υπολογίζει το βέλτιστο Warping Path δεχόμενη ως είσοδο τον συγκεντρωτικό πίνακα κόστους όπως αυτός προκύπτει από τη συνάρτηση `compute_accumulated_cost_matrix`. Ο αλγόριθμος ακολουθεί την ανάστροφη (backtracking) διαδικασία ξεκινώντας από το κελί  $(N, M)$  του πίνακα `table` (όπου  $N$  και  $M$  οι διαστάσεις του πίνακα) και καταλήγοντας στο κελί  $(1, 1)$ . Σε κάθε βήμα της επανάληψης υπολογίζεται το κόστος ανάμεσα στο τρέχον κελί και στα τρία γειτονικά του και επιλέγεται αυτή με τη μικρότερη τιμή κόστους. Η χρήση της ανάστροφης διαδικασίας είναι απαραίτητη για τον υπολογισμό του βέλτιστου Warping Path όπως φάνηκε και από το παράδειγμα της ενότητας **2.3.5 Παράδειγμα Εφαρμογής του Αλγορίθμου DTW**. Η

συνάρτηση επιστρέφει μία λίστα (path) που αποτελείται από ζεύγη τιμών, με κάθε ζεύγος να αντιπροσωπεύει την αντιστοίχιση ανάμεσα σε ένα CENS frame του ηχητικού σήματος X με ένα CENS frame του ηχητικού σήματος Y. Το βέλτιστο Warping Path στη συνέχεια μετατρέπεται και σε αντιστοίχιση χρονικών σημείων (αντί για frames), πολλαπλασιάζοντας το με το μέγεθος του hop και διαιρώντας με τον ρυθμό δειγματοληψίας. Στην **Εικόνα 5-7** φαίνεται η γραφική απεικόνιση ενός βέλτιστου Warping Path σε άξονες πραγματικού χρόνου, όπως αυτή προέκυψε από την παραπάνω διαδικασία.

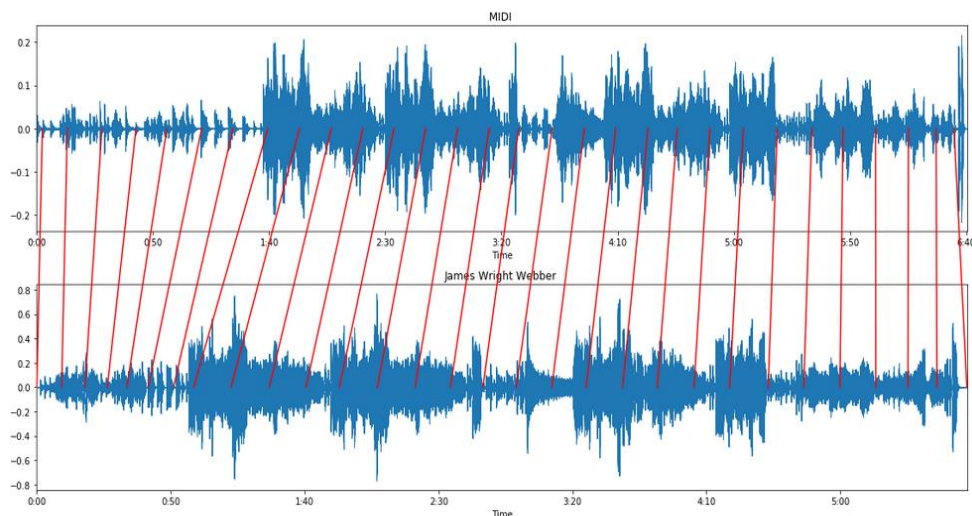


**Εικόνα 5-7**

*Το βέλτιστο Warping Path για δύο ηχητικά σήματα όπως υπολογίστηκε με τη χρήση της απόστασης συνημίτονου*

#### 5.4.1 Εμπειρική επαλήθευση του DTW

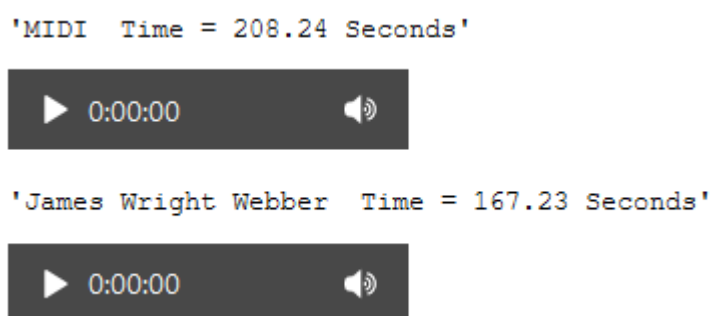
Προτού η υλοποίηση προχωρήσει στη φάση της οπτικοποίησης, και προκειμένου να διαπιστωθεί κατά πόσο η χρονική αντιστοίχιση που επέστρεψε ο αλγόριθμος DTW είναι αληθής, υλοποιήθηκαν δύο τρόποι εμπειρικής επαλήθευσης. Ο πρώτος τρόπος αποτελεί μία γρήγορη οπτική επαλήθευση που παρουσιάζει την αντιστοίχιση μεταξύ των κυματομορφών των δύο ηχητικών σημάτων, όπως φαίνεται στην εικόνα **Εικόνα 5-8**. Συνολικά παρουσιάζονται 30 σημεία αντιστοίχισης σε ίσες χρονικά αποστάσεις μεταξύ τους.



**Εικόνα 5-8**

*Αντιστοίχιση μεταξύ των δύο κυματομορφών όπως αυτή προέκυψε από τον αλγόριθμο DTW*

Η παραπάνω απεικόνιση αποδίδει μία συνολική εικόνα του κατά πόσο ο αλγόριθμος λειτούργησε σωστά και δεν οδήγησε σε αντιστοιχίσεις χωρίς νόημα ή αντιστοιχίσεις με μεγάλα χρονικά άλματα λόγω εκτεταμένων περιόδων σιγής ή επαναλαμβανόμενων θεμάτων. Για να επαληθευτεί ότι η αντιστοίχιση είναι ακριβής δημιουργήθηκε μια μέθοδος παράλληλης αναπαραγωγής, με χρήση της συνάρτησης `display.audio` της iPython, όπως φαίνεται στην **Εικόνα 5-9**. Δοθέντος ενός σημείου του βέλτιστου warping path (αριθμός από 0 ως το πλήθος των στοιχείων του) η αναπαραγωγή ξεκινάει το κάθε ηχητικό σήμα (παρτιτούρα σε MIDI και ερμηνεία προς σύγκριση) από τη χρονική στιγμή του καθενός που αντιστοιχεί στο συγκεκριμένο σημείο. Με τον τρόπο αυτό μπορούμε εύκολα να επαληθεύσουμε εμπειρικά τη σωστή λειτουργία του αλγορίθμου DTW, κάνοντας δοκιμές για διαφορετικά σημεία του warping path και ακούγοντας κατά πόσο οι δύο αναπαραγωγές ξεκινούν από το ίδιο σημείο του μουσικού έργου.



*Εικόνα 5-9*

*Αναπαραγωγή δύο ηχητικών σημάτων από ένα κοινό σημείο αντιστοίχισης*

## 5.5 Οπτικοποίηση της μουσικής έκφρασης

Για την υλοποίηση της εφαρμογής FlameCens, όπως αυτή περιγράφηκε στην παράγραφο **5.1** Η εφαρμογή FlameCens χρησιμοποιήθηκε η τεχνική γραφικών των σωματιδίων (particles) για τη δημιουργία της δυναμικής γραφικής απεικόνισης με μορφή «φλόγας». Συγκεκριμένα, με χρήση της βιβλιοθήκης PyGame, ορίστηκε μία επιφάνεια σχεδίασης με διαστάσεις 800X600 pixels η οποία ανανεώνεται με ρυθμό 60 καρτέ / δευτερόλεπτο. Ο τρόπος που τα σωματίδια λειτουργούν και σχηματίζουν τη φλόγα καθορίζεται από μία συνάρτηση – υποπρόγραμμα που λειτουργεί ως εξής:

Σε κάθε καρτέ εμφανίζονται 720 σωματίδια τα οποία δημιουργούνται μέσω μιας λίστας 720 θέσεων (1 θέση της λίστας για κάθε σωματίδιο) η οποία ανανεώνεται σε κάθε καρτέ. Το κάθε σωματίδιο δημιουργείται σε μία συγκεκριμένη θέση, με ένα συγκεκριμένο μέγεθος και μία συγκεκριμένη αρχική ταχύτητα και σε κάθε ανανέωση το μέγεθός του μειώνεται. Δημιουργείται έτσι για το κάθε σωματίδιο ένας κύκλος ζωής που διαρκεί όσο το μέγεθος του είναι θετικός αριθμός. Όταν το μέγεθός του μηδενιστεί, το σωματίδιο εξαφανίζεται από την οθόνη. Κάθε σωματίδιο έχει μία σειρά από χαρακτηριστικά, επίσης σε μορφή λίστας (το σύνολο των σωματιδίων επομένως αποτελεί μία λίστα από λίστες) τα οποία καθορίζουν τη θέση, την ταχύτητα και το χρώμα του σε κάθε ανανέωση. Οι τιμές των χαρακτηριστικών των σωματιδίων καθορίζονται από τα ηχητικά χαρακτηριστικά που ισχύουν σε κάθε χρονική στιγμή. Συγκεκριμένα το κάθε σωματίδιο αποτελείται από τα παρακάτω στοιχεία (σε αντιστοιχία με τη θέση τους στη λίστα του κάθε σωματιδίου):

0. Θέση X του σωματιδίου. Σε κάθε καρέ πολλαπλασιάζεται επί την τιμή της κατεύθυνσης κατά X (χαρακτηριστικό 4), τον συντελεστή των δυναμικών (χαρακτηριστικό 6) και την ταχύτητα.
1. Θέση Y του σωματιδίου. Σε κάθε καρέ πολλαπλασιάζεται επί την τιμή της κατεύθυνσης κατά Y (χαρακτηριστικό 3) , τον συντελεστή των δυναμικών (χαρακτηριστικό 6) και τη ταχύτητα.
2. Μέγεθος του σωματιδίου. Έχει μία συγκεκριμένη αρχική τιμή και σε κάθε καρέ μειώνεται μέχρι να μηδενιστεί
3. Κατεύθυνση κατά Y. Ορίζεται σταθερή σε -1 (κάθετη στον άξονα των X)
4. Κατεύθυνση κατά X. Σε κάθε καρέ ορίζεται ως ένας συντελεστής χρονικής απόκλισης από την παρτιτούρα, καθορίζοντας έτσι την κλίση της γραφικής απεικόνισης. Η τιμή σε κάθε καρέ ορίζεται από τη χρονική διαφορά ανάμεσα στο συγκεκριμένο σημείο του μουσικού έργου και του αντίστοιχου σημείου της παρτιτούρας.
5. Τύπος. Αν οριστεί 0 τότε για το κάθε σωματίδιο σχεδιάζεται μία έλλειψη με χρήση εντολών σχεδίασης (για χρήση στην παραλλαγή FlameCens-pitch). Αν οριστεί σε 1, τότε χρησιμοποιείται μία έτοιμη εικόνα bitmap με μορφή φλόγας.
6. Συντελεστής δυναμικών. Σε κάθε καρέ ορίζεται ως η διαφορά μεταξύ της έντασης (σε db) του συγκεκριμένου σημείου του μουσικού έργου με την ένταση του αντίστοιχου σημείου του MIDI αρχείου που κωδικοποιεί την παρτιτούρα (ή την διαφορετική ερμηνεία).
7. Χρώμα του σωματιδίου σε τιμή RGB (τρεις τιμές – για χρήση στην παραλλαγή FlameCens-pitch)
8. Δεν χρησιμοποιήθηκε προς στιγμή και αφήνεται για μελλοντική επέκταση. Η ιδέα είναι να κωδικοποιεί κάποια επιρροή στη μετατόπιση κατά X και Y (π.χ. συντελεστή βαρύτητας ή ανέμου) για τη δημιουργία πιο εντυπωσιακού οπτικού αποτελέσματος.

Όπως αναφέρθηκε, η παραλλαγή FlameCens-Pitch της εφαρμογής έχει τη δυνατότητα για δυναμική χρωματική αναπαράσταση της τονικότητας μόνο όμως για τη περίπτωση μονοφωνικής μουσικής (Ενότητα **5.5.2 Αντιστοίχιση τονικότητας με χρώματα**). Η παραλλαγή FlameCens-Pitch χρησιμοποιεί την ίδια συνάρτηση σωματιδίων όπου ο τύπος των σωματιδίων (Χαρακτηριστικό 5) ορίζεται υποχρεωτικά στη τιμή 0 (σχεδίαση έλλειψης από την εφαρμογή και όχι χρήση έτοιμης εικόνας). Η χρήση έτοιμης εικόνας αντί για σχεδίαση γεωμετρικού σχήματος μπορεί να οδηγήσει σε πιο εντυπωσιακά αποτελέσματα γραφικών, καθώς μία εικόνα bitmap μπορεί να περιλαμβάνει σχέδια διαβαθμισμένου χρώματος και απόδοσης (rendering). Μία έτοιμη εικόνα όμως δεν θα μπορούσε να αλλάζει δυναμικά χρώματα ανάλογα με την τονικότητα του ήχου (στη περίπτωση αυτή θα έπρεπε να χρησιμοποιηθούν πολλαπλές εικόνες, μία για κάθε τονικότητα που θα αναπαρασταθεί). Ο πλήρης κώδικας Pythοn της συνάρτησης δημιουργίας της οπτικοποίησης παρουσιάζεται στο **Παράρτημα 2 – Υλοποίηση οπτικοποίησης σε Python**.

### 5.5.1 Αντιστοίχιση χρονικής απόκλισης και δυναμικών

Κατά τη διαδικασία του αλγορίθμου DTW είχαν δημιουργηθεί δύο μορφές του βέλτιστου warping path, μία με αντιστοίχιση μεταξύ των CENS frames των δύο ακολουθιών και μία με αντιστοίχιση μεταξύ χρονικών σημείων (σε δευτερόλεπτα). Κάθε μία από αυτές είναι ένας

πίνακας δύο στηλών, με την πρώτη στήλη να αντιστοιχεί στην παρτιτούρα κωδικοποιημένη σε MIDI μορφή (ή στη διαφορετική ερμηνεία) και τη δεύτερη στην ερμηνεία που θα οπτικοποιηθεί. Για τη δημιουργία του συντελεστή χρονικής απόκλισης έγινε χρήση της δεύτερης μορφής (χρονικά σημεία). Συγκεκριμένα δημιουργήθηκε ένας πίνακας που περιλαμβάνει τις διαφορές ανάμεσα στα αντίστοιχα στοιχεία της δεύτερης και της πρώτης στήλης, παρακάμπτοντας τυχόν διπλές εγγραφές της πρώτης στήλης (είναι πιθανό ο αλγόριθμος DTW να αντιστοίχισε περισσότερες από μία τιμές της δεύτερης ακολουθίας σε κάποιο στοιχείο της πρώτης ακολουθίας). Αν δεν παρακάμπτονταν οι διπλές εγγραφές θα δημιουργούταν ένας πίνακας που θα περιλάμβανε αόριστες πολλαπλές διαφορές για το ίδιο χρονικό σημείο και θα είχε διαφορετικό μήκος από τον πίνακα των chroma frames της ίδιας ηχητικής ακολουθίας, δημιουργώντας έτσι πρόβλημα στην αντιστοίχιση της οπτικοποίησης. Η παράλειψη των στοιχείων αυτών δεν επηρεάζει την ακρίβεια της οπτικοποίησης καθώς πρόκειται για συνεχόμενα χρονικά σημεία (κανόνας μονοτονίας του DTW) με χρονική απόσταση μικρότερη του αντιληπτού ορίου. Κάθε στοιχείο του πίνακα που προκύπτει αποτελεί το συντελεστή χρονικής απόκλισης για τα σωματίδια που θα δημιουργηθούν στο αντίστοιχο χρονικό σημείο. Ο πίνακας έχει κανονικοποιηθεί σε τιμές από -1 ως 1, περιορίζοντας έτσι την κλίση της γραφικής απεικόνισης στις 45° δεξιά και αριστερά του χρονικού άξονα.

Για τον υπολογισμό του συντελεστή των δυναμικών, δημιουργείται ένας πίνακας για κάθε ένα από τα αρχεία που συγκρίνονται, ο οποίος αποτελείται από το άθροισμα ανά στήλη των τιμών του διακριτού μετασχηματισμού Fourier (STFT) του ηχητικού σήματος. Οι δύο πίνακες στη συνέχεια μετατρέπονται σε κλίμακα db (decibel), με χρήση της συνάρτησης *amplitude\_to\_db* της βιβλιοθήκης *librosa*, χρησιμοποιώντας ως ένταση αναφοράς τη μέγιστη τιμή και των δύο πινάκων. Ο τελικός πίνακας που δημιουργείται περιλαμβάνει τις διαφορές ανάμεσα στους δύο πίνακες (διαφορά στάθμης σε db). Κάθε στοιχείο του τελικού πίνακα αποτελεί τον συντελεστή των δυναμικών για τα σωματίδια που δημιουργούνται στο αντίστοιχο χρονικό σημείο. Για λόγους βελτιστοποίησης της γραφικής απεικόνισης, καθώς ο συντελεστής δυναμικών αποτελεί πολλαπλασιαστή της ταχύτητας των σωματιδίων, ο πίνακας προσαρμόζεται σε μία κλίμακα από 0 ως 8.

### 5.5.2 Αντιστοίχιση τονικότητας με χρώματα

Για την ανίχνευση της τονικότητας σε κάθε χρονική στιγμή δημιουργήθηκε ένας πίνακας ο οποίος για κάθε frame (στήλη) του χρωματογραφήματος περιέχει την τονική κλάση (γραμμή) με τη μέγιστη τιμή για το συγκεκριμένο frame. Καθώς ο πίνακας ενός χρωματογραφήματος περιέχει τη κατανομή ενέργειας της κάθε χρωματικής κλάσης σε κάθε frame, η χρωματική κλάση με τη μέγιστη τιμή ενέργειας θα αποτελεί τη θεμέλιο νότα, ενώ οι τονικότητες με μικρότερη ενέργεια τις αρμονικές της. Προφανώς αυτό ισχύει μόνο σε περιπτώσεις μονοφωνικής μουσικής, που είναι και η περίπτωση στην οποία η εφαρμογή λειτουργεί σωστά. Σε περιπτώσεις πολυφωνικής μουσικής, περισσότερες από μία νότες είναι παρούσες σε κάθε frame του χρωματογραφήματος, με αποτέλεσμα να μην είναι δυνατός ο διαχωρισμός τους με βάση την κατανομή ενέργειας, καθώς δεν είναι δυνατό να καθοριστεί αν μια τονική κλάση με μικρότερη ενέργεια είναι μία δυναμική της θεμελίου ή μια άλλη θεμέλιος νότα με χαμηλότερες δυναμικές. Η περίπτωση της πολυφωνικής μουσικής θα μπορούσε να καλυφθεί με την ανάπτυξη κάποιας μεθόδου αναγνώρισης συγχορδιών (chord recognition). Στη πρόσφατη βιβλιογραφία έχουν αναπτυχθεί διάφορες μέθοδοι αναγνώρισης

συγχορδιών (πχ. Müller 2015). Η αναγνώριση συγχορδιών δεν αποτέλεσε αντικείμενο της παρούσης εργασίας και αφήνεται για μελλοντική επέκταση.

## 5.6 Καταγραφή Εκφραστικών Αποκλίσεων

Για λόγους μουσικής ανάλυσης και μουσικολογικής έρευνας, η εφαρμογή FlameCens δημιουργεί ένα αρχείο καταγραφής των εκφραστικών αποκλίσεων και συγκεκριμένα της χρονικής απόκλισης και της διαφοράς των δυναμικών. Το αρχείο αποθηκεύεται σε μορφή .csv (Comma Separated Values) και περιλαμβάνει τόσες γραμμές όσες και τα audio frames για τα οποία υπολογίζονται τα ηχητικά χαρακτηριστικά CENS (μία γραμμή για κάθε CENS frame). Σε κάθε γραμμή καταγράφονται τρεις τιμές (**Εικόνα 5-10**), με την πρώτη να αντιστοιχεί το χρονικό σημείο (σε δευτερόλεπτα), τη δεύτερη στη χρονική απόκλιση (σε δευτερόλεπτα) και την τρίτη στην διαφορά δυναμικών (σε dB) την δεδομένη χρονική στιγμή.

1	Time(sec)	Tempo(sec)	Dynamics(db)
9592	222.679	-16.742	21.663
9593	222.703	-16.765	22.854
9594	222.726	-16.765	22.186
9595	222.749	-16.765	22.761
9596	222.772	-16.765	21.052
9597	222.795	-16.765	17.463
9598	222.819	-16.765	17.386
9599	222.842	-16.765	18.842
9600	222.865	-16.765	16.582
9601	222.888	-16.765	14.063
9602	222.912	-16.765	15.445
9603	222.935	-16.765	21.844
9604	222.958	-16.765	27.542
9605	222.981	-16.765	25.582
9606	223.004	-16.765	16.043
9607	223.028	-16.765	13.225
9608	223.051	-16.765	13.525
9609	223.074	-16.765	14.139
9610	223.097	-16.765	15.557

**Εικόνα 5-10**

*Τμήμα του αρχείου καταγραφής από τη Sonata 26 του Beethoven σε εκτέλεση του Arthur Rubinstein, συγκρινόμενη με την παρτιτούρα.*

Το αρχείο κάνει χρήση των πινάκων χρονικής απόκλισης και διαφοράς δυναμικών όπως αυτοί περιγράφονται στην παράγραφο **5.5.1 Αντιστοίχιση χρονικής απόκλισης και δυναμικών**. Συγκεκριμένα χρησιμοποιούνται οι αντίστοιχοι πίνακες με τις τιμές που έχουν πριν την κανονικοποίησή τους. Επισημαίνεται εδώ ότι η κανονικοποίηση των τιμών των αποκλίσεων πραγματοποιείται μονάχα για λόγους οπτικοποίησης με κινούμενη εικόνα (animation). Όλες οι τιμές έχουν στρογγυλοποιηθεί στα τρία δεκαδικά ψηφία.

## 6. Αξιολόγηση

Προκειμένου να αξιολογηθεί η απόδοση της εφαρμογής FlameCens πραγματοποιήθηκε μία σειρά από πειραματικές δοκιμές πάνω σε διάφορα είδη μουσικής (μονοφωνικές και πολυφωνικές εκτελέσεις κλασικών μουσικών έργων, αποσπάσματα σύγχρονης μουσικής, jazz κτλ.) τα αποτελέσματα των οποίων και παρουσιάζονται στο κεφάλαιο αυτό. Στόχος των πειραμάτων είναι να ελεγχθεί κατά πόσο η εφαρμογή αποδίδει ορθά και με κατανοητό τρόπο την αντιστοιχία (mapping) των εκφραστικών αποκλίσεων. Βασική προϋπόθεση για μία σωστή οπτικοποίηση αποτελεί για την εφαρμογή μας η σωστή χρονική αντιστοιχία από τον αλγόριθμο DTW, παράμετρος που ελέγχθηκε σε όλες τις πειραματικές δοκιμές με τη μέθοδο που παρουσιάστηκε στην ενότητα **5.4.1 Εμπειρική επαλήθευση του DTW**.

### 6.1 Αξιολόγηση της μουσικής έκφρασης

Η αξιολόγηση της μουσικής έκφρασης αποτελεί μία πολύπλοκη διαδικασία καθώς εμπλέκονται διάφοροι υποκειμενικοί παράγοντες, τόσο συναισθηματικοί όσο και αισθητικοί. Ταυτόχρονα, η ικανότητα του χρήστη να κατανοήσει εκφραστικές αποκλίσεις μπορεί να ποικίλει σημαντικά ανάλογα με το επίπεδο των μουσικών του γνώσεων.

Στις έρευνες πάνω στην οπτικοποίηση της πληροφορίας γενικότερα, αποτελεί θεμελιώδες ερώτημα ο ορισμός της «αποτελεσματικής οπτικοποίησης» και πως αυτή η αποτελεσματικότητα μπορεί να μετρηθεί. Πολλοί ερευνητές υποστηρίζουν ότι η αποτελεσματικότητα εξαρτάται βασικά από την αντιστοιχία ανάμεσα στην οπτικοποίηση και τα δεδομένα που αναπαριστά (Dastani 2002, Wattenberg & Fisher 2002, Tufte 2001). Μια οπτικοποίηση θεωρείται αποτελεσματική αν η δομή των δεδομένων και η αντιληπτή δομή της οπτικής αναπαράστασης ταυτίζονται. Μια άλλη προσέγγιση που υποστηρίζεται από αρκετούς ερευνητές θεωρεί ότι η αποτελεσματικότητα μιας οπτικοποίησης είναι εξαρτώμενη από το αντικείμενο – στόχο της (Casner 1991, Nowell 2002, Amar & Stasko 2005). Ο σχεδιαστής της οπτικοποίησης πρέπει να έχει στο νου του το συγκεκριμένο στόχο κατά την υλοποίησή της και αποτελεσματική οπτικοποίηση θεωρείται εκείνη που βελτιώνει την επιθυμητή αποτελεσματικότητα του αντικειμένου που οπτικοποιεί. Πολλές ψυχολογικές έρευνες υποστηρίζουν αυτήν τη δεύτερη άποψη καθώς και το γεγονός ότι η αποτελεσματικότητα μίας οπτικοποίησης εξαρτάται από τον βαθμό πολυπλοκότητας του περιεχομένου που οπτικοποιεί (Cox 1999, Freedman & Shah 2002, Vekiri 2002, Sweller 2002). Αντίστοιχα οι Shneiderman και Plaisant στην ερευνά τους (2006) τονίζουν ότι «η οπτικοποίηση της πληροφορίας αποτελεί συνήθως μέρος μιας δημιουργικής διεργασίας που απαιτεί από τους χρήστες να κάνουν υποθέσεις, να αναζητήσουν μοτίβα και εξαιρέσεις και στη συνέχεια να αναθεωρήσουν τις υποθέσεις τους». Οι Shneiderman και Plaisant προτείνουν μια μακροχρόνια και σε βάθος μελέτη και καταγραφή του πως ένα εργαλείο οπτικοποίησης χρησιμοποιείται και αξιοποιείται από εξειδικευμένους χρήστες προκειμένου να αξιολογηθεί η αποτελεσματικότητά του.

Η μουσική έκφραση και η μουσική γενικότερα αποτελεί μία πολύπλοκη και σύνθετη διεργασία που δεν είναι εύκολο να αξιολογηθεί με αυστηρά μετρήσιμα κριτήρια καθώς υπόκειται σε πλήθος υποκειμενικών παραγόντων, (εκφραστικών, καλλιτεχνικών και αισθητικών). Με βάση και όσα αναφέρθηκαν παραπάνω, η μουσική έκφραση ως ένα σύνολο δεδομένων προς οπτικοποίηση και η αποτελεσματικότητα της οπτικοποίησης αυτής είναι δύσκολο να μοντελοποιηθεί σε μία καθορισμένη και μετρήσιμη διαδικασία. Η αντιστοιχία μεταξύ των ηχητικών δεδομένων και του παραγόμενου οπτικού αποτελέσματος αποτελεί

σίγουρα απαραίτητο στοιχείο, δεν εγγυάται όμως από μόνο του την ικανοποιητική κατανόηση και ερμηνεία από τη μεριά του χρήστη.

Η αξιολόγηση μιας εφαρμογής μουσικής έκφρασης από κοινό, συνήθως με χρήση ερωτηματολογίου σε άτομα με κάποιο επίπεδο μουσικής παιδείας, αποτελεί μια πρακτική που έχει χρησιμοποιηθεί από πολλούς ερευνητές (Clarke 1993, Katyose et.al 2012, Williams 2019). Η μέθοδος αυτή, αποτελεί μεν μια ενδεικτική αξιολόγηση της εφαρμογής, υπόκειται όμως στην υποκειμενική κρίση των ακροατών και στην φύση των ερωτήσεων που θα τους υποβληθούν. Σε εφαρμογές οπτικοποίησης της μουσικής είτε πραγματοποιήθηκε αξιολόγηση επίσης με την μέθοδο του ερωτηματολογίου σε κοινό (π.χ. Fonteles et.al 2013) είτε παρέμεινε σε επίπεδο πειραματικής αξιολόγησης και καταγραφής των αποτελεσμάτων των πειραμάτων (Hiraga και Matsuda 2004, Nanayakkara et.al 2007, Klemenc et.al 2011).

Τέλος, σε μία εφαρμογή χρήσιμη είναι και η θεώρησή της με παραμέτρους αλληλεπίδρασης ανθρώπου – υπολογιστή (Human-Computer Interaction – HCI), όπως η ευχρηστία της και ο χρόνος απόκρισης της εφαρμογής.

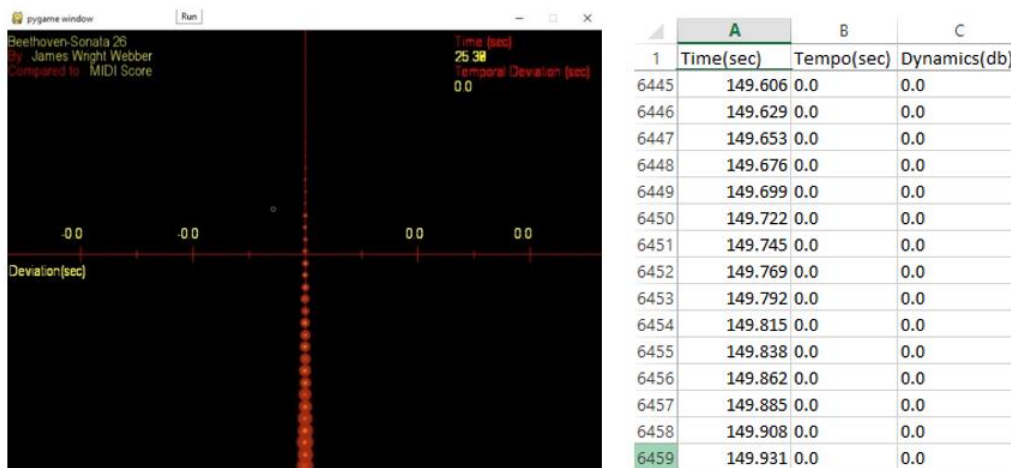
## 6.2 Πειραματική Αξιολόγηση

Προκειμένου να αξιολογηθεί η ορθή λειτουργία της εφαρμογής FlameCens πραγματοποιήθηκαν πειραματικές δοκιμές σε διάφορα μουσικά έργα όπου έγιναν συγκρίσεις τόσο μεταξύ πραγματικών ερμηνειών και της αντίστοιχης παρτιτούρας όσο και ταυτόσημων ερμηνειών του ίδιου μουσικού έργου από διαφορετικούς ερμηνευτές. Χρησιμοποιήθηκαν έργα κλασικής μουσικής τόσο πολυφωνικά (συνθέσεις για πιάνο) όσο και μονοφωνικά (συνθέσεις για σόλο βιολί). Επίσης έγιναν πειράματα και πάνω σε αποσπάσματα σύγχρονης rock και jazz μουσικής. Αξίζει να σημειωθεί εδώ ότι η σύγκριση ολόκληρων jazz κομματιών είναι δύσκολο να πραγματοποιηθεί, καθώς αυτά δεν βασίζονται σε αναπαραγωγή συγκεκριμένης παρτιτούρας, ενώ διαφορετικές διασκευές περιλαμβάνουν μεγάλο βαθμό αυτοσχεδιασμού αλλά και παραλλαγές στην ίδια τη δομή του κομματιού.

### 6.2.1 Αντιστοίχιση ηχητικού αρχείου με τον εαυτό του

Ως ένα πρώτο βήμα για την αξιολόγηση της ορθής λειτουργίας της εφαρμογής, ένα ηχητικό αρχείο συγκρίθηκε με τον εαυτό του. Εφόσον η εφαρμογή λειτουργεί σωστά, μία τέτοια δοκιμή θα πρέπει να μην καταγράφει καμία απόκλιση, τόσο στο πεδίο του χρόνου όσο και στις διαφορές των δυναμικών. Για το σκοπό αυτό χρησιμοποιήθηκε η ερμηνεία του James Wright Webber πάνω στη Sonata No.26 σε μι-ύφεση του Ludwig van Beethoven. Τα αποτελέσματα που προέκυψαν ήταν σε πλήρη αντιστοιχία με το επιθυμητό αποτέλεσμα. Κατά την οπτικοποίηση, η κλίση της φλόγας παραμένει σταθερή και πάνω στον κάθετο άξονα, αποτυπώνοντας έτσι τη μηδενική χρονική απόκλιση σε όλη τη διάρκεια του κομματιού. Ταυτόχρονα, το μέγεθος της φλόγας παραμένει σταθερό αποτυπώνοντας έτσι μηδενική διαφορά στις δυναμικές. Αντίστοιχα, στο αρχείο καταγραφής, οι τιμές της χρονικής απόκλισης και της διαφοράς των δυναμικών είναι μηδενικές για όλα τα χρονικά σημεία που καταγράφονται. Στην **Εικόνα 6-1** φαίνεται ένα στιγμιότυπο της οπτικοποίησης που προέκυψε από την παραπάνω δοκιμή καθώς και τμήμα του αρχείου καταγραφής.





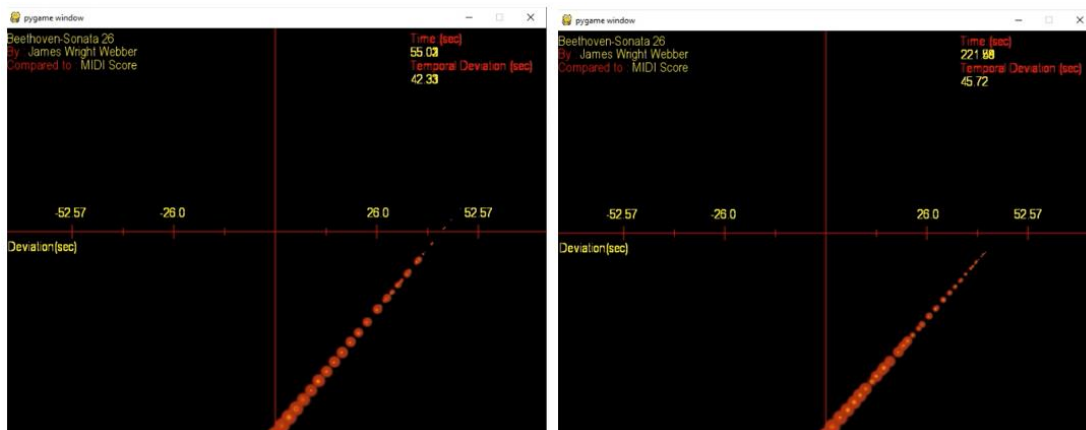
Εικόνα 6-1

Στιγμιότυπο της οπτικοποίησης και τμήμα του αρχείου καταγραφής για την Σονάτα No.26 σε μι-ύφεση του Beethoven σε εκτέλεση του James Wright Webber, συγκρινόμενη με τον εαυτό της.

### 6.2.2 Σύγκριση ερμηνείας με παρτιτούρα

Στο πρώτο πείραμα χρησιμοποιήθηκε η σονάτα για πιάνο No.26 σε μι-ύφεση μείζονα ματζόρε του Ludwig van Beethoven (Les Adieux). Χρησιμοποιήθηκε παρτιτούρα κωδικοποιημένη σε μορφή MIDI προερχόμενη από το MAESTRO dataset, μια συλλογή που περιλαμβάνει περισσότερες από 200 ώρες μουσικής για πιάνο κωδικοποιημένες σε μορφή MIDI σε αντιστοιχία με τη σημειογραφία τους με απόκλιση της τάξης των 3ms (διαθέσιμο στην ηλεκτρονική διεύθυνση <https://magenta.tensorflow.org/datasets/maestro>). Η χρονική διάρκεια του αρχείου MIDI είναι 6:40min. Αρχικά η παρτιτούρα συγκρίθηκε με την ερμηνεία του σύγχρονου Βρετανού πιανίστα James Wright Webber. Η συνολική διάρκεια της εκτέλεσης του είναι 5:47min. Στο πρώτο και πιο αργό μέρος του κομματιού, η εκτέλεση του Webber είναι γρηγορότερη από το tempo της παρτιτούρας, με αποτέλεσμα η φλόγα να κινείται σταθερά προς τα δεξιά αποτυπώνοντας την αύξηση της χρονικής απόκλισης και φτάνοντας σε μία χρονική απόκλιση 42sec. Μετά το 62<sup>ο</sup> δευτερόλεπτο, καθώς το κομμάτι περνάει στο δεύτερο πιο allegro μέρος, το tempo της εκτέλεσης του Webber ταυτίζεται με το tempo της παρτιτούρας με αποτέλεσμα η φλόγα να σταθεροποιείται. Μέχρι το τέλος του κομματιού προστίθεται επιπλέον χρονική απόκλιση που οφείλεται κυρίως σε εκφραστικές παρεκκλίσεις της άρθρωσης του εκτελεστή (παρατεταμένες νότες, παύσεις κτλ.) και φτάνουμε έτσι στη μέγιστη χρονική απόκλιση των 52,57sec από το αντίστοιχο σημείο της παρτιτούρας.

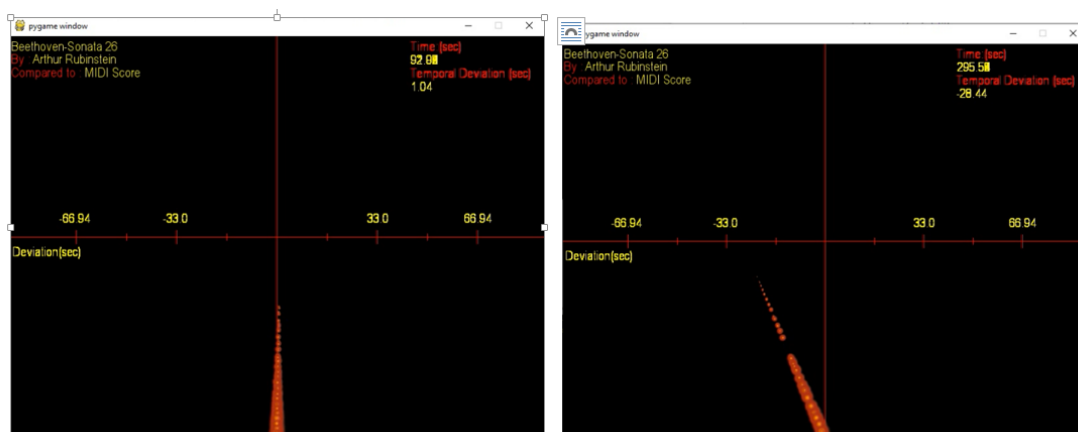
Όσο αφορά το ύψος των δυναμικών, ο Webber ερμηνεύει το κομμάτι με σταθερά υψηλές δυναμικές, χωρίς έντονες αυξομειώσεις στο παίξιμό του. Αυτό αποτυπώνεται στο μέγεθος της φλόγας η οποία παραμένει αρκετά μεγάλη και χωρίς μεγάλες αυξομειώσεις καθ' όλη τη διάρκεια του κομματιού. Στην Εικόνα 6-2 φαίνονται δύο ενδεικτικά στιγμιότυπα από την οπτικοποίηση της εκτέλεσης του.



**Εικόνα 6-2**

Στιγμιότυπα από την εκτέλεση της σονάτας No.26 του Beethoven από τον James Wright Webber, συγκρινόμενη με την παρτιτούρα

Στην συνέχεια η παρτιτούρα συγκρίθηκε με την εκτέλεση του διάσημου βιρτουόζου πιανίστα Arthur Rubinstein. Η συνολική διάρκεια της εκτέλεσης του Rubinstein είναι 7:36min και είναι πλούσια σε εκφραστικές αποκλίσεις. Στο πρώτο μέρος του κομματιού ο Rubinstein παραμένει πάνω στο tempo της παρτιτούρας με αποτέλεσμα η «φλόγα» να εμφανίζεται κοντά στον κάθετο άξονα, με μία μικρή απόκλιση προς τα δεξιά που οφείλεται κυρίως σε εκφραστικές παύσεις. Από το χρονικό σημείο 1:40 και για το υπόλοιπο του κομματιού ο Rubinstein επιλέγει ένα tempo ελαφρώς αργότερο από αυτό της σημειογραφίας. Η κλίση της φλόγας μετακινείται αργά και σταθερά προς τα αριστερά, για να φτάσει στην μέγιστη απόκλισή της στο τέλος του κομματιού. Όσο αφορά τις δυναμικές, το παίξιμο του Rubinstein περιέχει πολλές και έντονες μεταβολές εκφραστικών δυναμικών, παραμένοντας σε γενικές γραμμές σε χαμηλότερες εντάσεις από τις αντίστοιχες του Webber. Το γεγονός αυτό αποτυπώνεται στην οπτικοποίηση από τις μεταβολές στο μέγεθος της «φλόγας». Στην εικόνα **Εικόνα 6-3** φαίνονται δύο στιγμιότυπα από την οπτικοποίηση της εκτέλεσης του Arthur Rubinstein.



**Εικόνα 6-3**

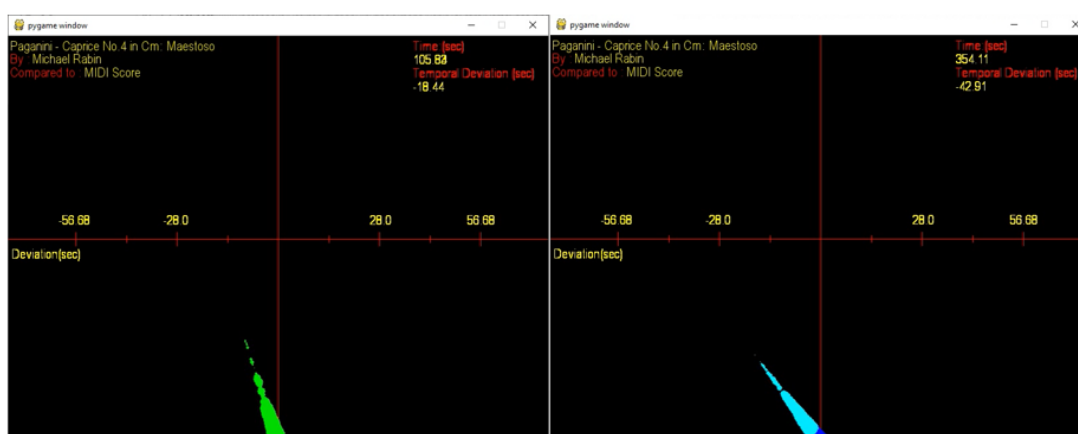
Στιγμιότυπα από την εκτέλεση της σονάτας No.26 του Beethoven από τον Arthur Rubinstein, συγκρινόμενη με την παρτιτούρα

Τα στιγμιότυπα αντιστοιχούν στα μουσικά σημεία στα οποία έχουν ληφθεί και τα στιγμιότυπα της εικόνας **Εικόνα 6-2**. Από την οπτική αντιπαραβολή των δύο εικόνων γίνεται αντιληπτή η διαφορά στην οπτικοποίηση των δύο διαφορετικών ερμηνειών. Καθώς η

εκτέλεση του Webber είναι γρηγορότερη από την παρτιτούρα στο πρώτο μέρος, η φλόγα αποκτά μια κλίση προς τα δεξιά (**Εικόνα 6-2** – αριστερά) η οποία παραμένει και στη συνέχεια καθώς ο Webber συνεχίζει την ερμηνεία του σε tempo αντίστοιχο με την παρτιτούρα (**Εικόνα 6-2** – δεξιά). Αντίθετα στην εκτέλεση του Rubinstein η φλόγα στο πρώτο μέρος βρίσκεται κοντά στον κάθετο άξονα (**Εικόνα 6-3**- αριστερά) ενώ στην συνέχεια αποκτά κλίση προς τα αριστερά καθώς το υπόλοιπο κομμάτι εκτελείται σε tempo αργότερο από την παρτιτούρα. Η διαφορά στο μέγεθος της φλόγας είναι επίσης εμφανής.

Για τον έλεγχο της μονοφωνικής έκδοσης της εφαρμογής επιλέχθηκε το έργο Caprice No.4 σε C ελάσσονα (Maestoso) του Niccolò Paganini, από το έργο του 24 Καπρίτσια για σόλο βιολί. Η παρτιτούρα κωδικοποιημένη σε μορφή MIDI προέρχεται από την ιστοσελίδα [classicalarchives.com](https://www.classicalarchives.com) (διαθέσιμο κατόπιν δωρεάν εγγραφής στην ηλεκτρονική διεύθυνση <https://www.classicalarchives.com/midi/composer/3114.html>). Αρχικά η παρτιτούρα συγκρίθηκε με την εκτέλεση του Αμερικανού βιολιστή Michael Rabin, μια εκτέλεση με έντονες εκφραστικές μεταβολές. Στην εκτέλεσή του ο Rabin παραμένει στο tempo της παρτιτούρας μέχρι το σημείο 0:58. Στη συνέχεια και μέχρι το χρονικό σημείο 2:30 που διαρκεί το δεύτερο και πιο έντονο μέρος του κομματιού, ο Rabin ερμηνεύει σε χαμηλότερο tempo από την παρτιτούρα με αποτέλεσμα η κλίση να μετατοπίζεται σταδιακά προς τα αριστερά. Στο τρίτο μέρος του κομματιού (2:30 ως 4:55) ο Rabin αλλάζει το tempo σε ελαφρώς πιο γρήγορο από την παρτιτούρα και η κλίση της φλόγας επανέρχεται προς τα δεξιά (χωρίς όμως να φτάνει στον κάθετο άξονα). Τέλος στο τελευταίο μέρος το tempo είναι αργότερο από την παρτιτούρα και η κλίση της φλόγας μετακινείται σταδιακά όλο και πιο αριστερά, φτάνοντας στη μέγιστη απόκλισή της στο τέλος του κομματιού.

Όσο αφορά τις δυναμικές, ο Rabin διατηρεί ένα σταθερό μέσο επίπεδο, αυξομειώνοντας σε σημεία όπου θέλει να δώσει έμφαση σε συγκεκριμένες νότες. Οι έντονες δοξαριές (attack των νοτών) και η έντονη χρήση του vibrato προκαλούν έντονο «τρεμοπαίξιμο» (flickering) της φλόγας. Στην **Εικόνα 6-4** φαίνονται δύο στιγμιότυπα από την οπτικοποίηση της εκτέλεσης του Rabin.



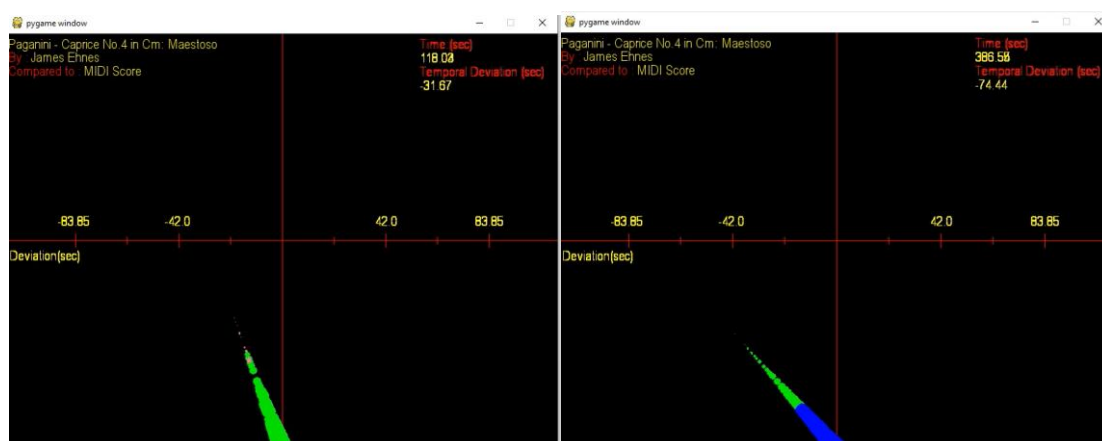
**Εικόνα 6-4**

*Στιγμιότυπα από την εκτέλεση του Caprice No.4 σε C μινόρε του Niccolò Paganini από τον Michael Rabin, συγκρινόμενη με την παρτιτούρα*

Στη συνέχεια έγινε σύγκριση της αντίστοιχης παρτιτούρας με την εκτέλεση του σύγχρονου Καναδού βιολιστή James Ehnes (διάρκεια εκτέλεσης 6:48). Ο Ehnes ακολουθεί σε όλη τη διάρκεια του κομματιού ένα ελαφρώς πιο αργό tempo από αυτό της παρτιτούρας. Η κλίση της «φλόγας» μετατοπίζεται συνεχώς και με αργό ρυθμό προς τα αριστερά και φτάνει στη

μέγιστη απόκλιση της στο τέλος του κομματιού. Εξάιρεση αποτελεί το σημείο από το 3:06 ως το 4:15 όπου ο Ehnes ακολουθεί το tempo της παρτιτούρας και η φλόγα έχει μικρές μετατοπίσεις και προς τα αριστερά και προς τα δεξιά.

Όσο αφορά της δυναμικές, ο Ehnes ακολουθεί ένα πιο «στρωτό» παίξιμο. Αυξομειώνει την έντασή του για να δώσει έμφαση στα σημεία που χρειάζεται αλλά σε γενικές γραμμές οι δυναμικές του είναι πιο ομοιόμορφες και με μικρότερη χρήση του vibrato. Αυτό έχει σαν αποτέλεσμα το μέγεθος της οπτικοποίησης να μην έχει τόσο πολλές και τόσο απότομες μεταβολές όσο στην περίπτωση του Rabin. Στην **Εικόνα 6-5** φαίνονται δύο στιγμιότυπα από την οπτικοποίηση της εκτέλεσης του Ehnes σε αντίστοιχα σημεία με αυτά της εκτέλεσης του Rabin που εμφανίζονται στην **Εικόνα 6-4**.



**Εικόνα 6-5**

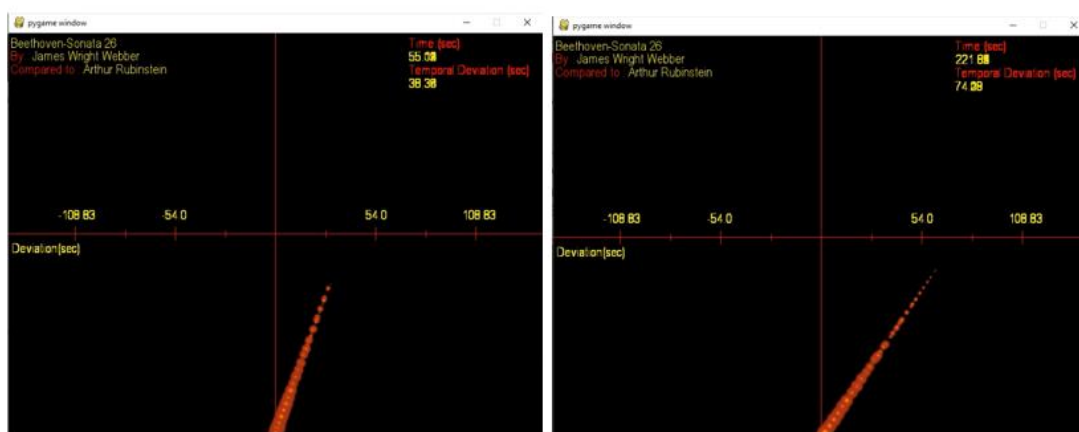
*Στιγμιότυπα από την εκτέλεση του Caprice No.4 σε C μόνωρε του Niccolò Paganini από τον James Ehnes, συγκρινόμενη με την παρτιτούρα*

### 6.2.3 Σύγκριση ανάμεσα σε διαφορετικές ερμηνείες του ίδιου έργου

Σε συνέχεια της πειραματικής αξιολόγησης, τα μουσικά έργα που χρησιμοποιήθηκαν στην προηγούμενη ενότητα συγκρίθηκαν μεταξύ τους. Στόχος ήταν να διαπιστωθεί η λειτουργικότητα της εφαρμογής στη σύγκριση δύο διαφορετικών ερμηνειών του ίδιου μουσικού έργου από διαφορετικούς καλλιτέχνες. Ταυτόχρονα, χρησιμοποιώντας τα ίδια μουσικά έργα, μας δίνεται η δυνατότητα να διασταυρωθεί κατά πόσο τα αποτελέσματα της μεταξύ τους σύγκρισης είναι σε συμφωνία με τα αποτελέσματα της επιμέρους τους σύγκρισης με την παρτιτούρα.

Αρχικά, για τη Sonata No.26 του Beethoven, η εκτέλεση του Webber συγκρίθηκε με την εκτέλεση του Rubinstein. Στο πρώτο μέρος του κομματιού (μέχρι το σημείο 0:57) η εκτέλεση του Webber είναι γρηγορότερη από την αντίστοιχη του Rubinstein καθώς ο Webber ερμηνεύει το μέρος αυτό σε tempo γρηγορότερο από της παρτιτούρας ενώ ο Rubinstein κοντά στο tempo της παρτιτούρας. Η κλίση της φλόγας μετακινείται συνεχώς προς τα δεξιά, περίπου με το ρυθμό που κινείται και στη σύγκριση με την παρτιτούρα. Στο υπόλοιπο του κομματιού, ο Webber σταθεροποιεί το tempo του κοντά στο tempo της παρτιτούρας ενώ ο Rubinstein ερμηνεύει το κομμάτι σε ένα tempo ελαφρώς αργότερο από την παρτιτούρα. Η κλίση της φλόγας μετακινείται συνεχώς προς τα δεξιά φτάνοντας στη μέγιστη απόκλιση στο τέλος του κομματιού

Όσο αφορά τις δυναμικές, ο Webber ερμηνεύει το κομμάτι με σταθερά υψηλές δυναμικές, ενώ ο Rubinstein ακολουθεί μια πιο εκφραστική ερμηνεία με μεγαλύτερες εναλλαγές στις δυναμικές. Αυτό έχει ως αποτέλεσμα το μέγεθος της φλόγας να παρουσιάζει αρκετές μεταβολές, αναπαριστώντας τη διαφορά ανάμεσα στις δυναμικές των δύο καλλιτεχνών σε κάθε χρονικό σημείο. Στην **Εικόνα 6-6** φαίνονται δύο στιγμιότυπα από την εκτέλεση του Webber συγκρινόμενη με αυτή του Rubinstein. Τα στιγμιότυπα βρίσκονται σε αντίστοιχα χρονικά σημεία με την **Εικόνα 6-2**. Στο δεξιό τμήμα της εικόνας παρατηρούμε ότι η χρονική απόκλιση αντιστοιχεί πλέον σε μεγαλύτερη χρονική διάρκεια καθώς στο συγκεκριμένο σημείο ο Rubinstein βρισκόταν χρονικά πίσω από την παρτιτούρα. Παρατηρείται επίσης ότι το μέγεθος της φλόγας έχει διαφορετικό μέγεθος, καθώς οι γενικά υψηλές δυναμικές του Webber συγκρίνονται πλέον με τις αντίστοιχες του Rubinstein οι οποίες έχουν αυξομειώσεις, σε αντίθεση με την παρτιτούρα σε μορφή MIDI. Σε γενικές γραμμές το πείραμα φαίνεται να είναι σε αντιστοιχία με τις συγκρίσεις των επιμέρους ερμηνειών με την παρτιτούρα, όπως αυτές παρουσιάστηκαν στην προηγούμενη ενότητα.



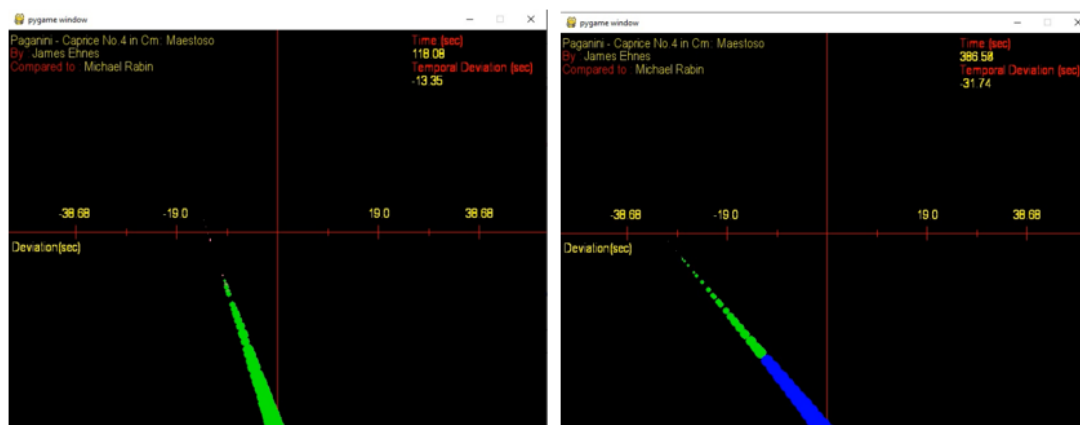
**Εικόνα 6-6**

*Στιγμιότυπα από την εκτέλεση της σονάτας No.26 του Beethoven από τον James Wright Webber, συγκρινόμενη με την αντίστοιχη εκτέλεση του Arthur Rubinstein*

Στη συνέχεια, η εκτέλεση του James Ehnes πάνω στο Caprice No.4 σε C μινόρε του Paganini συγκρίθηκε με την αντίστοιχη εκτέλεση του Michael Rabin. Στο πείραμα αυτό ο Ehnes ακολουθεί ένα σταθερά πιο αργό tempo από τη παρτιτούρα, ενώ ο Rabin έχει αυξομειώσεις κατά τη διάρκεια της εκτέλεσής του. Έτσι μέχρι το σημείο 1:10 η κλίση της φλόγας μετατοπίζεται σταδιακά προς τα αριστερά καθώς ο Rabin εκτελεί το αντίστοιχο μέρος πιο γρήγορα και κοντά στο tempo της παρτιτούρας. Δημιουργείται έτσι μια χρονική απόκλιση 12sec η οποία είναι και η αναμενόμενη, καθώς η εκτέλεση του Rabin ξεκινάει το αντίστοιχο δεύτερο μέρος του κομματιού στο χρονικό σημείο 0:58. Στη συνέχεια και μέχρι το σημείο 2:30 οι δύο συνθέτες βρίσκονται σε αντίστοιχο tempo (ελαφρώς αργότερο από την παρτιτούρα και για τους δύο) με αποτέλεσμα η κλίση της φλόγας να μη παρουσιάζει σημαντικές μεταβολές. Το δεύτερο μέρος του κομματιού ολοκληρώνεται με τον Ehnes να ερμηνεύει την τελευταία φράση στο ίδιο tempo ενώ αντίστοιχα ο Rabin την καθυστερεί σημαντικά επιμηκύνοντας τις νότες (αλλαγές στη άρθρωση). Έτσι στο σημείο αυτό η φλόγα μετατοπίζεται ελαφρώς δεξιά και φτάνει στο σημείο 2:38 σε μία απόκλιση -8sec (Η αντίστοιχη μουσική φράση στην εκτέλεση του Rabin ολοκληρώνεται στο σημείο 2:30 οπότε η χρονική απόκλιση αποδεικνύεται ορθή). Ακολουθεί το τρίτο μέρος του κομματιού όπου ο Ehnes παραμένει σταθερά σε tempo αργότερο της παρτιτούρας, ενώ ο Rabin επιταχύνει ερμηνεύοντας σε tempo μεγαλύτερο της παρτιτούρας. Η κλίση της φλόγας μετατοπίζεται

συνεχώς προς τα αριστερά και φτάνει στη μέγιστη απόκλιση των -38sec στο σημείο 5:33 όπου και ολοκληρώνεται το τρίτο μέρος (στην ερμηνεία του Rabin το τρίτο μέρος ολοκληρώνεται στο σημείο 4:55). Στη τελευταίο μέρος του κομματιού, ο Rabin μεταπηδά σε ένα αργότερο tempo με αποτέλεσμα η φλόγα να κινείται σταδιακά προς τα δεξιά πλέον, καθώς ο Ehnes παραμένει σταθερός στο tempo του. Η απόκλιση μεταξύ των δύο ερμηνειών μειώνεται για να καταλήξει στη τιμή -28sec στο τέλος του κομματιού.

Όσο αφορά τις δυναμικές, βλέπουμε πλέον το μήκος της φλόγας να έχει έντονες και απότομες διακυμάνσεις, καθώς οι γενικά ομαλές και σταθερές δυναμικές του Ehnes συγκρίνονται με τις πολλαπλές αυξομειώσεις και το δυνατό vibrato του Rabin. Τα αποτελέσματα της οπτικοποίησης και οι αποκλίσεις στα χρονικά σημεία που αναφέρθηκαν στην προηγούμενη παράγραφο βρίσκονται σε απόλυτη συμφωνία με τα αποτελέσματα της επιμέρους σύγκρισης των δύο ερμηνειών με την παρτιτούρα. Στην **Εικόνα 6-7** φαίνονται δύο στιγμιότυπα από την εκτέλεση του Ehnes συγκρινόμενη με αυτή του Rabin. Τα στιγμιότυπα βρίσκονται σε αντίστοιχα χρονικά σημεία με την **Εικόνα 6-5**. Παρατηρούμε ότι το μήκος της φλόγας και στις δύο περιπτώσεις είναι μεγαλύτερο, καθώς οι δυναμικές του Ehnes δεν συγκρίνονται πλέον με αυτές της παρτιτούρας αλλά με τις αντίστοιχες του Rabin, ο οποίος στα αντίστοιχα σημεία έχει αισθητά χαμηλότερες δυναμικές (**Εικόνα 6-4**). Η κλίση της φλόγας βρίσκεται συνεχώς προς τα αριστερά με μικρότερες όμως χρονικές αποκλίσεις από αυτές της σύγκρισης με την παρτιτούρα. Καθώς και οι δύο εκτελέσεις είναι πιο αργές από την παρτιτούρα με αυτή του Ehnes να είναι μεγαλύτερη χρονικά, το γεγονός αυτό είναι φυσιολογικό.



**Εικόνα 6-7**

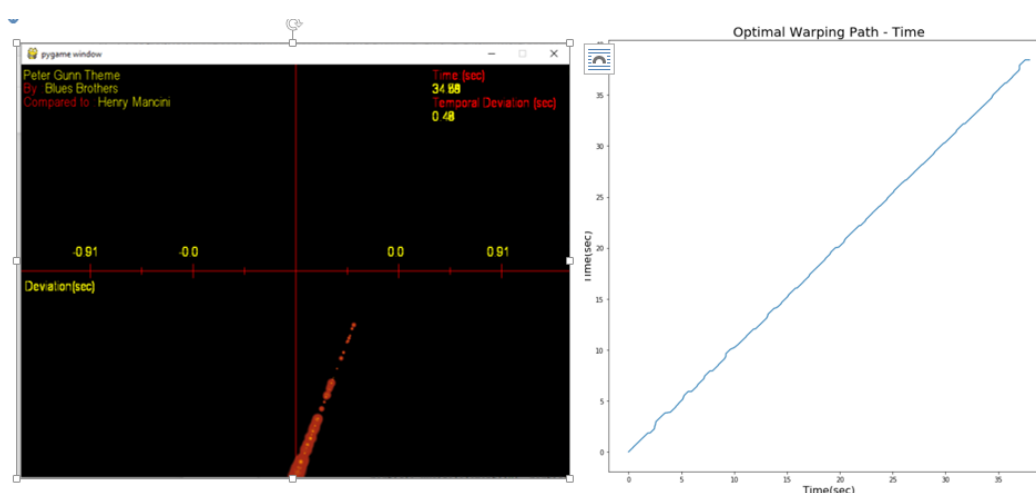
*Στιγμιότυπα από την εκτέλεση του Caprice No.4 σε C μόνωρε του Niccolò Paganini από τον James Ehnes, συγκρινόμενη με την αντίστοιχη εκτέλεση του Michael Rabin*

#### 6.2.4 Οπτικοποίηση για διαφορετικά μουσικά είδη

Από τα πρώτα στάδια ανάπτυξης της εφαρμογής και προκειμένου να ελεγχθεί ο ορθότητα της αντιστοίχισης του DTW, έγινε δοκιμή με τα πρώτα 19 μέτρα από το Peter Gunn Theme του Henry Mancini. Το πείραμα επαναλήφθηκε και μετά την ολοκλήρωση της εφαρμογής προκειμένου να δοκιμαστεί η λειτουργία της και σε άλλα μουσικά είδη πέρα της κλασσικής μουσικής. Συγκεκριμένα η εκτέλεση των Blues Brothers συγκρίθηκε με την εκτέλεση του ίδιου του συνθέτη. Στα πρώτα 19 μέτρα και οι δύο εκτελέσεις παραμένουν πάνω στο βασικό μουσικό θέμα, πριν περάσουν στο αυτοσχεδιαστικό μέρος των πνευστών. Στη συγκεκριμένη

δοκιμή έγινε χρήση και του παράγοντα tuning καθώς η εκτέλεση των Blues Brothers είναι ένα ημιτόνιο χαμηλότερα. Τα δύο τμήματα είναι παρόμοια καθώς έχουν χρονική απόκλιση μικρότερη του ενός δευτερολέπτου. Αυτό στην οπτικοποίηση προκαλεί πολύ έντονες μεταβολές στην κλίση της φλόγας καθώς η μετατόπιση κατά 45 μοίρες αντιστοιχίζεται από την εφαρμογή πάντα στη μέγιστη χρονική απόκλιση, όποια και αν είναι αυτή (**Εικόνα 6-8** αριστερά). Η αντιστοίχιση πάντως που πραγματοποίησε ο DTW είναι σωστή και αυτό μπορεί να εξακριβωθεί εμπειρικά από το αρχείο καταγραφή, από τη γραφική αναπαράσταση του βέλτιστου Warping Path (**Εικόνα 6-8** δεξιά) ή αναπαράγοντας τα δύο αρχεία από ένα κοινό σημείο αντιστοίχισης (Ενότητα 5.4.1 **Εμπειρική επαλήθευση του DTW**).

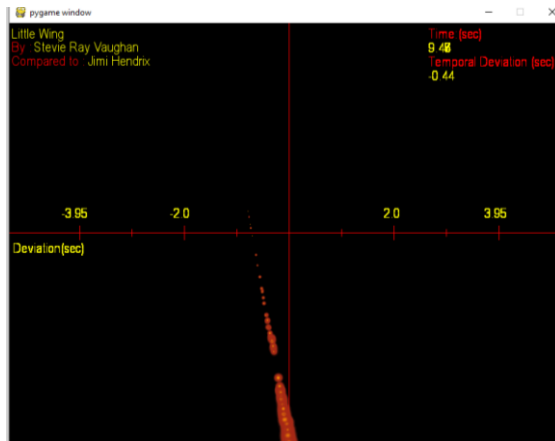
Όσο αφορά τις δυναμικές, το μέγεθος της φλόγας παρουσιάζει μεταβολές καθώς και οι δύο ερμηνείες περιέχουν πνευστά με έντονες δυναμικές. Η ερμηνεία των Blues Brothers όμως περιέχει staccato παίξιμο στα drums (σε αντίθεση με την jazz προσέγγιση των drums του Mancini) που αυξάνουν τη συνολική δυναμική του κομματιού ειδικά στα σημεία όπου δεν υπάρχουν πνευστά όργανα.



**Εικόνα 6-8**

Στιγμιότυπο από την οπτικοποίηση του Peter Gunn Theme των Blues Brothers συγκρινόμενο με την εκτέλεση του Henry Mancini και γραφική παράσταση του βέλτιστου Warping Path.

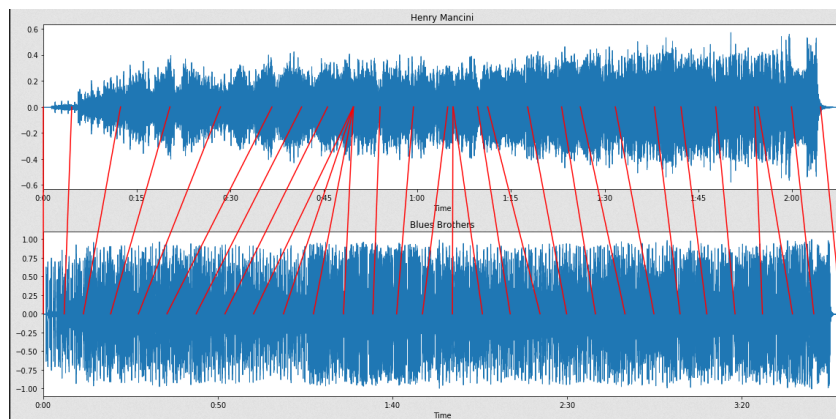
Παρόμοια δοκιμή σε rock μουσική επαναλήφθηκε με το κομμάτι Little Wing του Jimi Hendrix, από το οποίο χρησιμοποιήθηκε το αρχικό εισαγωγικό μέρος (μέχρι την έναρξη των στοίχων) το οποίο περιλαμβάνει διάφορες κιθαριστικές μουσικές φράσεις. Συγκρίθηκε το αντίστοιχο τμήμα από την ερμηνεία του Stevie Ray Vaughan με την πρωτότυπη εκτέλεση του Jimi Hendrix. Καθώς η χρονική διάρκεια είναι μικρή, η φλόγα κατά την οπτικοποίηση παρουσιάζει και πάλι έντονες μεταβολές στην κλίση της κατά τη διάρκεια μικρών χρονικών αποκλίσεων που οφείλονται κυρίως σε εκφραστικές παύσεις. Η αντιστοίχιση του DTW είναι πάντως ακριβής. Όσο αφορά το μέγεθος της φλόγας αυτή παρουσιάζει αυξωμειώσεις καθώς οι δύο ερμηνευτές επιλέγουν να τονίσουν με διαφορετικό τρόπο τις νότες ώστε να αποδώσουν την εκφραστική διάσταση που επιθυμούν. Ο Vaughan χαρακτηρίζεται από πιο έντονο attack στις νότες του και γενικότερα πιο καθαρό ήχο (τόσο λόγω στυλ όσο και λόγω μουσικής παραγωγής) ενώ στην εκτέλεση του Hendrix υπάρχει ένα χτύπημα στη «καμπάνα» του crash πιανινού στο τέλος κάθε μέτρου. Στην **Εικόνα 6-9** φαίνεται ένα στιγμιότυπο της οπτικοποίησης που προέκυψε από την παραπάνω δοκιμή.



Εικόνα 6-9

Στιγμιότυπο από την οπτικοποίηση του *Little Wing* του *Jimi Hendrix* συγκρινόμενο με την εκτέλεση του *HStevie Ray Vaughan*

Η σύγκριση ολόκληρων Jazz, Blues και Rock κομματιών είναι αρκετά δύσκολο να πραγματοποιηθεί επιτυχώς, καθώς οι διαφορετικές ερμηνείες σε αυτά τα είδη μουσικής δε βασίζονται σε κάποια δεδομένη παρτιτούρα και περιέχουν έντονα το στοιχείο της διασκευής και του αυτοσχεδιασμού. Ο DTW πρέπει να αντιστοιχίσει κάθε σημείο της ακολουθίας που συγκρίνεται με τουλάχιστον ένα σημείο της ακολουθίας που αποτελεί τη βάση σύγκρισης, σύμφωνα με όσα αναφέρθηκαν στην ενότητα **2.3 Ο Αλγόριθμος DTW**. Έτσι στο πρώτο μη συμβατό σημείο που θα συναντήσει ο DTW θα δημιουργήσει μία αντιστοίχιση χωρίς λογικό νόημα με όποιο σημείο θεωρήσει ως πιο «κοντινό» σύμφωνα με τον πίνακα συγκεντρωτικού κόστους. Το γεγονός αυτό συνήθως είναι καταστροφικό για όλη την υπόλοιπη διαδικασία καθώς ο DTW, τηρώντας τη συνθήκη μονοτονίας, θα αντιστοιχίσει όλα τα υπόλοιπα σημεία της πρώτης ακολουθίας σε μεταγενέστερα χρονικά σημεία της δεύτερης ακολουθίας, χάνοντας έτσι την όποια λογική αντιστοίχιση αυτά μπορεί να είχαν. Καθώς ο DTW ακολουθεί ανάστροφη (backtracking) τεχνική για την εύρεση του βέλτιστου Warring Path, η όποια σωστή αντιστοίχιση θα βρίσκεται στο τελευταίο τμήμα του μουσικού έργου. Στην **Εικόνα 6-10** φαίνεται μια γραφική αναπαράσταση της αντιστοίχισης ανάμεσα στις κυματομορφές των δύο ερμηνειών του Peter Gunn Theme όταν έγινε απόπειρα να αντιστοιχηθούν ολόκληρα τα μουσικά έργα. Τα σφάλματα στην αντιστοίχιση είναι εμφανή, καθώς πολλαπλά σημεία της ερμηνείας των Blues Brothers με σημαντική χρονική απόσταση μεταξύ τους έχουν αντιστοιχηθεί με το ίδιο σημείο της ερμηνείας του Mancini.



Εικόνα 6-10

Απόπειρα αντιστοίχισης ολόκληρων των ερμηνειών των *Blues Brother* και *Henry Mancini* πάνω στο *Peter Gunn Theme*



Με βάση όλες τις πειραματικές δοκιμές που πραγματοποιήθηκαν στο κεφάλαιο αυτό, μπορούμε να συμπεράνουμε ότι η εφαρμογή FlameCens λειτουργεί ικανοποιητικά όταν εφαρμόζεται σε διαφορετικές ερμηνείες κλασικών έργων που βασίζονται στην αναπαραγωγή καθορισμένης παρτιτούρας. Επίσης λειτουργεί και σε τμήματα διαφορετικών ερμηνειών άλλων ειδών μουσικής (Jazz, rock κτλ), όταν τα τμήματα αυτά εκφράζουν παρόμοιο μουσικό περιεχόμενο και όχι αυτοσχεδιαστικά ή διασκευασμένα μέρη. Στον **Πίνακας 7** παρουσιάζονται όλα τα μουσικά έργα ή τμήματα που χρησιμοποιήθηκαν κατά τις πειραματικές δοκιμές της εφαρμογής.

Μουσικό Έργο	Συνθέτης	Ερμηνείες
Σονάτα 26 σε μι-ύφεση μείζονα (Les Adieux)	Ludwig van Beethoven	Παρτιτούρα σε μορφή MIDI
		James Wright Webber
		Arthur Rubinstein
Caprice No.4 σε C ελάσσονα (Maestoso)	Niccolo Paganini	Παρτιτούρα σε μορφή MIDI
		James Ehnes
		Michael Rabin
Peter Gunn Theme (πρώτα 19 μέτρα)	Henry Mancini	Henry Mancini
		Blues Brothers
Little Wing	Jimi Hendrix	Jimi Hendrix
		Stevie Ray Vaughan

**Πίνακας 7**

*Μουσικά έργα ή τμήματα αυτών που χρησιμοποιήθηκαν κατά την πειραματική αξιολόγηση*

## 7. Συμπεράσματα

Ολοκληρώνοντας την παρούσα εργασία και λαμβάνοντας υπόψιν τα αποτελέσματα των πειραματικών δοκιμών της εφαρμογής FlameCens, στο παρόν κεφάλαιο παρουσιάζεται μία συνολική αποτίμηση και ανασκόπηση των βασικών σημείων της.

### 7.1 Επισκόπηση

Στόχος της συγκεκριμένης εργασίας ήταν εξ αρχής η οπτικοποίηση της μουσικής έκφρασης, η αποτύπωση δηλαδή εκείνων των χαρακτηριστικών της μουσικής που ο ερμηνευτής τροποποιεί προκειμένου να μεταδώσει μέσω της ερμηνείας του τη δική του μουσική προσέγγιση ενός συγκεκριμένου έργου. Συγκεκριμένα επιλέχθηκε να ερευνηθούν οι εκφραστικοί παράγοντες της χρονικής απόκλισης και της μεταβολής του ύψους των δυναμικών. Για το σκοπό αυτό αρχικά ερευνήθηκαν μέσω της σύγχρονης βιβλιογραφίας τα χαρακτηριστικά του ήχου που θα μπορούσαν να χρησιμοποιηθούν για το σκοπό αυτό, καθώς και ο αλγόριθμος αντιστοίχισης χρονικών ακολουθιών DTW (Ενότητα **2.3 Ο Αλγόριθμος DTW**). Η έρευνα επικεντρώθηκε στη χρήση των χρωματογραφικών χαρακτηριστικών και στην παραλλαγή τους που ονομάζεται CENS (Ενότητα **2.2.6 CENS Features**). Η δυνατότητα των χρωματογραφικών χαρακτηριστικών να αποτυπώνουν την τονικότητα τα καθιστά ιδανικά για τη διαδικασία χρονικής αντιστοίχισης δύο ηχητικών ακολουθιών, ενώ η CENS παραλλαγή τους επηρεάζεται λιγότερο από τη χροιά, τις δυναμικές και την άρθρωση. Τέλος, διερευνήθηκαν οι προϋπάρχουσες προσεγγίσεις στην υπολογιστική μοντελοποίηση της μουσικής έκφρασης και οι τρόποι με τους οποίους οι διάφοροι εκφραστικοί παράγοντες ποσοτικοποιούνται σε διαφορετικές ερευνητικές πρωτοβουλίες.

Στη συνέχεια, και αφού μελετήθηκαν διάφορες ερευνητικές εφαρμογές πάνω στην οπτικοποίηση εκφραστικών χαρακτηριστικών της μουσικής, αναπτύχθηκε η εφαρμογή FlameCens καθώς και η παραλλαγή της FlameCens-pitch. Η εφαρμογή πραγματοποιεί σύγκριση μεταξύ δύο αρχείων ήχου και χρονική αντιστοίχιση τους με χρήση του αλγορίθμου DTW. Στη συνέχεια καταγράφει τη μεταξύ τους χρονική απόκλιση και τη διαφορά στις δυναμικές για κάθε σημείο της αντιστοίχισης του DTW. Η εφαρμογή μπορεί να συγκρίνει τόσο διαφορετικές ερμηνείες του ίδιου μουσικού έργου από διαφορετικούς καλλιτέχνες, όσο και μια ερμηνεία με την αντίστοιχη παρτιτούρα η οποία εισάγεται κωδικοποιημένη σε μορφή MIDI. Προκειμένου να εξαχθούν τα κατάλληλα ηχητικά χαρακτηριστικά, η παρτιτούρα μετατρέπεται από μορφή MIDI σε μορφή .WAV από την ίδια την εφαρμογή, χωρίς να απαιτείται κάποιο πρόσθετο λογισμικό. Τα αποτελέσματα οπτικοποιούνται σε μία απεικόνιση με μορφή φλόγας η οποία μεταβάλλεται δυναμικά καθώς το κομμάτι αναπαράγεται. Η κλίση της φλόγας αναπαριστά τη χρονική απόκλιση μεταξύ των συγκρινόμενων αρχείων ενώ η μεταβολή του μεγέθους της αναπαριστά τις διαφορές στις δυναμικές. Στο παράθυρο προβολής της οπτικοποίησης υπάρχει ένας κάθετος άξονας που ορίζει τη μηδενική χρονική απόκλιση και ένας οριζόντιος άξονας, διαβαθμισμένος κατά περίπτωση, που υποδεικνύει τη τρέχουσα χρονική απόκλιση. Ταυτόχρονα εμφανίζονται και άλλες χρήσιμες πληροφορίες, όπως το τρέχον χρονικό σημείο του κομματιού, η τιμή της τρέχουσας χρονικής απόκλισης και (προαιρετικά) πληροφορίες σχετικά με τον τίτλο του έργου και τους ερμηνευτές που συγκρίνονται. Όλα τα δεδομένα της οπτικοποίησης καταγράφονται και σε αρχείο με μορφή .csv.

Στις ερευνητικές προσεγγίσεις που μελετήθηκαν γίνεται απόπειρα να οπτικοποιηθούν δομικοί και εκφραστικοί παράγοντες της μουσικής αναλύοντας ηχητικά χαρακτηριστικά ενός

μουσικού έργου (που αφορούν κυρίως στο tempo και στις δυναμικές) και απεικονίζοντας την εξέλιξη των χαρακτηριστικών αυτών ως προς το χρόνο. Η προσέγγιση αυτή, αποδίδει μεν την εκφραστική εξέλιξη του έργου, δεν αποδίδει όμως την ποσοτική και ποιοτική χρήση των εκφραστικών μέσων από το συγκεκριμένο ερμηνευτή καθώς δεν υπάρχει σύγκριση με άλλες αντίστοιχες ερμηνείες. Η προσέγγιση του Gasser (2005) καλύπτει την αντιστοίχιση ερμηνείας με τη σημειογραφία, περιορίζεται όμως σε δεδομένα εισόδου με μορφή MIDI (η ερμηνεία εισάγεται από ένα πιάνο Bosendorfer ελεγχόμενο από υπολογιστή) ενώ ταυτόχρονα δεν έχει τη δυνατότητα σύγκρισης μεταξύ δύο διαφορετικών ερμηνειών. Σε ένα μουσικό έργο κατά κανόνα θα αποτυπώνονται ερμηνευτικές διακυμάνσεις υποδεικνυόμενες από το συνθέτη, όπως μία αργή και χαμηλή σε δυναμικές εισαγωγή (Andante, Adagio, Lento) που θα ακολουθείται από ένα πιο γρήγορο και ζωηρό μέρος (Allegro, Presto), έγκειται όμως στον εκάστοτε ερμηνευτή ο τρόπος με τον οποίο οι διακυμάνσεις αυτές ποσοτικοποιούνται προκειμένου να αποδοθεί το ξεχωριστό μουσικό αποτέλεσμα που επιθυμεί. Η διαφορά που επιχειρείται από την εφαρμογή FlameCens αφορά στον γεγονός ότι η οπτικοποίηση των εκφραστικών παραγόντων πραγματοποιείται σε σύγκριση είτε με τη σημειογραφία, είτε με κάποια άλλη ερμηνεία του ίδιου μουσικού έργου. Η προσέγγιση αυτή μπορεί να έχει μεγάλο ενδιαφέρον στους τομείς της μουσικολογίας και της μουσικής εκπαίδευσης καθώς ένας μουσικός θα μπορεί να συγκρίνει την ερμηνεία του με τις αντίστοιχες κορυφαίων ερμηνευτών. Τέλος μπορεί να φανεί χρήσιμη και στον απλό χρήστη, εμπλουτίζοντας την εμπειρία της μουσικής ακρόασης.

## 7.2 Περιορισμοί και μελλοντικές προεκτάσεις

Κατά την ανάπτυξη της εφαρμογής, υλοποιήθηκε μια διαδικασία μετατροπής των MIDI αρχείων σε .WAV που απαιτήθηκε προκειμένου να εξαχθούν χαρακτηριστικά ήχου από την κωδικοποιημένη μορφή της παρτιτούρας. Η διαδικασία αυτή ενσωματώθηκε στην εφαρμογή FlameCens και μετατρέπει τα αρχεία MIDI σε πραγματικό χρόνο (play and write), κάνοντας χρήση μόνο βιβλιοθηκών της Python. Η διαδικασία υλοποιήθηκε με τον τρόπο αυτό ώστε η εφαρμογή να είναι ανεξάρτητη από πρόσθετο λογισμικό σύνθεσης (synthesizer όπως π.χ. το Fluidsynth) είναι όμως μια διαδικασία χρονοβόρα. Η ανάπτυξη μιας πιο γρήγορης διαδικασίας μετατροπής των MIDI αρχείων, πιθανά με απευθείας μετατροπή των μηνυμάτων MIDI σε κυματομορφή συγκεκριμένης δειγματοληψίας, θα ήταν μια χρήσιμη προσθήκη στην εφαρμογή. Καθώς ο χρόνος απόκρισης αποτελεί βασικό στοιχείο των επιδόσεων μιας εφαρμογής γενικότερα, μια υλοποίηση του αλγορίθμου DTW με μικρότερο υπολογιστικό κόστος θα ήταν επίσης σημαντική. Οι πιο χρονοβόρες και υπολογιστικά πολύπλοκες διαδικασίες της εφαρμογής είναι η εξαγωγή των χαρακτηριστικών CENS και ο υπολογισμός του συγκεντρωτικού πίνακα κόστους του DTW. Μια υλοποίηση του DTW με χρήση της ζώνης Sakoe-Chiba (Sakoe και Chiba 1978) ή του παραλληλόγραμμου Itakura (1975) πιθανό να βελτίωνε σημαντικά την συνολική απόδοση της εφαρμογής (με την προϋπόθεση βέβαια ότι κατά τη σύγκριση δύο ανάλογων ερμηνειών το βέλτιστο Warping Path δεν θα έχει μεγάλες αποκλίσεις από τη διαγώνιο).

Κατά την συνάρτηση της οπτικοποίησης ο πίνακας των χρονικών αποκλίσεων κανονικοποιείται ώστε η κλίση της φλόγας να περιορίζεται στις 45 μοίρες αριστερά και δεξιά του κάθετου άξονα, με τη μέγιστη κλίση να αντιστοιχεί πάντα στη μέγιστη χρονικά απόκλιση. Η επιλογή αυτή έγινε ώστε στη σύγκριση μεγάλων μουσικών έργων όπως οι κλασσικές συνθέσεις να μη δημιουργούνται τεράστιες οπτικές αποκλίσεις που θα μπορούσαν να οδηγήσουν τη φλόγα σε οριζόντια ή και αντίθετα κατεύθυνση (η σύγκριση κλασσικών έργων με την παρτιτούρα ήταν και το αρχικό έναυσμα για τη δημιουργία της εφαρμογής). Στη

σύγκριση όμως σύντομων τμημάτων, όπου η χρονική απόκλιση είναι μικρή, το γεγονός αυτό δημιουργεί έντονες και απότομες μεταπτώσεις της κλίσης για μικρές χρονικά αποκλίσεις. Το γεγονός αυτό δε δημιουργεί στο χρήστη μια αληθοφανή αίσθηση της χρονικής απόκλισης. Το πρόβλημα θα μπορούσε να εξομαλυνθεί αν η συνάρτηση μειώνει το συντελεστή της κλίσης όταν η μέγιστη χρονική απόκλιση είναι κάτω από ένα συγκεκριμένο όριο (π.χ. 3sec), θα απαιτείται όμως ταυτόχρονα και δυναμική επανασχεδίαση του οριζόντιου άξονα. Η προσθήκη της δυνατότητας αυτής βρίσκεται στο στάδιο της ανάπτυξης, δεν πρόλαβε όμως να ολοκληρωθεί στα όρια της παρούσης εργασίας.

Όπως αναλύθηκε στο κεφάλαιο της Υλοποίησης, η παραλλαγή της εφαρμογής FlameCens-ritch έχει τη δυνατότητα να ανιχνεύει την τονικότητα της τρέχουσας νότας και να αλλάζει ανάλογα το χρώμα της φλόγας στην οπτικοποίηση. Η δυνατότητα αυτή υπάρχει μόνο για μονοφωνικά μουσικά έργα. Μια ενδιαφέρουσα προσθήκη θα ήταν η δυνατότητα αυτή να επεκταθεί και για την περίπτωση της πολυφωνικής μουσικής ώστε να αποτελέσει μια καθολική δυνατότητα της εφαρμογής. Προκειμένου να πραγματοποιηθεί αυτό, απαιτείται η υλοποίηση κάποιας τεχνικής αναγνώρισης συγχορδιών (chord recognition – Müller, 2015). Στο ίδιο σύγγραμμα του Müller παρουσιάζεται και η τεχνική ανίχνευσης του tempo (tempo tracking). Η δυνατότητα αυτή επίσης θα μπορούσε να ενσωματωθεί στην εφαρμογή ώστε στην οπτικοποίηση, αντί για την απόλυτη χρονική απόκλιση, να παρουσιάζεται η απόκλιση του τοπικού tempo (σε bpm). Οι δύο παραπάνω δυνατότητες δεν έχουν διερευνηθεί ακόμη, υπάρχουν όμως ως ιδέες μελλοντικών επεκτάσεων, όπως και η δυνατότητα εντοπισμού των εκφραστικών μεταβολών της μουσικής άρθρωσης (articulation).

Τέλος, όπως αποτυπώθηκε στο κεφάλαιο της Πειραματικής αξιολόγησης, η εφαρμογή αδυνατεί να ανταποκριθεί στην περίπτωση που συγκρίνονται δύο διασκευές ενός έργου με αυτοσχεδιαστικά στοιχεία. Αυτό κατά κανόνα συμβαίνει σε μουσικά είδη όπως Jazz, Blues ή Rock στα οποία είτε δεν υπάρχει προσημειωμένη παρτιτούρα είτε η παρτιτούρα περιλαμβάνει μόνο στοιχειώδεις εντολές (π.χ. μελωδία και chord symbols). Η ανάλυση τέτοιου είδους κομματιών αποτελεί μια μεγάλη πρόκληση. Ένα πολύ-επίπεδο μοντέλο μηχανικής μάθησης που θα μπορούσε να μοντελοποιήσει μουσικά μέρη και μοτίβα (όπως αναφέρεται στην ενότητα **3.4.2 Το πολύ-επίπεδο μοντέλο**) θα μπορούσε πιθανά να αντιμετωπίσει μια τέτοια περίπτωση. Η ανάπτυξη ενός τέτοιου μοντέλου μπορεί να αποτελέσει ένα μελλοντικό, αρκετά εκτεταμένο θέμα έρευνας.

## 8. Αναφορές

- Amar, R.A. & Stasko, J.T. (2005) Knowledge Precepts for Design and Evaluation of Information Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 11, 432–442.
- Bresin, R. (2001). Articulation rules for automatic music performance. In: A. Schloss, & R. Dannenberg (Eds), *Proceedings of the 2001 International Computer Music Conference, Havana, Cuba* (pp. 294–297). San Francisco, CA: International Computer Music Association.
- Bresin, R., & Battel, G.U. (2000). Articulation strategies in expressive piano performance. *Journal of New Music Research*, 29, 211–224.
- Britannica.com (21 Μαΐου 2012). *Scientific Modelling*. Ανάκτηση στις 24 Απριλίου 2020 από το δικτυακό τόπο <https://www.britannica.com/science/scientific-modeling>
- Casner, S.M. (1991) A task-analytic approach to the automated design of graphic presentation. *ACM Transactions on Graphics* 10, 111–151.
- Clarke E. (1993). Imitating and Evaluating Real and Transformed Musical Performances. *Music Perception: An Interdisciplinary Journal*, Spring, 1993, Vol. 10, No. 3, pp. 317-341
- Clarke, E.F. (1985). Some aspects of rhythm and expression in performances of Erik Satie’s “Gnossienne No. 5”. *Music Perception*, 2, 299–328.
- Clarke, E.F. (1988). Generative principles in music performance. In: J.A. Sloboda (Ed.), *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition* (pp. 1–26). Oxford: Clarendon Press.
- Clarke, E.F., & Windsor, W.L. (2000). Real and simulated expression: A listening study. *Music Perception*, 17, 277–313.
- Collopy, F. (2009). Playing (with) color. *Glimpse J.Art+Sci.Seeing*, 2(3), 62-67
- Cooke, D. (1963) *The Language Of Music*. Oxford University Press
- Cox, R. (1999) Representation construction, externalised cognition and individual differences. *Learning and Instruction* 9, 343–363.
- Dastani, M. (2002). The Role of Visual Perception in DataVisualization. *Journal of Visual Languages and Computing* 13, 601–622
- Dixon, S., Goebel, W. & Widmer, G. (2005). The air worm: An interface for real-time manipulation of expressive music performance. *International Computer Music Conference, ICMC 2005*
- Dixon, S., Goebel, W., & Widmer, G. (2002). The performance worm: Real time visualization based on langner’s representation. In Nordahl, M. (editor) *Proceedings of the 2002 International Computer Music Conference*, pages 361–364, San Francisco, CA. International Computer Music Association.
- Feng, Y., Zhuang, Y., & Pan, Y. (2003). Music information retrieval by detecting mood via computational media aesthetics. In *Proceedings - IEEE/WIC International Conference on Web Intelligence, WI 2003* (pp. 235–241). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/WI.2003.1241199>

- Fonteles, J. H., Rodrigues, M. A. F., & Basso, V. E. D. (2013). Creating and evaluating a particle system for music visualization. *Journal of Visual Languages and Computing*, 24(6), 472–482. <https://doi.org/10.1016/j.jvlc.2013.10.002>
- Freedman, E.G. & Shah, P. (2002) Toward a Model of Knowledge-Based Graph Comprehension. In: *Hegarty, M., Meyer, B., Narayanan, N.H. (eds.) Diagrams 2002. LNCS (LNAI), vol. 2317, pp. 59–141.* Springer, Heidelberg .
- Friberg, A. (1995a). *A Quantitative Rule System for Musical Performance*. Doctoral dissertation, Department of Speech, Music and Hearing, Royal Institute of Technology, Stockholm.
- Friberg, A., Sundberg, J., & Frydén, L. (2000). Music from motion: Sound level envelopes of tones expressing human locomotion. *Journal of New Music Research*, 29, 199–210.
- Gabrielsson, A. (1999). The Performance of Music. In *The Psychology of Music* (pp. 501–602). Elsevier. <https://doi.org/10.1016/b978-012213564-4/50015-9>
- Gang, R., Lundberg, J., Bocko, M., & Headlam, D. (2010). The shape of musical sound: Real-time visualizations of expressiveness in music performance. In *Proceedings of Meetings on Acoustics* (Vol. 9). <https://doi.org/10.1121/1.3573496>
- Gasser, M. (2005) Interactive Visualization of Expressive Piano Performance. *Master Degree Thesis, Vienna University of Technology*
- Goebel, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America*, 110, 563–572.
- Hiraga, R., & Matsuda, N. (2004). Visualization of music performance as an aid to listener's comprehension. In *Proceedings of the Workshop on Advanced Visual Interfaces AVI* (pp. 103–106). <https://doi.org/10.1145/989863.989878>
- Hiraga, R., & Matsuda, N. (2004). Visualization of music performance as an aid to listener's comprehension. In *Proceedings of the Workshop on Advanced Visual Interfaces AVI* (pp. 103–106). <https://doi.org/10.1145/989863.989878>
- Hiraga, R., Mizaki, R., & Fujishiro, I. (2002). Performance visualization - A new challenge to music through visualization
- Honing, H. (2003). The final ritard: On music, emotion, and kinematic models. *Computer Music Journal*, 27, 66–72.
- Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. *IEEE Trans. Acoustics, Speech, and Signal Proc.*, Vol. ASSP-23, pp. 52-72.
- Jeong, YS. & Jeong, M. & Omitaomu, O. (2010) Weighted dynamic time warping for time series classification. *Pattern Recognition Vol.44 Issue.9*, 2231-2240.
- Katayose H., Hashida M., de Poli G. & Hirata K. (2012). On Evaluating Systems for Generating Expressive Music Performance: The Rencon Experience. *Journal of New Music Research*.
- Klemenc, B., Chiuha, P., Subelj, & Bajec, M. (2011) Visual and Aural: Visualization of Harmony in Music with Colour. *Transactions on Internet Research*, vol. 7, no. 1
- Kornfeld J. (2001). *Music notation and theory for intelligent beginners*.
- Laurier, C., Herrera, P. (2008) Mood Cloud : A Real-Time Music Mood Visualization Tool, *Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain*
- librosa.github.io (2019) librosa 0.7.2 documentation. Ανάκτηση 12 Μαρτίου 2020 από τον δικτυακό τόπο: <https://librosa.github.io/librosa/index.html>

- Mazzola, G. (1990). *Geometrie der Töne. Elemente der Mathematischen Musiktheorie*. Basel: Birkhäuser Verlag.
- McFee, B., Lostanlen, V., McVicar, M., Metsai, A., Balke, S., Thorne, C., Raffel, C., et al. *Librosa: v0.7.2*. Zenodo, 2020. doi:10.5281/zenodo.3606573.
- McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., Nieto, O. (2015). *Librosa: Audio and music signal analysis in Python*. In *Proceedings of the 14<sup>th</sup> Python In Science conference (Scipy 2015)*.
- Müller, M & Mattes ,H & Kurth ,F.(2006) An Efficient Multiscale Approach to Audio Synchroniza-tion. In *Proceedings of the 7th International Conference on Music Information Retrieval(ISMIR)*, pages 192–197.Victoria, Canada, 2006
- Müller, M. & Ewert, S. (2011). *Chroma Toolbox: MATLAB Implementations For Extracting Variants of Chroma-Based Audio Features*. Proceedings of the International Society for Music Information Retrieval Conference: 215–220.
- Müller, M., Kurth, F., & Clausen, M. (2005a). Chroma-based statistical audio features for audio matching. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (pp. 275–278). <https://doi.org/10.1109/ASPAA.2005.1540223>
- Müller, M., Kurth, F., & Clausen, M. (2005b). Audio matching via chroma-based statistical features. In *ISMIR 2005 - 6th International Conference on Music Information Retrieval* (pp. 288–295).
- Müller, M & Klapuri, A. (2013). *Music Signal Processing*. Στο *Academic Press Library in Signal Processing* (σσ. 713-756). Academic Press
- Müller, M. (2007). *Information Retrieval for Music and Motion*. Berlin , MA: Springer Science & Business Media.
- Müller, M. (2015). *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Basingstoke, England: Springer.
- MusicInformationRetrieval.com (2020) *DTW*. Ανάκτηση στις 11 Φεβρουαρίου 2020 από τον δικτυακό τόπο <https://musicinformationretrieval.com/dtw.html>
- Nanayakkara, S. C., Taylor, E., Wyse, L., & Ong, S. H. (2007). Towards building an experiential music visualizer. In *2007 6th International Conference on Information, Communications and Signal Processing, ICICS*. <https://doi.org/10.1109/ICICS.2007.4449609>
- Nowell, L., Schulman, R. & Hix, D. (2002) Graphical Encoding for Information Visualization: An Empirical Study. In: *Proceedings of the IEEE Symposium on Information Visualization 2002 (InfoVis)*.
- Palmer, C. (1996). Anatomy of a performance: Sources of musical expression. *Music Perception*, 13, 433–453.
- Randel M. (2003). "Set theory", *The Harvard Dictionary of Music*, p.776. Harvard.
- Reeves, W.T. (1983) Particle systems—a technique for modeling a class of fuzzy objects. In Proceedings of the 10th Annual Conference on Computer Graphics and Interactive Techniques, Detroit, Michigan, United States, 1983, pp. 359-375.
- Repp, B.H. (1998). A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, 104, 1085–1100.

- Repp, B.H. (1999). A microcosm of musical expression: II. Quantitative analysis of pianists' dynamics in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, 105, 1972–1988.
- Roads, C., & Strawn, J. (1996). *The Computer Music Tutorial*. Cambridge, MA: MIT Press.
- Sakoe, H. & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoustics, Speech, and Signal Proc.*, Vol. ASSP-26. pp. 43-49
- Schoonderwaldt, E., & Friberg, A. (2001). Towards a rule-based model for violin vibrato. In: C.L. Buyoli, & R. Loureiro (Eds.), *Workshop on Current Research Directions in Computer Music, November 15 17, 2001 (pp. 61–64)*. Barcelona, Spain: Audiovisual Institute, Pompeu Fabra University
- Schoonderwaldt, E., Friberg, A., Bresin, R., & Juslin, P. (2002) A system for improving the communication of emotion in music performance by feedback learning. *Journal of the Acoustical Society of America*, 111(5).
- Senin, P. (2009). *Dynamic Time Warping Algorithm Review*. Honolulu, USA
- Shaffer, L.H., Clarke, E.F., & Todd, N.P.M. (1985). Metre and rhythm in pianoplaying. *Cognition*, 20, 61–77.
- Shneiderman B. & Plaisant C. (2006). Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In: *Proceedings of BELIV'06: BEyond time and errors - novel EvaLuation methods for Information Visualization. A workshop of the AVI 2006 International Working Conference*
- Sundberg, J. (2000). Four years of research on music and motion. *Journal of New Music Research*, 29, 183–185
- Sundberg, J., Askenfelt, A., & Frydén, L. (1983). Musical performance. A synthesis-by-rule approach. *Computer Music Journal*, 7, 37–43.
- Sundberg, J., Friberg, A., & Frydén, L. (1991). Common secrets of musicians and listeners – An analysis-bysynthesis study of musical performance. In: P. Howell, R. West, & I. Cross (Eds.), *Representing Musical Structure (pp. 161–197)*. London: Academic Press.
- Sweller, J. (2002) Visualisation and Instructional Design. In: *Proceedings of the International Workshop on Dynamic Visualizations and Learning*
- Taylor, R., Boulanger, P., & Torres, D. (2006). Real-Time Music Visualization Using Responsive Imagery. *8th International Conference on Virtual Reality*, 26–30
- Timmers, R., Ashley, R., Desain, P., Honing, H., & Windsor, L.W. (2002). Timing of ornaments in the theme of Beethoven's *Paisello Variations: Empirical data and a model*. *Music Perception*, 20, 3–33.
- Tobudic, A., & Widmer, G. (2003a). Playing mozart phrase by phrase. In: K.D. Ashley, & D.G. Bridge (Eds.), *Proceedings of the 5th International Conference on Case-based Reasoning (ICCBR'03), Trondheim, Norway (pp. 552–566)*. Berlin: Springer.
- Tobudic, A., & Widmer, G. (2003b). Relational ibl in music with a new structural similarity measure. In: T. Horváth, & A. Yamamoto (Eds.), *Proceedings of the 13th International Conference on Inductive Logic Programming (ILP'03), Szeged, Hungary (pp. 365–382)*. Berlin: Springer.
- Todd, N.P.M. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America*, 97, 1940–1949.



- Todd, N.P.M. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America*, 97, 1940–1949.
- Tufte, E.R. (2001). *The Visual Display of Quantitative Information*, 2nd edn. Graphics Press.
- Vekiri, I. (2002) What Is the Value of Graphical Displays in Learning? *Educational Psychology Review* 14, 261–312.
- Wattenberg, M., Fisher, D. (2004). Analyzing perceptual organization in information graphics. *Information Visualization* 3, 123–133
- Widmer G. & Goebel W. (2004) Computational Models of Expressive Music Performance: The State of the Art. *Journal of New Music Research*, Vol. 33, No.3, σσ, 203-216.
- Widmer, G. (2000). Large-scale induction of expressive performance rules: first quantitative results. In: I. Zannos (Ed.), *Proceedings of the 2000 International Computer Music Conference, Berlin, Germany (pp. 344–347)*. San Francisco, CA: International Computer Music Association.
- Widmer, G. (2002). Machine discoveries: A few simple, robust local expression principles. *Journal of New Music Research*, 31, 37–50.
- Widmer, G. (2003). Discovering simple rules in complex data: A meat-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146, 129–148.
- Widmer, G., & Tobudic, A. (2003). Playing Mozart by analogy: Learning multi-level timing and dynamics strategies. *Journal of New Music Research*, 32, 259–268.
- Widmer, G., Dixon, S.E., Goebel, W., Pampalk, E., & Tobudic, A. (2003). In: search of the Horowitz factor. *AI Magazine*, 24, 111–130.
- Wikipedia.org (2020a) *Music Alignment*. Ανάκτηση 19 Ιανουαρίου 2020 από τον δικτυακό τόπο [https://en.wikipedia.org/wiki/Music\\_alignment](https://en.wikipedia.org/wiki/Music_alignment)
- Wikipedia.org (2020b) *Chroma Feature*. Ανάκτηση 18 Απριλίου 2020 από τον δικτυακό τόπο [https://en.wikipedia.org/wiki/Chroma\\_feature](https://en.wikipedia.org/wiki/Chroma_feature)
- Wikipedia.org (2020c) *Dynamic Time Warping*. Ανάκτηση 19 Ιανουαρίου 2020 από τον δικτυακό τόπο [https://en.wikipedia.org/wiki/Dynamic\\_time\\_warping](https://en.wikipedia.org/wiki/Dynamic_time_warping)
- Wikipedia.org (2020d). *Information visualization*. Ανάκτηση 5 Μαΐου 2020 από τον δικτυακό τόπο [https://en.wikipedia.org/wiki/Information\\_visualization](https://en.wikipedia.org/wiki/Information_visualization)
- Wikipedia.org (2020e), *Pygame*. Ανάκτηση 9 Ιουνίου 2020 από τον δικτυακό τόπο <https://en.wikipedia.org/wiki/Pygame>
- Williams D. A. H. (2019). Evaluating BCI for Musical Expression: Historical Approaches, Challenges and Benefits. *Brain Art* 2019.
- Windsor, W.L., & Clarke, E.F. (1997). Expressive timing and dynamics in real and artificial musical performances: Using an algorithm as an analytical tool. *Music Perception*, 15, 127–152.
- Zanon, P., & De Poli, G. (2003a). Estimation of parameters in rule systems for expressive rendering in musical performance. *Computer Music Journal*, 27, 29–46.

## Παράρτημα 1 – Υλοποίηση αλγορίθμου DTW σε Python

Στην παρακάτω υλοποίηση έχουμε ως δεδομένα εισόδου δύο πίνακες chr1 και chr2 που αντιστοιχούν στα CENS features δύο αρχείων ήχου.

```
#-----Cost Matrix-----
xt=np.transpose(chr1)
yt=np.transpose(chr2)
C = scipy.spatial.distance.cdist(yt,xt, 'cosine')

#-----Accumulated Cost Matrix-----
@jit(nopython=True)
def compute_accumulated_cost_matrix(C):
    N, M = C.shape
    D = np.zeros((N, M))
    D[:, 0] = np.cumsum(C[:, 0])
    D[0, :] = np.cumsum(C[0, :])
    for n in range(1, N):
        for m in range(1, M):
            D[n, m] = C[n, m] + min(D[n-1, m], D[n, m-1], D[n-1, m-1])
    return D

D = compute_accumulated_cost_matrix(C)

#-----Warping Path-----
@jit(nopython=True)
def compute_optimal_warping_path(D):

    N, M = D.shape
    n = N - 1
    m = M - 1
    P = [(n, m)]
    while n > 0:
        if m == 0:
            cell = (n - 1, 0)
        else:
            val = min(D[n-1, m-1], D[n-1, m], D[n, m-1])
            if val == D[n-1, m-1]:
                cell = (n-1, m-1)
            elif val == D[n-1, m]:
                cell = (n-1, m)
            else:
                cell = (n, m-1)
        P.append(cell)
        n, m = cell
    P.reverse()
    return np.array(P)

path = compute_optimal_warping_path(D)
```

## Παράρτημα 2 – Υλοποίηση οπτικοποίησης σε Python

Τα δεδομένα εισόδου της συνάρτησης particles είναι:

S : Ο κανονικοποιημένος πίνακας με τις διαφορές στις δυναμικές των δύο αρχείων ήχου  
fs : Ο ρυθμός δειγματοληψίας των δύο αρχείων ήχου  
dev, dev\_sec : Οι πίνακες των χρονικών αποκλίσεων ανάμεσα στα δύο αρχεία ήχου, σε κανονικοποιημένη και μη κανονικοποιημένη μορφή  
point : Η μέγιστη χρονική απόκλιση  
namea, nameb : Οι ερμηνευτές του πρώτου και του δεύτερου αρχείου αντίστοιχα  
title : Ο τίτλος του μουσικού έργου

---

```
FPS = 60
```

```
WINDOWWIDTH = 800
```

```
WINDOWHEIGHT = 600
```

```
FIRE_YELLOW = pygame.image.load('fire_yellow.png')
```

```
linecolor=(255,0,0)
```

```
def particles(S,fs,dev,dev_sec,point,namea,nameb,title):
```

```
    global FPSCLOCK, DISPLAYSURF
```

```
    pygame.init()
```

```
    FPSCLOCK = pygame.time.Clock() #time counter
```

```
    DISPLAYSURF = pygame.display.set_mode((WINDOWWIDTH, WINDOWHEIGHT))
```

```
    pygame.mouse.set_visible(0)
```

```
    # particle_xysize Elements:
```

```
    # Its a List of Lists, where particle_xysize[element][0,1,2,3,4,5,6..]
```

```
    # 0=x
```

```
    # 1=y
```

```
    # 2=size(squared-same x,y size)
```

```
    # 3,4=direction
```

```
    # 5=type
```

```
    # 6=dynamics aspect
```

```
    # 7=color(RGB)
```

```
    #8=influence
```

```
    particles = 720
```

```
    particle_xysize = []
```

```
    while particles > 0:
```

```
        particle_xysize.append([0,0,0,0,0,0,0,(0,0,0),(0,0)])
```

```
        particles -= 1
```

```
    velocity = []
```

```
    for particle in particle_xysize:
```

```
        velocity.append(5)
```

```
    init_x = 400
```

```
    init_y = 600
```

```

pygame.mixer.music.load(file2)
pygame.mixer.music.play(0)
color2=(255,255,0)
t=0
a = datetime.datetime.now()
while True:

    # Get Events of Game Loop
    for event in pygame.event.get():
        if event.type == QUIT or (event.type == KEYUP and event.key == K_ESCAPE):
            pygame.mixer.music.stop()
            pygame.quit()
            sys.exit()

    # Drawing Axis
    DISPLAYSURF.fill((0, 0, 0))
    pygame.draw.line(DISPLAYSURF,linecolor,(400,0),(400,600))
    pygame.draw.line(DISPLAYSURF,linecolor,(0,300),(800,300))
    pygame.draw.line(DISPLAYSURF,linecolor,(100,290),(100,310))
    pygame.draw.line(DISPLAYSURF,linecolor,(175,295),(175,305))
    pygame.draw.line(DISPLAYSURF,linecolor,(250,290),(250,310))
    pygame.draw.line(DISPLAYSURF,linecolor,(325,295),(325,305))
    pygame.draw.line(DISPLAYSURF,linecolor,(700,290),(700,310))
    pygame.draw.line(DISPLAYSURF,linecolor,(625,295),(625,305))
    pygame.draw.line(DISPLAYSURF,linecolor,(550,290),(550,310))
    pygame.draw.line(DISPLAYSURF,linecolor,(475,295),(475,305))

    #Displaying Info
    myFont = pygame.font.SysFont("Helvetica", 20)
    song_title=myFont.render(title, 1,color2)
    label2=myFont.render("By : ", 1,linecolor)
    label3=myFont.render("Compared to : ", 1,linecolor)
    pianist = myFont.render(nameb, 1,color2)
    compared = myFont.render(namea, 1,color2)
    DISPLAYSURF.blit(song_title, (3, 3))
    DISPLAYSURF.blit(label2, (3, 22))
    DISPLAYSURF.blit(pianist, (35, 22))
    DISPLAYSURF.blit(label3, (3, 42))
    DISPLAYSURF.blit(compared, (112, 42))

    # Draw Elements
    for element in range(len(particle_xysize)):
        width = particle_xysize[element][2]
        height = particle_xysize[element][2]
        particle_x = particle_xysize[element][0]
        particle_y = particle_xysize[element][1]
        dynamics = particle_xysize[element][6]
        influence = particle_xysize[element][8]

```

```

color = particle_xysize[element][7]

particle_x += velocity[element] * particle_xysize[element][4] * dynamics
particle_y += velocity[element] * particle_xysize[element][3] * dynamics

if particle_xysize[element][5] == 0:
    pygame.draw.ellipse(DISPLAYSURF, color, (particle_x - width , particle_y - height ,
width*2, height*2))
elif particle_xysize[element][5] == 1:
    fire_yellow = pygame.transform.scale(FIRE_YELLOW, (int(width * 2), int(height *
2)))
    DISPLAYSURF.blit(fire_yellow,[particle_x - width,particle_y - height])

if particle_xysize[element][2] > 0:
    particle_xysize[element][2] -= 0.5
    velocity[element] += 2
else:
    while True:
        b = datetime.datetime.now()
        c = b - a
        t=c.total_seconds()

        #Displaying Time and Deviation
        randNumLabel = myFont.render("Time:(sec)", 1, linecolor)
        timeDisplay = myFont.render(str(round(t,2)), 1,color2)
        DISPLAYSURF.blit(randNumLabel, (600, 3))
        DISPLAYSURF.blit(timeDisplay, (600, 23))
        idx=librosa.time_to_frames(t,sr=fs, hop_length=512)
        val=S[idx]
        min_dev = myFont.render(str(-1*point), 1, color2)
        half=round(point/2)
        half_minus = myFont.render(str(-1*half), 1, color2)
        DISPLAYSURF.blit(min_dev, (75, 260))
        DISPLAYSURF.blit(half_minus, (230, 260))
        max_dev = myFont.render(str(point), 1, color2)
        half_plus = myFont.render(str(half), 1, color2)
        DISPLAYSURF.blit(max_dev, (680, 260))
        DISPLAYSURF.blit(half_plus, (535, 260))
        text3 = myFont.render("Deviation(sec)", 1, color2)
        DISPLAYSURF.blit(text3, (5, 310))
        text1=myFont.render("Temporal Deviation (sec)",1,linecolor)
        DISPLAYSURF.blit(text1, (600, 43))
        timeDev = myFont.render(str(round(dev_sec[idx-1],2)), 1,color2)
        DISPLAYSURF.blit(timeDev, (600, 63))

```

```
if idx==(len(S)-1):
    pygame.mixer.music.stop()
    pygame.quit()
    sys.exit()

particle_xysize[element][3] = -1
particle_xysize[element][4] = dev[idx]
particle_xysize[element][5] = 1
particle_xysize[element][7] = (80,100,150)
particle_xysize[element][6] = val
particle_xysize[element][8] = (1,1)
particle_xysize[element][2] = random.randint(1, 15)
velocity[element] = 0
particle_xysize[element][0], particle_xysize[element][1] = init_x, init_y
break

pygame.display.update()
FPSLOCK.tick(FPS)
```