



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ

Εξέταση Τεχνικών Αυτόματης Μεταγραφής  
Ακουστικού Σήματος Κιθάρας σε Συμβολική  
Αναπαράσταση με Χρήση Μεθόδων Ψηφιακής  
Επεξεργασίας Σήματος και Μηχανικής  
Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΣΤΕΦΑΝΟΥ-ΙΩΑΝΝΗ Ε. ΚΟΥΤΟΥΠΗ

Επιβλέπων: Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2021

---





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας

Εξέταση Τεχνικών Αυτόματης Μεταγραφής  
Ακουστικού Σήματος Κιθάρας σε Συμβολική  
Αναπαράσταση με Χρήση Μεθόδων Ψηφιακής  
Επεξεργασίας Σήματος και Μηχανικής  
Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΣΤΕΦΑΝΟΥ-ΙΩΑΝΝΗ Ε. ΚΟΥΤΟΥΠΗ

Επιβλέπων: Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 16 Ιουλίου 2021.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....  
Πέτρος Μαραγκός  
Καθηγητής  
Ε.Μ.Π.

.....  
Αλέξανδρος Ποταμιάνος  
Αναπληρωτής Καθηγητής  
Ε.Μ.Π.

.....  
Βασίλης Κατσούρος  
Ερευνητής Α΄  
Διευθυντής Ερευνών Ι.Ε.Λ.

Αθήνα, Ιούλιος 2021





Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.  
Στέφανος-Ιωάννης Ε. Κουτούπης, 2021.

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

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

(Υπογραφή)

.....  
Στέφανος-Ιωάννης Ε.

Κουτούπης

16 Ιουλίου 2021



## Περίληψη

---

Βασικός στόχος αυτής της εργασίας είναι η παρουσίαση ενός συστήματος αυτόματης μεταγραφής μονοφωνικών μουσικών εκτελέσεων κιθάρας σε ταμπλατούρα. Το πρόβλημα της μεταγραφής αντιμετωπίζεται με τη χρήση επιμέρους υποσυστημάτων επεξεργασίας για την ανίχνευση έναρξης νότας (onset), την αναγνώριση τονικού ύψους και την ταξινόμηση ενεργής χορδής. Κύρια συνεισφορά της εργασίας αποτελεί ο συνδυασμός αναρμονικής ανάλυσης και περιορισμών εκτελεσιμότητας για την αναγνώριση χορδής κιθάρας. Ο υπολογισμός του συντλεστή αναρμονικότητας για επιμέρους στιγμιότυπα νοτών επιτρέπει ένα πρώτο στάδιο αναγνώρισης χορδής με βάση το ηχητικό σήμα. Προηγείται μια σύντομη φάση συλλογής ολιγάριθμων δειγμάτων νοτών από το χρήστη για την μετέπειτα προσαρμογή του ταξινομητή. Έτσι, καθίσταται δυνατό να προσαρμοστεί το σύστημα σε κάθε κιθάρα παρά τα διαφορετικά φυσικά χαρακτηριστικά της. Παρατίθενται τέσσερις διαφορετικές μέθοδοι, με κλιμακωτή ακρίβεια και απαιτούμενο χρόνο προσαρμογής, ανάλογα με τη χρήση που επιθυμεί κάθε εκτελεστής. Ως δεύτερο στάδιο, παρουσιάζεται ένας γενετικός αλγόριθμος που λαμβάνει υπόψη φυσικούς περιορισμούς και περιορισμούς εκτελεσιμότητας για τη βελτίωση της ακρίβειας αναγνώρισης χορδής. Ο γενετικός αλγόριθμος αποτελεί το τελευταίο σε σειρά υποσύστημα και αποδεικνύεται ιδιαίτερα επωφελής για την επίδοση του συστήματος συνολικά, ενώ δεν προϋποθέτει καμία πρότερη ρύθμιση από τον χρήστη. Ακόμη, αξιοποιούνται υποσυστήματα για την ανίχνευση έναρξης νοτών με χρήση Χρονικών Συνελικτικών Δικτύων (TCN) ενώ το τονικό ύψος εκτιμάται με το προ-εκπαιδευμένο Συνελικτικό Νευρωνικό Δίκτυο (CNN) CREPE. Η προτεινόμενη μέθοδοι αποτιμώνται σε επισημειωμένες (annotated) ηχογραφήσεις πραγματικών εκτελέσεων κιθάρας αλλά και σε μεμονωμένες ηχογραφήσεις νοτών.

## Λέξεις Κλειδιά

μεταγραφή ακουστικού σήματος κιθάρας σε ταμπλατούρα, γενετικός αλγόριθμος, αναρμονικότητα, περιορισμοί εκτελεσιμότητας





# Abstract

---

The main goal of this thesis is to introduce a modular approach towards guitar tablature transcription in the case of monophonic (no chords) audio recordings. Ideas were employed from previous work for the onset detection and pitch estimation. The tablature transcription problem is tackled by employing different sub-modules for the onset detection, the pitch estimation and the string classification. The main contribution is the combination of a novel method for guitar string classification based on inharmonicity analysis and a module incorporating playability constraints. Inharmonicity computation of the distinct note instances facilitates a first step for string classification. An agile few-sample adaptation phase is introduced as a significant phase before the classification. Relying on just a small set of samples of different note recordings, we make it possible for guitar players to adapt the system to their guitar on the spot, regardless of the instrument's particular physical characteristics. Four distinct methods are proposed, with different accuracy and time needed for adaptation, in order to meet the various needs that each performer demands. As a second stage, physical limitations and common standards of human performers are incorporated by employing a genetic algorithm which significantly improves accuracy. The genetic algorithm is the last module and has proven to be very beneficial for the system's accuracy while not needing any training or adaptation. Moreover, a method for guitar audio onset detection using Temporal Convolutional Neural Networks (TCNs) is employed based on previous work, while pitch is estimated with the CREPE pre-trained Convolutional Neural Network. The proposed methods are evaluated on annotated guitar performances and isolated note recordings.

## Keywords

Guitar tablature transcription, genetic algorithm, inharmonicity, playability constraints



*στον Στέλιο Καζατζίδη*



## Ευχαριστίες

---

Θα ήθελα καταρχήν να ευχαριστήσω τον καθηγητή κ. Πέτρο Μαραγκό για την επίβλεψη αυτής της διπλωματικής εργασίας. Επίσης ευχαριστώ ιδιαίτερα τον υποψήφιο διδάκτορα Γρηγόρη Μπάστα για την καθοδήγησή του και την εξαιρετική συνεργασία που είχαμε χωρίς την οποία τίποτα από τα παρακάτω δεν θα ήταν δυνατό. Ακόμα, θα ήθελα να ευχαριστήσω τους Μάξιμο Καλιακάτσο-Παπακώστα και Βασίλη Κατσούρο για την πολύτιμη συνεισφορά και τις συμβουλές που πρόσφεραν καθ' όλη τη διάρκεια της διπλωματικής. Τέλος θα ήθελα να ευχαριστήσω την εξεταστική επιτροπή της διπλωματικής μου για το χρόνο της και την τόσο σύντομη ανταπόκριση.

Αθήνα, Ιούλιος 2021

*Στέφανος-Ιωάννης Ε. Κουτούπης*



# Περιεχόμενα

---

<b>Περίληψη</b>	<b>1</b>
<b>Abstract</b>	<b>3</b>
<b>Ευχαριστίες</b>	<b>7</b>
<b>1 Εισαγωγή</b>	<b>15</b>
1.1 Σκοπός, Συνεισφορές και Διάρθρωση της Εργασίας	17
<b>2 Προϋπάρχουσες Μελέτες</b>	<b>19</b>
2.1 Παραγωγή Εκτελέσιμης Ταμπλατούρας	19
2.2 Εξαγωγή ακριβούς ταμπλατούρας από ηχητικό αρχείο	21
<b>3 Θεωρητικό Υπόβαθρο</b>	<b>23</b>
3.1 Μουσική	23
3.2 Δομή της κιθάρας	24
3.2.1 Τυπικές Μέθοδοι Αναπαράστασης	25
3.3 Η χορδή ως Στάσιμο κύμα	25
3.3.1 Αναρμονικότητα	26
3.4 Music Information Retrieval (MIR)	31
3.4.1 Onset Detection	31
3.4.2 Εκτίμηση Τονικού Ύψους (pitch estimation)	31
3.5 Επεξεργασία Σήματος και Αναπαραστάσεις	33
3.6 Μηχανική Μάθηση	34
3.6.1 Επιβλεπόμενη Μάθηση	34
3.6.2 Μη επιβλεπόμενη μάθηση	35
3.6.3 Ενισχυτική Μάθηση	35
3.6.4 Νευρωνικά Δίκτυα	35
3.7 Γενετικοί αλγόριθμοι	36
<b>4 Περιγραφή Μεθόδου</b>	<b>39</b>
4.1 Συνοπτική Παρουσίαση	39
4.2 Υπολογισμός Αναρμονικότητας	41
4.2.1 Ταξινομητής χορδής βασισμένος σε αναρμονική ανάλυση	42
4.2.2 Εισάγοντας Περιορισμούς Εκτελεσιμότητας - Γενετικός Αλγόριθμος	44

<b>5</b>	<b>Αξιολόγηση</b>	<b>49</b>
5.1	Σύνολα Δεδομένων (Datasets)	49
5.2	Προσαρμογή στην Κιθάρα	50
5.3	Αποτελέσματα και Αξιολόγηση	50
5.3.1	Πείραμα σε μεμονωμένα ηχογραφημένα στιγμιότυπα νοτών	50
5.3.2	Αναγνώριση Χορδής σε Ολόκληρα κομμάτια κιθάρας	51
5.3.3	Αξιολόγηση Γενετικού Αλγορίθμου	55
5.3.4	Αποτελέσματα Αξιολόγησης Αυστηρής Μεταγραφής Ταμπλατούρας	59
5.4	Μετατροπή σε MIDI και μουσική σύνθεση	60
<b>6</b>	<b>Επίλογος</b>	<b>61</b>
6.1	Συμπεράσματα	61
6.2	Μελλοντικές Επεκτάσεις	61
	<b>Παραρτήματα</b>	<b>63</b>
	<b>Α' Εκτίμηση τονικότητας με το CREPE και ανίχνευση ενάρξεων με TCN</b>	<b>65</b>
	Α'.1 Εκτίμηση τονικότητας με το CREPE	65
	Α'.2 Αναγνώριση έναρξης νότας με TCN	66
	<b>Β' Τεχνολογίες που χρησιμοποιήθηκαν</b>	<b>67</b>
	<b>Βιβλιογραφία</b>	<b>72</b>
	<b>Συντομογραφίες - Αρκτικόλεξα - Ακρωνύμια</b>	<b>73</b>
	<b>Απόδοση ξενόγλωσσων όρων</b>	<b>75</b>



## Κατάλογος Σχημάτων

---

1.1	Ταστιέρα μιας κιθάρας. . . . .	15
1.2	Μια παρτιτούρα και ταμπλατούρα για την ίδια μελωδία. Σε αντίθεση με την παρτιτούρα, η ταμπλατούρα μπορεί να παρουσιάζει δυο τελείως διαφορετικές αποτυπώσεις για την ίδια μελωδία. . . . .	16
3.1	Νότες πάνω στη κιθάρα για κανονικό κούρδισμα . . . . .	24
3.2	Παράδειγμα Παρτιτούρας . . . . .	25
3.3	Παράδειγμα Ταμπλατούρας . . . . .	25
3.4	Συχνοτική ανάλυση μιας χορδής χτυπημένη στο κέντρο της. Οι τρόποι ταλάντωσης μονής τάξεως, αθροίζονται σε φάση και πλάτος για να προκύψει τελικά η μορφή της χορδής. Πηγή :[1] . . . . .	26
3.5	Οι διακεκομμένες γραμμές αντιπροσωπεύουν τις τιμές των αρμονικών. Καθώς η τάξη των <i>partial</i> μεγαλώνει, εξίσου αυξάνεται και η απόκλισή τους από τις τιμές των αρμονικών. . . . .	27
3.6	Αλγόριθμος MAT αριστερά και δεξιά σχήμα για τις 3 πρώτες επαναλήψεις [2] . . . . .	29
3.7	Αλγόριθμος και διάγραμμα για την εκτίμηση του συντελεστή αναρμονικότητας με την μέθοδο που παρουσιάζεται στο [3] (διάγραμμα από το ίδιο) . . . . .	30
3.8	αποτύπωση της απόκλισης των αρμονικών σε σχέση με τις θεωρητικές τιμές τους για $\beta = 10^{-4}$ , $f_0 = 120\text{Hz}$ . Οι κάθετες γραμμές αποτυπώνουν το μήκος του παραθύρου αναζήτησης. . . . .	32
3.9	Αριστερά η περιβάλλουσα πλάτους για πιάνο δεξιά για βιολί. Πηγή [4] . . . . .	33
3.10	Φασματογράφημα ενός κομματιού κιθάρας 20 δευτερολέπτων . . . . .	34
3.11	Διασταύρωση δυο γονέων, A-B και C-D όπου προκύπτουν τα παιδιά A-D και C-D . . . . .	36
3.12	Μετάλλαξη ενός ατόμου . . . . .	37
3.13	Παράδειγμα ενός πιθανού τρόπου διασταύρωσης και μεταλλαγής σε περίπτωση ταμπλατούρας. . . . .	37
4.1	Απο άκρη-σε-άκρη σύστημα μεταγραφής ακριβούς ταμπλατούρας . . . . .	39
4.2	Εξέλιξη του συντελεστή αναρμονικότητας $\beta$ για κάθε <i>midi</i> νότα βασισμένη στη σχέση (1.2) για της μεθόδους προσαρμογής 1Fret (συμπαγής) και 3Fret (διακεκομένη). . . . .	40
4.3	Παράδειγμα αντιστοίχισης στιγμιότυπου νότας με βάση την αναρμονικότητα. Η νότα E4 που εξετάζεται, τελικά αντιστοιχείζεται στη χορδή με τη πλησιέστερη αναρμονικότητα δηλαδή την B. . . . .	41

4.4	Περιπτώσεις ασαφών νοτών: (α) θορυβώδες, (β) απόκλιση προς τα πίσω, (ς) χαμηλής έντασης αρμονικές και (δ) λανθασμένη θεμελιώδης. . . . .	42
4.5	Εξέλιξη του $n'$ σε σχέση με το τάστο $n$ σύμφωνα με τον υπολογισμό του $\beta$ για κάθε χορδή. Για κάθε πραγματικό ζεύγος χορδής τάστου, το $n'$ υπολογίζεται από τη διάμεσο των αντίστοιχων $\beta$ όλων των στιγμιότυπων νοτών που υπάρχουν στο <i>GuitarSet dataset</i> . . . . .	44
4.6	Γράφημα που αναπαριστά όλες τις πιθανές ταμπλατούρες που αντιστοιχούν στην ακολουθία των νοτών 50 - 52 - 55 (αρίθμηση MIDI. . . . .	45
4.7	Ο γενετικός αλγόριθμος αντιστοιχεί τα ασαφή στιγμιότυπα σε χορδές και διορθώνει τις μη διαισθητικές περιπτώσεις. . . . .	46
5.1	Πίνακες σύγχυσης για τις ηχογραφήσεις με μαγνήτη ( <i>pickup</i> ) . . . . .	53
5.2	Πίνακες σύγχυσης για τις ηχογραφήσεις με μικρόφωνο ( <i>microphone</i> ) . . . . .	54
5.3	Γράφημα που παρουσιάζει τη βελτίωση που προσδίδει ο γενετικός αλγόριθμος όταν ένα ποσοστό των στιγμιότυπων νοτών έχουν αντιστοιχηθεί στις σωστές χορδές και όλες οι υπόλοιπες είναι χαρακτηρισμένες ασαφείς. . . . .	55
5.4	Γράφημα που παρουσιάζει τη βελτίωση που προσδίδει ο γενετικός αλγόριθμος όταν ένα ποσοστό των στιγμιότυπων νοτών έχουν αντιστοιχηθεί στις σωστές χορδές και όλες οι υπόλοιπες έχουν αντιστοιχηθεί τυχαία. . . . .	56
5.5	Γράφημα που παρουσιάζει την απόδοση του γενετικού αλγόριθμου στην περίπτωση που το 10% των στιγμιότυπων έχει αποδοθεί σε λάθος τόνο και αναλόγως το ποσοστό ασαφών. . . . .	57
5.6	Γράφημα που παρουσιάζει την απόδοση του γενετικού αλγόριθμου στην περίπτωση που το 25% των στιγμιότυπων έχει αποδοθεί σε λάθος τόνο και αναλόγως το ποσοστό ασαφών. . . . .	57
5.7	Πίνακες σύγχυσης του γενετικού αλγόριθμου για τις ηχογραφήσεις με μαγνήτη ( <i>pickup</i> ) . . . . .	58
5.8	Εικόνα από το <i>FL studio</i> όπου εμφανίζεται κάθε χορδή κωδικοποιημένη σε διαφορετικό κανάλι <i>MIDI</i> . . . . .	60
A'.1	Αρχιτεκτονική του <i>CREPE</i> [5] . . . . .	65
A'.2	Αρχιτεκτονική του <i>Temporal Convolutional Network (TCN)</i> . . . . .	66

## Κατάλογος Πινάκων

---

5.1	<i>Ακρίβεια του συστήματος ηχητικής ταξινόμησης με αναρμονική ανάλυση στα δεδομένα από το dataset των Hjerrild και Christensen [6]. . . . .</i>	51
5.2	<i>Ακρίβεια της μεθόδου βασισμένη σε αναρμονική ανάλυση, για τις μονοφωνικές εκτελέσεις από το GuitarSet dataset. . . . .</i>	51
5.3	<i>Πίνακας αποτελεσμάτων για τη σύγκριση της μεθόδου ηχητικής ταξινόμησης λίγων δειγμάτων με έναν Naive Bayes ταξινομητή σε απομονωμένα στιγμιότυπα νοτών κομμένα από τις ηχογραφήσεις του GuitarSet Dataset(Η ακρίβεια των αποτελεσμάτων, διαφέρει με τους άλλους πίνακες καθώς δεν εξετάζεται το σύνολο των τάσεων και τονικών υψών). . . . .</i>	54
5.4	<i>Αποτελέσματα για την ολοκληρωμένη αυστηρή μεταγραφή ταμπλατούρας για τις μεθόδους προσαρμογής με την μεγαλύτερη ακρίβεια. . . . .</i>	59



## Κεφάλαιο 1

### Εισαγωγή

---

**Η** αυτόματη μεταγραφή μουσικής συνιστά το πρόβλημα της εξαγωγής τυπικής μουσικής γραφής από ένα μουσικό σήμα. Ενώ η παρτιτούρα αποτελεί το κυρίαρχο εργαλείο για σημειογραφία στη δυτική μουσική, στην περίπτωση της κιθάρας, η ταμπλατούρα έχει αποκτήσει ιδιαίτερη δημοτικότητα ειδικά σε αυτοδίδαχτους και νέους κιθαρίστες.



Σχήμα 1.1: Ταστιέρα μιας κιθάρας.

Η ταμπλατούρα αναπαριστά ακολουθίες μουσικών νοτών υποδεικνύοντας τις θέσεις των δαχτύλων πάνω στη ταστιέρα (βλ. Σχήμα 1.1). Πιο συγκεκριμένα, αναπαριστά πάνω σε 6 γραμμές τη σειρά των χορδών-τάστων που αντιστοιχούν στις μουσικές νότες ενός κομματιού (βλ. Σχήμα 1.2). Αντίθετα, οι παρτιτούρες δεν περιλαμβάνουν την πληροφορία της χορδής και του τάστου, αλλά του τονικού ύψους, ενώ παράλληλα διατηρούν πληροφορίες για την ρυθμική δομή του κομματιού. Οι παραπάνω ιδιότητες καταστούν τις ταμπλατούρες πολύτιμες σε ποικίλες περιπτώσεις. Η ανάγνωσή τους είναι ευκολότερη για κάποιο άτομο χωρίς μουσική εκπαίδευση, ακόμη και αν η κατανόησή τους εξαρτάται άμεσα από την μουσική αντίληψη του αναπαριστώμενου κομματιού. Επιπροσθέτως, εφόσον καταγράφεται η αναπαράσταση σε χορδές-τάστα, μπορεί να βοηθήσει για την αναπαραγωγή εκφραστικών και μη τεχνικών, αλλά και για εκπαιδευτικούς σκοπούς.

Η ανάπτυξη ενός συστηματικού τρόπου για αυστηρή μεταγραφή ταμπλατούρας από μια δοσμένη ηχογράφιση κιθάρας, αποτελεί στόχο μεγάλης σημασίας αλλά και ταυτόχρονα μια μεγάλη πρόκληση. Λόγω της γεωμετρίας της κιθάρας, ίδιου τονικού ύψους νότες μπορούν να



Σχήμα 1.2: Μια παρτιτούρα και ταμπλατούρα για την ίδια μελωδία. Σε αντίθεση με την παρτιτούρα, η ταμπλατούρα μπορεί να παρουσιάζει δυο τελείως διαφορετικές αποτυπώσεις για την ίδια μελωδία.

παιχτούν σε παραπάνω από μια θέσεις στην ταστιέρα - δεδομένου κανονικού κουρδίσματος - όπως φαίνεται στο Σχήμα 3.1. Αυτή η παρατήρηση αναδεικνύει και τον κύριο στόχο για την αυστηρή μεταγραφή ταμπλατούρας, αφού ο προσδιορισμός του τονικού ύψους δεν επαρκεί, και αποδεικνύεται αναγκαία η αναζήτηση πιο σύνθετων φασματικών χαρακτηριστικών.

Το ηχόχρωμα ενός στιγμιότυπου νότας οποιουδήποτε μουσικού οργάνου μπορεί να αναλυθεί ως μια υπέρθεση ημιτόνων με διάφορα πλάτη και συχνότητες. Οι συχνότητες αυτές θεωρούνται αρμονικές όταν βρίσκονται ιδανικά στα ακέραια πολλαπλάσια της θεμελιώδους συχνότητας ( $f_0$ ). Σε πολλά νυχτά έγχορδα όργανα (νυχτά ονομάζονται τα έγχορδα μουσικά όργανα που παράγεται ο ήχος τους με το τράβηγμα των χορδών σε αντίθεση με τα τοξωτά που παράγεται με τη χρήση δοξαριού) όμως, όπως η κιθάρα και το πιάνο, παρατηρείται μια απόκλιση των μερικών συχνοτήτων (partials) από τις αντίστοιχες αρμονικές συχνότητες (βλ. Σχήμα 3.5). Όπως αναφέρεται στο [7], αυτή η απόκλιση εξαρτάται από την ακαμψία (stiffness) της χορδής και του τεντώματος (stretching) κατά το χτύπημα. Η συνιστώσα του τεντώματος κατά το χτύπημα, είναι μικρότερης σημασίας, πολύ σύντομα, καθώς η επιρροή της αρχικής διέγερσης, σβήνει γρήγορα [8]. Η τιμή της  $k$ -οστής μερικής συχνότητας (partial) μπορεί να υπολογιστεί ως:

$$f_k = k \cdot f_0 \cdot \sqrt{1 + \beta \cdot k^2}, \quad (1.1)$$

όπου το  $\beta$  ονομάζεται συντελεστής αναρμονικότητας (inharmonic coefficient). Ο συντελεστής διαφέρει μεταξύ τάσεων  $n$  της ίδιας χορδής  $s$  και ακολουθεί την παρακάτω σχέση:

$$\beta(s, n) = \beta(s, 0) \cdot 2^{\frac{n}{8}}. \quad (1.2)$$

Στα πλαίσια της διπλωματικής, ο υπολογισμός του  $\beta$  χρησιμοποιείται ως χαρακτηριστικό (feature) για την αναγνώριση της χορδής. Για την εξαγωγή χαρακτηριστικών συναφή με την αναγνώριση χορδής, οι περισσότερες προσεγγίσεις βασίζονται σε ένα μεγάλο δείγμα ηχογραφημένων νοτών μόνο από μια ή λίγες κιθάρες. Η αντιμετώπιση τέτοιων περιπτώσεων, περιορίζεται στα συγκεκριμένα όργανα που χρησιμοποιούνται για την εκπαίδευση τέτοιων συστημάτων. Μοντέλα λίγων δειγμάτων (few sample models) και μοντέλα βασισμένα σε προσομοίωση (simulation-based models) για αναγνώριση χορδής έχουν προσφάτως προταθεί για

την αντιμετώπιση του προβλήματος στα [6] και [8] αντίστοιχα. Στην παρούσα διπλωματική εργασία παρουσιάζεται μια προσέγγιση λίγων δειγμάτων και εύκολα προσαρμόσιμη μέθοδος και διερευνώνται μερικές περιπτώσεις προσαρμογής της με διαφορετικά αποτελέσματα στην επίδοση.

Σε ένα μουσικό κομμάτι, οι θέσεις στην ταστιέρα των ακολουθιακά παιγμένων νοτών δεν πρέπει να αντιμετωπίζονται ως ανεξάρτητα συμβάντα. Ο εκτελεστής τείνει να χρησιμοποιεί ευκολότερες δακτυλοθεσίες και θέσεις στην ταστιέρα ενσωματώνοντας στο παίξιμό του κριτήρια όπως η ελαχιστοποίηση απόστασης, η διατήρηση θέσης και σχήματος δακτυλισμού κ.λπ. Ένα σύστημα με στόχο να προβλέψει ακριβή ταμπλατούρα, μπορεί να ευνοηθεί από την θεμελίωση τέτοιων περιορισμών και την εισαγωγή κριτηρίων εκτελεσιμότητας, απλότητας, άνεσης κ.λπ. Έτσι, ενσωματώθηκαν φυσικοί περιορισμοί στην συνάρτηση καταλληλότητας ενός γενετικού αλγορίθμου (GA), ο οποίος έχει στόχο να αντιστοιχίσει τα ασαφή στιγμιότυπα νοτών (βλ. κεφάλαιο 5) σε χορδές αλλά και να τροποποιεί τις προβλέψεις του ταξινομητή αναρμονικής ανάλυσης, με στόχο να βελτιώσει την επίδοση συνολικά. Τέλος, για την συμπλήρωση του συστήματος σε μια ρεαλιστική συνθήκη απ' άκρη σ' άκρη (end to end) αλγόριθμου πρόβλεψης ταμπλατούρας εισήχθησαν δυο αυτόνομα συστήματα για εκτίμηση έναρξης νότας (onset) και τονικού ύψους (pitch). Για την περίπτωση του τονικού ύψους χρησιμοποιήθηκε το δημοσιευμένο προ-εκπαιδευμένο σύστημα CREPE [5], ενώ για την ανίχνευση έναρξης χρησιμοποιήθηκε μια δημοσιευμένη προσαρμογή του μοντέλου οπτικής ανίχνευσης ενάρξεων μέσω TCN [9] για το σκοπό της ηχητικής αναγνώρισης. Με την συμπερίληψη των παραπάνω υποσυστημάτων, καθίσταται πλέον το σύστημα ικανό, για πλήρη μεταγραφή μονοφωνικού ηχητικού σήματος κιθάρας σε ταμπλατούρα μόνο από το σήμα ήχου.

## 1.1 Σκοπός, Συνεισφορές και Διάρθρωση της Εργασίας

Σκοπός της παρούσας διπλωματικής, είναι η μελέτη τεχνικών πρόβλεψης ενεργής χορδής ως κομμάτι ενός συστήματος για την αυτόματη μεταγραφή ηχητικού σήματος κιθάρας σε ταμπλατούρα. Η μελέτη βασίστηκε κυρίως σε πραγματικές ερμηνείες σε κιθάρα, καταλλήγοντας έτσι σε μια ρεαλιστική αποτύπωση των δυνατοτήτων των προτεινόμενων μεθόδων. Η αναγνώριση χορδών, που αποτέλεσε και το βασικό πεδίο μελέτης, έγινε με βάση το υπό εξέταση σήμα και μελετήθηκε σε ποικίλες περιπτώσεις. Η εισαγωγή φυσικών περιορισμών και περιορισμών εκτελεσιμότητας, μέσω ενός γενετικού αλγόριθμου, μελετήθηκε ξεχωριστά καθώς παρουσίασε αναπάντεχα καλά αποτελέσματα. Ως συνεισφορές της παρούσας διπλωματικής συγκαταλέγουμε τα ακόλουθα:

- Μελετήθηκε και αποτιμήθηκε μια μέθοδος με αναρμονική ανάλυση για παραγωγή ακριβούς ταμπλατούρας η οποία απαιτεί ελάχιστο χρόνο και φόρτο για την προσαρμογή
- Η ένταξη φυσικών περιορισμών στη διαδικασία μεταγραφής μέσω του γενετικού αλγόριθμου, στρατηγική που με τη σειρά της ανοίγει ένα καινούριο πεδίο μελέτης για σύνδεση των, παλιότερα διακριτών μεταξύ τους, ηχητικών προσεγγίσεων μεταγραφής με τις προσεγγίσεις που βασίζονται στην εκτελεσιμότητα του κομματιού
- την υλοποίηση κώδικα για τη δημιουργία ενός συστήματος μεταγραφής ταμπλατούρας

από άκρη σε άκρη, με χρήση διακριτών επιμέρους υποσυστημάτων επεξεργασίας, με έμφαση στην ανάπτυξη ενός ταξινομητή χορδής βάσει του υπολογισμού συντελεστή αναρμονικότητας και υλοποίησης ενός γενετικού αλγόριθμου

Όσον αφορά στη διάρθρωση της διπλωματικής εργασίας:

- Στο κεφάλαιο 2 παρουσιάζονται μελέτες σχετικά με την παραγωγή εύκολα εκτελέσιμης ταμπλατούρας και δαχτυλοθεσίας τόσο από συμβολικές αναπαραστάσεις όσο και από ήχο, καθώς και με την αυστηρή μεταγραφή ταμπλατούρας από ηχητικό σήμα.
- Στο κεφάλαιο 3 παρουσιάζονται έννοιες σχετικές με το υπό εξέταση πρόβλημα και θεωρητικά εργαλεία τα οποία χρησιμοποιήθηκαν κατά τη διαδικασία της εκπόνησης της διπλωματικής εργασίας. Εκτενής αναφορά γίνεται στην αναρμονικότητα του ήχου και στις μαθηματικές της εκφράσεις αξιοποιώντας την έννοια του συντελεστή αναρμονικότητας. Παρουσιάζονται επίσης ορισμένοι αλγόριθμοι εκτίμησής του από τη βιβλιογραφία, καθώς αποτελεί το βασικό στοιχείο που εξετάζει η παρούσα υλοποίηση.
- Στο κεφάλαιο 4 παρουσιάζεται η μέθοδος που σχεδιάστηκε και υλοποιήθηκε για την πρόβλεψη ενεργής χορδής καθώς και η διασύνδεσή της με άλλα υποσυστήματα για την τελική σύσταση του απ' άκρη σ' άκρη συστήματος μεταγραφής κιθάρας σε ταμπλατούρα. Ακόμα, γίνεται αναφορά στις διαφορετικές διατάξεις συλλογής δεδομένων (δοσμένα από τον εκάστοτε χρήστη του συστήματος) για την προσαρμογή του υπολογιστικού μοντέλου του συστήματος ενώ παρουσιάζεται ο γενετικός αλγόριθμος που σχεδιάστηκε.
- Στο κεφάλαιο 5 γίνεται αποτίμηση των μεθόδων που χρησιμοποιήθηκαν, σχολιάζονται τα αποτελέσματα και συγκρίνονται με άλλες σχετικές μελέτες. Επιχειρείται παράλληλα η αυτόνομη αξιολόγηση των επιμέρους υποσυστημάτων που παρουσιάστηκαν δίνοντας μια πιο συνολική εικόνα της επίδοσης του συστήματος.
- Τέλος, στο κεφάλαιο 6 παρουσιάζονται τα συμπεράσματα που αντλήθηκαν και πιθανές μελλοντικές κατευθυντήριες γραμμές.



## Κεφάλαιο 2

### Προϋπάρχουσες Μελέτες

---

Στο παρόν κεφάλαιο παρουσιάζονται μελέτες σχετικές με το υπό εξέταση πρόβλημα και τις μεθόδους που χρησιμοποιήθηκαν για την αντιμετώπισή του. Αρχικά εκτίθενται μελέτες σχετικά με την αυτόματη παραγωγή εκτελέσιμης ταμπλατούρας (δηλαδή ταμπλατούρα που μπορεί να παιχτεί με βολικό τρόπο σε κιθάρα), είτε από συμβολική αναπαράσταση, είτε από ηχητικό σήμα, χωρίς όμως να διερευνώνται χαρακτηριστικά του ήχου που προσιδιάζουν μοναδικά σε κάθε χορδή. Έπειτα, παρουσιάζονται μελέτες που αντιμετωπίζουν το πρόβλημα της αυστηρής μεταγραφής ταμπλατούρας από σήμα ήχου. Ιδιαίτερη έμφαση δίνεται σε προσεγγίσεις που χρησιμοποιούν τον συντελεστή αναρμονικότητας.

#### 2.1 Παραγωγή Εκτελέσιμης Ταμπλατούρας

Ως μεταγραφή ταμπλατούρας εννοούμε τη διαδικασία μέσω της οποίας μια σειρά από παιγμένες νότες, αναπαρίστανται με συστηματικό τρόπο σε 6 οριζόντιες γραμμές που αντιστοιχούν στις χορδές της κιθάρας. Αν θεωρήσουμε ότι κάθε νότα μπορεί να παιχτεί σε 3 θέσεις περίπου στη ταστιέρα (στην πραγματικότητα 1 έως 4 θέσεις για κιθάρα με 17 τάστα και κανονικό κούρδισμα),  $N$  νότες αντιστοιχούν σε  $3^N$  διαφορετικές ταμπλατούρες. Παρόλα αυτά μόνο μερικές από αυτές είναι ρεαλιστικές και θα μπορούσαν να παιχτούν.

Έχουν μελετηθεί διάφοροι τρόποι εξαγωγής ταμπλατούρας. Μεταξύ αυτών έχουν χρησιμοποιηθεί αλγόριθμοι γραφημάτων και γενετικοί αλγόριθμοι, ενώ πολλές φορές συμπεριλαμβάνεται η μελέτη της δακτυλοθεσίας. Είναι σημαντικό να τονιστεί, πως στο συγκεκριμένο πρόβλημα δεν μπορούμε να θεωρήσουμε ότι υπάρχει βέλτιστη λύση, μιας και δεν μπορούμε να θεωρήσουμε ένα αυστηρό σύστημα αξιωμάτων-κανόνων που χαρακτηρίζει μια ακολουθία ως βέλτιστη.

Μερικές έρευνες σχετικά με την εξαγωγή ταμπλατούρας, επικεντρώθηκαν στον προσδιορισμό των συνδιασμών χορδών-τάστων μιας ακολουθίας νοτών βασισμένες κυρίως σε περιορισμούς εκτελεσιμότητας (playability constraints). Τέτοιες μέθοδοι είναι καταλληλότερες για πολυφωνικές εκτελέσεις (συγχορδίες) και έτσι συχνά συμπεριλαμβάνουν ένα σύστημα εκτίμησης πολλαπλών τονικών υψών (multi-pitch estimation module). Οι Barbancho et al. [10], πρότειναν ένα σύστημα για αναγνώριση συγχορδίας και δακτυλοθεσίας χρησιμοποιώντας Hidden Markov Models (HMMs) κωδικοποιώντας την πιθανότητα μετάβασης από τη μία θέση στην άλλη.

Στο [11], οι μεταβάσεις μεταξύ θέσεων (δηλ. ζευγών τάστου-χορδής), έχουν κωδικοποιη-

θεί χρησιμοποιώντας κατευθυνόμενα ακυκλικά γραφήματα με βάρη (WDAGs) και ο αλγόριθμος  $A^*$  χρησιμοποιείται για να προσδιοριστεί η βέλτιστη ακολουθία.

Στα [12] και [13], μια αναπαράσταση βασισμένη σε γραφήματα (WDAGs) χρησιμοποιήθηκε μαζί με ένα δυναμικό αλγόριθμο για να βρεθεί το μακρύτερο μονοπάτι. Στις τελευταίες περιπτώσεις είναι ξεκάθαρο ότι οι λεπτομέρειες σχετικά με το σήμα που παράχθηκε δεν λαμβάνονται υπόψη αφού η πληροφορία είναι κωδικοποιημένη σε MIDI.

Άλλες μελέτες έχουν επίσης διερευνήσει το πρόβλημα της δαχτυλοθεσίας και παραγωγής ταμπλατούρας δεδομένης μιας σειράς από νότες (υπό τη μορφή παρτιτούρας, MIDI ή με κάποιο άλλο τρόπο τυπικής καταγραφής), είτε με τη χρήση γραφημάτων και δυναμικού προγραμματισμού όπως στα [14], [15], [16], [17], βαθιάς μάθησης [18] ή αλγόριθμους βελτιστοποίησης [19], [20], [21].

Γενετικοί αλγόριθμοι έχουν χρησιμοποιηθεί για την εξαγωγή εκτελέσιμης ταμπλατούρας από συμβολική αναπαράσταση. Το 2015 οι Ramos et al. [20] παρουσίασαν μια συγκριτική μελέτη μεταξύ γενετικών αλγόριθμων για την αυτόματη μεταγραφή παρτιτούρας σε ταμπλατούρα. Τα πιο ενθαρρυντικά αποτελέσματα δίνονται με αλγόριθμους Ant Colony Optimization και GA 2-opt, GA sub-population 2-opt. Η συνάρτηση καταλληλότητας (fitness function) που χρησιμοποιείται σε όλες τις περιπτώσεις είναι η ευκλείδεια απόσταση που προκύπτει μεταξύ διαδοχικών νοτών, εάν μοντελοποιηθεί η ταστιέρα της κιθάρας ως ένα καρτεσιανό επίπεδο.

Οι Tuohy και Potter [19] πρότειναν έναν απλό γενετικό αλγόριθμο που αξιοποιεί φυσικούς περιορισμούς που χαρακτηρίζουν το παίξιμο της κιθάρας, σχεδιάζοντας μια συνάρτηση καταλληλότητας που αποτιμά την πολυπλοκότητα των δαχτυλοθεσιών σε μια ταμπλατούρα. Όλα τα στοιχεία της ταμπλατούρας αναγνωρίζονται ως συγχορδίες. Έτσι ακόμα και οι μεμονωμένες νότες αντιμετωπίζονται ως συγχορδίες με μόνο μια νότα. Για τη συνάρτηση καταλληλότητας, βασιζόμενοι στη δουλειά των Heijink και Meulnbroek [22], δομούν 2 κατηγορίες κριτηρίων: *κίνηση χεριού*, όπως το πόσες φορές πατιούνται διαφορετικές χορδές, την απόσταση μεταξύ διαδοχικών νοτών-συγχορδιών, την απόσταση μεταξύ των γύρω νοτών κ.λπ. και *διαχείριση χεριού* που εισάγει κριτήρια για την αξιολόγηση της δυσκολίας διατήρησης συγκεκριμένης θέσης στην ταστιέρα (π.χ. κάποια θέση μιας συγχορδίας μπορεί να απαιτεί μεγάλο άνοιγμα χεριών και ως εκ τούτου να μη συνηθίζεται).

Οι Tuohy και Potter στο [23], προτείνουν έναν καταναμημένο γενετικό αλγόριθμο (Distributed Genetic Algorithm - DGA), εξερευνώντας έτσι το χώρο αναζήτησης αυτόνομα προς διαφορετικές κατευθύνσεις. Η συνάρτηση καταλληλότητας προκύπτει από 14 διαφορετικά κριτήρια τα οποία χωρίζονται σε 3 κατηγορίες: πώς κινείται το χέρι, πώς τοποθετείται και από γνωσιακούς παράγοντες. Ακόμα προτείνεται ένας μετα-γενετικός (όπως οι ίδιοι τον ορίζουν, δηλαδή ως ένας γενετικός αλγόριθμος που χρησιμοποιεί έναν άλλο τέτοιο αλγόριθμο για την αποτίμηση της συνάρτησης καταλληλότητας), για τον προσδιορισμό των παραμέτρων του DGA, και των επιμέρους βαρών της συνάρτησης καταλληλότητας (fitness function). Τέλος, συμπεριλαμβάνεται ένα νευρωνικό δίκτυο για την αντιστοίχιση δαχτύλων σε κάθε νότα, ενώ εφαρμόζονται δύο ακόμα γενετικοί αλγόριθμοι για την βελτιστοποίηση των παραμέτρων του DGA και του περιβάλλοντος εκμάθησης του νευρωνικού δικτύου.

Νευρωνικά δίκτυα έχουν επίσης χρησιμοποιηθεί για να αντιμετωπίσουν το πρόβλημα της εκτίμησης ταμπλατούρας από ηχητικό σήμα. Στο [24], Multi-layer Perceptrons χρησιμοποιήθηκαν σε δυο φάσεις, πρώτα για μια αρχική ταξινόμηση για τον προσδιορισμό των ενεργών

χορδών και της θέσης του χεριού στην ταστιέρα, και ακολούθως για την πρόβλεψη των ζευγών χορδής-τάστου που παίχτηκαν. Ακόμη, συνελκτικά Νευρωνικά Δίκτυα (CNNs) έχουν χρησιμοποιηθεί [25] για να προσδιορίσουν μια σειρά από συγχορδίες και τις αντίστοιχες αναπαραστάσεις σε ταμπλατούρα από φασματικές αναπαραστάσεις παραγόμενες με τον constant-Q μετασχηματισμό.

## 2.2 Εξαγωγή ακριβούς ταμπλατούρας από ηχητικό αρχείο

Η αυστηρή μεταγραφή σήματος κιθάρας σε ταμπλατούρα, μπορεί εκ πρώτης όψεως να φαντάζει εύκολο πρόβλημα. Παρ' όλα αυτά, σε αντίθεση με όργανα όπως το πιάνο, στην κιθάρα παρατηρούνται ίδιοι τόνοι σε πάνω από μία θέσεις. Έτσι, η σωστή αναγνώριση του τόνου δεν επαρκεί για την σωστή αναγνώριση της αντίστοιχης χορδής. Είναι λοιπόν απαραίτητη η αναγνώριση περαιτέρω χαρακτηριστικών του σήματος κάθε νότας πλέον της έναρξης νότας (onset) και τονικού ύψους (pitch) τα οποία θα ήταν αρκετά για την μεταγραφή ενός σήματος πιάνου σε piano roll. Μερικές μελέτες εστιάζουν στο να καταγράφουν αυτά τα χαρακτηριστικά εμμέσως, όπως σε περιπτώσεις νευρωνικών δικτύων, ενώ άλλες άμεσα, όπως σε μελέτες όπου μετράται η αναρμονικότητα. Παρακάτω, καταγράφονται μερικές από τις μελέτες σε σχέση με την μεταγραφή σε ακριβή ταμπλατούρα και την αναγνώριση ενεργής χορδής.

Οι Kehling et al. [26] σχεδίασαν ένα σύστημα για μεταγραφή ταμπλατούρας όπου προσδιορίζουν έναρξη (onset), λήξη (offset), τονικό ύψος (pitch), χορδή και εκφραστικό στυλ (finger style, picked, muted). Εξήγαγαν 774 χαρακτηριστικά (features) στα οποία εφαρμόστηκε μείωση διάστασης (dimensionality reduction) ξέχωρα για κάθε επιμέρους πρόβλημα πριν την εκπαίδευση μιας Μηχανής Διανουσμάτων Υποστήριξης (Super Vector Machine SVM).

Πρόσφατα, χρησιμοποιήθηκαν Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks CNNs) με παρόμοιο τρόπο από τους Wiggins και Kim [27] για αναγνώριση χορδής-τάστου σε ηχογραφήσεις κιθάρας από το GuitarSet dataset [28]. Σημαντικό είναι πως σε αντίθεση με άλλες περιπτώσεις, χρησιμοποιούν πραγματικά κομμάτια και όχι μεμονωμένα ηχογραφημένα στιγμιότυπα νοτών. Παρόλα αυτά, η ακρίβειά τους δεν μετριέται με βάση τον αριθμό των στιγμιότυπων νοτών, αλλά ανά παράθυρο χρόνου (frame-by-frame). Τέτοιου είδους μετρικές δεν αντιπροσωπεύουν την ακρίβεια της μεταγραφής σε ταμπλατούρα του συστήματος. Γενικά, η έλλειψη ομοιομορφίας στον αριθμό των στιγμιότυπων νοτών που αντιστοιχούν σε κάθε χορδή για κάθε διαφορετική νότα, τείνει να επηρεάζει την ταξινόμηση χορδών καταλήγοντας σε μεροληπτικές (biased) εξόδους.

Ο Abecker [29] διερεύνησε την αναγνώριση χορδής κιθάρας και μπάσου. Εξήγαγε 48 διαφορετικά χαρακτηριστικά (συμπεριλαμβανομένου του συντελεστή αναρμονικότητας) και εφάρμοσε μείωση διάστασης (dimensionality reduction) σε αυτές καταλήγοντας σε 5 και 3 χαρακτηριστικά για τη κιθάρα και το μπάσο αντίστοιχα. Ακολούθως, ένα SVM εκπαιδεύτηκε σε αυτά για ταξινόμηση.

Οι Michelson et al. [30], χρησιμοποιούν επίσης τον συντελεστή αναρμονικότητας, για τον καθορισμό της θέσης στην ταστιέρα. Ο συντελεστής, υπολογίζεται σύμφωνα με τον αλγόριθμο Median Adjustive Trajectories (MAT) των Hodgkinson et al. [2], πάνω σε μεμονω-

μένες ηχογραφήσεις νοτών μεγάλης διάρκειας. Στην πρώτη μέθοδο, εφαρμόζεται λογαριθμική μετατροπή στην εξίσωση  $\beta(s, n) = \beta(s, 0)2^{\frac{sn}{6}}$ , εφαρμόζεται γραμμική παλινδρόμηση (linear regression) των λογαριθμικών αναρμονικοτήτων και προσδιορίζεται η ενεργή χορδή σύμφωνα με την μέση τιμή και διακύμανση που μεγιστοποιεί την πιθανότητα της μετρούμενης αναρμονικότητας. Η δεύτερη μέθοδος, αναπτύχθηκε για λόγους αξιολόγησης. Κάθε θέση στη ταστιέρα μοντελοποιείται από μια κανονική κατανομή και χαρακτηρίζεται ως ενεργή χορδή αυτή που μεγιστοποιεί την πιθανότητα της μετρούμενης αναρμονικότητας.

Οι Geib et al. [31] πρότειναν ένα καινούριο τρόπο για αναγνώριση ενεργής χορδής βασιζόμενοι στην ιδιότητα που ονόμασαν String Inverse Frequencies (SIF). Αν και αρκετά εφευρετικός και αντισυμβατικός, απαιτεί πρόσθετα εργαλεία για την εφαρμογή του, καθώς βασίζεται στον προσδιορισμό των συχνοτήτων που παράγονται στο κομμάτι της χορδής από την τοποθέτηση του δακτύλου προς τα κλειδιά.

Οι Barbancho et al. [3] ανέπτυξαν μια μεθοδολογία για ταυτόχρονη αναγνώριση χορδής και τονικού ύψους. Συνέθεσαν μια μέθοδο για να εκτιμήσουν τον συντελεστή αναρμονικότητας από απομονωμένες ηχογραφήσεις κιθάρας μέσω της αναγνώρισης των μερικών συχνοτήτων. Πιο συγκεκριμένα αφού προσδιοριστούν οι τιμές του συντελεστή αναρμονικότητας για κάθε χορδή, για να ταξινομηθεί μια καινούρια νότα, χρησιμοποιείται ένα σύστημα ψηφοφορίας ανάμεσα σε όλες τις πιθανές θέσεις (κάθε θέση αντιστοιχεί και σε διαφορετικό συντελεστή αναρμονικότητας) για όλες τις πιθανές θεμελιώδεις συχνότητες και επιλέγεται αυτό που προσεγγίζει καταλλήλότερα τον αρχικά μετρημένο συντελεστή αναρμονικότητας.

Οι Hjerrild και Christensen [6] πρότειναν μια εναλλακτική μέθοδο για ταυτόχρονη εκτίμηση τονικότητας ( $\omega_0$ ) και συντελεστή αναρμονικότητας ( $\beta$ ) βασιζόμενοι σε μεγιστοποίηση πιθανότητας (likelihood maximization) και grid search. Το παραγόμενο σήμα από την παλλόμενη χορδή μοντελοποιήθηκε ως ένα σύνθετο σήμα με μια αναρμονική ημιτονοειδή και επιπρόσθετο λευκό γκαουσιανό θόρυβο. Ένας Maximum a Posteriori (MAP) ταξινομητής χρησιμοποιήθηκε για αναγνώριση χορδής, μοντελοποιώντας την συνάρτηση πυκνότητας πιθανότητας των  $\beta$  και  $\omega_0$  κάθε ζεύγους χορδής-τάστου ως μια διμεταβλητή κανονική κατανομή, χρησιμοποιώντας μόνο ένα τάστο ανά χορδή (δηλαδή 9-10 δείγματα ανά χορδή). Το σχετικό σύνολο δεδομένων (Dataset) δεν συμπεριλάμβανε εκτελέσεις κομματιών σε κιθάρα, αλλά μόνο απομονωμένες ηχογραφήσεις στιγμιότυπων νοτών.

Ακόμη οι Hjerrild et. al [8] πρότειναν μια μέθοδο για αναγνώριση ενεργής χορδής, δεδομένων των φυσικών χαρακτηριστικών της χορδής της υπό εξέταση κιθάρας, με στόχο να αποφευχθεί πλήρως η εξάρτηση του ταξινομητή χορδής από ηχογραφημένες νότες.

Τέλος, οι Servaityte et al [32], χρησιμοποιούν Harmonic Percussive Separation (HPS) και ακολούθως Non-Negative Matrix Factorization (NMF) και Non-Negative Matrix Factor Deconvolution (NMF-D) στο αρμονικό (harmonic) και κρουστό (percussive) μέρος αντίστοιχα, επιχειρώντας όχι απλά αναγνώριση χορδής, αλλά το διαχωρισμό του σήματος ανά χορδή (String Separation). Παρόλα αυτά, όπως αναφέρουν τα αποτελέσματά τους συγκρινόμενα με το απλό NMF σχήμα που χρησιμοποίησαν για σύγκριση και αξιολόγηση δεν είναι ενθαρρυντικά. Είναι προφανές, πως σε περίπτωση επίλυσης του παραπάνω προβλήματος, αρκεί η σωστή αναγνώριση τόνου, ώστε να γίνει αυστηρή μεταγραφή σε ταμπλατούρα.

## Κεφάλαιο 3

### Θεωρητικό Υπόβαθρο

---

Στο παρόν κεφάλαιο, παρουσιάζεται το απαραίτητο θεωρητικό υπόβαθρο ώστε να μπορεί να γίνει κατανοητό τόσο το υπό εξέταση πρόβλημα, όσο και οι λύσεις που εξετάστηκαν. Πρώτα, αναφέρονται συνοπτικά σημαντικά στοιχεία σχετικά με τη μουσική. Ακολουθεί μια περιγραφή της δομής της κιθάρας και των φυσικών χαρακτηριστικών της, ενώ ιδιαίτερη έμφαση δίνεται στην αναμονικότητα και στον τρόπο υπολογισμού της. Ακολούθως, εξηγούνται τα σχετικά προβλήματα που εμπíπτουν στο πεδίο της Ανάκτησης Μουσικής Πληροφορίας (Music information Retrieval, MIR) καθώς και εργαλεία απαραίτητα για την αντιμετώπισή τους, όπως τεχνικές μηχανικής μάθησης και νευρωνικά δίκτυα. Κλείνοντας γίνεται μια λεπτομερής αναφορά στους γενετικούς αλγορίθμους, οι οποίοι χρησιμοποιήθηκαν και στην παρούσα διπλωματική.

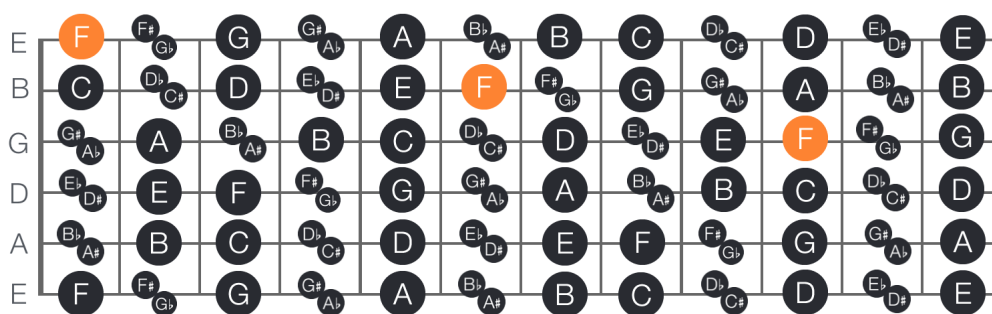
#### 3.1 Μουσική

Έννοιες στη μουσική οι οποίες χαρακτηρίζουν μια αισθητηριακή αντίληψη εγχολπώνουν ένα σύνολο φυσικών ιδιοτήτων του μουσικού σήματος. Παραδείγματος χάριν, όταν ξεχωρίζουμε έναν τόνο από ένα άλλο, στην πραγματικότητα απλά διακρίνουμε δύο διαφορετικές θεμελιώδεις συχνότητες. Αναπόφευκτα λοιπόν, σε μια διερεύνηση μουσικών σημάτων οφείλουν να καταγραφούν τέτοιες έννοιες, να επεξηγηθεί η φυσική τους σημασία αλλά και να τονιστεί η πιθανή αμφισημία τους. Παρακάτω, σε αυτήν την ενότητα παρουσιάζονται έννοιες και ορισμοί ενώ γίνεται μια απόπειρα να συνδεθούν οι μουσικοί όροι με τη φυσική τους σημασία. Ακολούθως, καταγράφονται πληροφορίες συγκεκριμένα για τη δομή και τα χαρακτηριστικά της κιθάρας, και τέλος επεξηγούνται προβλήματα άντλησης μουσικής πληροφορίας, τα οποία αντιμετώπισε η παρούσα διπλωματική άμεσα ή έμμεσα.

Τα βασικά υποκειμενικά χαρακτηριστικά της αισθητηριακής αντίληψης της μουσικής όπως αναφέρονται στο [33], είναι το τονικό ύψος (pitch), η ένταση (loudness), η διάρκεια (duration) και το ηχόχρωμα (timbre).

Το τονικό ύψος αναφέρεται στη διαισθητική κατανομή των ήχων σε κλίμακα από τα χαμηλά στα ψηλά ως προς τη συχνότητα. Η θεμελιώδης συχνότητα  $F_0$  είναι ο σχετικός φυσικός όρος, οριζόμενος για περιοδικά ή σχεδόν περιοδικά σήματα μόνο. Η ένταση όπως την αντιλαμβάνεται ένας άνθρωπος έχει μια πιο περίπλοκη σύνδεση με τα φυσικά χαρακτηριστικά από ό,τι θα περιμέναμε. Στη μουσική επεξεργασία, αντιστοιχείται για λόγους ευκολίας, με την μέση τετραγωνική ισχύ του σήματος ενώ ακολούθως μεταφέρεται σε λογαριθμική κλίμακα για





Σχήμα 3.1: Νότες πάνω στη κιθάρα για κανονικό κούρδισμα

να αντιμετωπίσει το μεγάλο εύρος συχνοτήτων που παρουσιάζεται. Η διάρκεια διατηρεί ίδια έννοια τόσο αισθητηριακά όσο και από φυσικής σημασίας. Το ηχόχρωμα είναι άμεσα συνδεδεμένο με την αναγνώριση της πηγής που παράγει το σήμα που ακούγεται. Η έννοιά του δεν είναι εύκολα επεξηγήσιμη καθώς διαφορετικά όργανα μπορεί να έχουν ίδιο τονικό ύψος ένταση και διάρκεια αλλά να είναι διαχωρίσιμα από ένα ανθρώπινο αυτί. Εξαρτάται κυρίως, από την κατανομή ενέργειας ενός σήματος στις διαφορετικές συχνότητες, και από την εξέλιξη της στο χρόνο. Η έναρξη (onset) αναφέρεται στη βιβλιογραφία πολλές φορές με διαφορεόμενο τρόπο. Διαισθητικά, μπορεί να θεωρηθεί ως η έναρξη ενός μουσικού γεγονότος, η αρχή ενός μουσικού σήματος. Τεχνικά, συσχετίζεται/συγγέεται πολλές φορές με το attack το οποίο αναφέρεται στη χρονική περίοδο κατά την οποία αυξάνει απότομα η περιβάλλουσα ενέργειας του σήματος. [4].

Στη δυτική μουσική, τα τονικά ύψη είναι οργανωμένα στη φυσική κλίμακα η οποία αποτελεί μια λογαριθμική αναπαράσταση των θεμελιωδών συχνοτήτων (συγκερασμένο σύστημα). Κάθε οκτάβα αποτελείται από 12 επιμέρους ημιτόνια, κάθε ένα από τα οποία διατηρεί λόγο με το γειτονικό του ίσο με  $2^{1/12}$ . Αναπαρίσταται με ένα διαφορετικό αριθμό (π.χ.  $E_5$  υποδηλώνοντας τη νότα Μι της πέμπτης οκτάβας). Ίδιες νότες διαφορετικής οκτάβας, έχουν σχέση συχνοτήτων  $\frac{1}{2}$ , με άλλα λόγια η πρώτη αρμονική της χαμηλότερης θεμελιώδους αντιστοιχεί στην θεμελιώδη της δεύτερης. Οι 12 νότες αποτελούνται από τις Ντο, Ντο#, Ρε, Ρε#, Μι, Φα, Φα#, Σολ, Σολ#, Λα, Λα#, Σι. Κάθε αυξημένη νότα (#) μπορεί να γραφτεί και ως ελαττωμένη τοποθετώντας ύφεση αντί για δίεση (πχ. Ντο# - Ρεb).

## 3.2 Δομή της κιθάρας

Η κιθάρα, αποτελεί ένα από τα πιο διαδεδομένα όργανα σε πληθώρα μουσικά είδη και περιοχές. Είναι ένα νυκτό, έγχορδο, συγκερασμένο όργανο, το οποίο συνήθως διαθέτει 6 χορδές, κουρδισμένες σε Μι (E)-Λα (A)-Ρε (D)-Σολ (G)-Σι (B)-Μι (e) (από την μπάσα προς την ψηλότερη). Ο ήχος της προκύπτει είτε μέσω της αντήχησης του ηχείου της (ακουστική, κλασική) είτε μέσω ενίσχυσης. Ο χωρισμός του οργάνου με τέτοιο τρόπο ώστε να παράγονται οι ζητούμενοι τόνοι, γίνεται με μεταλλικά ελάσματα (τάστα) τοποθετημένα στο μπράτσο της κιθάρας. Η απόσταση μεταξύ διαδοχικών τάστων ισούται με ένα ημιτόνιο, ενώ δυο ημιτόνια στη σειρά αποτελούν ένα τόνο 3.1.

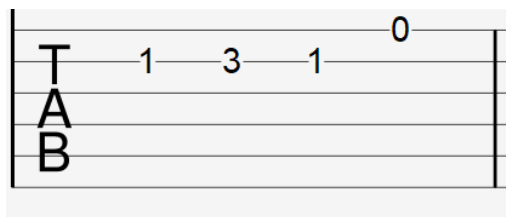
### 3.2.1 Τυπικές Μέθοδοι Αναπαράστασης

**Παρτιτούρα** ονομάζεται η συμβολική αναπαράσταση μιας μουσικής μελωδίας σε πεντάγραμμο. Περιέχει πληροφορίες τόσο για το τονικό ύψος, (pitch) όσο και για την αξία (τη διάρκεια δηλαδή της νότας), και τη θέση της στο χρόνο σε σχέση με τις υπόλοιπες. Επί της ουσίας, από μια παρτιτούρα θα μπορούσαμε να αντλήσουμε πληροφορίες για τις ενάρξεις (onset), τα πέρατα (offset), τα τονικά ύψη (pitch), το ρυθμό κ.α.



Σχήμα 3.2: Παράδειγμα Παρτιτούρας

Η **ταμπλατούρα** αποτελεί συμβολική αναπαράσταση της μελωδίας ενός κομματιού για συγκεκριμένα έγχορδα. Αποτυπώνεται πάνω στην ταστιέρα του αντίστοιχου οργάνου, (δηλαδή στην περίπτωση μιας κιθάρας σε 6 χορδές ενός μπάσου 4 κ.ο.κ) και καταγράφει τα τάστα και τις χορδές στις οποίες παίχτηκε η κάθε νότα δηλώνοντας εμμέσως το τονικό ύψος. Κάθε στιγμιότυπο νότας σημειώνεται πάνω σε μία από τις 6 γράμμες της ταμπλατούρας με έναν αριθμό που αντιστοιχεί στο τάστο που πρέπει να πατηθεί στη συγκεκριμένη χορδή. Αριθμούμε τα τάστα ξεκινώντας από το 0 που αντιστοιχεί στην ανοιχτή χορδή και ο αριθμός αυξάνει πλησιάζοντας προς το σώμα της κιθάρας 3.1. Δεν περιλαμβάνει τις αξίες, ενώ η θέση στο χρόνο της κάθε νότας αποτυπώνεται μόνο κατά σειρά σε σχέση με τις υπόλοιπες. Παρόλα αυτά, επειδή μπορεί να διαβαστεί από οποιονδήποτε δεν γνωρίζει ανάγνωση παρτιτούρας, χρησιμοποιείται αρκετά.

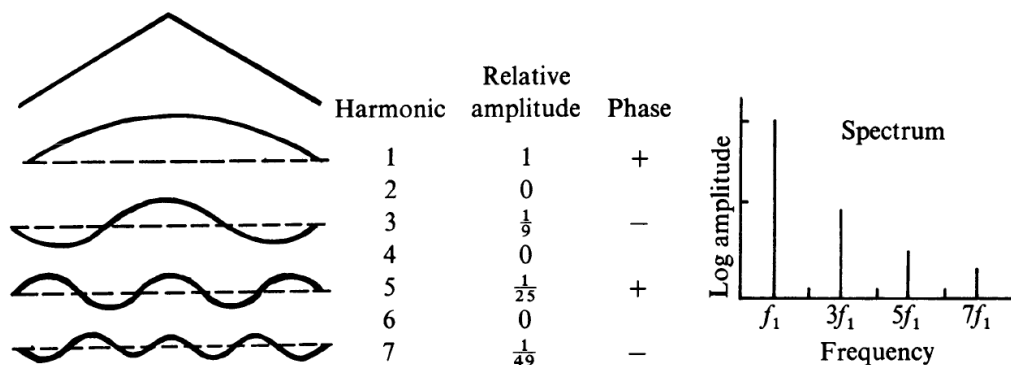


Σχήμα 3.3: Παράδειγμα Ταμπλατούρας

Τέλος, το πρωτόκολλο **midi** (Musical Instrument Digital Interface), (Musical Instrument Digital Interface) αποτελεί τρόπο κωδικοποίησης μουσικής πληροφορίας ώστε να μπορεί να αναγνωριστεί από μουσικά όργανα και υπολογιστές. Η πληροφορία που μεταφέρεται αποτελείται από εντολές όπως το τονικό ύψος, διάρκεια κ.λπ.

## 3.3 Η χορδή ως Στάσιμο κύμα

Με το κτύπημα της χορδής, ένα κύμα ταξιδεύει πάνω στη χορδή, όταν φτάσει στο ένα της άκρο, ένα αντίρροπο όμοιο κύμα παράγεται. Η υπέρθεση των δύο παράγει ένα στάσιμο κύμα. Το συχνοτικό της περιεχόμενο αποτελείται από την θεμελιώδη συχνότητα και τις αρμονικές



Σχήμα 3.4: Συχνοτική ανάλυση μιας χορδής χτυπημένη στο κέντρο της. Οι τρόποι ταλάντωσης μονής τάξεως, αθροίζονται σε φάση και πλάτος για να προκύψει τελικά η μορφή της χορδής. Πηγή :[1]

της. Η **θεμελιώδης συχνότητα μιας ταλαντευόμενης χορδής**, εξαρτάται κυρίως, από τη διάμετρό της, το υλικό της, την τάση υπό την οποία είναι τεντωμένη και το μήκος της χορδής. Το πάτημα ενός τάστου, αλλάζει το μήκος και ως εκ τούτου αλλάζει η θεμελιώδης συχνότητα. Παρ' όλα αυτά, δεν είναι οι μοναδικές παράμετροι που επηρεάζουν το συχνοτικό περιεχόμενο, με χαρακτηριστικό παράδειγμα το σημείο στο οποίο χτυπιέται μια χορδή το οποίο επηρεάζει τα πλάτη των αρμονικών που παράγονται. Αν για παράδειγμα, χτυπηθεί η χορδή ακριβώς στη μέση της, σχεδόν εκμηδενίζονται τα πλάτη των άρτιας τάξεως αρμονικών και το παραγόμενο σήμα προκύπτει από τις μονής τάξεως αρμονικές 1η, 3η, 5η κ.ο.κ (βλ. Σχήμα 3.4 [1]).

Εξ ορισμού στο συγκεκριμένο σύστημα και ως εκ τούτου εκ κατασκευής της κιθάρας για απόσταση ενός τάστου στην ίδια χορδή, ή **σχέση μεταξύ των θεμελιωδών συχνοτήτων** προκύπτει ως:

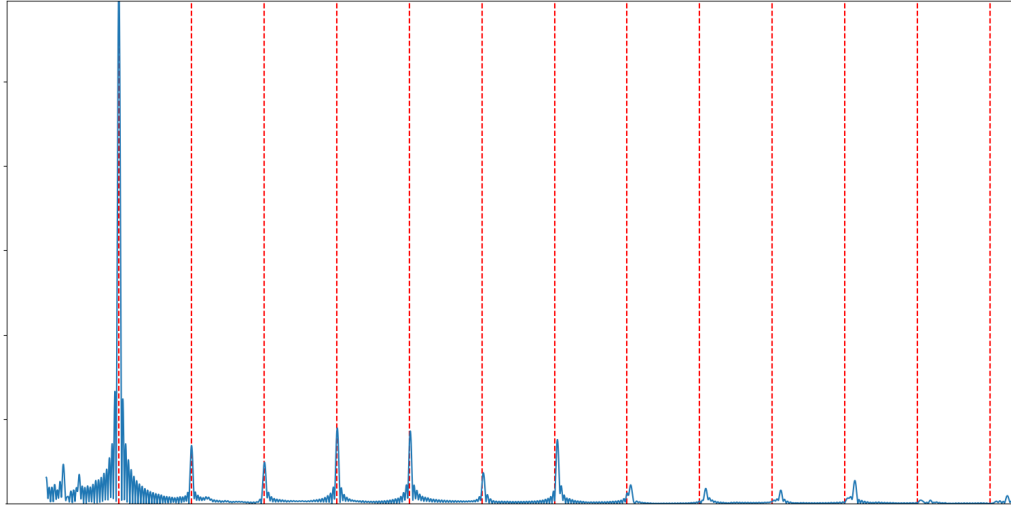
$$f(s, n) = f(s, n - 1) \cdot 2^{1/12} \quad (3.1)$$

### 3.3.1 Αναρμονικότητα

Σε πολλά μουσικά όργανα, σε μεγάλο βαθμό στα κρουστά, σε μικρότερο σε νυκτά έγχορδα (πιάνο, κιθάρα, μπάσο), και λίγο έως καθόλου σε τοξωτά έγχορδα ή πνευστά (βιολί, κλαρίνο), παρατηρείται ότι οι αρμονικές τείνουν να αποκλίνουν από τα ακέραια πολλαπλάσια της θεμελιώδους συχνότητας. Ο βαθμός στον οποίο αυτό συμβαίνει, αποτυπώνεται από το συντελεστή αναρμονικότητας (inharmonicity coefficient). Όπως αναφέρεται στο [7], αυτή η απόκλιση εξαρτάται από την ακαμψία (stiffness) της χορδής και του τεντώματος κατά το χτύπημα (stretching when plucked), με τη δεύτερη συνιστώσα να είναι μικρότερης σημασίας πολύ σύντομα μετά το χτύπημα της χορδής, καθώς η επιρροή της αρχικής διέγερσης σβήνει γρήγορα [8]. Στην περίπτωση της κιθάρας, η αναρμονικότητα εξαρτάται από τα φυσικά χαρακτηριστικά της χορδής και για δεδομένη διάμετρο χορδής  $d$ , μήκος χορδής  $l$ , και τάση  $T$ , προκύπτει ως:

$$\beta = \frac{\pi \cdot Q \cdot d^4}{64 \cdot l^2 \cdot T} \quad (3.2)$$





Σχήμα 3.5: Οι διακεκομμένες γραμμές αντιπροσωπεύουν τις τιμές των αρμονικών. Καθώς η τάξη των partial μεγαλώνει, εξίσου αυξάνεται και η απόκλισή τους από τις τιμές των αρμονικών.

όπου το  $Q$  είναι σταθερά που ονομάζεται Young's modulus και σχετίζεται με την αντίσταση ελαστικών σωμάτων στην παραμόρφωση. Εύκολα μπορεί να αποδειχτεί ότι μεταξύ δυο τάσεων  $n, k$  στην ίδια χορδή  $s$ , ο συντελεστής inharmonicity  $\beta$  προκύπτει από την παραπάνω σχέση ως:

$$\beta(s, n) = \beta(s, k) \cdot 2^{\frac{n-k}{6}} \quad (3.3)$$

Για θεμελιώδη συχνότητα  $f_0$ , η συχνότητα της  $k$ -οστής αρμονικής, προκύπτει από τη σχέση:

$$f_k = k \cdot f_0 \sqrt{1 + \beta \cdot k^2} \quad (3.4)$$

Κύρια δυσκολία στον προσδιορισμό της ενεργής χορδής αποτελεί το γεγονός ότι η ίδια νότα (θεμελιώδης συχνότητα) παρατηρείται σε πάνω από ένα σημείο στη ταστιέρα. Αυτό σημαίνει, ότι τα φάσματα ίδιων νοτών σε διαφορετικές θέσεις, θα είναι δυσδιάκριτα. Για την επίλυση του συγκεκριμένου προβλήματος, από τις σημαντικότερες ενδείξεις που μπορούν να δώσουν μια φυσική εξήγηση σε ένα τέτοιο πρόβλημα, είναι ο συντελεστής αναρμονικότητας ο οποίος εκφράζει τον βαθμό απόκλισης των αρμονικών μιας παλλόμενης χορδής από το αναμενόμενο, δίνοντας σημαντική πληροφορία η οποία εξαρτάται από τα φυσικά χαρακτηριστικά μιας χορδής. Συνεπώς, ο ακριβής υπολογισμός του, συντελεί στην αναγνώριση της σωστής χορδής που αντιστοιχεί σε κάθε νότα.

Ο υπολογισμός του συντελεστή αναρμονικότητας, εκ πρώτης όψεως φαίνεται σχετικά εύκολο πρόβλημα. Από την εξίσωση  $f_k = k f_0 \sqrt{1 + \beta k^2}$  βλέπουμε ότι γνωρίζοντας την θεμελιώδη συχνότητα  $f_0$  και μια οποιαδήποτε αρμονική της, θα μπορούσαμε να υπολογίσουμε το συντελεστή  $\beta$ , με απλή αντικατάσταση. Παρόλα αυτά, για μικρής τάξεως αρμονικές, οι αποκλίσεις είναι πολύ μικρές (σε σχέση με τον ακριβή προσδιορισμό των αρμονικών), ενώ για υψηλής τάξεως (όπου οι αποκλίσεις είναι σχετικά μεγάλες) οι αρμονικές εντοπίζονται δυσκολότερα

χωρίς πρότερη γνώση του συντελεστή αναρμονικότητας [2]. Έτσι καταλήγει αναγκαία, η θεμελίωση πιο εκλεπτυσμένων μεθόδων που θα συμπεριλαμβάνουν πολλαπλές αρμονικές στον υπολογισμό του συντελεστή.

### Αλγόριθμος Median Adjustive Trajectories (MAT)[2]

Οι Hodgkinson et al. [2], παρουσίασαν τον αλγόριθμο Median Adjustive Trajectories (MAT) για τον προσδιορισμό του συντελεστή αναρμονικότητας  $\beta$  μιας ταλαντευόμενης χορδής. Βασίζομενοι στο γεγονός ότι οι χαμηλής τάξεως μερικές (partials) έχουν πολύ μικρές διαφορές από τις αναμενόμενες και είναι εύκολα να εντοπιστούν, ενώ τα μεγάλης τάξης έχουν μεγάλες διαφορές αλλά είναι δυσκολότερο, χρησιμοποιούν τις σχέσεις:

$$f_0 = \frac{f_m}{m\sqrt{1 + \beta m}} \quad (3.5)$$

$$\beta = \frac{(f_k \frac{m}{k})^2 - f_m^2}{k^2 f_m^2 - m^2 (f_k \frac{m}{k})^2} \quad (3.6)$$

ώστε να προσδιορίσουν με επαναληπτικό τρόπο, ταυτόχρονα τόσο οι μερικές συχνότητες (partials) όσο και το συντελεστή  $\beta$ . Πιο συγκεκριμένα, για κάθε επανάληψη, υπολογίζεται πρώτα το  $\beta$  ως η διάμεσος των β' από ζεύγη αρμονικών μέσω της εξίσωσης 3.6. Ακολούθως χρησιμοποιείται για να υπολογιστεί η διάμεσος των θεμελιωδών συχνοτήτων που αντιστοιχούν σε κάθε αρμονική από την εξίσωση 3.5 και τέλος εντοπίζεται η επόμενη αρμονική που βρίσκεται κοντά στη συχνότητα  $f_k$  όπως προκύπτει από την εξίσωση 3.7. Ο αλγόριθμος και σχηματική αναπαράσταση των 3 πρώτων επαναλήψεων παρουσιάζονται στο Σχήμα 3.6.

$$f_k = k f_0 \sqrt{1 + \beta k^2} \quad (3.7)$$

### Αλγόριθμος Hjerrild και Christensen[6]

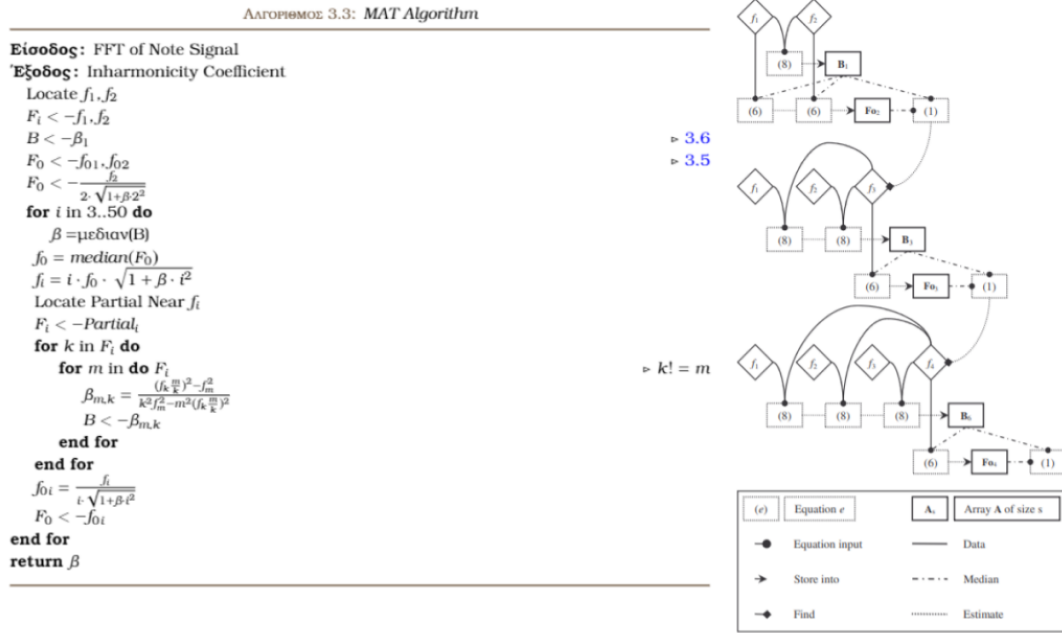
Οι [Hjerrild et al. [6] [8], χρησιμοποιούν το δικό τους τρόπο εκτίμησης του συντελεστή αναρμονικότητας. Αφού μοντελοποιήσουν την παλλόμενη χορδή συμπεριλαμβάνοντας την αναρμονικότητα, υπολογίζουν τη θεμελιώδη συχνότητα και το συντελεστή αναρμονικότητας, μεγιστοποιώντας μια συνάρτηση πιθανότητας.

### Αλγόριθμος Abesser[29]

Ο Abesser [29] προσδιορίζει τις συχνότητες των πρώτων 15 partials. Από την εξίσωση 3.7, καταλήγει στην προσέγγιση

$$(f_k/f_0)^2 = \sum_{i=1}^4 p_i k^i \quad (3.8)$$

οπου το  $p_4$  αντιστοιχεί στο συντελεστή αναρμονικότητας  $\beta$ .



Σχήμα 3.6: Αλγόριθμος MAT αριστερά και δεξιά σχήμα για τις 3 πρώτες επαναλήψεις [2]

### Αλγόριθμος Υπολογισμού Αναρμονικότητας Barbancho et al. [3]

Οι Barbancho et al [3], για τον υπολογισμό του συντελεστή αναρμονικότητας, μετρούν την απόκλιση των αρμονικών από τις αναμενόμενες θεωρητικές τιμές και εφαρμόζουν τη μέθοδο ελαχίστων τετραγώνων για τον προσδιορισμό του. Αναλυτικότερα, θεωρώντας ότι η μετρούμενη θεμελιώδη συχνότητα μιας νότας  $\hat{f}_0(s, n)$  διαφέρει από την πραγματική τιμή  $f_0(s, n)$  κατά ένα παράγοντα  $\epsilon_0(s, n)$  άρα:

$$\hat{f}_0(s, n) = f_0(s, n) + \epsilon_0(s, n) \quad (3.9)$$

αν θεωρηθεί αντίστοιχα ένα μοντέλο για τη μετρούμενη αρμονική ως:

$$\hat{f}_k(s, n) = k \cdot f_0(s, n) \cdot \sqrt{1 + \beta(s, n) \cdot k^2} + \epsilon_k(s, n) \quad (3.10)$$

τότε μπορούμε να κατασκευάσουμε ένα μοντέλο για τις αποκλίσεις των αρμονικών  $c_k(s, n)$  ως:

$$\begin{aligned} c_k(s, n) &= \hat{f}_k(s, n) - k \cdot \hat{f}_0(s, n) \\ &= k \cdot f_0(s, n) \cdot (\sqrt{1 + \beta(s, n) \cdot k^2} - 1) \\ &\quad - k \cdot \epsilon_0(s, n) + \epsilon_k(s, n) \end{aligned} \quad (3.11)$$

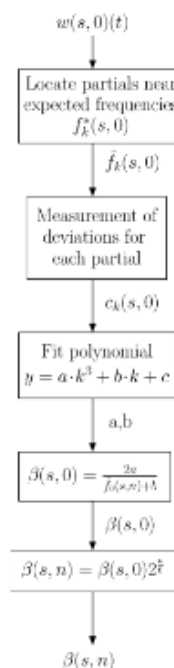
αντικαθιστώντας με τους δυο πρώτους όρους της σειράς Taylor στο  $\sqrt{1 + \beta \cdot k^2}$  τότε η παραπάνω εξίσωση μπορεί να ξαναγραφτεί ως:

$$c_k(s, n) = \frac{1}{2} \cdot \beta(s, n) \cdot f_0(s, n) \cdot k^3 - k \cdot \epsilon_0(s, n) + \epsilon_k(s, n) \quad (3.12)$$

---

**Είσοδος:** FFT of Note Signal  
**Έξοδος:** Inharmonicity Coefficient  
**for**  $k$  **in** 2..50 **do**  
 $f_k = k \cdot f_0 \cdot \sqrt{1 + \beta(s, n) \cdot k^2}$   
 Locate Partial Near  $f_k$   
 $c_k$  Append  $f_k$   
 $a, b, c = \text{LstSqr}(c_k)$   
 $\beta = \frac{2 \cdot a}{\hat{f}_k(s, n) + b}$   
**end for**  
**return**  $\beta$

---



Σχήμα 3.7: Αλγόριθμος και διάγραμμα για την εκτίμηση του συντελεστή αναρμονικότητας με την μέθοδο που παρουσιάζεται στο [3] (διάγραμμα από το ίδιο)

Ακολούθως, η μέθοδος ελαχίστων τετραγώνων μπορεί να χρησιμοποιηθεί στη συνάρτηση  $y = a \cdot k^3 + b \cdot k + c$  για τον υπολογισμό των  $a, b$  και τελικά την εύρεση μιας προσέγγισης του  $\beta$  από την εξίσωση:

$$\beta(s, n) = \frac{2 \cdot a}{\hat{f}_k(s, n) + b} \quad (3.13)$$

Για τον σωστότερο εντοπισμό των αρμονικών συχνοτήτων, σε κάθε βήμα του αλγορίθμου πριν τον εντοπισμό της επόμενης αρμονικής, συνεκτιμάται η υπολογιζόμενη αναρμονικότητα, τοποθετώντας το παράθυρο αναγνώρισης στη θέση που προκύπτει από τη συνάρτηση 3.7. Το παράθυρο έχει μέγεθος το μισό από τη θεμελιώδη συχνότητα του υπό εξέταση στιγμιότυπου.

### Τροποποιημένος αλγόριθμος Barbancho et al. [3]

Στη παρούσα διπλωματική, αρχικά δοκιμάστηκε η υλοποίηση του αλγορίθμου των Barbancho et al. [3] αποδίδοντας όμως μέτρια αποτελέσματα. Επειδή τα υπό εξέταση σήματα προκύπτουν από πραγματικές ηχογραφήσεις κιθάρας και συνεπώς περιέχουν αρκετό θόρυβο ή/και πολλαπλές νότες την ίδια στιγμή, είναι πιθανή η λανθασμένη αναγνώριση θορύβου ή άλλης συχνότητας ως αρμονική, με συνέπεια λανθασμένο υπολογισμό του  $\beta$  και μετέπειτα λανθασμένη αναγνώριση και τον επόμενων αρμονικών. Έτσι, τροποποιήθηκε ο παραπάνω αλγόριθμος ώστε να μην επανατοποθετεί το παράθυρο αναγνώρισης σε κάθε βήμα βάση του συντελεστή αναρμονικότητας που έχει υπολογιστεί, αλλά να ελέγχει μόνο για τις 13 πρώτες αρμονικές στις θέσεις γύρω από το  $f_k$  όπως προκύπτει από την εξίσωση  $f_k = k \cdot f_0$  με μήκος παραθύρου  $\frac{f_0}{2}$ . Σημαντικό είναι να τονιστεί πως στη περίπτωση αυτή, ο αλγόριθμος έχει σημαντικά μικρότερη πολυπλοκότητα καθώς αποφεύγεται η εφαρμογή της μεθόδου ελαχίστων τετραγώνων επαναληπτικά σε κάθε βήμα. Παρ' όλα αυτά, αδυνατεί να αναγνωρίσει υψηλής

---

 ΑΛΓΟΡΙΘΜΟΣ 3.1: *Inharmonicity Computation non-Iterative*


---

**Είσοδος:** FFT of Note Signal  
**Έξοδος:** Inharmonicity Coefficient

```

for  $k$  in 2..14 do
   $f_k = k \cdot f_0 \cdot \sqrt{1 + \beta(s, n) \cdot k^2}$ 
  Locate Partial Near  $k \cdot f_0$ 
   $c_k$  Append  $f_k$ 
end for
 $a, b, c = LstSqrS(c_k)$ 
 $\beta = \frac{2 \cdot a}{f_k(s, n) + b}$ 
return  $\beta$ 

```

---

τάξεως αρμονικές καθώς φεύγουν εκτός παραθύρου (βλ. Σχήμα 3.8).

## 3.4 Music Information Retrieval (MIR)

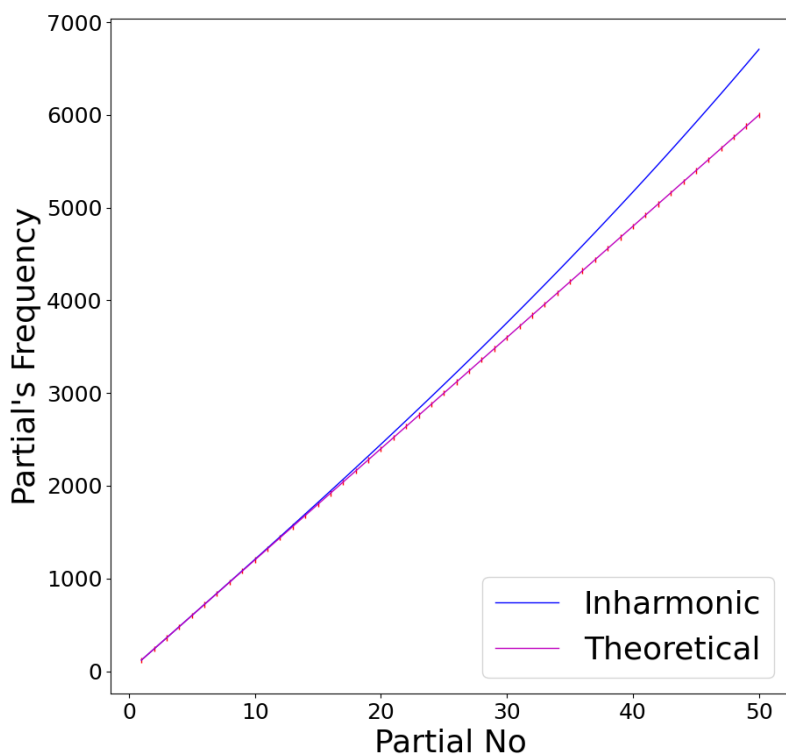
Ως Music Information Retrieval αναφέρεται το επιστημονικό πεδίο μελέτης και εξαγωγής μουσικής πληροφορίας. nset detection, η pitch estimation, source separation, genre classification, είναι μερικά από τα πεδία ενδιαφέροντος. Η παρούσα διπλωματική εντάσσεται και αυτή στο μωσαϊκό αυτών των προβλημάτων και συνδέεται άρρηκτα με κάποια από αυτά τα προβλήματα στα οποία θα αναφερθούμε παρακάτω.

### 3.4.1 Onset Detection

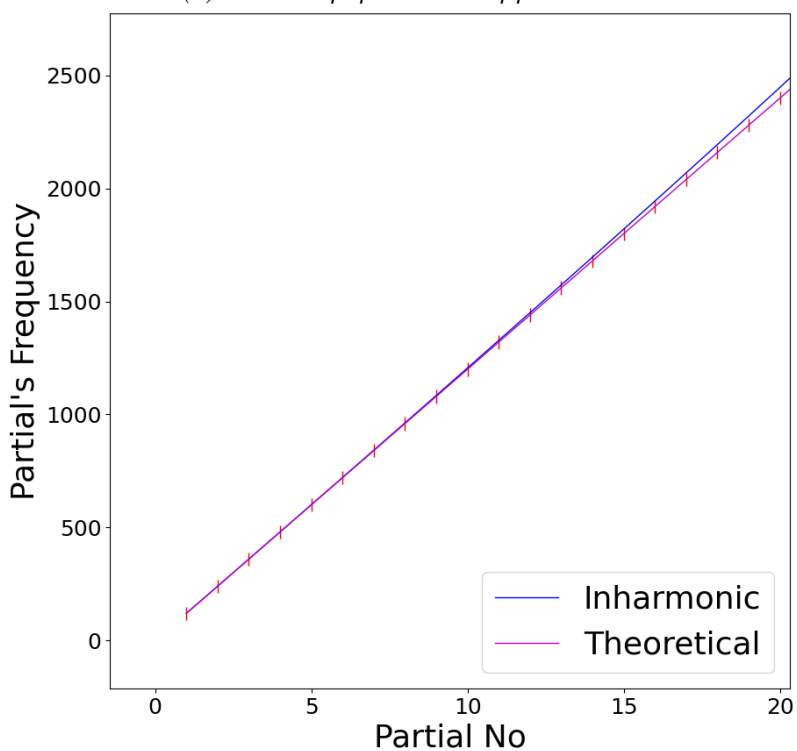
Ως έναρξη (Onset) αναφέρεται η χρονική στιγμή έναρξης ενός συγκεκριμένου μουσικού γεγονότος. Αυτό μπορεί να είναι μια νότα σε μια κιθάρα, μια συγχορδία, ένα χτύπημα σε ένα κρουστό ή κάποιο άλλο μουσικό γεγονός. Η σωστή αναγνώριση αυτής της χρονικής στιγμής, είναι κρίσιμης σημασίας σε πολλές εφαρμογές σχετικές με άντληση μουσικής πληροφορίας. Η γενικά ιδέα, έγκειται στην αναγνώριση απότομων αλλαγών. Σε πολλές περιπτώσεις αυτό μπορεί να επιτευχθεί εντοπίζοντας τις περιοχές όπου η περιβάλλουσα πλάτους αυξάνει. Παρόλα αυτά, σε άλλες περιπτώσεις, όπου οι μεταβάσεις είναι πιο ομαλές, το πρόβλημα γίνεται σημαντικά πιο περίπλοκο [4]. Χαρακτηριστικό παράδειγμα φαίνεται και στην εικόνα 3.9.

### 3.4.2 Εκτίμηση Τονικού Ύψους (pitch estimation)

Ένα περιοδικό σήμα μπορεί να αναλυθεί ως μια υπέρθεση συχνοτήτων, ακέραια πολλαπλάσια της θεμελιώδους. Το πρόβλημα του εντοπισμού της θεμελιώδους συχνότητας ( $f_0$ ) αναφέρεται ως *εκτίμηση τονικού ύψους* (pitch estimation) [34]. Στην σύνθετη εκδοχή του, απαιτείται η αναγνώριση πολλαπλών θεμελιωδών συχνοτήτων, όπως στην περίπτωση μιας συγχορδίας, ένα πρόβλημα αυξημένης δυσκολίας. Τέτοιες περιπτώσεις αναφέρονται στη βιβλιογραφία ως αναγνώριση πολλαπλών τονικών υψών (multi-pitch estimation).

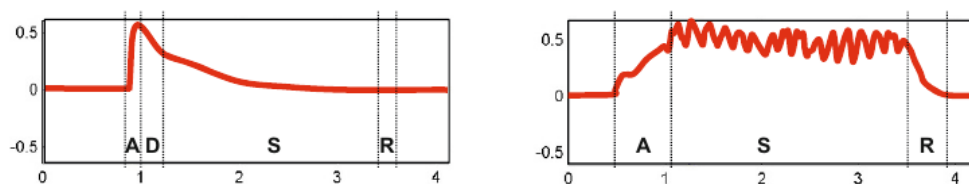


(α) απόκλιση πρώτων 50 αρμονικών



(β) σε μεγέθυνση

Σχήμα 3.8: αποτύπωση της απόκλισης των αρμονικών σε σχέση με τις θεωρητικές τιμές τους για  $\beta = 10^{-4}$ ,  $f_0 = 120\text{Hz}$ . Οι κάθετες γραμμές αποτυπώνουν το μήκος του παραθύρου αναζήτησης.



Σχήμα 3.9: Αριστερά η περιβάλλουσα πλάτους για πιάνο δεξιά για βιολί. Πηγή [4]

### 3.5 Επεξεργασία Σήματος και Αναπαραστάσεις

#### DFT-FFT

Ο διακριτός μετασχηματισμός Fourier (DFT) αποτελείται από μια ακολουθία αντί συνάρτησης μιας συνεχούς μεταβλητής, που αντιστοιχεί σε ισαπέχοντα (στο πεδίο της συχνότητας) δείγματα του μετασχηματισμού Fourier του σήματος [35].

Ο ευθύς και αντίστροφος μετασχηματισμός μπορεί να υπολογιστεί από τις παρακάτω εξισώσεις:

$$X[k] = \sum_{n=0}^{N-1} x[n] W_N^{kn} \quad (3.14)$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] W_N^{-kn} \quad (3.15)$$

όπου:

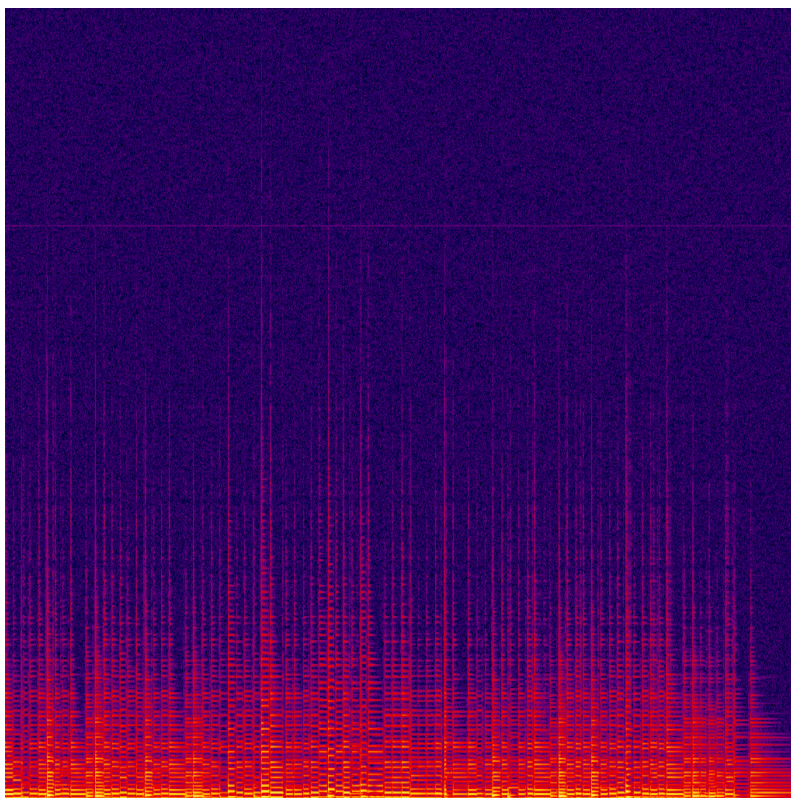
$$W_N = e^{-j(2\pi/N)} \quad (3.16)$$

Παρόλα αυτά, ο υπολογισμός του διακριτού μετασχηματισμού Fourier, αποτελεί ένα απαιτητικό και κοστοβόρο πρόβλημα. Έτσι αναπτύχθηκαν αλγόριθμοι οι οποίοι εκμεταλλεύονται την περιοδικότητα και τη συμμετρία του συντελεστή  $W_N$  ώστε να μειώσουν σημαντικά την πολυπλοκότητα των υπολογισμών. Η οικογένεια αυτών των αλγορίθμων ονομάζεται Fast Fourier Transform (FFT από δω και στο εξής), και βασίζονται στο διαδοχικό "πάσιμό" του σε όλο και μικρότερους μετασχηματισμούς Fourier [35]. Παράδειγμα αποτύπωσης σε γράφημα του FFT ενός σήματος, φαίνεται και στο σχήμα 3.5.

#### Φασματογράφημα - STFT

Το φασματογράφημα (spectrogram) αποτελεί ένα εργαλείο που μας προσφέρει μια οπτική αναπαράσταση του συχνοτικού περιεχομένου ενός σήματος και το πως αλλάζει στο χρόνο. Κλασική αναπαράσταση που παρουσιάζεται και στο σχήμα 3.10, είναι ο ένας άξονας να αντιστοιχεί στις συχνότητες, ο άλλος στο χρόνο και το χρώμα στην ένταση. Συνήθως χρησιμοποιείται ο Short Time Fourier Transform (STFT) για το υπολογισμό του, ο οποίος αποτελείται από τους συνενωμένους μετασχηματισμούς Fourier κομματιών του αρχικού σήματος στο χρόνο. Τα κομμάτια μπορεί να είναι μερικώς επικαλυπτόμενα.





Σχήμα 3.10: Φασματογράφημα ενός κομματιού κιθάρας 20 δευτερολέπτων

## 3.6 Μηχανική Μάθηση

Το αντικείμενο της μηχανικής μάθησης είναι η διερεύνηση τρόπων, για τον προγραμματισμό υπολογιστών ώστε να βελτιώνονται αυτόματα με τη γνώση [36]. Η μηχανική μάθηση, αποτελεί απαραίτητο εργαλείο στις μελέτες σχετικά με την εξαγωγή μουσικής πληροφορίας, ενώ χρησιμοποιούνται και για μουσική σύνθεση. Αναλόγως τον τρόπο εκπαίδευσης του αλγορίθμου χωρίζονται σε 3 βασικές υποκατηγορίες.

### 3.6.1 Επιβλεπόμενη Μάθηση

Ως επιτηρούμενη μάθηση, εννοούμε τα μοντέλα μηχανικής μάθησης, στα οποία διατίθενται είσοδοι ταιριασμένες με τις αντίστοιχες εξόδους. Επιχειρείται έτσι, να προσδιορίσει μια συνάρτηση μεταξύ εισόδων εξόδων από τα δεδομένα εκπαίδευσης που του διατίθενται. Στόχος είναι, μέσω αυτής, να μπορέσει μεταγενέστερα να προβλέψει εξόδους, για εισόδους που δεν έχει ξανασυναντήσει [37].

#### Naives Bayes Classifier

Ο Naives Bayes Classifier είναι ένας απλός αλγόριθμος επιτηρούμενης μάθησης. Κάθε είσοδος περιγράφεται από ένα σύνολο χαρακτηριστικών  $\langle a_1, a_2, \dots, a_n \rangle$  ενώ οι εξόδους, είναι ένα πεπερασμένο σύνολο  $V$ . Ο αλγόριθμος εκπαιδεύεται σε δεδομένα ζευγών εισόδων-εξόδων. Τελικά, αντιστοιχεί τις υπό εξέταση εισόδους με χαρακτηριστικά στην πιθανότερη έξοδο βάση μιας συνάρτησης  $U_{MAP}$  όπου:



$$U_{MAP} = \operatorname{argmax} P(u_j | a_1, a_2 \dots a_n) \quad (3.17)$$

Από το θεώρημα του Bayes παίρνουμε ότι:

$$\begin{aligned} U_{MAP} &= \operatorname{argmax} \frac{P(a_1, a_2 \dots a_n | u_j) \cdot P(u_j)}{P(a_1, a_2 \dots a_n)} \\ &= \operatorname{argmax} P(a_1, a_2 \dots a_n | u_j) \cdot P(u_j) \end{aligned} \quad (3.18)$$

Τέλος, στον Naives Bayes υποθέτουμε ότι τα γνωρίσματα  $\langle a_1, a_2, \dots a_n \rangle$  είναι ανεξάρτητα. Άρα  $P(a_1, a_2 \dots a_n | u_j) = \prod_i P(a_i | u_j)$  άρα καταλήγουμε:

$$U_{NB} = \operatorname{argmax} P(u_j) \cdot \prod_i P(a_i | u_j) \quad (3.19)$$

που τελικά αποτυπώνει την πιθανότητα που αποδίδει ο Naive Bayes στην έξοδο  $u_j$  δεδομένων γνωρισμάτων εισόδου  $\langle a_1, a_2, \dots a_n \rangle$ . [36]

### 3.6.2 Μη επιβλεπόμενη μάθηση

Ως μη επιτηρούμενη μάθηση, εννοούμε τα μοντέλα στα οποία δίνονται μόνο δεδομένα εισόδου. Ακολούθως ο αλγόριθμος επιχειρεί να αναγνωρίσει τη δομή τους ή επαναλαμβανόμενα μοτίβα σε αυτά. Πχ, να αναγνωρίσει συστάδες που παρουσιάζονται ή συσχέτιση μεταξύ των δεδομένων.

### 3.6.3 Ενισχυτική Μάθηση

Ως ενισχυτική μάθηση, αναφέρονται τα μοντέλα τα οποία μαθαίνουν από άμεση αλληλεπίδραση με το περιβάλλον. Σε αντίθεση με την επιτηρούμενη μάθηση, το σύστημα δεν έχει ένα 'δάσκαλο' να το οδηγεί αν κάτι είναι σωστό ή λανθασμένο, αλλά ποινές και αμοιβές, αναλόγως τις επιλογές που κάνει.

### 3.6.4 Νευρωνικά Δίκτυα

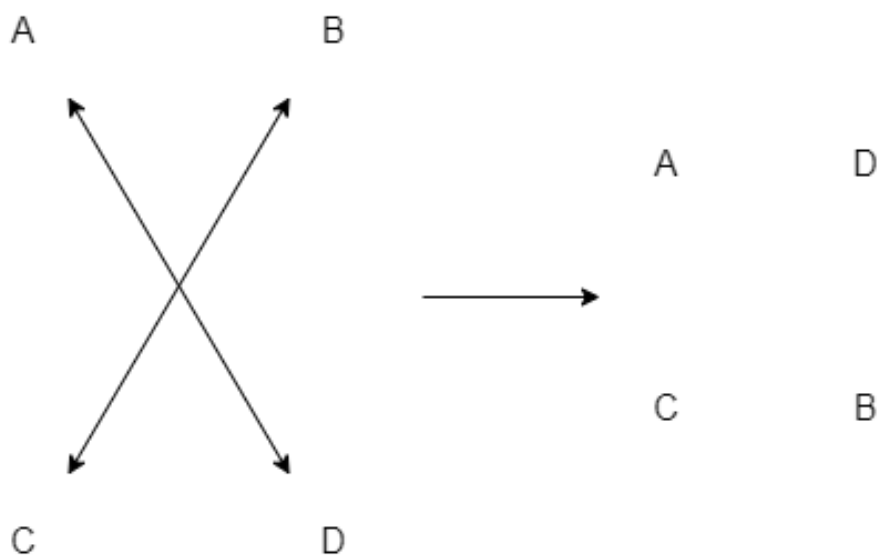
Τα νευρωνικά δίκτυα επιχειρούν να προσομοιάσουν τις συνάψεις του ανθρωπίνου εγκεφάλου. Ένα feedforward neural network μπορεί να θεωρηθεί ως μια μη γραμμική μαθηματική συνάρτηση, που αντιστοιχεί εισόδους σε εξόδους [38]. Αποτελείται από πολλά επίπεδα σύνδεσης νευρώνων μέσω ακμών. Κατά τη διαδικασία της εκμάθησης, αλλάζουν τα βάρη των ακμών, βάσει των ζευγών εισόδων-εξόδων από τα δεδομένα εκπαίδευσης. Συνήθως, οι νευρώνες ομαδοποιούνται σε επίπεδα, όπου κάθε επίπεδο, επιτελεί ένα συγκεκριμένο σκοπό. Π.χ. ένα επίπεδο μπορεί να χρησιμοποιείται για να αναγνωρίσει κάθετες ακμές σε μια εικόνα, ενώ ένα άλλο οριζόντιες, με τελικό στόχο να αναγνωρίζει το σύστημα εικόνες από τετράγωνα σχήματα.

### 3.7 Γενετικοί αλγόριθμοι

Όπως και με τα νευρωνικά δίκτυα, έτσι και οι γενετικοί αλγόριθμοι είναι εμπνευσμένοι από τη βιολογία. Πιο συγκεκριμένα, επιχειρούν να μιμηθούν τη διαδικασία της εξέλιξης. Αν και υπάρχουν πολλές παραλλαγές που εγχολπώνουν διαφορετικές διαδικασίες της εξέλιξης, η κυριότερη αποτελείται από τις φάσεις της διασταύρωσης, μετάλλαξης και φυσικής επιλογής ενός αρχικού πληθυσμού. Αξιοποιούνται ώστε να εξερευνηθεί ένα μεγάλο πεδίο αναζήτησης. Κάθε άτομο, μοντελοποιείται ως ένα χρωμόσωμα, μια συμβολοσειρά εν ολίγοις, η οποία αντιπροσωπεύει μια πιθανή λύση του προβλήματος. Ακολούθως, έχοντας έναν αρχικό πληθυσμό, τα άτομα διασταυρώνονται, μεταλλάσσονται και από αυτά επιλέγονται πιθανοτικά τα καλύτερα για την επόμενη γενιά, μέχρις ότου παρατηρηθεί σύγκλιση ή εξαντληθούν οι πόροι του συστήματος που έχουμε ορίσει. Συνεπώς, τα «γονίδια» των ατόμων που θεωρούνται καταλληλότερα, θα περάσουν στις επόμενες γενιές κατά τη φυσική επιλογή και τελικά η μέθοδος θα συγκλίνει σε μια πιθανή λύση του προβλήματος.

#### Διασταύρωση

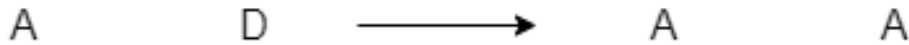
Κατά τη διαδικασία της διασταύρωσης, επιλέγονται (συνήθως) ζευγάρια ατόμων από τον πληθυσμό. Ακολούθως, συνδυάζονται μέρη των χρωμοσωμάτων τους, ώστε να παράξουν δυο νέα άτομα.



Σχήμα 3.11: Διασταύρωση δυο γονέων, A-B και C-D όπου προκύπτουν τα παιδιά A-D και C-D

#### Μετάλλαξη

Στη μετάλλαξη ενός ατόμου, αλλάζει ένα μέρος του χρωμοσώματός του. Η εξερεύνηση του search space δεν είναι εξαντλητική. Η μετάλλαξη, είναι σημαντική ώστε να μειωθούν οι πι-

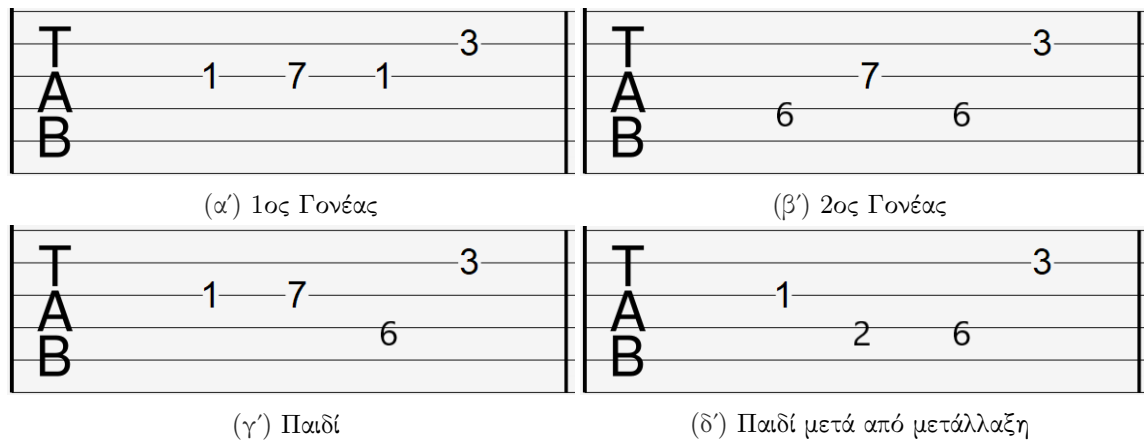


Σχήμα 3.12: Μετάλλαξη ενός ατόμου

θανότητες εμφάνισης τοπικών μεγίστων, και να διευρύνουμε κατά τρόπο τυχαίο την αναζήτησή μας.

### Επιλογή

Η επιλογή είναι η διαδικασία του αλγορίθμου, κατά την οποία επιλέγονται τα πιο κατάλληλα άτομα από τον τρέχοντα πληθυσμό για να διαμορφώσουν την επόμενη γενιά. Πιο συγκεκριμένα, μέσω μιας συνάρτησης καταλληλότητας (η οποία επιβραβεύει θεμιτά χαρακτηριστικά, τιμωρεί ανθέμιτα) ταξινομούνται τα άτομα και επιλέγονται αυτά που διαθέτουν τα στοιχεία που επιθυμούμε στην επόμενη γενιά. Η συνάρτηση καταλληλότητας τελικά, μεγιστοποιείται ή ελαχιστοποιείται στην τελική λύση που θα βρούμε, η οποία δεν είναι αναγκαίο να είναι η βέλτιστη.



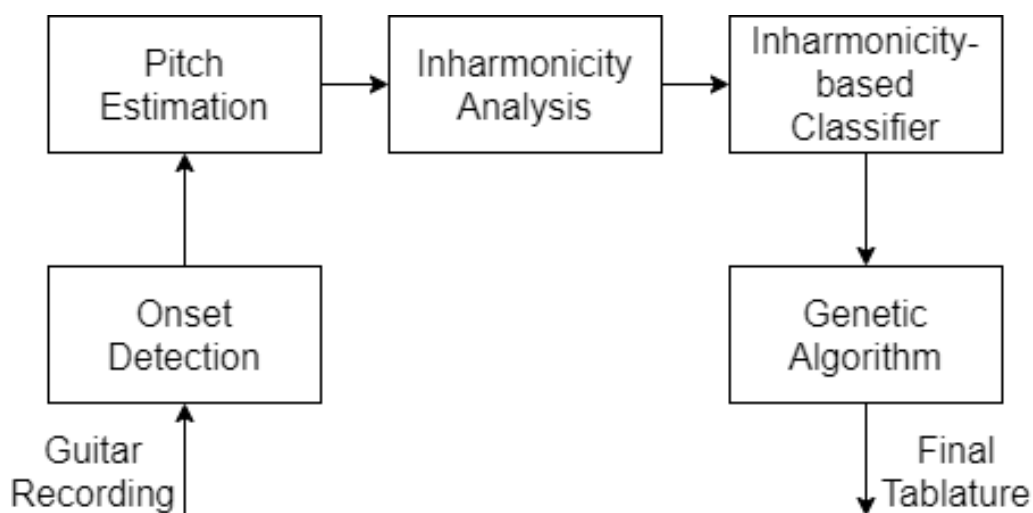
Σχήμα 3.13: Παράδειγμα ενός πιθανού τρόπου διασταύρωσης και μεταλλαγής σε περίπτωση ταμιλατούρας.



## Περιγραφή Μεθόδου

---

Σε αυτό το κεφάλαιο, παρουσιάζονται τα επιμέρους υποσυστήματα που σχεδιάστηκαν και υλοποιήθηκαν για την αυστηρή μεταγραφή ταμπλατούρας. Πιο συγκεκριμένα, τα επιμέρους υποσυστήματα για την αναρμονική ανάλυση, την αναγνώριση ενεργής χορδής με βάση τον συντελεστή αναρμονικότητας και τον γενετικό αλγόριθμο παρουσιάζονται εκτενώς παρακάτω, ενώ τα ήδη ανεπτυγμένα υποσυστήματα για αναγνώριση έναρξης νότας και τονικού ύφους [5] που χρησιμοποιήθηκαν για την απ΄ άκρη σ΄ άκρη μεταγραφή ταμπλατούρας παρουσιάζονται στα παραρτήματα Α και Β αντίστοιχα.



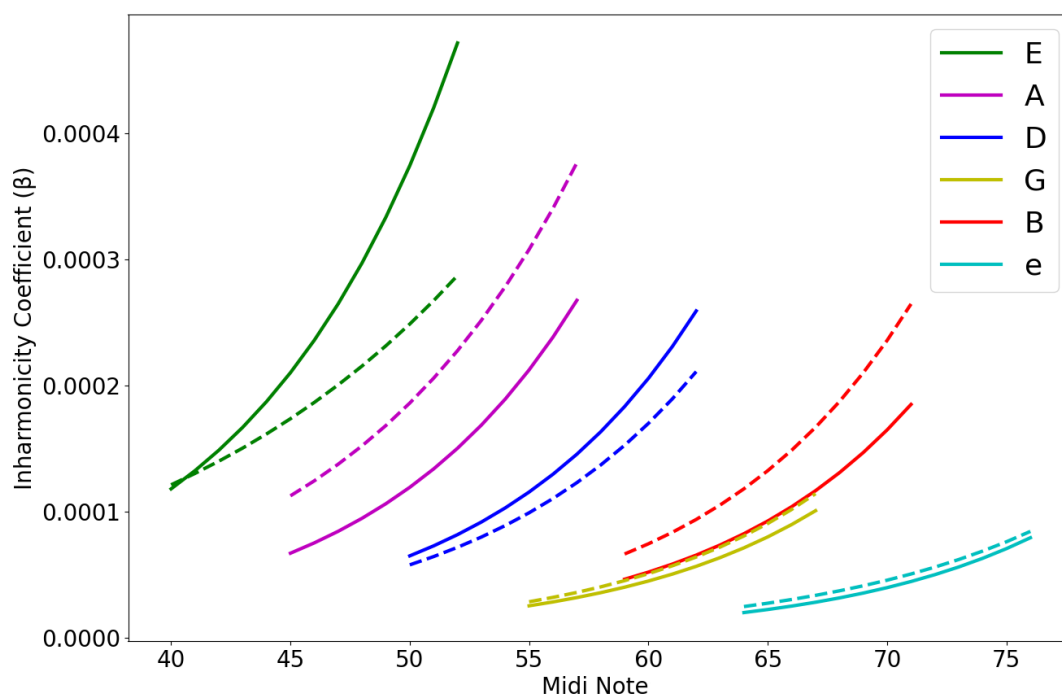
Σχήμα 4.1: Απο άκρη-σε-άκρη σύστημα μεταγραφής ακριβούς ταμπλατούρας

### 4.1 Συνοπτική Παρουσίαση

Τον πυρήνα του συστήματος που παρουσιάζεται αποτελεί ένας αλγόριθμος αναγνώρισης ενεργής χορδής βασισμένος σε αναρμονική ανάλυση. Χρειάζεται μόνο μερικά στιγμιότυπα νοτών για να προσαρμοστεί στην κάθε κιθάρα, ενώ συμπληρώνεται από ένα επόμενο στάδιο που εγχολπώνει τις φυσικές δυσκολίες και περιορισμούς του παιξίματος κιθάρας, για να βελτιώσει την ακρίβεια των αποτελεσμάτων.

Κατά τη φάση της προσαρμογής του αλγορίθμου αναγνώρισης, ο συντελεστής αναρμονικότητας  $\beta$  υπολογίζεται μόνο για ένα μικρό υποσύνολο νοτών, (από 1 έως 3 τάστα κάθε

μιας από τις 6 χορδές) χρησιμοποιώντας μια παραλλαγή του αλγορίθμου που παρουσίασαν οι Barbancho et al. [3] για μονοφωνικές ηχογραφήσεις (στο κεφάλαιο 3 παρουσιάζεται τόσο η αρχική πρόταση των Barbancho et al. όσο και η τροποποιημένη εκδοχή που χρησιμοποιήθηκε). Ακολούθως, γενικεύεται η εκτίμηση των  $\beta$  για όλα τα τάστα σύμφωνα με αυτά τα λίγα δείγματα. Έτσι, αντιστοιχείται μια τιμή του  $\beta$  σε κάθε τάστο, που προσεγγίζει την συμπεριφορά της αναρμονικότητάς του και κατ' επέκταση ολόκληρης της χορδής όπως φαίνεται στο σχήμα 4.2.

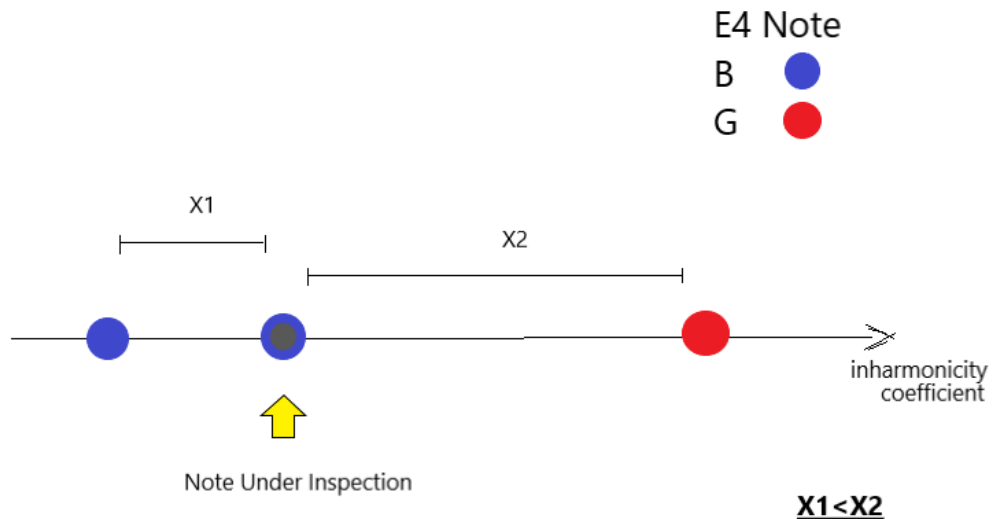


Σχήμα 4.2: Εξέλιξη του συντελεστή αναρμονικότητας  $\beta$  για κάθε midi νότα βασισμένη στη σχέση (1.2) για της μεθόδους προσαρμογής 1Fret (συμπαγής) και 3Fret (διακεκομμένη).

Ομοίως κατά την φάση της ταξινόμησης, υπολογίζονται οι  $\hat{\beta}$  συντελεστές για κάθε ανιχνευμένο στιγμιότυπο νότας σε μια ηχογραφημένη παράσταση και ανατίθεται σε κάθε ένα, το ζεύγος χορδής τάστου με το πλησιέστερο  $\beta$  4.3. Τέλος, η αρχική εκτίμηση της ταμπλατούρας τροφοδοτείται σε ένα γενετικό αλγόριθμο, ο οποίος αναθέτει την κατάλληλη χορδή στα ασαφή (inconclusive) στιγμιότυπα νοτών και βελτιώνει τα αποτελέσματα συνολικά, τροποποιώντας κάποιες προβλέψεις.

Οι υπομονάδες για ανίχνευση έναρξης νότας (onset detection) και εκτίμηση τονικού ύψους (pitch estimation), είναι πολύ σημαντικές, καθώς ο προσδιορισμός ενεργής χορδής βασίζεται στην γνώση των χρονικών στιγμών και των αντίστοιχων τονικών υψών. Μετά την προσαρμογή του ταξινομητή, σύμφωνα με τους υπολογισμένους συντελεστές αναρμονικότητας, το σύστημα είναι έτοιμο να δεχθεί νέες ηχογραφήσεις προς επεξεργασία όπως παρουσιάζεται στο σχήμα 4.1.

Πρωτίστως, εκτιμούνται οι στιγμές έναρξης νοτών (onset times) με το TCN μοντέλο [9] (κώδικας στο [39]), και το τονικό ύψος με το προ-εκπαιδευμένο μοντέλο CREPE [5]. Μετά συμπεραίνεται το ζεύγος χορδής-τάστου μέσω της αναρμονικής ανάλυσης και εφαρμόζεται το τελικό στάδιο διορθώσεων που ενσωματώνει τα φυσικά χαρακτηριστικά και τις δυσκολίες του



Σχήμα 4.3: Παράδειγμα αντιστοίχισης στιγμιοτύπου νότας με βάση την αναρμονικότητα. Η νότα E4 που εξετάζεται, τελικά αντιστοιχείζεται στη χορδή με τη πλησιέστερη αναρμονικότητα δηλαδή την B.

παιξίματος κιθάρας μέσω του προτεινόμενου γενετικού αλγόριθμου. Τα επιμέρους στοιχεία που αναπτύχθηκαν για την μεταγραφή ταμπλατούρας παρουσιάζονται λεπτομερώς παρακάτω.

## 4.2 Υπολογισμός Αναρμονικότητας

Ο αλγόριθμος υπολογισμού της αναρμονικότητας, λειτουργεί πάνω σε κομμάτια διάρκειας 60ms που αντιστοιχούν σε 1000 νότες το λεπτό, ή δεκατάεκτα παιζόμενα σε 250bpm, ένα παράθυρο που είναι επαρκές για τις περισσότερες εκτελέσεις σε κιθάρα.

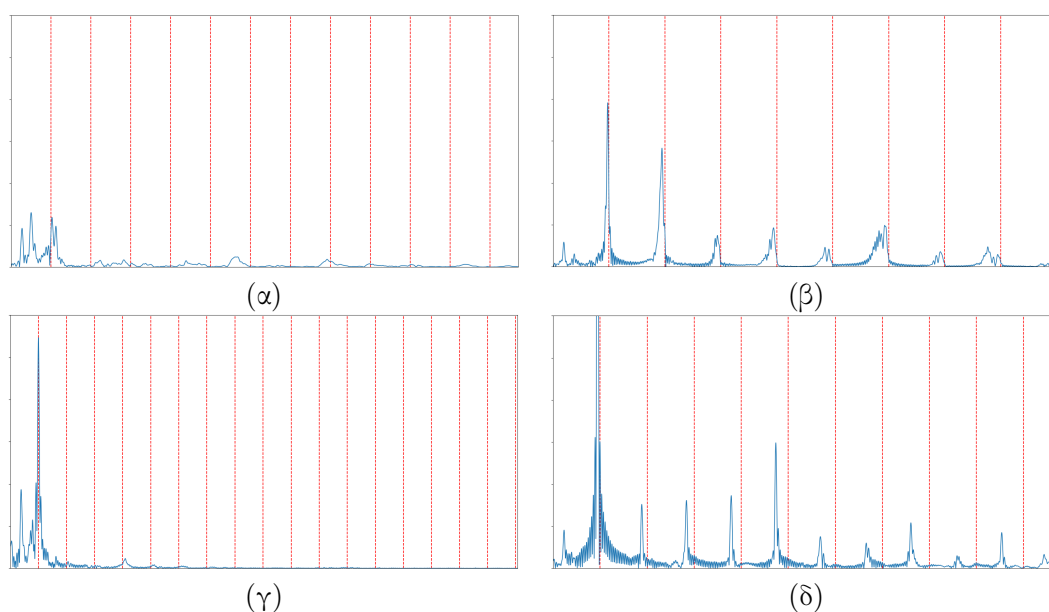
Ένας φαινομενικά προφανής τρόπος να υπολογιστεί ο συντελεστής αναρμονικότητας θα ήταν η επίλυση της εξίσωσης (1.1) ως προς  $\beta$ , δεδομένης της τιμής της θεμελιώδους και μιας μόνο μερικής συχνότητας. Αυτό όμως θα αποτύγχανε να προσδώσει ικανή ακρίβεια για το υπό εξέταση πρόβλημα. Πιο συγκεκριμένα, υψηλής τάξεως αρμονικές προσφέρουν μεγαλύτερη ακρίβεια αφού η διαφορά από την αναμενόμενη συχνότητα είναι σχετικά μεγαλύτερη. Παρόλα αυτά η αναγνώριση υψηλής τάξεως αρμονικών αυξάνει σε δυσκολία χωρίς πρότερη γνώση του συντελεστή  $\beta$ . Από την άλλη, οι αποκλίσεις των χαμηλής τάξεως αρμονικών είναι τόσο μικρές που συνήθως δεν φανερώνουν επαρκείς πληροφορίες για τον ακριβή υπολογισμό της αναρμονικότητας. Για αυτό, γίνεται αναγκαίο να συνυπολογιστούν πολλαπλές αρμονικές στην ανάλυση για να βελτιωθεί η εκτίμηση του  $\beta$ .

Στην παρούσα εργασία, εφαρμόζεται μια παραλλαγή του αλγορίθμου που παρουσιάστηκε από τις Barbancho et al. [3] (βλ. Κεφάλαιο 3.3.1). Αντί για έως 50 μερικές συχνότητες λαμβάνονται υπόψη μόνο οι 13 πρώτες, χωρίς την επαναληπτική επανατοποθέτηση του παραθύρου που χρησιμοποιείται για την αναγνώρισή τους, καθώς μεγάλης τάξεως μερικές συχνότητες είναι δύσκολο να υπολογιστούν χωρίς επανατοποθέτηση παραθύρου, ενώ οι ενδιαμέσοι υπολογισμοί του συντελεστή αναρμονικότητας για την επανατοποθέτηση παραθύρου είναι πιθανότερο να είναι λανθασμένοι για θορυβώδη σήματα. Επίσης, η συγκεκριμένη υλοποίηση, προϋποθέτει πρότερη γνώση της θεμελιώδους συχνότητας, σε αντίθεση με το [3] που γίνεται ταυτόχρο-

νος προσδιορισμός χορδής και θεμελιώδους συχνότητας. Η μέθοδος ελαχίστων τετραγώνων χρησιμοποιείται για να υπολογίσουμε τον συντελεστή αναρμονικότητας από τις μετρημένες αποκλίσεις των μερικών συχνοτήτων. Για την αναγνώριση τους, εφαρμόζεται FFT σε παράθυρα των 60ms με zero-padding ώστε να καταλήξει στα  $2^{18}$  δείγματα (όπως σε παλιότερες μελέτες [3]), και χρησιμοποιείται ένα παράθυρο για αναγνώριση κορυφών (peak/partial detection) ίσο με  $f_0/2$ .

Πριν την παραπάνω διαδικασία, επαναυπολογίζεται η συχνότητα  $f_0$ , γνωστή από τις επισημειώσεις (annotations) ή την υπομονάδα εκτίμησης τονικού ύψους, βρίσκοντας την ψηλότερη κορυφή σε ένα παράθυρο  $\pm 1.6Hz$  και υπολογίζοντας το σταθμισμένο μέσο όρο σε παράθυρο ίδιου μεγέθους.

Όταν ο υπολογισμένος συντελεστής αναρμονικότητας δεν είναι έγκυρος, ( $\beta$  μικρότερο του  $10^{-7}$ ) χαρακτηρίζονται αυτά τα στιγμιότυπα νοτών ως ασαφή (inconclusive). Ο FFT μερικών χαρακτηριστικών ασαφών στιγμιοτύπων παρουσιάζεται στο σχήμα 4.4.



Σχήμα 4.4: Περιπτώσεις ασαφών νοτών: (α) θορυβώδεις, (β) απόκλιση προς τα πίσω, (γ) χαμηλής έντασης αρμονικές και (δ) λανθασμένη θεμελιώδης.

#### 4.2.1 Ταξινομητής χορδής βασισμένος σε αναρμονική ανάλυση

Ως ένα πρώτο βήμα προς την εύρεση ακριβούς ταμπλατούρας, μια απλή μέθοδος για ανίχνευση ενεργής χορδής προτείνεται βάση του μετρούμενου συντελεστή αναρμονικότητας. Κατά τη φάση της πρόβλεψης χορδών θεωρούνται δεδομένα, από τα προηγούμενα στάδια, οι ενάρξεις και τα τονικά ύψη όλων των στιγμιοτύπων. Ακολουθώντας, σε κάθε υπό διερεύνηση στιγμιότυπο νότας, υπολογίζεται ο συντελεστής αναρμονικότητάς του και ανατίθεται στο ζεύγος χορδής-τάστου, του οποίου η αναμενόμενη τιμή είναι πλησιέστερα στην μετρούμενη τιμή αναρμονικότητας (βλ. Σχήμα 4.3). Όταν η μέθοδος αδυνατεί να συμπεράνει την διεγερμένη χορδή, το στιγμιότυπο νότας χαρακτηρίζεται ως ασαφές.

Όπως αναφέρεται παραπάνω, ο συντελεστής αναρμονικότητας μιας νότας που παίζεται σε μια χορδή  $s$  και ένα τάστο  $n$ , μπορεί να υπολογιστεί από το  $\beta$  της ανοιχτής χορδής με την



εξίσωση (1.2). Εφόσον αλλάζοντας τον εξοπλισμό (π.χ. χορδές) μπορεί να επηρεαστεί η αναρμονικότητα του οργάνου, είναι σημαντικό να μπορεί το μοντέλο να προσαρμοστεί γρήγορα. Κατά τη φάση της προσαρμογής του μοντέλου στην εκάστοτε κιθάρα, εκτιμάται ο συντελεστής αναρμονικότητας κάθε ζεύγους χορδής-τάστου για όλη την ταστιέρα της κιθάρας. Προς αυτό το σκοπό, σχεδιάστηκαν τέσσερις παρόμοιες μέθοδοι για την προσαρμογή του μοντέλου ταξινόμησης χορδής με αναρμονική ανάλυση, οι οποίες απαιτούν τον υπολογισμό της αναρμονικότητας έως 3 θέσεων σε κάθε χορδή.

### Μέθοδος Ενός Τάστου (1Fret)

Στην απλούστερη εκδοχή, μετريέται ο συντελεστής  $\beta(s, 0)$  κάθε ανοιχτής χορδής και ακολούθως εκτιμώνται τα  $\beta(s, n)$  για κάθε τάστο  $n$  μέσω της εξίσωσης (1.2). Αυτή η μέθοδος επιτρέπει την ταχύτερη διαδικασία προσαρμογής αφού απαιτεί τα ελάχιστα στιγμιότυπα νοτών για κάθε χορδή.

### Μέθοδος Τριών Τάστων (3Fret)

Είναι σύνηθες μια ταστιέρα να είναι ελαφρώς σκεβρωμένη (σκέβρωμα ονομάζεται η καμπύλωση της ταστιέρας της κιθάρας συνήθως λόγω υγρασίας και θερμοκρασίας) ενώ μπορεί να υπάρχουν και άλλες ιδιαιτερότητες σχετιζόμενες με το κάθε όργανο. Ενδέχεται λοιπόν, σε μερικές περιπτώσεις, όπως το χέρι μετακινείται στην ταστιέρα προς το σώμα της κιθάρας, η εξίσωση (1.2) να μην είναι απόλυτα ακριβής. Για τη διερεύνηση της παραπάνω υπόθεσης υπολογίστηκαν οι διάμεσοι συντελεστές αναρμονικότητας όλων των στιγμιοτύπων νοτών στο GuitarSet dataset [28] (σχετικά με το GuitarSet dataset, βλ. κεφάλαιο 5.1) για κάθε θέση στην ταστιέρα και σχεδιάστηκαν στο γράφημα 4.5. Ακολούθως, για κάθε τάστο  $n$  υπολογίστηκε το τροποποιημένο  $n'$  που αντιστοιχεί στις υπολογισμένες διάμεσες τιμές  $\beta$  λύνοντας την εξίσωση (1.2) ως προς  $n$ :

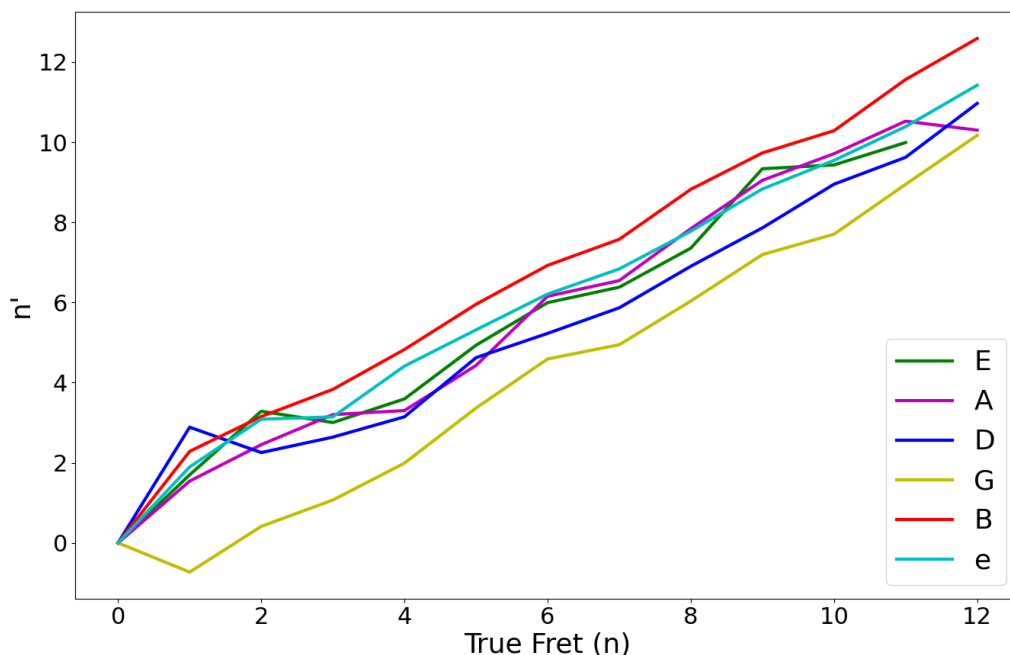
$$n' = 6 \cdot (\log_2(\beta(s, n)) - \log_2(\beta(s, 0))). \quad (4.1)$$

Αποτυπώθηκε η σχέση μεταξύ του πραγματικού τάστου  $n$  σε σχέση με το  $n'$  για κάθε χορδή στο σχήμα 4.5. Στο γράφημα, παρατηρείται ότι οι μετρούμενες τιμές, αποκλίνουν από την αναμενόμενη  $y = x$  καμπύλη, με μη τυχαίο τρόπο, και υποδεικνύουν μια γραμμική σχέση μεταξύ των  $n'$  και  $n$  όπως εκφράζεται παρακάτω.

$$n' = a \cdot n + b, \text{ όπου } a, b \in \mathbb{R}. \quad (4.2)$$

Λαμβάνοντας αυτά υπόψη, σχεδιάσαμε τρεις ακόμη μεθόδους για τον υπολογισμό του  $\beta$  οι οποίες συμπεριλαμβάνουν παραπάνω από ένα τάστο σε κάθε χορδή. Βέβαια, τα τάστα αναπαρίστανται από ακέραιες τιμές, συνεπώς το  $n'$  δεν αντιπροσωπεύει κάποιο φυσικό χαρακτηριστικό του οργάνου, αλλά αποκτά νόημα μόνο σε θεωρητικό/μαθηματικό πλαίσιο. Αντικαθιστώντας το  $n$  με  $n'$  στην εξίσωση (1.2) μπορεί να υπολογιστεί ο συντελεστής αναρμονικότητας από τα 2 τάστα  $(i, j)$  της επιλογής μας και της ανοιχτής χορδής ως:

$$a = \frac{\log_2(\beta(s, i)) - \log_2(\beta(s, j))}{i - j} \quad (4.3)$$



Σχήμα 4.5: Εξέλιξη του  $n'$  σε σχέση με το τάστο  $n$  σύμφωνα με τον υπολογισμό του  $\beta$  για κάθε χορδή. Για κάθε πραγματικό ζεύγος χορδής τάστου, το  $n'$  υπολογίζεται από τη διάμεσο των αντίστοιχων  $\beta$  όλων των στιγμιοτύπων νοτών που υπάρχουν στο GuitarSet dataset.

$$b = 6 \cdot (\log_2(\beta(s, i)) - \log_2(\beta(s, 0))) - i \cdot a \quad (4.4)$$

### Μέθοδος Δύο Τάστων (2Fret)

Όμοια, σχεδιάστηκαν 2 ακόμη μέθοδοι που απαιτούν το  $\beta$  μόνο ενός τάστου και της ανοιχτής χορδής για να γίνει η προσαρμογή.

Οι μέθοδοι 2FretA και 2FretB προκύπτουν από την πρότερη ανάλυση, για  $b = 0$  ή  $a = 1$  στις εξισώσεις (4.2) αντίστοιχα. Σε κάθε περίπτωση υπολογίζονται τα αντίστοιχα  $a$  και  $b$  ως:

$$a = 6 \cdot \frac{\log_2 \beta(s, i) - \log_2 \beta(s, 0)}{i}, \quad (4.5)$$

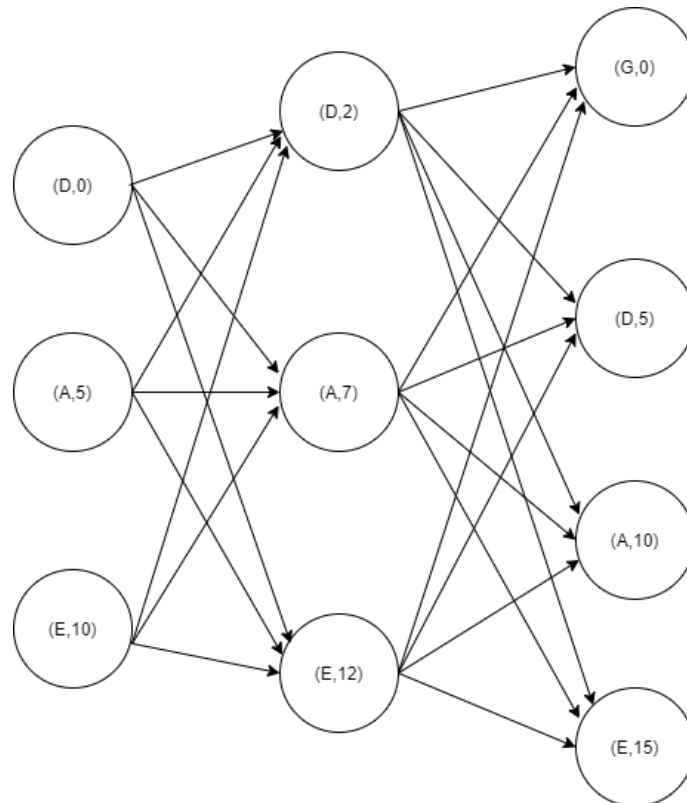
$$b = 6 \cdot (\log_2(\beta(s, i)) - \log(\beta(s, 0))) - i. \quad (4.6)$$

Οι συγκεκριμένες δύο μέθοδοι, προτείνονται ως μια εναλλακτική, οι οποίες μπορούν να εξισορροπήσουν το χρόνο που διαθέτει ο κιθαρίστας ώστε να εκπαιδεύσει το μοντέλο και την ακρίβεια που επιθυμεί.

### 4.2.2 Εισάγοντας Περιορισμούς Εκτελεσιμότητας - Γενετικός Αλγόριθμος

Υποθέτοντας ότι κάθε νότα μπορεί να παιχτεί σε τρεις διαφορετικές θέσεις, τότε μια σειρά από  $N$  νότες αντιστοιχεί σε  $3^N$  διαφορετικές ταμπλατούρες, αν και μόνο μερικές από αυτές θα μπορούσαμε να θεωρήσουμε ότι "παίζονται". Στόχος της παρούσας ανάλυσης είναι να α-

ξιοποιηθούν οι περιορισμοί εκτελεσιμότητας που προκύπτουν από την γεωμετρία της κιθάρας ώστε να βελτιωθεί η ακρίβεια των αποτελεσμάτων. Επιπροσθέτως, θεωρώντας ότι ο ταξινομητής αναρμονικής ανάλυσης, έχει επιτύχει ένα ικανό ποσοστό ακρίβειας, αυτές οι προβλέψεις μπορούν να αξιοποιηθούν ώστε να αναδείξουν ταμπλατούρες οι οποίες μοιάζουν περισσότερο με το αποτέλεσμα του πρότερου σταδίου αναρμονικής ανάλυσης. Έτσι, συντίθεται τελικά η αναγνώριση χορδής από δύο συμπληρωματικά υποσυστήματα, αυτό της αναρμονικής ανάλυσης και αυτό του γενετικού αλγόριθμου με τους περιορισμούς εκτελεσιμότητας.



Σχήμα 4.6: Γράφημα που αναπαριστά όλες τις πιθανές ταμπλατούρες που αντιστοιχούν στην ακολουθία των νοτών 50 - 52 - 55 (αρίθμηση MIDI).

Εφόσον ο ταξινομητής δεν μπορεί να αποδώσει σε κάθε στιγμιότυπο νότας κάποια χορδή (π.χ ασαφή στιγμιότυπα), είναι αναγκαίο να βρεθεί ένας τρόπος να αντιστοιχούνται σε χορδές όσο το δυνατόν καλύτερα. Έτσι, με τους περιορισμούς που εισήχθησαν μέσω του γενετικού αλγόριθμου, επιτυγχάνεται τόσο να αντιστοιχηθούν τα ασαφή στιγμιότυπα, αλλά και να βελτιωθεί η ακρίβεια του συστήματος συνολικά.

Ακόμη, ταμπλατούρες που παίζονται από έναν άνθρωπο, οφείλουν να διατηρούν μια λογική δομή. Για παράδειγμα, μεγάλες αποστάσεις σε τάστα είναι παράλογες όταν δεν είναι απαραίτητες. Προς αυτό το σκοπό, ο γενετικός αλγόριθμος πριμοδοτεί ταμπλατούρες σχετικές με την έξοδο του ταξινομητή αναρμονικής ανάλυσης, ενώ παράλληλα λαμβάνει υπόψη την εκτελεσιμότητα της ταμπλατούρας. Έτσι ένας γενετικός αλγόριθμος μπορεί ταυτόχρονα να αντιστοιχίσει χορδές στις ασαφείς νότες, και να διορθώσει προφανέστατα λανθασμένες (βλ. Σχήμα 4.7).

(α) Ταξινόμηση Βασισμένη  
σε Αναρμονική Ανάλυση

(β) Επιδιόρθωση με  
Γενετικό Αλγόριθμο

Σχήμα 4.7: Ο γενετικός αλγόριθμος αντιστοιχεί τα ασαφή στιγμιότυπα σε χορδές και διορθώνει τις μη διαισθητικές περιπτώσεις.

### Fitness Function

Η ταμπλατούρα μπορεί να μοντελοποιηθεί ως ένα δισδιάστατο επίπεδο, με όρια τον αριθμό των χορδών και των τάστων για την κάθετη και οριζόντια συνιστώσα αντίστοιχα. Η συνάρτηση καταλληλότητας που χρησιμοποιείται εισάγει φυσικούς περιορισμούς σχετικούς με το παίξιμο μιας ταμπλατούρας λαμβάνοντας παράλληλα υπόψη τις προβλέψεις από την αναρμονική ανάλυση. Τελικά, προκύπτει από το άθροισμα 6 διαφορετικών συνιστωσών βασισμένες σε ιδέες από το [19] (οι οποίοι όμως, χρησιμοποίησαν κάποιες από αυτές τις συνιστώσες, για την παραγωγή εκτελέσιμης ταμπλατούρας από συμβολική γραφή και όχι για την εύρεση ακριβούς ταμπλατούρας από σήμα ήχου).

Επιγραμματικά οι 6 συνιστώσες είναι:

- **depress factor:** Ο αριθμός που χρειάστηκε να πατηθεί ένα διαφορετικό τάστο-χορδή. Διαισθητικά, 'τιμωρεί' εάν αλλάξουμε χορδή-τάστο ενώ παίζουμε την ίδια νότα.
- **distance from average position:** Η διαφορά κάθε νότας της ταμπλατούρας με το μέσο όρο τάστου και χορδής σε ένα διάστημα έως  $\pm 3$  νότες, εφόσον δεν μεσολαβεί σχετικά μεγάλο χρονικό διάστημα μεταξύ τους.
- **fret distance:** Η απόσταση σε αριθμό τάστων κάθε νότας, από την προηγούμενή της.
- **string distance:** Η απόσταση σε αριθμό χορδών, κάθε νότας από την προηγούμενή της.
- **open string:** Επειδή τα προηγούμενα κριτήρια, μπορεί να 'τιμωρίσουν' έμμεσα περιπτώσεις ανοιχτών χορδών, έχει προστεθεί ένας συμπληρωματικός παράγοντας ώστε να αναπληρώσει. Στην γενική περίπτωση, οι ανοιχτές θέσεις, βολεύουν για να παιχτεί μια ταμπλατούρα, αλλά εφόσον εμείς επιθυμούμε να βρούμε το πως παίχτηκε η συγκεκριμένη και όχι μόνο μια καλή απόδοση, πρέπει να μετριάσει προσεχτικά η συγκεκριμένη παράμετρος.
- **ομοιότητα με προηγούμενο στάδιο:** Το συγκεκριμένο κριτήριο, αποτελεί κομβική προσθήκη, το οποίο εξ όσων γνωρίζω δεν συμπεριλαμβάνεται σε προηγούμενες

μελέτες. Στόχος του είναι να πριμοδοτεί ταμπλατούρες οι οποίες έχουν περισσότερα κοινά σημεία (δηλαδή ζεύγη τάστα-χορδών για το ίδιο στιγμιότυπο νότας), με την ταμπλατούρα από το προηγούμενο στάδιο αναρμονικής ανάλυσης.

---

ΑΛΓΟΡΙΘΜΟΣ 4.1: *Depress Factor*

---

**Είσοδος:** *Tablature*  
**Έξοδος:** *DepressWeight*  
**for** *i* **in** *Tablature* **do**  
    **if**  $fret(i) \neq fret(i + 1)$  **then**  
         $DepressWeight+ = 1$   
    **return**  $DepressWeight/Size(Tablature)$

---



---

ΑΛΓΟΡΙΘΜΟΣ 4.2: *Fret Factor*

---

**Είσοδος:** *Tablature*  
**Έξοδος:** *FretWeight*  
**for** *i* **in** *Tablature* **do**  
     $FretWeight+ = |fret(i + 1) - fret(i)|$   
    **return**  $FretWeight/Size(Tablature)$

---



---

ΑΛΓΟΡΙΘΜΟΣ 4.3: *String Factor*

---

**Είσοδος:** *Tablature*  
**Έξοδος:** *StringWeight*  
**for** *i* **in** *Tablature* **do**  
     $StringWeight+ = |string(i + 1) - string(i)|$   
    **return**  $StringWeight/Size(Tablature)$

---

### Παρουσίαση Γενετικού Αλγορίθμου

Ο γενετικός αλγόριθμος που σχεδιάστηκε αρχικοποιείται με 40000 τυχαίες πιθανές ταμπλατούρες, δεδομένων των τονικών υψών και της ταμπλατούρας που προέκυψε από το στάδιο της αναρμονικής ανάλυσης. Αναλυτικότερα, από την ταμπλατούρα που προβλέφθηκε από το προηγούμενο στάδιο, αντιστοιχούνται σε τυχαίες πιθανές θέσεις 1) οι ασαφείς νότες, 2) κάθε άλλη με πιθανότητα 0.3. Ο γενετικός αλγόριθμος εκτελείται έως ότου περάσουν 300 γενιές ή οι 500 καλύτερες ταμπλατούρες είναι όμοιες. Αποτελείται από μια φάση διασταύρωσης με πιθανότητα 0.5 (δηλαδή το ποσοστό των γονέων που διασταυρώνονται και αντικαθίστανται από τα δύο παιδιά τους) όπου χρησιμοποιείται η Two-point Crossover συνάρτηση. Ακολούθως στα παιδιά εφαρμόζεται με πιθανότητα 0.2 η συνάρτηση μετάλλαξης (πιθανότητα να επιλεχθεί μια ταμπλατούρα για μετάλλαξη) η οποία με τη σειρά της αλλάζει με πιθανότητα 0.1 κάθε στιγμιότυπο νότας με μια τυχαία θέση στη ταστιέρα (πιθανότητα να επιλεχθεί ένα στιγμιότυπο νότας για αλλαγή κατά τη φάση της μετάλλαξης). Τέλος, για την φάση της επιλογής,

---

 ΑΛΓΟΡΙΘΜΟΣ 4.4: *Open String Factor*


---

**Είσοδος:** *Tablature*  
**Έξοδος:** *OpenStringWeight*  
**for**  $i$  **in** *Tablature* **do**  
   **if**  $fret(i) > 0$  **then**  
      $OpenStringWeight += 1$   
**return**  $OpenStringWeight / Size(Tablature)$

---



---

 ΑΛΓΟΡΙΘΜΟΣ 4.5: *Average Factor*


---

**Είσοδος:** *Tablature*  
**Έξοδος:** *AveragePositionWeight*  
**for**  $i$  **in** *Tablature* **do**  
   compute average fret-string of up to 6 surrounding notes,  
   such that between adjacent notes, time difference is less than 1 sec  
    $AveragePositionWeight += \sqrt{(avg_{string} - string(i))^2 + (avg_{fret} - fret(i))^2}$   
**return**  $AveragePositionWeight / Size(Tablature)$

---

επιλέγονται τα άτομα με τις καλύτερες τιμές της συνάρτησης καταλληλότητας όπως παρουσιάστηκε παραπάνω, 3000 από την νέα γενιά, και 100 από την προηγούμενη, δομώντας πλέον την επόμενη γενιά.

## Κεφάλαιο 5

### Αξιολόγηση

---

Στο παρόν κεφάλαιο παρουσιάζονται συνοπτικά τα δεδομένα πάνω στα οποία διεξάχθηκε η σχετική πειραματική διαδικασία και ακολούθως τα πειράματα που εφαρμόστηκαν για την αξιολόγηση των προτεινόμενων μεθόδων. Πρώτα, παρουσιάζεται η αξιολόγηση του ταξινομητή με βάση την αναρμονικότητα σε μεμονωμένα στιγμιότυπα νοτών από το σύνολο δεδομένων (Dataset) των Hjerrild και Christensen [6]. Ακολουθεί ένα συγκριτικό σενάριο μεταξύ της προτεινόμενης μεθόδου (χωρίς την εφαρμογή του γενετικού αλγόριθμου) και ενός Naive Bayes ταξινομητή πάνω στα δεδομένα του GuitarSet dataset [28]. Έπειτα διεξάγεται ένα πείραμα για ολόκληρα κομμάτια σε κιθάρα, όπου θεωρούνται γνωστά η έναρξη νότας (onset) και το τονικό ύψος (pitch), δίνοντας έμφαση τόσο στην ενδιάμεση ακρίβεια που προσφέρει το σύστημα βασιζόμενο στην αναρμονικότητα, όσο και στη βελτίωση που παρέχει ο γενετικός αλγόριθμος. Το ποσοστό βελτίωσης του γενετικού αλγόριθμου, μελετάται ξέχωρα και παρουσιάζονται για μερικές ενδεικτικές παραμετροποιήσεις τα ποσοστά βελτίωσης που μπορεί να πετύχει. Τέλος, παρουσιάζεται η ακρίβεια που μπορεί να παρέχει το σύστημα από άκρη σε άκρη συμπεριλαμβάνοντας τα δυο νευρωνικά δίκτυα για εκτίμηση έναρξης νότας (onset) και τονικού ύψους (pitch).

#### 5.1 Σύνολα Δεδομένων (Datasets)

Τα δεδομένα που αξιοποιήθηκαν για αξιολόγηση του συστήματός, προέκυψαν ως ένα υποσύνολο 52 κομματιών από τα 360 του συνόλου, από το GuitarSet dataset [28]. Εφόσον το σύστημα δουλεύει για μονοφωνική εκτέλεση (όχι συγχορδίες), επιλέχθηκαν τα ηχητικά αρχεία όπου όλες οι διαδοχικές νότες είχαν αναμεταξύ τους τουλάχιστον 60ms διαφορά στις ενάρξεις τους. Τα δεδομένα είναι πλούσια όσον αφορά τα εκφραστικά στυλ, τις μελωδίες, ρυθμούς, αλλά και τεχνικές παίξιματος, σχηματίζοντας ένα ολοκληρωμένο σύνολο για αξιολόγηση του συστήματος.

Το dataset αποτελείται από 4 διαφορετικές ηχογραφήσεις κάθε κομματιού: μια με μικρόφωνο, μία με εξαφωνικό μαγνήτη σε έξι κανάλια - ένα για κάθε χορδή, μία με εξαφωνικό μαγνήτη όπου τα 6 κανάλια είναι debleded ούτως ώστε να μειωθούν θόρυβος και παράσιτα, και ένα όπου τα 6 κανάλια είναι συγχωνευμένα σε ένα. Χρησιμοποιούνται μόνο οι ηχογραφήσεις της πρώτης και της τέταρτης περίπτωσης, δηλαδή με μικρόφωνο και με εξαφωνικό μαγνήτη με τα 6 κανάλια συγχωνευμένα σε ένα (microphone και pickup αντίστοιχα από εδώ και στο εξής). Ελέγχθηκε επίσης ο προτεινόμενος ταξινομητής χορδής με αναρμονική ανάλυση (χωρίς το

στάδιο του γενετικού αλγορίθμου) στο dataset που ανέπτυξαν και διέθεσαν οι Hjerrild και Christensen [6] το οποίο περιέχει ηχογραφήσεις απομονωμένων νοτών για όλες τις χορδές και για όλα τα τάστα έως το 12ο σε δυο κιθάρες, μια ηλεκτρική (Les Paul Firebrand) και μια ακουστική (Martin DR).

## 5.2 Προσαρμογή στην Κιθάρα

Για τον αρχικό υπολογισμό των  $\beta$  κάθε χορδής της κιθάρας και την προσαρμογή σε αυτή, επιλέχθηκαν σχετικά καθαρές ηχογραφήσεις για κάθε μια από τις 6 χορδές, από την ανοιχτή χορδή, το τρίτο και το δωδέκατο τάστο. Άλλωστε, σε ένα ρεαλιστικό σενάριο προσαρμογής μιας κιθάρας, θα απαιτούνταν από τον κιθαρίστα καθαρές ηχογραφήσεις μεμονωμένων νοτών. Ακολούθως υπολογίζεται η διάμεσος τιμή των  $\beta$  για κάθε θέση και χρησιμοποιείται από τον ταξινομητή αναρμονικής ανάλυσης. Σε σχήματα προσαρμογής όπου περιλαμβάνονται στιγμιότυπα νοτών από τάστα πέραν της ανοιχτής χορδής, οι διορθωτικοί συντελεστές  $a$  και  $b$  υπολογίζονται για κάθε χορδή όπως περιγράφεται στο κεφάλαιο 4.

## 5.3 Αποτελέσματα και Αξιολόγηση

### 5.3.1 Πείραμα σε μεμονωμένα ηχογραφημένα στιγμιότυπα νοτών

Πρώτα, παρουσιάζονται τα αποτελέσματα της προτεινόμενης μεθόδου στο dataset από τους Hjerrild και Christensen [6]. Το dataset αποτελείται από μεμονωμένες νότες, έως το 12ο τάστο. Για την ακουστική κιθάρα, κάθε ζεύγος χορδής-τάστου έχει ηχογραφηθεί 10 φορές, ενώ για την ηλεκτρική 9. Υπολογίσαμε τη διάμεση αναρμονικότητα όλων των διαθέσιμων στιγμιότυπων από την ανοιχτή χορδή το 3ο και το 12ο τάστο σύμφωνα με την προτεινόμενη μέθοδο προσαρμογής. Η ακρίβεια της ταξινόμησης των τάστων ελέγχθηκε στα τάστα που δεν χρησιμοποιήθηκαν για την προσαρμογή στην αντίστοιχη περίπτωση.

Για την ακουστική κιθάρα (Martin), η μέθοδος 3Fret παράγει τα καλύτερα αποτελέσματα ενώ για την ηλεκτρική (Firebrand), η 2FretA (βλ. Πίνακα 5.1. Σημαντική παρατήρηση είναι ότι η 1Fret μέθοδος, δίνει χειρότερα αποτελέσματα από όλες τις υπόλοιπες που αξιοποιούν πλέον του ενός τάστου για κάθε χορδή. Συγκρίνοντας τα αποτελέσματα με την inharmonic pitch estimation μέθοδο και τον προτεινόμενο MAP-optimal classifier των Hjerrild και Christensen [6] παρατηρείται ότι αν και καταφέρνουν να πάρουν καλύτερα αποτελέσματα για την ακουστική κιθάρα, η 2FretA μέθοδος παράγει σχεδόν ίδια ακρίβεια με τη δική τους για την Firebrand. Επιπροσθέτως διερευνώντας την πτώση σε ακρίβεια του μοντέλου για την ακουστική κιθάρα (Martin), παρατηρήθηκε ότι προκύπτει από ανακριβή υπολογισμό του συντελεστή αναρμονικότητας. Αυτό, υποδεικνύει ότι το σχήμα υπολογισμού αναρμονικότητας που αναπτύχθηκε, μπορεί να βελτιωθεί και να τροποποιηθεί ώστε να παράγει ακριβέστερα αποτελέσματα.

Όσον αφορά τη μέθοδο ταξινόμησης που προτείνεται στο [6], δεν υπάρχουν επαρκείς ενδείξεις ότι μπορεί να εξαχθεί η κατανομή των  $\beta$  μόνο από μερικές απομονωμένες ηχογραφημένες νότες και να γενικευτεί για τα  $\beta$  όλων των στιγμιότυπων νοτών προς ταξινόμηση ενός πραγματικού κομματιού σε κιθάρα. Αυτό συμβαίνει επειδή τα  $\beta$  των δειγμάτων που συλλέχθηκαν κατά το στάδιο προσαρμογής του ταξινομητή, έχουν πολύ μικρή διασπορά (καθώς αποτελούν



Μέθοδος Προσαρμογής	Martin	Firebrand
1Fret	81.2%	92.3%
2FretA	85%	<b>97%</b>
2FretB	88.9%	93.4%
3Fret	<b>89.7%</b>	96.1%
MAP-optimal [6]	100%	97.1%

Πίνακας 5.1: Ακρίβεια του συστήματος ηχητικής ταξινόμησης με αναρμονική ανάλυση στα δεδομένα από το dataset των Hjerrild και Christensen [6].

ένα μικρό σύνολο απομονωμένων ηχογραφημένων νοτών) συγκριτικά με τα  $\beta$  όταν μετρούνται σε κανονικά κομμάτια. Αυτός είναι και ο λόγος που στη παρούσα διπλωματική προτείνεται και αξιοποιείται μια πιο απλή μέθοδος ταξινόμησης, η οποία δεν προϋποθέτει γνώση σχετικά με την κατανομή των  $\beta$ .

### 5.3.2 Αναγνώριση Χορδής σε Ολόκληρα κομμάτια κιθάρας

Για την αξιολόγηση της προτεινόμενης μεθόδου σε ρεαλιστικές συνθήκες, δηλαδή πραγματικά μουσικά κομμάτια, ελέγχθηκαν οι τέσσερις μέθοδοι που προτείνονται παραπάνω για ταξινόμηση χορδής με αναρμονική ανάλυση, ενώ χρησιμοποιήθηκε ένας γενετικός αλγόριθμος ώστε να συμπεριληφθούν οι περιορισμοί εκτελεσιμότητας (playability constraints). Τα δεδομένα στα οποία έγινε η αξιολόγηση, αποτελούνται από ένα υποσύνολο 52 κομματιών του GuitarSet, που περιλαμβάνει μόνο μονοφωνικά κομμάτια όπως ορίστηκαν και παραπάνω (βλ. διαδοχικές νότες με διαφορά onset μεγαλύτερη των 60ms). Ελέγχθηκαν τόσο οι ηχογραφήσεις από μικρόφωνο όσο και από μαγνήτη. Τα αποτελέσματα υποδεικνύουν ότι η απλή μέθοδος που προτείνεται, μαζί με τον γενετικό αλγόριθμο αποτελούν ισχυρά εργαλεία προς την επίλυση του υπό εξέταση προβλήματος.

Μέθοδος Προσαρμογής	Ακρίβεια Ταξινομητή Χορδής	Ακρίβεια Χορδής Γενετικού Αλγορίθμου
Μαγνήτης		
1Fret	<b>71.1%</b>	<b>81.6%</b>
2FretA	67.2%	78.9%
2FretB	68.1%	78.7%
3Fret	67.4%	78.5%
Μικρόφωνο		
1Fret	<b>66.6%</b>	<b>78.4%</b>
2FretA	64%	77.2%
2FretB	65.1%	78.2%
3Fret	64.6%	77.1%

Πίνακας 5.2: Ακρίβεια της μεθόδου βασισμένη σε αναρμονική ανάλυση, για τις μονοφωνικές εκτελέσεις από το GuitarSet dataset.

Όπως φαίνεται στον Πίνακα 5.2, οι 4 μέθοδοι προσαρμογής, αντιστοιχούν σε ακρίβεια πρόβλεψης από 64% έως 71.1% όταν τα onset και τα τονικά ύψη θεωρούνται γνωστά. Στην

περίπτωση του μαγνήτη παρατηρούνται ελαφρώς καλύτερα αποτελέσματα συγκριτικά με το μικρόφωνο (3-5%). Εν μέρει, αυτό προκύπτει από το γεγονός ότι στο μικρόφωνο αντιστοιχούν περισσότερα ασαφή στιγμιότυπα νοτών, 11% εν αντιθέσει με το μαγνήτη 8%. Επιπροσθέτως, θόρυβοι (π.χ. σύρσιμο δαχτύλων στις χορδές) είναι πιο συχνοί σε περιπτώσεις μικροφώνου από ό,τι μαγνήτη, γεγονός που έχει ευθεία σύνδεση με τις ασαφείς νότες.

Σε αυτό το πείραμα, η απλούστερη μέθοδος προσαρμογής 1Fret έδειξε τα καλύτερα αποτελέσματα τόσο για το μαγνήτη όσο και για το μικρόφωνο. Οι 2Fret και 3Fret παρουσιάζουν χειρότερα αποτελέσματα επειδή τα στιγμιότυπα νοτών που χρησιμοποιήθηκαν για την προσαρμογή ήταν αποκομμένα από ολόκληρα κομμάτια κιθάρας και όχι απομονωμένες νότες όπως σε ένα πραγματικό σενάριο προσαρμογής. Προς αυτό συνηγορούν και τα αποτελέσματα του Πίνακα 5.1, όπου παρουσιάζεται ένα ιδανικό σενάριο προσαρμογής σε απομονωμένες νότες. Τελικά, τα λίγα δείγματα που αντλήθηκαν από το 3ο και 12ο τάστο δεν είναι αρκετά αντιπροσωπευτικά ώστε να δομήσουν μια στέρεη βάση για να παραχθούν ακριβείς εκτιμήσεις των  $\beta$  και να επιτευχθεί μια καλή προσέγγιση της αναρμονικής συμπεριφοράς ολόκληρης της ταστιέρας.

Ο γενετικός αλγόριθμος βελτιώνει κάθε φορά τα αποτελέσματα κατά τουλάχιστον 10%, με την μέθοδο 1Fret να κατέχει τα μεγαλύτερα ποσοστά επιτυχίας. Είναι ξεκάθαρο, ότι μεγαλύτερο ποσοστό επιτυχίας του ταξινομητή αναρμονικής ανάλυσης, οδηγεί σε μεγαλύτερη ακρίβεια συνολικά μετά την έξοδο του γενετικού αλγόριθμου.

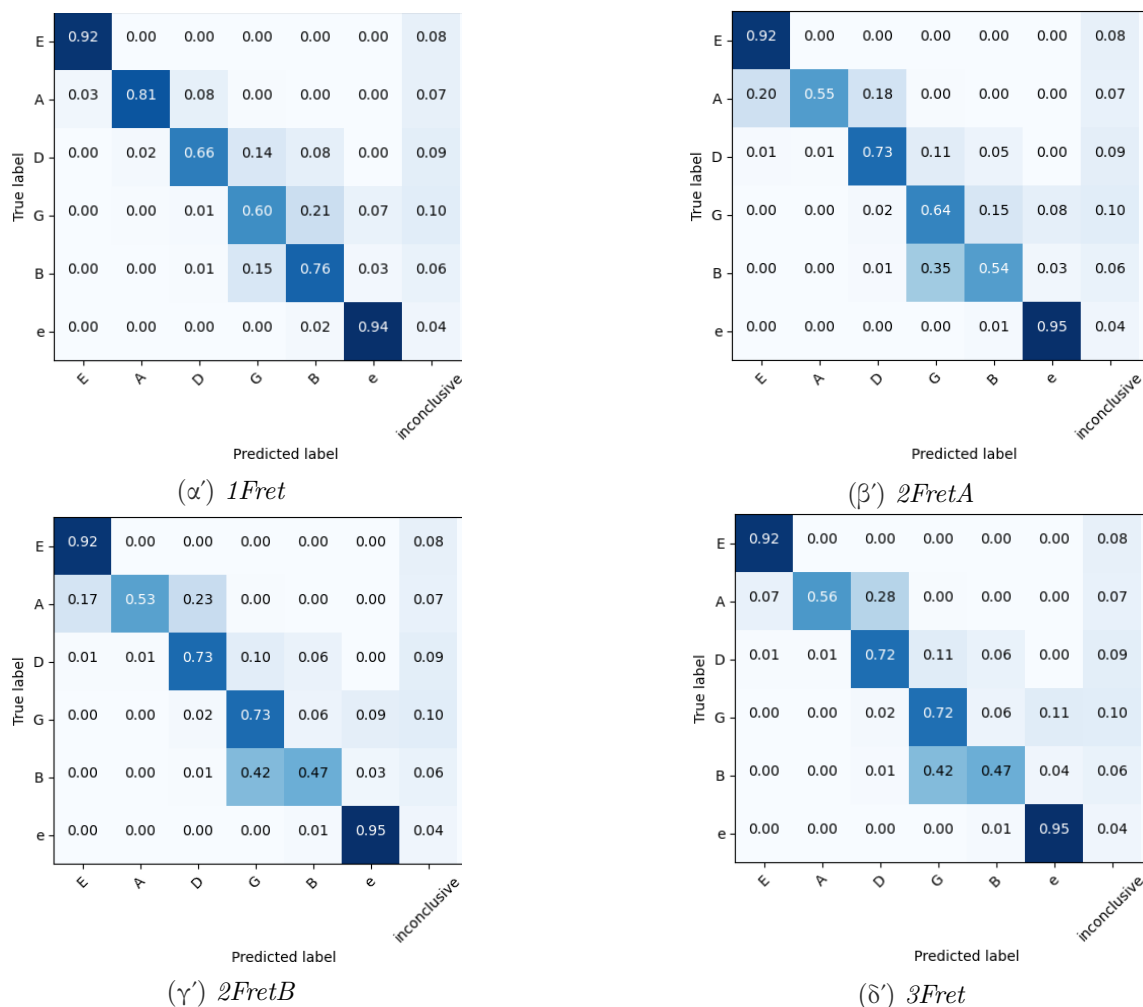
## Πίνακες Σύγχυσης

Κοιτώντας λίγο πιο προσεκτικά τα αποτελέσματα, μπορούν να εξαχθούν μερικά ακόμη συμπεράσματα για την ακρίβεια του συστήματος, τις δυνατότητες και τις αδυναμίες του. Πιο συγκεκριμένα, τόσο στην περίπτωση ηχογραφήσεων από μικρόφωνο (Βλ. Σχήμα 5.2) όσο και από μαγνήτη (Βλ. Σχήμα 5.1), τα αποτελέσματα φαίνονται να έχουν παρόμοια μορφή. Δηλαδή, η σύγχυση μεταξύ χορδών, φαίνεται να υπάρχει με τρόπο όμοιο και στις δύο περιπτώσεις.

Η μεγαλύτερη σύγχυση μεταξύ χορδών υπάρχει για τις χορδές Σι (B) και Σολ (G) οι οποίες μπερδεύονται μεταξύ τους, ενώ αναλόγως τη μέθοδο προσαρμογής, παρατηρείται ότι πριμοδοτείται η μία έναντι της άλλης χορδής. Το ίδιο ισχύει και σε μικρότερο βαθμό για τη σχέση μεταξύ της χορδής Λα (A) και Ρε (D). Σημαντική παρατήρηση αποτελεί επίσης το γεγονός ότι η λανθασμένη αναγνώριση χορδής παρατηρείται σε σημαντικό βαθμό σε ζεύγη. Δηλαδή, χαμηλό ποσοστό σε μια χορδή προκύπτει από την σύγχυση μεταξύ δύο χορδών, και όχι από την λανθασμένη μοντελοποίηση μιας χορδής όπου τελικά τα στιγμιότυπα θα κατανέμονταν ομοιόμορφα σε όλες τις πιθανές θέσεις. Αυτό συμβαίνει καθώς, σε μια κιθάρα με κανονικό κούρδισμα, για τις χορδές Μι (E), Λα (A), Ρε (D), Σι (B), στο πέμπτο τους τάστο και για τη χορδή Σολ (G) στο τέταρτο, βρίσκεται το ίδιο τονικό ύψος με την ανοιχτή της από κάτω χορδής. Συνεπώς, μεταξύ των πρώτων ισχύει ότι  $Pitch(string, fret) = Pitch(string + 1, fret + 5)$ . Αντιστοίχως για τη Σολ χορδή  $Pitch(string, fret) = Pitch(string + 1, fret + 4)$ . Έτσι, στην απλή περίπτωση της μεθόδου 1Fret, για τις χορδές Μι (E), Λα (A), εάν ισχύει η σχέση

$$\beta(E, 0) = \frac{\beta(A, 0)}{2^{\frac{5}{8}}}, \quad (5.1)$$

τότε για όλα τα τονικά ύψη των δύο χορδών, ο συντελεστής αναρμονικότητας είναι σχεδόν

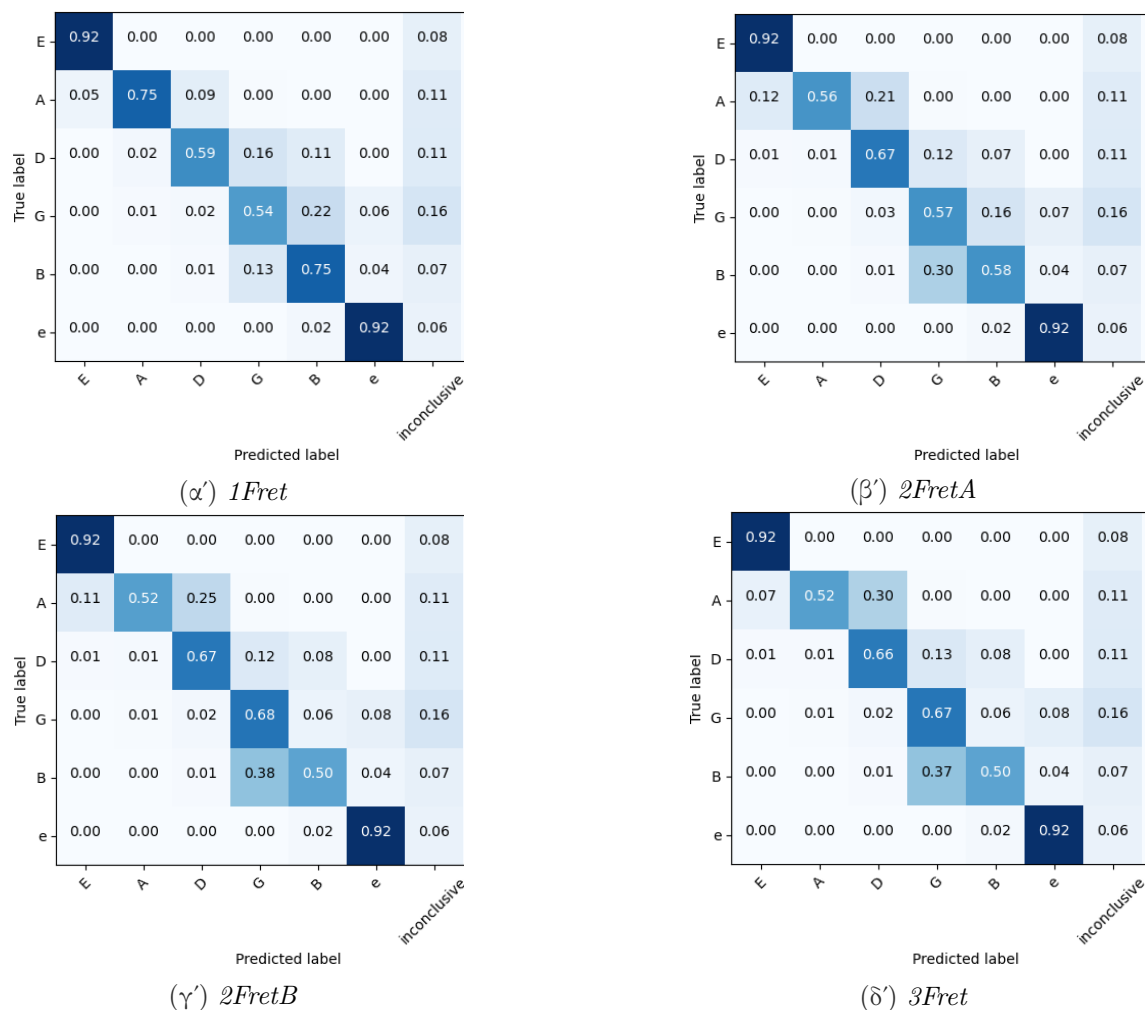


Σχήμα 5.1: Πίνακες σύγκρισης για τις ηχογραφήσεις με μαγνήτη (pickup)

ίδιος. Εδώ φανεράνεται και η παγίδα που μπορεί να υποπέσει ένα τέτοιο σύστημα, καθώς και η ανάγκη για όσο το δυνατόν ακριβέστερο υπολογισμό του συντελεστή αναρμονικότητας. Εν μέρει, αυτό εξηγεί γιατί οι παραπάνω χορδές μερδεύονται συχνότερα, ενώ οι χορδές Mi (E), μι (E) σχεδόν ποτέ. Όταν οι συντελεστές αναρμονικότητας είναι ευδιάκριτοι, γίνεται και σωστότερη η διάκριση μεταξύ των χορδών.

### Σύγκριση με Naive Bayes Ταξινομητή

Με στόχο να συγκριθεί το προτεινόμενο σχήμα με έναν ταξινομητή που δεν ακολουθεί την ίδια λογική προσαρμογής με λίγα δεδομένα, χρησιμοποιείται ένας naive Bayes ταξινομητής εκπαιδευμένος σε 40 στιγμιότυπα νοτών για κάθε ζεύγος χορδής-τάστου (χρησιμοποιείται πάλι ο συντελεστής αναρμονικότητας με ίδιο αλγόριθμο υπολογισμού). Είναι σημαντικό να τονιστεί ότι σε ολόκληρο το dataset, δεν υπάρχει επαρκής αριθμός στιγμιότυπων νοτών από τις MIDI νότες 51, 52 όπως και από σχεδόν όλα τα τάστα μεγαλύτερα του 12ου (αφού υπήρχαν λιγότερα από 40 στιγμιότυπα για κάθε κλάση) για την εκπαίδευση του συστήματος. Έτσι, σε αυτό το πείραμα, περιορίζονται οι έλεγχοι και τα αποτελέσματα για τις MIDI νότες 40 έως 76 χωρίς τις 51, 52 και χωρίς να λαμβάνονται υπόψιν στιγμιότυπα νοτών από τάστα μεγαλύτερα του 12ου.



Σχήμα 5.2: Πίνακες σύγκρισης για τις ηχογραφήσεις με μικρόφωνο (microphone)

Μέθοδος Προσαρμογής	Ακρίβεια σε Ηχογραφήσεις Μικροφώνου	Ακρίβεια σε Ηχογραφήσεις Μαγνήτη
Naive Bayes	66%	71%
1Fret	64.9%	70.7%
2FretA	65.7%	71.7%
2FretB	<b>67.3%</b>	<b>73.3%</b>
3Fret	66.9%	72.8%

Πίνακας 5.3: Πίνακας αποτελεσμάτων για τη σύγκριση της μεθόδου ηχητικής ταξινόμησης λήγων δειγμάτων με έναν Naive Bayes ταξινομητή σε απομονωμένα στιγμιότυπα νοτών κομμένα από τις ηχογραφήσεις του GuitarSet Dataset (Η ακρίβεια των αποτελεσμάτων, διαφέρει με τους άλλους πίνακες καθώς δεν εξετάζεται το σύνολο των τάσεων και τονικών υψών).

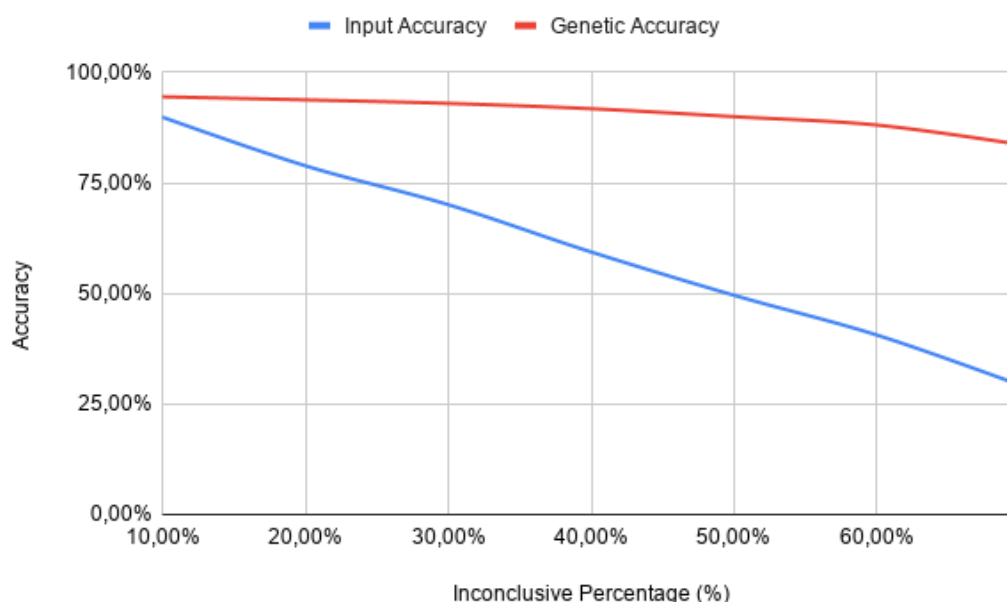
Τα αποτελέσματα του Πίνακα 5.3, δείχνουν ότι η προτεινόμενη μέθοδος παράγει πολύ κοντινή ακρίβεια με αυτή του Naive Bayes classifier ακόμα και αν στην πρώτη περίπτωση χρειάζονται μόνο από 5 έως 15 στιγμιότυπα από κάθε χορδή (30 έως 90 συνολικά) για να εκπαιδευτεί σε αντίθεση με τα 2880 που απαιτεί ο αντίστοιχος μασεσιανός ταξινομητής. Είναι άξιο αναφοράς, ότι το GuitarSet Dataset, δεν μπορεί να παρέχει επαρκή δεδομένα για

την πλήρη εκπαίδευση ενός τέτοιου σχήματος. Η μέθοδος που παρουσιάζεται σε αυτή την εργασία, δεν περιορίζεται από αυτό το γεγονός, καθώς απαιτεί πολύ λίγα στιγμιότυπα για να προσαρμοστεί το σύστημα μεταγραφής.

### 5.3.3 Αξιολόγηση Γενετικού Αλγορίθμου

Παρακάτω, παρουσιάζονται αποτελέσματα σχετικά με τη βελτίωση που μπορεί να παρέχει ο γενετικός αλγόριθμος υπό διαφορετικές συνθήκες. Ελέγχθηκε το κατά πόσο ασαφείς νότες, λανθασμένες χορδές, αλλά και λανθασμένα αναγνωρισμένα τονικά ύψη επηρεάζουν τον γενετικό αλγόριθμο, δίνοντας ιδιαίτερα ενθαρρυντικά αποτελέσματα.

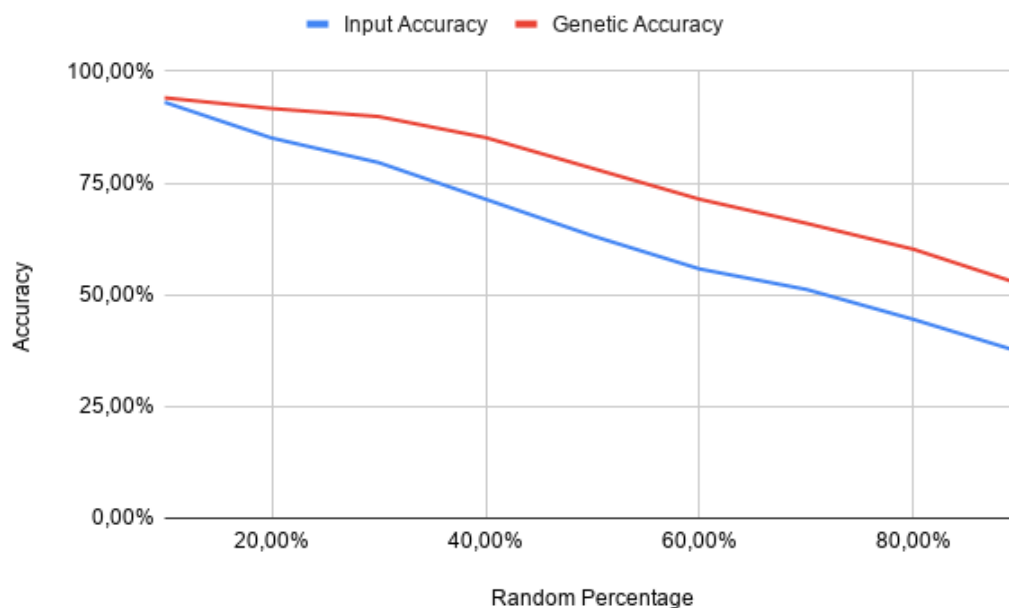
#### Συσχέτιση με Ασαφείς Νότες



Σχήμα 5.3: Γράφημα που παρουσιάζει τη βελτίωση που προσδίδει ο γενετικός αλγόριθμος όταν ένα ποσοστό των στιγμιότυπων νοτών έχουν αντιστοιχηθεί στις σωστές χορδές και όλες οι υπόλοιπες είναι χαρακτηρισμένες ασαφείς.

Πρώτον, η απόκριση του γενετικού αλγορίθμου σε σχέση με τις ασαφείς νότες μελετήθηκε. Προς αυτό το σκοπό, παρακάμφθηκε η εξέταση του ηχητικού αρχείου και ως έξοδος του προηγούμενου σταδίου τροφοδοτήθηκε στον γενετικό αλγόριθμο ταμπλατούρα όπου ένα ποσοστό των στιγμιότυπων των νοτών της είχε χαρακτηριστεί ασαφές, και όλες οι υπόλοιπες χορδές σωστές. Τα αποτελέσματα δείχνουν ότι γνωρίζοντας ένα αρκετά μικρό ποσοστό των σωστών χορδών (άλλα με μεγάλη ακρίβεια), καταστά τον γενετικό αλγόριθμο ικανό να παρουσιάσει συντριπτική βελτίωση των αποτελεσμάτων (βλ. Σχήμα 5.3). Δηλαδή, μεγάλη μείωση του αριθμού των γνωστών χορδών-τάσεων, επιφέρει πολύ μικρή μείωση στην ακρίβεια της ταξινόμησης χορδής μετά το στάδιο του γενετικού. Έτσι, ανοίγεται ο δρόμος για περαιτέρω μελέτη σχετικών αλγορίθμων και προβλημάτων για την αυστηρή μεταγραφή ταμπλατούρας, καθώς επίσης δείχνει ότι σημαντικότερο είναι το υψηλό precision και όχι το recall για το

πρώτο στάδιο ηχητικής ταξινόμησης.



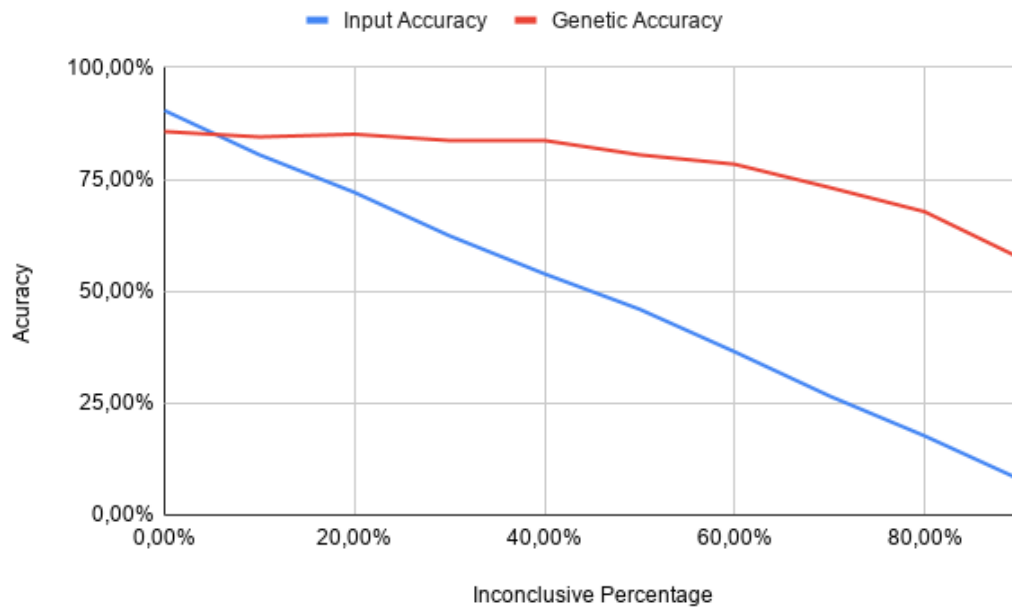
Σχήμα 5.4: Γράφημα που παρουσιάζει τη βελτίωση που προσδίδει ο γενετικός αλγόριθμος όταν ένα ποσοστό των στιγμιότυπων νοτών έχουν αντιστοιχηθεί στις σωστές χορδές και όλες οι υπόλοιπες έχουν αντιστοιχηθεί τυχαία.

### Συσχέτιση με Λανθασμένες Χορδές

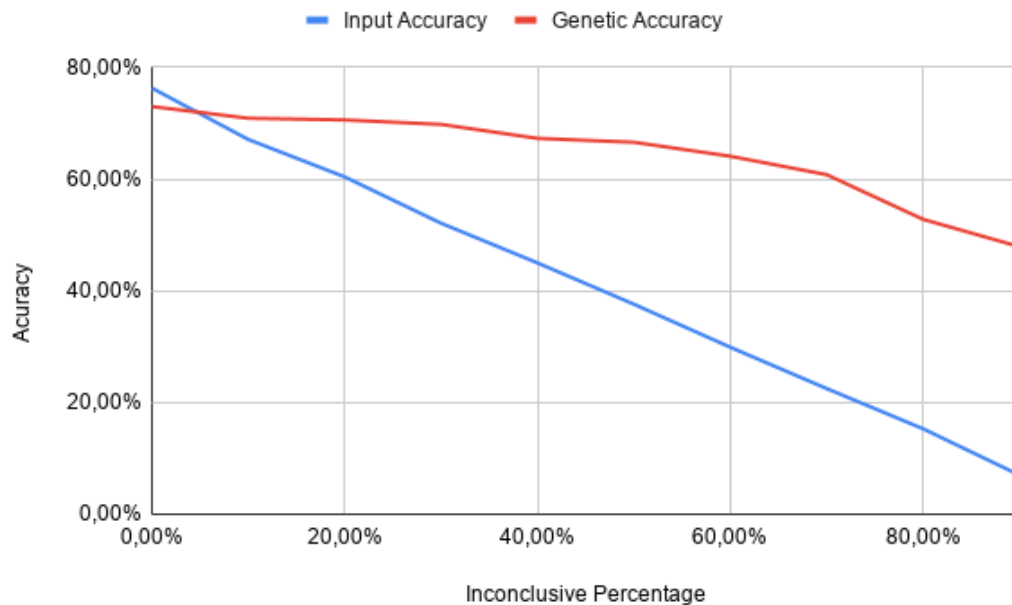
Ακολούθως, μελετήσαμε την απόδοση του γενετικού αλγορίθμου σε περιπτώσεις όπου ένα ποσοστό των στιγμιότυπων νοτών έχει αντιστοιχηθεί τυχαία σε τονικά ύψη. Αυτό βέβαια, δεν σημαίνει ότι απαραίτητως θα αντιστοιχηθούν και σε λάθος χορδή, αν και μένει πλέον στην τύχη. Στο Σχήμα 5.4 παρουσιάζονται τα αποτελέσματα για διαφορετικά ποσοστά τυχαίων νοτών. Συγκρίνοντας με το Σχήμα 5.3, μπορούμε εύκολα να καταλήξουμε ότι η σημείωση κάποιων στιγμιότυπων νοτών ως ασαφή, των οποίων η χορδή δεν μπορεί να προβλεφθεί με σιγουριά, είναι ευνοϊκό για τον γενετικό αλγόριθμο, ακόμη και σε περιπτώσεις όπου η αρχική είσοδος έχει μικρότερη ακρίβεια. Π.χ. στην περίπτωση όπου όλες είναι σωστές ή ασαφείς, για ακρίβεια 29,3% καταλήγουμε σε 83,7% μετά το γενετικό αλγόριθμο, ενώ στη δεύτερη περίπτωση, όπου όλες είναι ή σωστές ή αντιστοιχισμένες τυχαία, με 63,2% καταλήγουμε σε 78,3% ακρίβεια.

### Συσχέτιση με Λανθασμένες Νότες

Ακόμη μελετήθηκε η επιρροή λανθασμένων νοτών στο αποτέλεσμα του γενετικού αλγορίθμου. Καθώς ένας εξαντλητικός έλεγχος του προβλήματος είναι δύσκολος και απαιτεί πολύ χρόνο, επιλέχθηκε να προσομοιαστούν οι περιπτώσεις αναλόγως την ακρίβεια, αλλά και των λαθών στην έξοδο του συνελικτικού δικτύου εκτίμησης τονικού ύψους ΉΡΕΠΕ. το συνελικτικό δίκτυο, παρουσίασε ακρίβεια 93% και όπως ήταν αναμενόμενο, κάνει λάθη συνήθως 1 ή 2 ημιτόνια, ή μια ολόκληρη οκτάβα. Οι λανθασμένες νότες κατά προσέγγιση αποδίδονταν



Σχήμα 5.5: Γράφημα που παρουσιάζει την απόδοση του γενετικού αλγόριθμου στην περίπτωση που το 10% των στιγμιότυπων έχει αποδοθεί σε λάθος τόνο και αναλόγως το ποσοστό ασαφών.



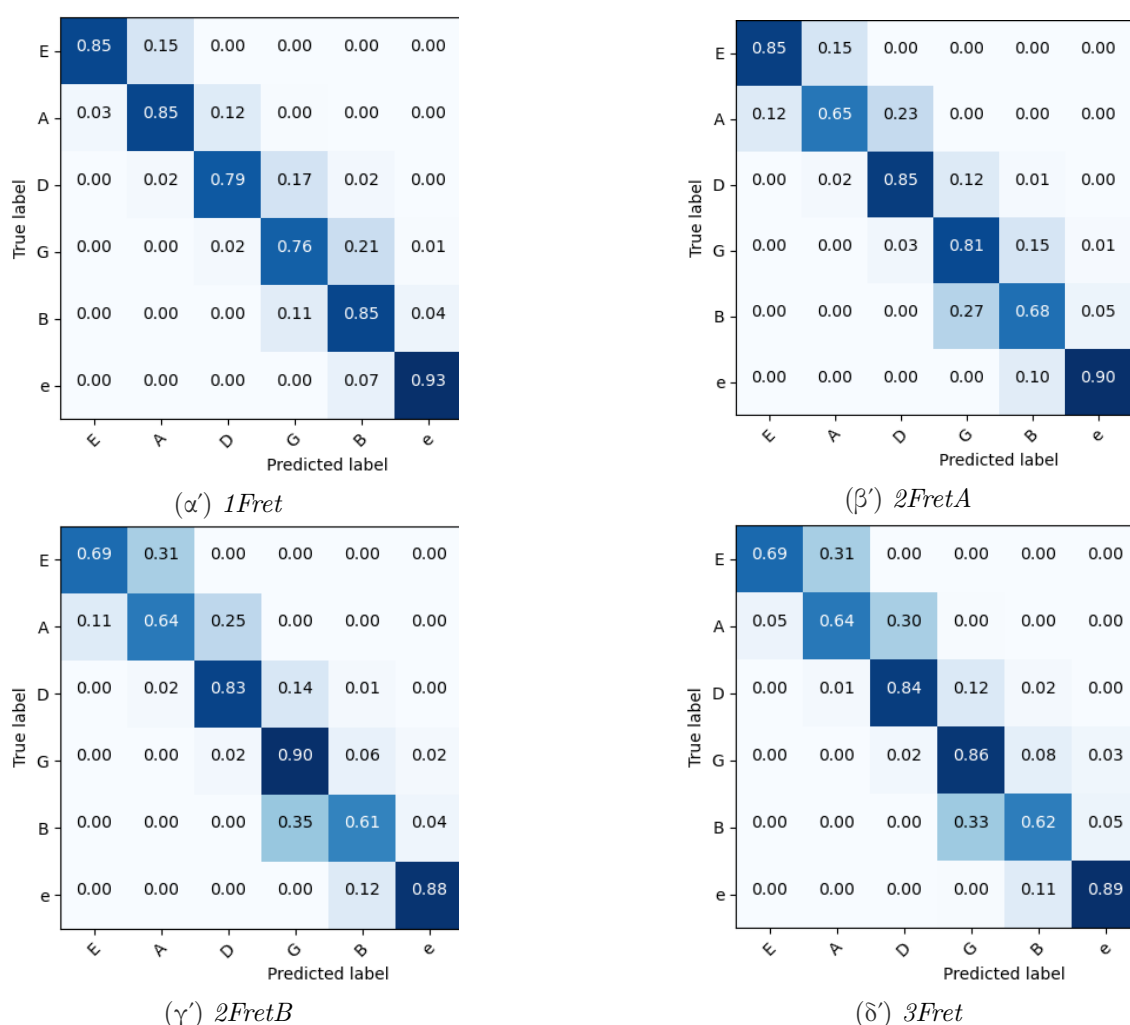
Σχήμα 5.6: Γράφημα που παρουσιάζει την απόδοση του γενετικού αλγόριθμου στην περίπτωση που το 25% των στιγμιότυπων έχει αποδοθεί σε λάθος τόνο και αναλόγως το ποσοστό ασαφών.

κατά  $1/8$  σε κάποια οκτάβα,  $2/8$  σε νότα αυξημένη κατά ένα ημιτόνιο,  $3/8$  μειωμένη κατά 1 ημιτόνιο,  $1/8$  αυξημένη κατά 2 ημιτόνια, και  $1/8$  μειωμένη κατά 2 ημιτόνια. Επίσης το 65% των στιγμιότυπων που τους αποδίδεται λανθασμένη τονικότητα, καταλήγουν και σε ασαφή υπολογισμό του συντελεστή αναρμονικότητας, έτσι κάθε λανθασμένη τονικότητα χαρακτηρίστηκε ως ασαφής. Είναι σημαντικό να τονιστεί, ότι τα δεδομένα πιθανότατα να μην είναι



αρκετά για να βγάλουμε ασφαλή συμπεράσματα για αυτές τις κατανομές. Παρόλα αυτά, με σκοπό να ελεγχθεί κατά πόσο λανθασμένα αναγνωρισμένοι τόνοι επηρεάζουν τον γενετικό αλγόριθμο, θεωρήθηκε ότι επαρκούν για μια πρώτη εκτίμηση. Όπως φαίνεται στα σχήματα 5.5, 5.6 ο γενετικός αλγόριθμος ανταποκρίνεται ακόμα και σε περιπτώσεις όπου ένα ποσοστό των τονικοτήτων έχει χαρακτηριστεί λανθασμένα. Αν και δεν ανταποκρίνεται πλήρως στην πραγματικότητα, καθώς δεν είναι απαραίτητο ο λανθασμένος τόνος να αντιστοιχεί σε ασαφή νότα, μπορούμε να αντλήσουμε με ασφάλεια το συμπέρασμα ότι μπορεί να λειτουργήσει επαρκώς καλά και σε περιπτώσεις με μεγαλύτερο ποσοστό λανθασμένων τόνων.

### Πίνακας Σύγκρισης



Σχήμα 5.7: Πίνακες σύγκρισης του γενετικού αλγόριθμου για τις ηχογραφήσεις με μαγνήτη (pickup)

Τέλος, όπως φαίνεται στα σχήματα την εφαρμογή του γενετικού αλγόριθμου 5.1 και μετά 5.7, ο γενετικός αλγόριθμος τείνει να διατηρεί κατά μια έννοια τις συγχύσεις μεταξύ των χορδών που παρατηρήθηκαν και από την έξοδο του σταδίου της αναμονικής ανάλυσης. Μπορούμε να συμπεράνουμε λοιπόν, ότι «σέβεται» την έξοδο του προηγούμενου σταδίου και ότι δεν καταφέρνει απλώς να βρει μια αρκετά καλή ταμπλατούρα η οποία τυχαίνει να είναι αυτή



που παίχτηκε.

### 5.3.4 Αποτελέσματα Αξιολόγησης Αυστηρής Μεταγραφής Ταμπλατούρας

Οι μέθοδοι προσαρμογής που απέδωσαν τα καλύτερα αποτελέσματα στην αναγνώριση ενεργής χορδής στο GuitarSet dataset ελέγχθηκαν αρχικά στην περίπτωση που το τονικό ύψος εξάγεται χρησιμοποιώντας το CREPE [5] νευρωνικό δίκτυο, με τις ενάρξεις νοτών να θεωρούνται γνωστές. Ακολούθως, ελέγχθηκε η ακρίβεια των μεθόδων όταν το τονικό ύψος εξάγεται με το CREPE και οι ενάρξεις με το TCN [9]. Εκτός από την ακρίβεια αναγνώρισης ενεργής χορδής, εφαρμόζουμε μια μετρική Tablature Disambiguation Rate (TDR) προτεινόμενη από τους Wiggins και Kim [27] για αξιολόγηση της ακρίβειας μεταξύ των σταδίων και τη σύνδεση της εκτίμησης τονικότητας και του υπολογισμού onset με την αναγνώριση ενεργής χορδής. Το TDR μπορεί να υπολογιστεί ως:

$$TDR_{OS} = \text{correct}_{strings} / \text{correct}_{onsets} \quad (5.2)$$

$$TDR_{PS} = \text{correct}_{strings} / \text{correct}_{pitches} \quad (5.3)$$

$$TDR_{OP} = \text{correct}_{pitches} / \text{correct}_{onsets}. \quad (5.4)$$

Μέθοδος Προσαρμογής	Ακρίβεια Ταξινομητή Χορδής	Ακρίβεια Χορδής Γενετικού Αλγορίθμου	TDR <sub>O-S</sub>	TDR <sub>O-P</sub>	TDR <sub>P-S</sub>
Μικρόφωνο					
1Fret (TCN)	62.4%	70.8%	75.2%	93.4%	80.5%
1Fret (Γνωστές ενάρξεις)	64.5%	72.2%	-	-	78.8%
Μαγνήτης					
1Fret (TCN)	65.8%	72.1%	77.7%	94.4%	82.3%
1Fret (Γνωστές ενάρξεις)	69.6%	76.5%	-	-	82.1%

Πίνακας 5.4: Αποτελέσματα για την ολοκληρωμένη αυστηρή μεταγραφή ταμπλατούρας για τις μεθόδους προσαρμογής με την μεγαλύτερη ακρίβεια.

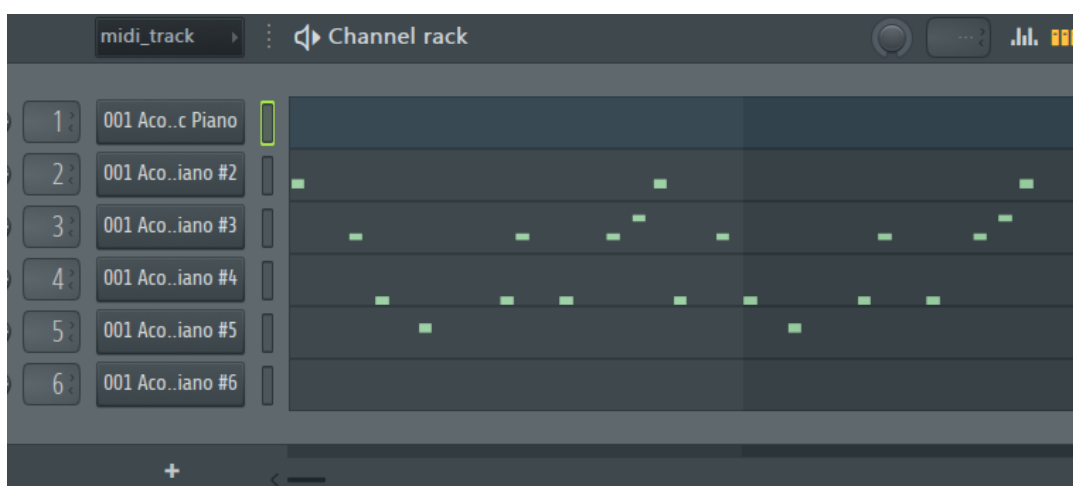
Όπως παρουσιάζεται στον Πίνακα 5.4, η μέθοδος φτάνει ποσοστό 72.2% και 76.5% ακρίβειας για την περίπτωση της ηχογράφησης με μικρόφωνο και μαγνήτη αντίστοιχα για πλήρη μεταγραφή σε ταμπλατούρα. Αν και ο γενετικός αλγόριθμος δεν παρουσιάζει τα ίδια ποσοστά βελτίωσης όπως όταν το τονικό ύψος και οι ενάρξεις θεωρούνται γνωστές, συνεχίζει να βελτιώνει τα αποτελέσματα κατά 7-8%, παγιώνοντας την σπουδαιότητά ενός υποσυστήματος που εγχολπώνει τις φυσικές δυσκολίες- και ευκολίες- του παιχνιδιού της κιθάρας, ενώ αναδεικνύει την πρακτική αξία του σε ένα από άκρη σε άκρη σύστημα μεταγραφής ηχητικού σήματος κιθάρας σε ταμπλατούρα.

Μελετήθηκε επίσης και ο ανοιχτά δημοσιευμένος κώδικας των Wiggins και Kim για αυστηρή μεταγραφή ταμπλατούρας [27] και αξιολογήθηκε για το υποσύνολο των μονοφωνικών κομματιών (βλ. Κεφάλαιο 5.1). Μετρήθηκε το F-measure για κάθε μονοφωνική ηχογράφηση

από μικρόφωνο και υπολογίστηκε η μέση τιμή τους στα 0.766. Παρ' όλα αυτά, δεν μπορεί να γίνει έγκυρη και ουσιαστική σύγκριση με το μοντέλο τους, καθώς οι μετρήσεις τους δεν αντιπροσωπεύουν την ακρίβεια σε στιγμιότυπα νοτών που παίχτηκαν, αλλά την ακρίβεια σωστής αναγνώρισης σε παράθυρα των 10ms. Αυτή η μέθοδος αξιολόγησης, δεν αντιπροσωπεύει την ακρίβεια της αυστηρής μεταγραφής ταμπλατούρας. Διαδοχικές νότες που παίζονται στο ίδιο τάστο και χορδή αντιμετωπίζονται με τον ακριβώς ίδιο τρόπο, όπως μια νότα που παίζεται στο ίδιο τάστο και χορδή για την ίδια διάρκεια. Ακόμη, η ακρίβεια που μετριέται, πριμοδοτείται από μεγάλης διάρκειας στιγμιότυπα νοτών, καθώς είναι πιο εύκολες να αναγνωριστούν και συνεισφέρουν περισσότερο στην μετρούμενη ακρίβεια. Τέλος, μέθοδοι που βασίζονται σε πολλά δείγματα για την προσαρμογή σε μόνο ένα όργανο, υστερούν σημαντικά, καθώς είναι αναγκαία πολλά δεδομένα για την προσαρμογή τους σε κάθε διαφορετικό όργανο.

## 5.4 Μετατροπή σε MIDI και μουσική σύνθεση

Για λόγους επίδειξης μιας πιθανής χρησιμότητας του προτεινόμενου συστήματος, σχεδιάστηκε ένα υποσύστημα, το οποίο αξιοποιεί την αναγνώριση ενεργής χορδής και καταγράφει σε διαφορετικό κανάλι midi - αναλόγως τη χορδή που αναγνωρίστηκε - τη μελωδία που παίχτηκε στη κιθάρα. Στόχος, είναι να χρησιμοποιηθεί για μουσική σύνθεση. Έχοντας ξεχωρίσει τις νότες αναλόγως τις χορδές, μπορούμε να εισάγουμε διαφορετικά εφέ σε κάθε κανάλι παράγοντας έτσι ένα τελείως καινούριο ήχο. Σε εξέλιξη της παρούσας εργασίας, εφόσον μετατραπεί το παρόν σύστημα σε πραγματικού χρόνου (βασικός περιορισμός είναι ο γενετικός αλγόριθμος που εφαρμόζεται στο σύνολο της ταμπλατούρας άρα προϋποθέτει πρότερη γνώση όλης της δομής του κομματιού), μπορεί να χρησιμοποιηθεί μια τέτοια λειτουργία για ζωντανές παραστάσεις.



Σχήμα 5.8: Εικόνα από το FL studio όπου εμφανίζεται κάθε χορδή κωδικοποιημένη σε διαφορετικό κανάλι MIDI

Η μετατροπή σε midi έγινε με απλή καταγραφή στο αντίστοιχης χορδής κανάλι κάθε νότας του κομματιού. Χρησιμοποιήθηκε η βιβλιοθήκη MidiUtil της python και ακολούθως το DAW Fruit Loops Studio για την περαιτέρω επεξεργασία και εισαγωγή εφέ σε κάθε διαφορετικό κανάλι.

## Κεφάλαιο 6

### Επίλογος

---

#### 6.1 Συμπεράσματα

Στη παρούσα διπλωματική εργασία μελετήθηκε το πρόβλημα της αυτόματης μεταγραφής ηχητικού σήματος κιθάρας σε ταμπλατούρα. Για την μεταγραφή ταμπλατούρας είναι αναγκαία η αναγνώριση έναρξης νότας (onset), η εκτίμηση τονικού ύψους (pitch) και τέλος, η αναγνώριση ενεργής χορδής. Η παρούσα εργασία, επικεντρώθηκε στην αναγνώριση ενεργής χορδής η οποία αποτελεί και το πιο πολυσύνθετο και λιγότερο μελετημένο υποπρόβλημα από τα παραπάνω. Μελετήθηκαν τόσο περιπτώσεις όπου η γνώση για το τονικό ύψος και τα onset έχουν βασιστεί σε καταγεγραμμένες επισημειώσεις, όσο και σε περιπτώσεις που εξάγονται από το ηχητικό σήμα χρησιμοποιώντας ήδη δημοσιευμένα συστήματα (CREPE [5], TCN [9]). Έτσι μπόρεσε να αποτιμηθεί η προτεινόμενη μέθοδος υπό διάφορες συνθήκες και παραμετροποιήσεις, δίνοντας μια σφαιρική εικόνα για την ακρίβεια και τις δυσκολίες που προκύπτουν.

Η αναρμονικότητα αποδείχθηκε μια σημαντική παράμετρος για την αναγνώριση ενεργής χορδής, ενώ ο γενετικός αλγόριθμος φαίνεται να παρουσιάζει πολύ θετικά αποτελέσματα για διάφορες ρυθμίσεις. Διαφαίνεται ότι ακόμη και με πολύ λιτή παραμετροποίηση, ένα μικρό σύνολο σωστών προβλέψεων - αλλά με απόλυτη ακρίβεια - του επιτρέπει να βελτιώσει συντριπτικά την πρόβλεψη ταμπλατούρας κατά ποσοστά που αγγίζουν το 40 – 50%. Τελικά, συμπεραίνουμε πως ο συνδιασμός της ηχητικής ταξινόμησης με την ταξινόμηση βάση κριτηρίων εκτελεσιμότητας, που αποτελεί και την κύρια συνεισφορά της παρούσας εργασίας, επιτυγχάνει πολύ ενθαρρυντικά αποτελέσματα.

Εν κατακλείδι, το σχήμα που προτείνεται, δίνει πολύ ενθαρρυντικά αποτελέσματα για το εν λόγω πρόβλημα, ενώ απαιτεί ελάχιστο χρόνο και κόπο για την προσαρμογή του σε κάθε κιθάρα. Ακόμη, ο διαχωρισμός σε επιμέρους υποσυστήματα, αποτελεί μια προσέγγιση εύκολα επεκτάσιμη, καθώς κάθε επιμέρους κομμάτι μπορεί να αντικατασταθεί με ένα αντίστοιχο υποσύστημα, ενώ πρόσθετα υποσυστήματα μπορούν να προστεθούν για να βελτιώσουν την ακρίβεια κάθε σταδίου (πχ βίντεο συνιστώσα).

#### 6.2 Μελλοντικές Επεκτάσεις

Η εκτενής μελέτη του γενετικού αλγόριθμου, αλλά και οποιοδήποτε παρόμοιου συστήματος που μπορεί να αξιοποιήσει τις 'προτιμήσεις' και τους φυσικούς περιορισμούς ενός κιθαρίστα, θα βοηθήσει σημαντικά στην ακριβέστερη και ταχύτερη επίλυση σχετικών προβλημάτων. Συ-

νεπώς, ανάπτυξη και παραμετροποίηση ενός καλύτερου γενετικού αλγόριθμου ή ακόμα και η εκπαίδευση ενός νευρωνικού δικτύου θα ήταν ένα σημαντικό βήμα μπροστά. Επιπροσθέτως, το προτεινόμενο σύστημα απαιτεί χρόνο για την εξαγωγή της ταμπλατούρας καθώς και ο γενετικός εφαρμόζεται σε ολόκληρη την ταμπλατούρα. Έτσι για την καταγραφή ταμπλατούρας ενός κομματιού, είναι απαραίτητο να ηχογραφηθεί ολόκληρο πριν την επεξεργασία του. Σημαντικό λοιπόν θα ήταν η τροποποίηση των παραπάνω ώστε να λειτουργούν σε πραγματικό χρόνο (on-line), έτσι ώστε να μπορούν να χρησιμοποιηθούν για εκπαιδευτικούς σκοπούς, αλλά και σε ζωντανές παραστάσεις μουσικής.

Η εξαγωγή και αξιοποίηση κι άλλων ηχιτικών χαρακτηριστικών μιας χορδής είναι μια ενδιαφέρουσα κατεύθυνση έρευνας ενώ σημαντική είναι η γενίκευση της προτεινόμενης μεθόδου για πολυφωνικό ήχο. Αξίζει να τονιστεί, πως παραμένει ανοιχτό να προσδιοριστεί ο αριθμός στιγμιοτύπων νοτών και οι κατάλληλες συνθήκες που απαιτούνται, ώστε να προσεγγιστεί καλύτερα η συμπεριφορά του συντελεστή αναρμονικότητας κάθε χορδής από λίγα τάστα και έτσι να βελτιωθούν τα αποτελέσματα συνολικά. Τέλος, στις μελλοντικές επεκτάσεις, προσδοκάται μεταξύ άλλων η χρήση του παραπάνω εργαλείου για μουσική σύνθεση. Πιο συγκεκριμένα, με τη μετατροπή σε MIDI κάθε χορδής σε διαφορετικό κανάλι είναι δυνατόν να χρησιμοποιηθεί για μουσική παραγωγή και σύνθεση.

# Παραρτήματα

---



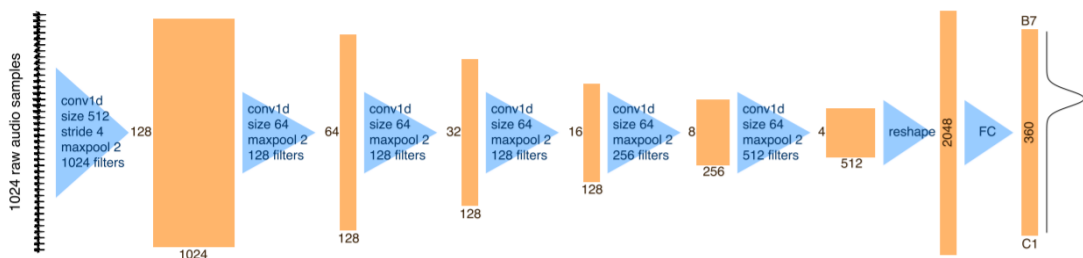
## Παράρτημα **A'**

# Εκτίμηση τονικότητας με το CREPE και ανίχνευση ενάρξεων με TCN

### A'.1 Εκτίμηση τονικότητας με το CREPE

Το CREPE (Convolutional REpresentation for Pitch Estimation) [5], αποτελεί ένα νευρωνικό δίκτυο, που λειτουργεί κατευθείαν στο πεδίο του χρόνου. Η είσοδος του είναι ένα κομμάτι των 1024 δειγμάτων/σημείων από την κυματομορφή στο πεδίο του χρόνου, με 16kHz ρυθμό δειγματοληψίας. Υπάρχουν 6 συνελικτικά επίπεδα του νευρωνικού τα οποία συνδέονται με το επίπεδο εξόδου χρησιμοποιώντας τη σιγμοειδή ως συνάρτηση ενεργοποίησης (activation function) και καταλήγοντας έτσι σε μια έξοδο 360 διαστάσεων  $\hat{y}$ . Κάθε μια από τις 360 διαστάσεις αντιστοιχεί ακολουθώντας σε ένα τονικό ύψος εκφρασμένο σε cents (μέγεθος που αντιπροσωπεύει μουσικά διαστήματα σε σχέση με ένα τονικό ύψος  $f_{ref}$  σε Hz και υπολογίζεται ως:

$$c(f) = 1200 \cdot \log_2 \frac{f}{f_{ref}} \quad (\text{A'.1})$$



Σχήμα A'.1: Αρχιτεκτονική του CREPE [5]

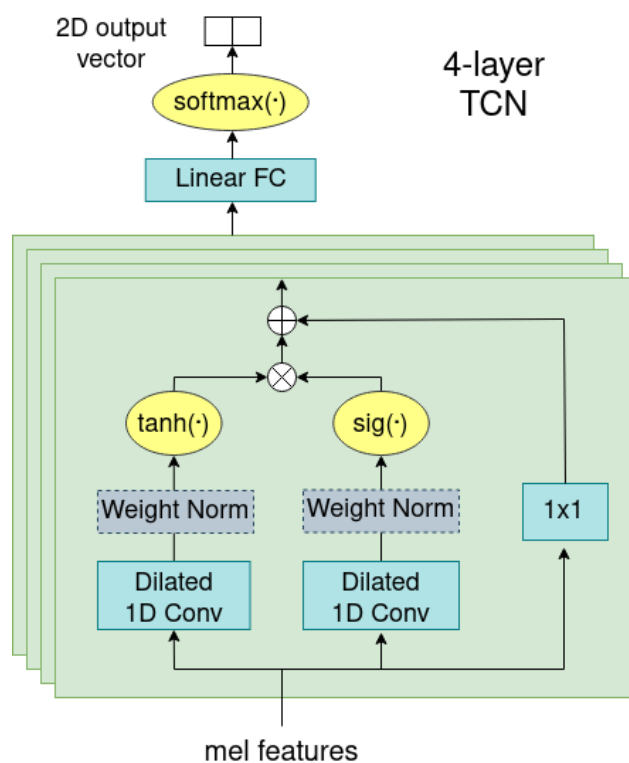
Στην προκειμένη περίπτωση έχει επιλεγεί  $f_{ref} = 10\text{Hz}$ . 100 cents είναι ίσα με ένα ημίτονο. Οι 360 τιμές αποτυπώνονται με τα σύμβολα  $c_1, c_2, c_3, \dots, c_{360}$  καλύπτουν 6 οκτάβες με διαφορά αναμεταξύ τους 20 cent. Η εκτιμώμενη συχνότητα  $\hat{c}$  προκύπτει από τον υπολογισμό του σταθμισμένου μέσου της εξόδου:

$$\hat{c} = \frac{\sum_{i=1}^{360} \hat{y}_i \cdot c_i}{\sum_{i=1}^{360} \hat{y}_i} \quad (\text{A'.2})$$

$$\hat{f} = f_{ref} \cdot 2^{\hat{c}/1200} \tag{A'.3}$$

## A'.2 Αναγνώριση έναρξης νότας με TCN

Η ανίχνευση έναρξεων νοτών, έχει επιτευχθεί με τη εφαρμογή της έρευνας των Bastas et al. [9] (κώδικας στο [39]). Για την αναγνώριση έναρξης νότας, το μοντέλο που χρησιμοποιήθηκε επεξεργάζεται το σήμα εισόδου των 44.1kHz και δημιουργεί τα mel φασματογραφήματα για κάθε ηχογράφιση με μήκος αναπήδησης (hop size) των 10ms, μήκος παραθύρων (frame length) των 40ms και 40 φασματικές λωρίδες (frequency bands). Το μοντέλο αντλεί ιδέες από την αρχιτεκτονική Wavenet [40] και το TCN που παρουσιάστηκε στο [41]. Παρόμοια αρχιτεκτονική επίσης έχει προταθεί για συνδυασμένη χρήση εικόνας ήχου για αναγνώριση έναρξης νότας για τοξωτά έγχορδα όργανα στο [42].



Σχήμα A'.2: Αρχιτεκτονική του Temporal Convolutional Network (TCN).



## Παράρτημα Β'

### Τεχνολογίες που χρησιμοποιήθηκαν

---

Για τις ανάγκες τις παρούσας διπλωματικής αναπτύχθηκε λογισμικό γραμμένο στη γλώσσα Python. Χρησιμοποιήθηκαν βιβλιοθήκες για μηχανική μάθηση (sci-kit learn), ανάλυση και αποτύπωση δεδομένων (pandas, matplotlib), σχεδιασμό γενετικών αλγορίθμων (DEAP [43]), επεξεργασία σήματος ήχου (librosa[44]), ανάκτηση μουσικής πληροφορίας (madmom[45]) και άλλα. Κύρια πηγή δεδομένων αποτέλεσε το Guitarset Dataset[28], το οποίο αποτελείται από ένα σύνολο ηχογραφήσεων κιθάρας συνοδείας και μελωδικής (solo). Συμπεριλαμβάνει ηχογραφήσεις από 6 διαφορετικούς εκτελεστές, σε 5 είδη (Jazz, Funk, Rock, Bossa Nova, Singer-Songwriter) και διάφορες κλίμακες. Περιλαμβάνει πληροφορία σχετικά με τα υπό εξέταση κομμάτια για ρυθμό, χρονική έναρξη νότας, το πέρας της, τονικό ύψος, κλειδί, συγχορδία και άλλα, κωδικοποιημένα και καταγεγραμμένα με το πρωτόκολλο JAMS [46]. Αποτελεί μια πλούσια συλλογή που περιλαμβάνει πλήθος εκφραστικών τρόπων και μουσικών γεγονότων. Περιλαμβάνει 4 μέρη με διαφορετικές συνθήκες ηχογράφησης, το πρώτο περιέχει εξακάλυπτες ηχογραφήσεις με μαγνήτη (ένα κανάλι για κάθε χορδή), το δεύτερο όπου ακολούθως τα κομμάτια έχουν γίνει debleeded, το τρίτο αποτελείται από μονοκάλυπτες οι οποίες προκύπτουν από το σύντηξη των επι μέρους 6 καναλιών του πρώτου σεναρίου, και το τρίτο, αποτελείται από μονοκάλυπτες ηχογραφήσεις με μικρόφωνο. Σε όλες τις περιπτώσεις, τα κομμάτια είναι ταυτόσημα. Στα αρνητικά του, εντοπίζεται πως αδυνατεί να αποτελέσει μια πλήρης χαρτογράφηση της έκτασης των δυνατοτήτων και ιδιαιτεροτήτων της κιθάρας, αφενός διότι δεν χρησιμοποιούνται όλα τα τάστα της, αφετέρου διότι αποτελείται από ηχογραφήσεις μόνο μιας κιθάρας. Η σχετική αντίστοιξη είναι αποθηκευμένη σε αρχεία JAMS. Το JAMS από τους Humphrey et al [46] είναι ένα JSON-based σχήμα, καταγραφής μουσικής αντίστοιξης για μελέτες σχετικές με την εξαγωγή μουσικής πληροφορίας. Χρησιμοποιήθηκε για τις ανάγκες τις διπλωματικής σε διάφορα στάδιά της για την ανάγνωση της σχετικής μουσικής πληροφορίας που συνδέεται με κάθε κομμάτι του Guitarset Dataset [28]. Ακόμα, σημαντικό είναι να τονιστεί η αξία του ως ένα τυπικό τρόπο καταγράφης μουσικών παρατηρήσεων και πληροφορίας με οργανωμένη και εύκολα αναγνωρίσιμη μορφή.



## Βιβλιογραφία

---

- [1] Neville H Fletcher και Thomas D Rossing. *The physics of musical instruments*. Springer Science & Business Media, 2012.
- [2] Matthieu Hodgkinson, Jian Wang, Joseph Timoney και Victor Lazzarini. *Handling inharmonic series with median-adjustive trajectories*. *DAFx 09 proceedings of the 12th International Conference on Digital Audio Effects, Politecnico di Milano, Como Campus, Sept. 1-4, Como, Italy*. Dept. of Electronic Engineering, Queen Mary Univ. of London, 2009.
- [3] Isabel Barbancho, Lorenzo J Tardon, Simone Sammartino και Ana M Barbancho. *Inharmonicity-based method for the automatic generation of guitar tablature*. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6), 2012.
- [4] Meinard Müller. *Fundamentals of music processing: Audio, analysis, algorithms, applications*. Springer, 2015.
- [5] Jong Wook Kim, Justin Salamon, Peter Li και Juan Pablo Bello. *Crepe: A convolutional representation for pitch estimation*. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018.
- [6] Jacob Møller Hjerrild και Mads Græsbøll Christensen. *Estimation of guitar string, fret and plucking position using parametric pitch estimation*. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019.
- [7] Thomas D Rossing και others. *The science of string instruments*. Springer, 2010.
- [8] Jacob Møller Hjerrild, Silvin Willemsen και Mads Græsbøll Christensen. *Physical Models for Fast Estimation of Guitar String, Fret and Plucking Position*. *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2019.
- [9] V. Katsouros P. Maragos G. Bastas, A. Gkiokas. *Convolutional Networks for Visual Onset Detection in the Context of Bowed String Instrument Performances*. 2021.
- [10] Ana M Barbancho, Anssi Klapuri, Lorenzo J Tardón και Isabel Barbancho. *Automatic transcription of guitar chords and fingering from audio*. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(3), 2011.
- [11] Gregory Burlet και Ichiro Fujinaga. *Robotaba Guitar Tablature Transcription Framework*. *ISMIR*, 2013.

- [12] Kazuki Yazawa, Daichi Sakaue, Kohei Nagira, Katsutoshi Itoyama και Hiroshi G Okuno. *Audio-based guitar tablature transcription using multipitch analysis and playability constraints*. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013.
- [13] Kazuki Yazawa, Katsutoshi Itoyama και Hiroshi G Okuno. *Automatic transcription of guitar tablature from audio signals in accordance with player's proficiency*. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014.
- [14] Aleksander Radisavljevic και Peter F Driessen. *Path difference learning for guitar fingering problem*. ICMC, τόμος 28, 2004.
- [15] Daniele Radicioni και Vincenzo Lombardo. *Guitar fingering for music performance*. ICMC, 2005.
- [16] Masanobu Miura, Isao Hirota, Nobuhiko Hama και Masazo Yanagida. *Constructing a system for finger-position determination and tablature generation for playing melodies on guitars*. *Systems and Computers in Japan*, 35(6), 2004.
- [17] Gen Hori, Hirokazu Kameoka και Shigeki Sagayama. *Input-output HMM applied to automatic arrangement for guitars*. *Information and Media Technologies*, 8(2), 2013.
- [18] Yu Hua Chen, Yu Hsiang Huang, Wen Yi Hsiao και Yi Hsuan Yang. *Automatic composition of guitar tabs by transformers and groove modeling*. *arXiv preprint arXiv:2008.01431*, 2020.
- [19] Daniel R Tuohy και Walter D Potter. *A genetic algorithm for the automatic generation of playable guitar tablature*. ICMC, 2005.
- [20] Joao Victor Ramos, Andre Stylianos Ramos, Carlos N Silla και Danilo Sipoli Sanches. *Comparative study of genetic algorithm and ant colony optimization algorithm performances for the task of guitar tablature transcription*. 2015 Brazilian Conference on Intelligent Systems (BRACIS). IEEE, 2015.
- [21] Joao Victor Ramos, André Stylianos Ramos, Carlos N Silla και Danilo Sipoli Sanches. *An Evaluation of Different Evolutionary Approaches Applied in the Process of Automatic Transcription of Music Scores into Tablatures*. 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2016.
- [22] Hank Heijink και Ruud GJ Meulenbroek. *On the complexity of classical guitar playing: functional adaptations to task constraints*. *Journal of motor behavior*, 34(4), 2002.
- [23] Daniel R Tuohy και Walter D Potter. *Generating guitar tablature with LHF notation via DGA and ANN*. *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 2006.

- [24] Thierry Gagnon, Steeve Larouche και Roch Lefebvre. *A neural network approach for preclassification in musical chords recognition. The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, τόμος 2. IEEE, 2003.
- [25] Eric J Humphrey και Juan P Bello. *From music audio to chord tablature: Teaching deep convolutional networks to play guitar. 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2014.
- [26] Christian Kehling, Jakob Abeßer, Christian Dittmar και Gerald Schuller. *Automatic Tablature Transcription of Electric Guitar Recordings by Estimation of Score-and Instrument-Related Parameters. DAFX*, 2014.
- [27] Andrew Wiggins και Youngmoo Kim. *Guitar Tablature Estimation with a Convolutional Neural Network. ISMIR*, 2019.
- [28] Qingyang Xi, Rachel M Bittner, Johan Pauwels, Xuzhou Ye και Juan Pablo Bello. *GuitarSet: A Dataset for Guitar Transcription. ISMIR*, 2018.
- [29] Jakob Abeßer. *Automatic string detection for bass guitar and electric guitar. International Symposium on Computer Music Modeling and Retrieval*. Springer, 2012.
- [30] Jonathan Michelson, Richard Stern και Thomas Sullivan. *Automatic guitar tablature transcription from audio using inharmonicity regression and bayesian classification. Audio Engineering Society Convention 145*. Audio Engineering Society, 2018.
- [31] Tobias Geib, Maximilian Schmitt και Björn Schuller. *Automatic guitar string detection by string-inverse frequency estimation. Gesellschaft für Informatik, Bonn*, 2017.
- [32] Dalia Senvaityte, Johan Pauwels και Mark Sandler. *Guitar String Separation Using Non-Negative Matrix Factorization and Factor Deconvolution. Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, 2019.
- [33] Anssi Klapuri. *Introduction to music transcription. Signal processing methods for music transcription*. Springer, 2006.
- [34] Mads Græsbøll Christensen και Andreas Jakobsson. *Multi-pitch estimation. Synthesis Lectures on Speech & Audio Processing*, 5(1), 2009.
- [35] Alan V Oppenheim. *Discrete-time signal processing*. Pearson Education India, 1999.
- [36] Tom M Mitchell και others. *Machine learning*. 1997.
- [37] Pierre Lison. *An introduction to machine learning. Language Technology Group (LTG)*, 1(35), 2015.
- [38] Chris M Bishop. *Neural networks and their applications. Review of scientific instruments*, 65(6), 1994.
- [39] *Visual onset detection from bowed strings. [Online]*. Available: <https://github.com/gbastas/avOnset.git>.

- [40] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior και Koray Kavukcuoglu. *Wavenet: A generative model for raw audio*. *arXiv preprint arXiv:1609.03499*, 2016.
- [41] Shaojie Bai, J Zico Kolter και Vladlen Koltun. *An empirical evaluation of generic convolutional and recurrent networks for sequence modeling*. *arXiv preprint arXiv:1803.01271*, 2018.
- [42] Grigoris Bastas, Aggelos Gkiokas, Vassilis Katsouros και Petros Maragos. *Improving Audio Onset Detection for String Instruments by Incorporating Visual Modality*. *MML 2020*.
- [43] Félix Antoine Fortin, François Michel De Rainville, Marc André Gardner, Marc Parizeau και Christian Gagné. *DEAP: Evolutionary Algorithms Made Easy*. *Journal of Machine Learning Research*, 13, 2012.
- [44] B McFee, V Lostanlen, A Metsai, M McVicar, S Balke, C Thomé, C Raffel, F Zalkow, A Malek, K Lee Dana και others. *viktorandreevichmorozov, J. Moore, R. Bittner, S. Hidaka, Z. Wei, nullmightybofo, D. Herenú, F.-R. Stöter, P. Friesch, A. Weiss, M. Vollrath, and T. Kim, "librosa/librosa: 0.8. 0," Jul, 2020*.
- [45] Sebastian Böck, Filip Korzeniowski, Jan Schlüter, Florian Krebs και Gerhard Widmer. *madmom: a new Python Audio and Music Signal Processing Library*. *Proceedings of the 24th ACM International Conference on Multimedia*, Amsterdam, The Netherlands, 2016.
- [46] Eric J Humphrey, Justin Salamon, Oriol Nieto, Jon Forsyth, Rachel M Bittner και Juan Pablo Bello. *JAMS: A JSON Annotated Music Specification for Reproducible MIR Research*. *ISMIR*, 2014.

## Συντομογραφίες - Αρχικόλεξα - Ακρωνύμια

---

βλ.	βλέπε
κ.λπ.	και λοιπά
κ.ο.κ	και ούτω καθεξής
TCN	Temporal Convolutional Network
CNN	Convolutional Neural Network
GA	Genetic Algorithm





## Απόδοση ξενόγλωσσων όρων

---

### Απόδοση

συντελεστής αναρμονικότητας  
πίνακας σύγχυσης  
ανάκτηση μουσικής πληροφορίας  
έναρξη  
τονικό ύψος  
αναγνώριση έναρξης  
εκτίμηση τονικού ύψους  
μερική συχνότητα  
αρμονική συχνότητα  
χαρακτηριστικό  
αναρμονική ανάλυση  
στιγμιότυπο νότας  
ασαφής νότα  
επισημειωμένο  
ακαμψία (χορδής)  
κανονικό κούρδισμα  
μοντέλα λίγων δειγμάτων  
μοντέλα βασισμένα σε προσωμοίωση  
φυσικοί περιορισμοί  
παιζόμενη ταμπλατούρα  
εκτελεσιμότητα  
συνάρτηση ενεργοποίησης  
γνωσιακός

### Ξενόγλωσσος όρος

inharmonic coefficient  
confusion matrix  
music information retrieval  
onset  
pitch  
onset detection  
pitch estimation  
partial frequency  
harmonic frequency  
feature  
inharmonic analysis  
note instance  
inconclusive note  
annotated  
(string's) stiffness  
standard tuning  
few-sample models  
simulation-based models  
physical constraints  
playable tablature  
playability  
activation function  
cognitive

