



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

## Αυτόματη αναγνώριση μουσικών σημάτων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

**ΑΝΤΙΓΟΝΗΣ Ε.ΤΣΙΑΜΗ**

**Επιβλέπων :** Κωνσταντίνος Παπαοδυσσεύς  
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβρης, 2012





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

## Αυτόματη αναγνώριση μουσικών σημάτων

### ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

**ΑΝΤΙΓΟΝΗΣ Ε.ΤΣΙΑΜΗ**

**Επιβλέπων :** Κωνσταντίνος Παπαοδυσσεύς  
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 5<sup>η</sup> Νοεμβρίου 2012.

.....  
Κωνσταντίνος Παπαοδυσσεύς,  
Αναπλ. Καθηγητής Ε.Μ.Π.

.....  
Γεώργιος Καμπουράκης,  
Αναπλ. Καθηγητής Ε.Μ.Π.

.....  
Ηλίας Κουκούτσης  
Επικ. Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβρης, 2012

.....

## **ΑΝΤΙΓΟΝΗ ΤΣΙΑΜΗ**

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

Copyright © Αντιγόνη Τσιάμη, 2012.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

## Περίληψη

Η παρούσα διπλωματική συνιστά μία μελέτη ενός πολύ σημαντικού προβλήματος στο χώρο της ψηφιακής επεξεργασίας μουσικών συνθέσεων. Συγκεκριμένα στην παρούσα διπλωματική μελετάται η αυτόματη ταυτοποίηση μιας μουσικής σύνθεσης ανεξαρτήτως του εκτελεστή και του μουσικού οργάνου εκτέλεσης. Κατά τη διάρκεια της μελέτης αυτής διαπιστώθηκε ότι η απεικόνιση, τόσο στο πεδίο της συχνότητας όσο και στο πεδίο του χρόνου, μιας μουσικής σύνθεσης παιγμένης από διαφορετικούς εκτελεστές κατά κανόνα εμφανίζει πολύ σημαντικές διαφορές.

Εντούτοις, ο ακροατής συνήθως αναγνωρίζει και ταυτοποιεί τη μουσική σύνθεση ανεξαρτήτως εκτελεστή και οργάνου εκτέλεσης. Αυτό το γεγονός σημαίνει ότι υπάρχει ένας πυρήνας σε όλες τις διαφορετικές εκτελέσεις μιας μουσικής σύνθεσης, ο οποίος παραμένει αναλλοίωτος μετά την υποσυνείδητη επεξεργασία που υφίσταται στον ανθρώπινο εγκέφαλο. Αυτός ο πυρήνας εκτιμούμε ότι πρέπει να έχει μαθηματική υπόσταση είτε στο πεδίο του χρόνου είτε στο πεδίο της συχνότητας. Η παρούσα διπλωματική επιθυμεί να συμβάλει στον εντοπισμό αυτού του πυρήνα. Ειδικότερα, σε αυτή την εργασία εφαρμόστηκαν τεχνικές που σχετίζονται με το συχνοτικό περιεχόμενο μιας μουσικής σύνθεσης και εντοπίστηκαν κριτήρια ομοιότητας μεταξύ συνθέσεων παιγμένων από συμφωνική ορχήστρα αφενός και από σόλο όργανα αφετέρου. Οι τεχνικές αυτές περιελάμβαναν α) εντοπισμό των σημαντικότερων κορυφών του φάσματος κάθε μουσικής σύνθεσης, β) κατάλληλη σταχυολόγηση των κορυφών με χρήση της διαδικασίας masking, γ) κατάταξη των κορυφών που τελικά επελέγησαν σε ειδικά επιλεγμένα διαστήματα που προσομοιάζουν τη μουσική εκθετική κλίμακα, δ) πολλαπλή εφαρμογή της διαδικασίας μετά από ομοιοθεσία στη συχνότητα, ε) μερικές νέες πρωτοποριακές τεχνικές. Η προσέγγιση εφαρμόστηκε σε τρεις κλασικές μουσικές συνθέσεις και τα αποτελέσματα κρίνονται ιδιαίτερος ενθαρρυντικά.

**Λέξεις-κλειδιά:** αυτόματη ταυτοποίηση μουσικών συνθέσεων, ψηφιακή επεξεργασία μουσικής, masking, ομοιοθεσία στο πεδίο της συχνότητας



## Abstract

The present dissertation deals with the study of the automatic identification of a musical composition, independently of the performer and the instruments used for the performance. The study manifested that there are substantial differences between two performances of the same musical composition, both in the time and the frequency domain.

However, as a rule, a listener identifies a musical composition independently of the performer and the employed instruments. This fact suggests that there is a common kernel in all different performances of a musical composition that listeners recognize as the same one. This kernel must reflect on the time and on the frequency domain. The purpose of the present work is to contribute towards the understanding of the various mathematical characteristics that may form this kernel. More specifically, a number of algorithms and related techniques have been tested and/or developed associated with:

- a) the determination of the dominant peaks in the spectrum of a musical composition
- b) further selection of the dominant peaks by application of a proper masking procedure
- c) classification of the selected peaks in proper intervals that simulate the exponential music scale
- d) multiple application of the previous steps after “stretching”, i.e. after scaling in the frequency domain
- e) a number of novel approaches associated with the octave, the masking procedure and the number of selected peaks

Using the aforementioned techniques and algorithms, proper criteria have been developed for the determination of similarities between musical composition performed by an orchestra in one hand and by two solo instruments on the other.

The methodology has been applied in three distinct classical compositions offering statistically satisfactory and promising results.

**Key-words:** automatic identification of musical compositions, digital music processing, masking, scaling in the frequency domain





## Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον καθηγητή μου κ. Κωνσταντίνο Παπαοδυσσέα για την πολύτιμη καθοδήγηση και υποστήριξή του καθ' όλη τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας. Επίσης, για την ευκαιρία που μου έδωσε να ασχοληθώ με ένα πολύ ενδιαφέρον θέμα, το οποίο συνδυάζει τη μουσική με την έρευνα και την τεχνολογία. Νιώθω πολύ τυχερή για τη συνεργασία μου και για τη γνωριμία μου με αυτόν τον σπάνιο άνθρωπο και ερευνητή.

Ένα μεγάλο ευχαριστώ στο Φώτη Γιαννόπουλο για την συνεργασία και την πολύτιμη βοήθειά του, καθώς και για τις ατελείωτες ώρες που αφιέρωσε ώστε να μπορέσουμε να δώσουμε λύσεις σε διάφορα προβλήματα.

Θα ήθελα να ευχαριστήσω τους αγαπημένους φίλους και ανθρώπους μου για όλη την υποστήριξη και τη συμπαράστασή τους όλα αυτά τα χρόνια, την κοινή πορεία μας και τα όνειρά μας, την Έλενα, τη Νίκη, τη Μάρθα και το Γιώργο.

Τέλος, θα ήθελα να ευχαριστήσω τη μητέρα μου Γιώτα, που ήταν πάντα δίπλα μου σε όλα και τον πατέρα μου Λευτέρη, για όλη την υποστήριξη που μου παρείχαν σε όλα τα χρόνια των σπουδών μου. Ένα τεράστιο ευχαριστώ στον αδερφό μου Τάσο, που αποτέλεσε το μεγαλύτερο έμπρακτο και ψυχολογικό μου στήριγμα. Η καθοδήγηση και η βοήθειά του ήταν πολύτιμες και αναντικατάστατες.

## Περιεχόμενα

1	Εισαγωγή και στόχος της παρούσας διπλωματικής.....	13
2	Προϋπάρχουσες δημοσιευμένες εργασίες πάνω στην αυτόματη αναγνώριση και καταγραφή μουσικών συνθέσεων .....	17
3	Τα ενδογενή σημαντικά προβλήματα στην ταυτοποίηση μουσικών συνθέσεων ανεξαρτήτως εκτελεστή και οργάνου.....	21
3.1	Μουσική σύνθεση: Rondo Allegro από το K525 .....	23
3.1.1	Όλο το σήμα .....	23
3.1.2	Τμήμα των σημάτων που ταιριάζει ακουστικά .....	27
3.2	Μουσική σύνθεση: Romanze Andante από το K525 .....	33
3.2.1	Όλο το σήμα .....	33
3.2.2	Τμήμα των σημάτων που ταιριάζει ακουστικά .....	37
3.3	Μουσική σύνθεση: Adagio από το K622 (Concerto for Clarinet) .....	43
3.3.1	Όλο το σήμα .....	43
3.2.2	Τμήμα των σημάτων που ταιριάζει ακουστικά .....	48
4	Θεμελιώδεις τεχνικές και αλγόριθμοι στους οποίους βασίστηκε η μέθοδος που εφαρμόστηκε .....	55
4.1	Μαθηματικό υπόβαθρο και αναδρομικός FFT.....	55
4.1.1	Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Transform-DFT) .....	56
4.1.2	Αλγόριθμος προσαρμοστικού/αναδρομικού (adaptive) DFT.....	73
4.2	Διαχωρισμός σε νότες .....	78
4.3	Η έννοια του masking, για τον εντοπισμό των ακουστικά σημαντικών κορυφών και αποθορυβοποίηση.....	80
4.4	Η έννοια του stretch (ομοιοθεσία) για την αντιστάθμιση της διαφοράς ρυθμού και τονικότητας .....	85
5	Πρωτότυπες επεκτάσεις των τεχνικών και αλγορίθμων.....	93
5.1	Αλλαγή οκτάβας.....	93
5.2	Αριθμός κορυφών N ανά παράθυρο.....	94

5.3	Εύρος του masking .....	95
6	Η μεθοδολογία που εφαρμόστηκε και μερικά πρώτα αποτελέσματα .....	97
6.1	Συνοπτική περιγραφή της διαδικασίας ταυτοποίησης μουσικών συνθέσεων που αναπτύχθηκε.....	97
6.2	Μερικά πρώτα αποτελέσματα .....	99
7	Συμπεράσματα και μελλοντικές επεκτάσεις.....	101
8	Βιβλιογραφία.....	103



# ***1***

## ***Εισαγωγή και στόχος της παρούσας διπλωματικής***

Ένα από τα σημαντικότερα θεωρητικά και πρακτικά προβλήματα στο χώρο της ψηφιακής επεξεργασίας μουσικών συνθέσεων είναι η αυτόματη ταυτοποίηση μιας μουσικής σύνθεσης ανεξαρτήτως του εκτελεστή και του μουσικού οργάνου εκτέλεσης.

Η θεωρητική αξία της επίλυσης του συγκεκριμένου προβλήματος είναι μάλλον προφανής: Εάν κάποιος επιτύχει να ταυτοποιεί αυτόματα τις μουσικές συνθέσεις ανεξαρτήτως του εκτελεστή και του οργάνου εκτέλεσης, αυτό σημαίνει ότι έχει εντοπισθεί ο μαθηματικός πυρήνας του ακουστικού περιεχομένου των μουσικών συνθέσεων. Πέρα από αυτό το τεράστιο θεωρητικό ενδιαφέρον, μια τέτοια αυτόματη ταυτοποίηση θα είχε και σημαντικές πρακτικές εφαρμογές, μερικές από τις οποίες αναφέρονται παρακάτω:

Ίσως η πλέον σημαντική από αυτές τις εφαρμογές έχει να κάνει με την προστασία της πνευματικής ιδιοκτησίας, δεδομένου ότι θα επιτρέπει την ταυτοποίηση τμημάτων μουσικών συνθέσεων που περιέχονται αυτούσια σε άλλες, ανεξαρτήτως του εκτελεστή

και των οργάνων εκτέλεσης.

Ένα τέτοιο σύστημα ταυτοποίησης στην ενδελέχειά του μπορεί να χρησιμοποιηθεί ακόμα και εμπορικά στην ανεύρεση επιθυμητών μουσικών συνθέσεων. Συγκεκριμένα, σε ένα τέτοιο σύστημα μπορεί ένας χρήστης να σφυρίξει ή ακόμα και να παίξει με απλό όργανο τη μουσική σύνθεση που έχει συγκρατήσει και το σύστημα του προσφέρει απάντηση ποια είναι αυτή.

Άλλη πολύ σημαντική πρακτική εφαρμογή είναι ότι η επίλυση αυτού του προβλήματος μπορεί να ανοίξει δρόμο για την αυτόματη καταγραφή μιας μουσικής σύνθεσης σε παρτιτούρα, γεγονός που αν πραγματοποιηθεί λογικά θα έχει και τεράστια εμπορική επιτυχία. Ειδικότερα, η ταυτοποίηση μιας μουσικής σύνθεσης ανεξαρτήτως εκτέλεσης και οργάνου εκτέλεσης μπορεί σχεδόν αμέσως να οδηγήσει στην καταγραφή ενός βασικού συνόλου/σκελετού νοτών, ο οποίος περιγράφει το βασικό ακουστικό πυρήνα της σύνθεσης.

Σε έναν πολύ μακροπρόθεσμο ανοιχτό ορίζοντα, μπορεί κανείς να φανταστεί ότι ένα τέτοιο σύστημα αυτόματης ταυτοποίησης μπορεί να βοηθήσει σπουδαστές μουσικής στην καλύτερη εκμάθηση των οργάνων.

Η παρούσα διπλωματική συνιστά μία προσπάθεια προς την προαναφερθείσα κατεύθυνση ταυτοποίησης μουσικών συνθέσεων οι οποίες έχουν παιχτεί αφ' ενός από συμφωνική ορχήστρα και από ένα πλήθος σόλο οργάνων αφετέρου. Συγκεκριμένα η δομή της παρούσας εργασίας έχει ως εξής:

Στο κεφάλαιο 2 έγινε μια προσπάθεια να συγκεντρωθεί ένα πλήθος δημοσιεύσεων από επιστημονικά περιοδικά με αντικείμενο την επεξεργασία μουσικών σημάτων και ειδικότερα τομείς όπως η αναγνώριση/ταυτοποίηση μουσικών συνθέσεων, η αυτόματη καταγραφή νοτών, δακτυλισμών κ.α. Σε κάθε παράγραφο του κεφαλαίου αυτού αναφέρεται συνοπτικά ο σκοπός και η τεχνική που χρησιμοποιήθηκε σε κάθε δημοσίευση.

Στο κεφάλαιο 3 περιγράφονται οι ενδογενείς, σημαντικές δυσκολίες του προβλήματος που αντιμετωπίστηκε. Πιο συγκεκριμένα, αναφέρονται οι θεμελιακές διαφορές που έχουν δύο εκτελέσεις της ίδιας μουσικής σύνθεσης, από ορχήστρα και από σόλο όργανο, τόσο στο πεδίο του χρόνου όσο και στο πεδίο της συχνότητας. Επίσης,

μέσα από παραδείγματα από συγκεκριμένες μουσικές συνθέσεις, με δύο ή τρεις διαφορετικές εκτελέσεις, που οπτικοποιούνται σε διαγράμματα τόσο στο πεδίο του χρόνου όσο και στις συχνότητες γίνεται ακόμη πιο φανερή η συνθετότητα του προβλήματος.

Στο κεφάλαιο 4 περιγράφονται το μαθηματικό υπόβαθρο και οι τεχνικές που χρησιμοποιήθηκαν στην παρούσα διπλωματική για τη διαδικασία της αυτόματης ταυτοποίησης μουσικών συνθέσεων. Το μαθηματικό υπόβαθρο αφορά κυρίως τον Διακριτό Μετασχηματισμό Fourier (DFT) και τον προσαρμοστικό/αναδρομικό DFT που αποτέλεσε το κύριο εργαλείο για την επεξεργασία των μουσικών σημάτων. Επίσης, παρουσιάζονται οι τεχνικές masking και η τεχνική της ομοιοθεσίας στο χρόνο (stretch) και ο χωρισμός του ακουστικού πεδίου σε διαστήματα ώστε να προσομοιαστεί κατά το δυνατόν η μουσική εκθετική κλίμακα.

Στο κεφάλαιο 5 περιγράφονται κάποιες πρωτότυπες τεχνικές που χρησιμοποιήσαμε κατά την εκπόνηση της παρούσας διπλωματικής, όπως για παράδειγμα α) η αλλαγή οκτάβας, η δυνατότητα δηλαδή να αναγνωρίζονται δύο διαφορετικές εκτελέσεις της ίδιας μουσικής σύνθεσης ως ίδιες ακόμα και αν βρίσκονται σε διαφορετικές οκτάβες, β) μια τεχνική επιλογής των κορυφών που συμμετέχουν στην ταυτοποίηση και γ) το εύρος του masking, ώστε να επιτυγχάνονται τα καλύτερα δυνατά ποσοστά ταυτοποίησης.

Στο κεφάλαιο 6 παρατίθεται ολόκληρη η μεθοδολογία αυτόματης ταυτοποίησης μουσικών συνθέσεων που χρησιμοποιήθηκε, σε βήματα, καθώς και κάποια πρώτα αποτελέσματα που προέκυψαν από την εφαρμογή αυτής της μεθοδολογίας σε τρεις διαφορετικές μουσικές συνθέσεις.

Τέλος, στο κεφάλαιο 7 αναφέρονται τα συμπεράσματα που προέκυψαν από την εφαρμογή της διαδικασίας αυτόματης ταυτοποίησης μουσικών συνθέσεων στην παρούσα διπλωματική καθώς και κάποιες μελλοντικές επεκτάσεις που μπορούν να γίνουν. Αυτές αφορούν τόσο τη βαθύτερη μελέτη των κριτηρίων ταυτοποίησης όσο και βελτίωση των τεχνικών και των παραμέτρων που χρησιμοποιήθηκαν.



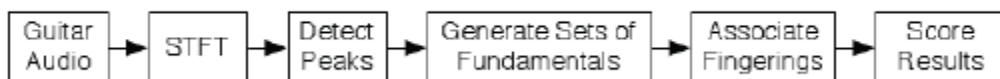


# 2

## *Προϋπάρχουσες δημοσιευμένες εργασίες πάνω στην αυτόματη αναγνώριση και καταγραφή μουσικών συνθέσεων (state of the art)*

Σε αυτή την ενότητα παρουσιάζονται συνοπτικά κάποιες δημοσιευμένες εργασίες που αφορούν έρευνα που έχει πραγματοποιηθεί στον τομέα της αναγνώρισης και ταυτοποίησης μουσικών συνθέσεων, της καταγραφής μουσικών συνθέσεων, της αναγνώρισης νοτών, δακτυλισμών κ.α.

Στο πρώτο άρθρο [1] επιτελείται αυτόματη καταγραφή νοτών κιθάρας σε πολυφωνική μουσική σύνθεση επί τη βάσει της εξής διαδικασίας:



Οι συγγραφείς ισχυρίζονται 33% τέλεια απόδοση της θέσης του χεριού που παίζει.

Σε επόμενη δημοσίευση [2] γίνεται αναγνώριση των strong και weak beats που εμφανίζονται σε μια μουσική σύνθεση. Επιπλέον βρίσκει νότες που εμφανίζονται πολλαπλώς σε μία σύνθεση.

Στο άρθρο [3] πραγματοποιείται αυτόματη αναγνώριση μεμονωμένων μονοφωνικών τμημάτων μουσικών συνθέσεων. Η κεντρική ιδέα της προτεινόμενης μεθόδου βασίζεται σε έναν αλγόριθμο εντοπισμού θεμελιωδών συχνοτήτων. Ο αλγόριθμος αυτός παράγει μία ακολουθία θεμελιωδών συχνοτήτων, τόσο στις γνωστές όσο και στην άγνωστη μουσική σύνθεση. Η ακολουθία διακριτοποιείται με χρήση κατάλληλων σχετικών καταωφλίων. Η παραγόμενη ακολουθία κβαντισμένων συχνοτήτων εισάγεται ως είσοδος σε ένα σύστημα αναγνώρισης.

Στο άρθρο [4] παρουσιάζεται η εξής ιδέα: Υπάρχει μία στάνταρ τυποποίηση μετατροπής μιας μελωδίας σε ένα σύνολο αριθμητικών ακολουθιών (standard musical instrument digital interface files). Σε αυτό το άρθρο δοκιμάζονται διάφορες στατιστικές μέθοδοι κατάταξης δεδομένων (Bayesian, k-neighbors) για την επεξεργασία αυτών των στάνταρ χαρακτηριστικών. Επί τη βάση αυτών των αποτελεσμάτων επιλέγονται καλά χαρακτηριστικά για την προτυποποίηση των συνθέσεων. Τέλος, ελέγχεται η αποδοτικότητα των σχετικών αλγορίθμων κατάταξης σε σχέση με άλλα μοντέλα περιγραφής.

Σε αυτό το άρθρο [5] επιχειρείται η αυτόματη αναγνώριση του είδους (style) μουσικών συνθέσεων. Η αναγνώριση βασίζεται σε τρία χαρακτηριστικά της σύνθεσης: α) στη χροιά/ηχοχρωματική υφή β) το ρυθμικό περιεχόμενο και γ) η τονικότητα. Αυτά τα τρία χαρακτηριστικά δίνονται ως είσοδος σε αλγόριθμους εκπαίδευσης συστημάτων στατιστικής αναγνώρισης. Η επιτυχία του συστήματος είναι 61%, έχοντας δοκιμαστεί σε υπαρκτές συλλογές μουσικών συνθέσεων.

Σε αυτή τη δημοσίευση [6] χρησιμοποιούνται μέθοδοι αναζήτησης σε μεγάλες βάσεις

δεδομένων για την ταυτοποίηση μουσικών συνθέσεων. Προς το σκοπό αυτό κάθε μουσική σύνθεση μετατρέπεται σε ένα σύνολο χαρακτηριστικών με δύο διαφορετικούς τυπικούς τρόπους: α) polyphonic score data και β) pulse-code modulation ηχητικού υλικού.

Στο άρθρο [7] επιχειρείται η ταυτοποίηση τμημάτων μουσικών συνθέσεων με χρήση hidden Markov models.

Σε αυτή τη δημοσίευση [8] κατ' αρχήν δοκιμάζεται η επίδοση διαφόρων νευρωνικών δικτύων για την αυτόματη αναγνώριση τόνων σε μουσικές συνθέσεις, επί τη βάση της παράστασης τους στο πεδίο της συχνότητας. Η επίδοση των νευρωνικών δικτύων βελτιώνεται με τη χρήση κατάλληλων δικτύων ταλαντωτών.

[9]:Εδώ δοκιμάζονται δύο μέθοδοι ανίχνευσης του βασικού ρυθμού (beat) μιας μουσικής σύνθεσης. Η πρώτη μέθοδος χρησιμοποιεί Bayesian στατιστική, ενώ η δεύτερη βασίζεται σε gradient descent. Πιο αναλυτικά στην πρώτη μέθοδο, επιλέγεται ένα σύνολο παραμέτρων που σχετίζονται με την εξέλιξη του ρυθμού και επ' αυτών πραγματοποιείται στατιστική επεξεργασία. Η δεύτερη μέθοδος που χρησιμοποιεί gradient descent σε παράθυρα beats συγκλίνει πιο αργά ταλαντούμενη.

Σε αυτό το άρθρο [10] επιχειρείται η αυτόματη καταγραφή συγχορδιών παιγμένων από κιθάρα ως και ο εντοπισμός της ακολουθίας των σχετικών δακτυλισμών. Η μέθοδος χρησιμοποιεί hidden Markov models όπου οι κρυμμένες καταστάσεις (hidden states) αντιστοιχούν στις διαφορετικές δακτυλικές τοποθετήσεις. Επιπλέον, τα παρατηρούμενα ακουστικά χαρακτηριστικά αποκτώνται από έναν εκτιμητή πολλαπλών θεμελιωδών συχνοτήτων.

[11]:Σε αυτή τη δημοσίευση χρησιμοποιείται η θεμελιώδης συχνότητα της φωνής ως θεμελιώδης μονάδα για την αυτόματη καταγραφή μουσικών συνθέσεων.

Η μεθοδολογία και το σύστημα που προτείνεται στο άρθρο [12] πραγματοποιεί αυτόματη αναγνώριση μουσικών συνθέσεων οι οποίες αναπαράγονται ή και αναμεταδίδονται από πηγές που μπορεί να εισάγουν μέχρι και ισχυρή παραμόρφωση στο πεδίο της συχνότητας. Η διαδικασία της αναγνώρισης βασίζεται στη σύγκριση της άγνωστης μουσικής ηχογράφησης, που προέρχεται από μια πηγή όπως ραδιόφωνο, τηλεόραση κ.α., με ένα σύνολο από μοντέλα ηχογραφήσεων, σύμφωνα με κάποια συγκεκριμένα χαρακτηριστικά των σημάτων. Για την εξαγωγή των χαρακτηριστικών αυτών από τις μουσικές ηχογραφήσεις χρησιμοποιήθηκαν πρωτότυποι αλγόριθμοι εξαγωγής χαρακτηριστικών. Η διαδικασία αυτή εφαρμόζεται σε όλο το σύνολο των μουσικών ηχογραφήσεων-μοντέλων, δημιουργώντας έτσι μια βάση δεδομένων με τα χαρακτηριστικά των μοντέλων. Κάθε φορά που υπάρχει μια ηχογράφηση προς αναγνώριση, εφαρμόζεται σε αυτήν η ίδια διαδικασία που εφαρμόζεται και στα μοντέλα και ακολούθως τα χαρακτηριστικά που προκύπτουν συγκρίνονται με τα περιεχόμενα της βάσης δεδομένων διαμέσου ενός συνόλου κριτηρίων. Η μεθοδολογία αυτή οδήγησε στην υλοποίηση ενός συστήματος που παρουσίασε υψηλό ποσοστό επιτυχούς αναγνώρισης μουσικών ηχογραφήσεων, της τάξης του 94%.

# 3

## *Τα ενδογενή σημαντικά προβλήματα στην ταυτοποίηση μουσικών συνθέσεων ανεξαρτήτως εκτελεστή και οργάνου*

Ας υποθεθεί ότι ένας ακροατής ακούει μια μουσική σύνθεση, έστω  $\Sigma$ , εκτελεσμένη από συμφωνική ορχήστρα. Ας υποθεθεί ότι στη συνέχεια, ο ίδιος ακροατής ακούει τη συγκεκριμένη σύνθεση  $\Sigma$  εκτελεσμένη από ένα σόλο όργανο. Παρ' ότι πρόκειται για την ίδια μουσική σύνθεση, οι δύο εκτελέσεις είναι πολύ διαφοροποιημένες, ακόμα και στην περίπτωση που μπορεί να έχουν την ίδια τονικότητα και η σόλο εκτέλεση να μην περιέχει λάθη.

Συνήθως η σόλο εκτέλεση εμφανίζει και διάφορα προβλήματα, όπως για παράδειγμα τονικά λάθη, δηλαδή μεμονωμένα σημεία της μελωδίας που δεν είναι ίδια με την αρχική μουσική σύνθεση  $\Sigma$ . Επίσης, οι δύο εκτελέσεις της  $\Sigma$  έχουν σχεδόν με απόλυτη βεβαιότητα διαφορετικό ρυθμό παιχνιδιού. Είναι ακόμη πιθανό οι δύο εκτελέσεις να βρίσκονται σε διαφορετική τονικότητα, δηλαδή στο πεδίο της

συχνότητας το ίδιο τμήμα της σύνθεσης να απεικονίζεται σε διαφορετικές συχνότητες στη μία και την άλλη περίπτωση. Βέβαια, ακόμα και να βρίσκονται στην ίδια τονικότητα, μπορεί να βρίσκονται σε διαφορετικές οκτάβες.

Επιπλέον, η σόλο σύνθεση εκτός των παραπάνω προβλημάτων που αναφέραμε τα οποία είναι σχετικά με την εκτέλεση της σύνθεσης (υποκειμενικές) παρουσιάζει και ενδογενείς αδυναμίες και μειονεκτήματα σε σχέση με την εκτέλεση της συμφωνικής ορχήστρας. Κάποια από αυτά είναι:

- η έλλειψη βάθους
- η έλλειψη πλούτου
- φτωχότερο ηχόχρωμα

που είναι και αναμενόμενα διότι στην εκτέλεση από συμφωνική ορχήστρα έχουμε τη συμβολή πολλών και ποικίλων οργάνων, ενώ αντιθέτως στην περίπτωση της σόλο σύνθεσης, το βάθος, ο πλούτος και το ηχόχρωμα καθορίζονται από ένα και μόνο όργανο.

Εντούτοις, ο ακροατής και στις δύο περιπτώσεις αναγνωρίζει και ταυτοποιεί τη μουσική σύνθεση. Αυτό το γεγονός σημαίνει ότι υπάρχει ένας πυρήνας σε όλες τις διαφορετικές εκτελέσεις μιας μουσικής σύνθεσης, ο οποίος παραμένει αναλλοίωτος μετά την υποσυνείδητη επεξεργασία που υφίσταται στον ανθρώπινο εγκέφαλο. Αυτός ο πυρήνας εκτιμούμε ότι πρέπει να έχει μαθηματική υπόσταση είτε στο πεδίο του χρόνου είτε στο πεδίο της συχνότητας. Η παρούσα διπλωματική επιθυμεί να συμβάλει στον εντοπισμό αυτού του πυρήνα.

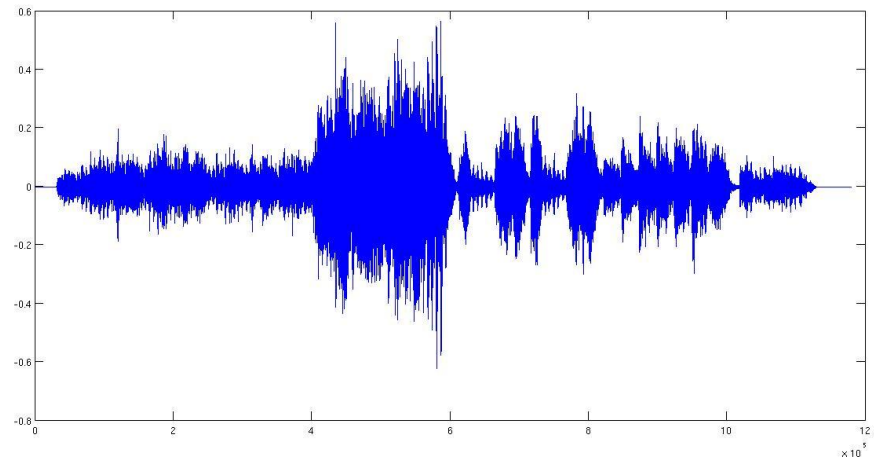
Στις παρακάτω ενότητες θα παρουσιαστούν παραδείγματα διαφορετικών εκτελέσεων της ίδιας μουσικής σύνθεσης για τρεις διαφορετικές μουσικές συνθέσεις τόσο στο πεδίο του χρόνου όσο και στο πεδίο της συχνότητας ώστε να καταστεί πιο σαφής η μεγάλη δυσκολία του προβλήματος. Επιπλέον, για καθεμία από τις μουσικές συνθέσεις θα παραθέσουμε διαγράμματα που αφορούν όλο το σήμα, καθώς και τμήματα του σήματος που είναι κοινά ακουστικά στις διάφορες εκτελέσεις.

### 3.1 Μουσική σύνθεση: Rondo Allegro από το K525

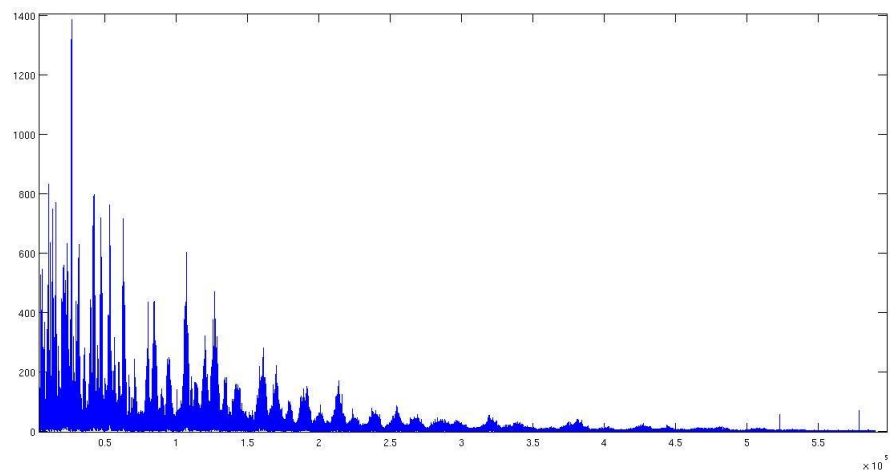
#### 3.1.1 Όλο το σήμα

Α) Εκτέλεση από ορχήστρα

Πεδίο του χρόνου

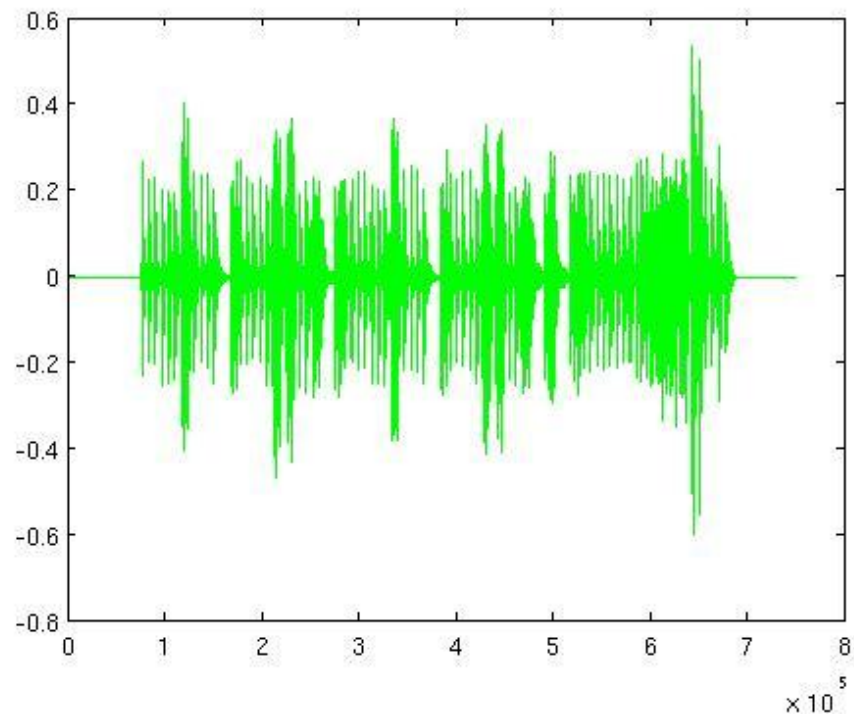


Πεδίο της συχνότητας

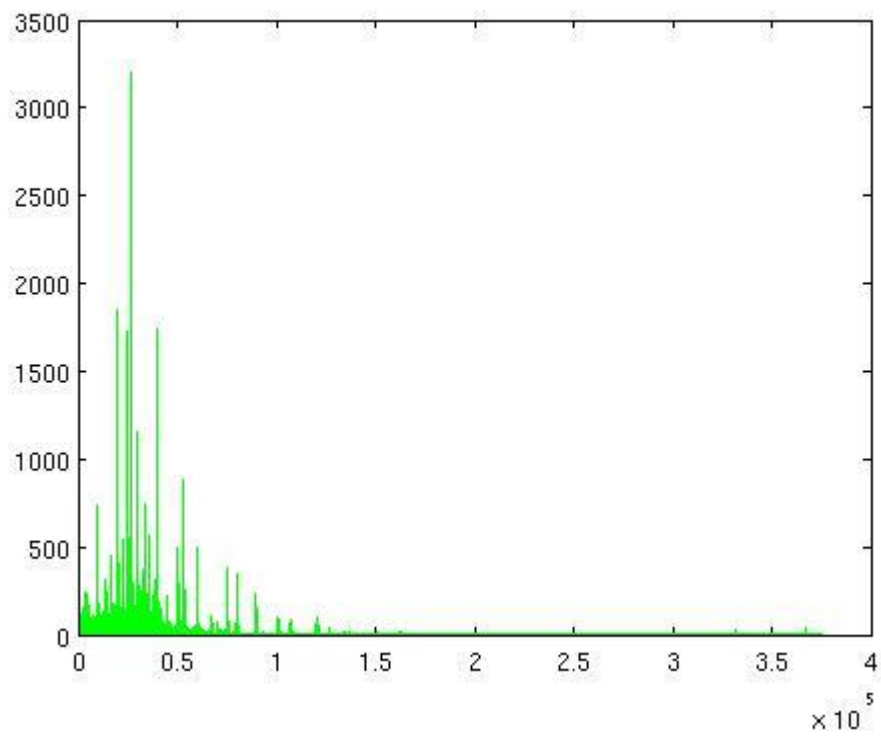


Β) Εκτέλεση από σόλο πιάνο στην ίδια τονικότητα χωρίς λάθη

Πεδίο του χρόνου

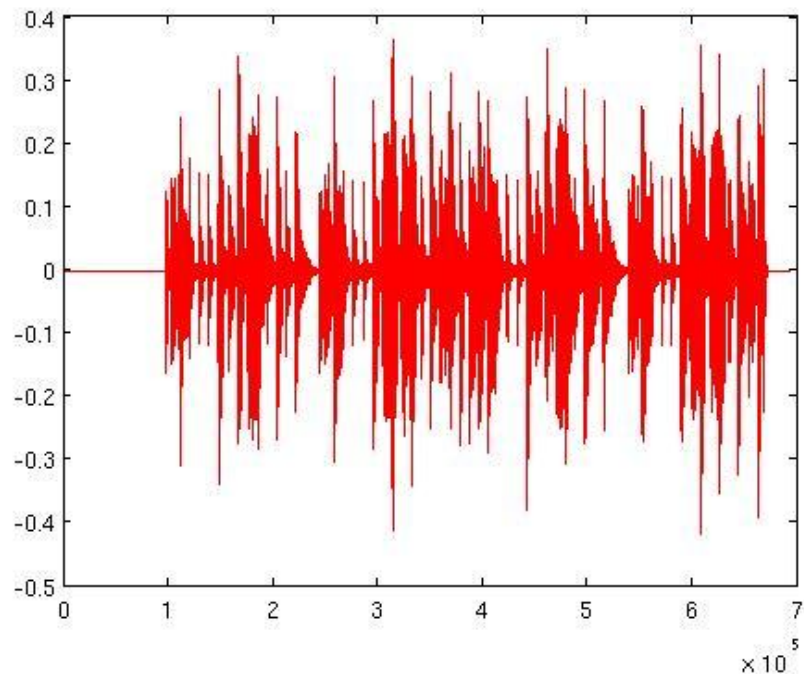


Πεδίο της συχνότητας

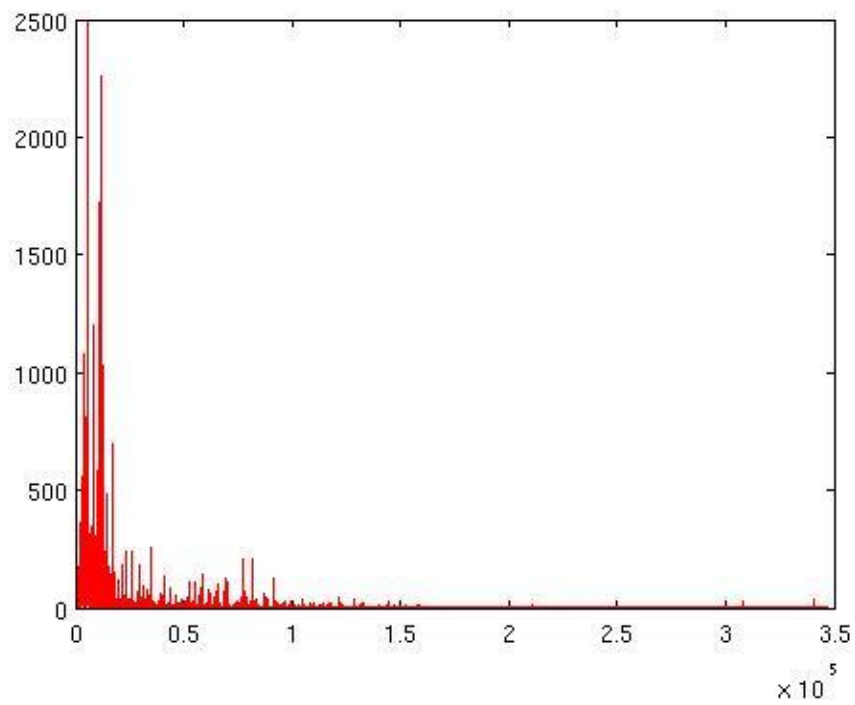




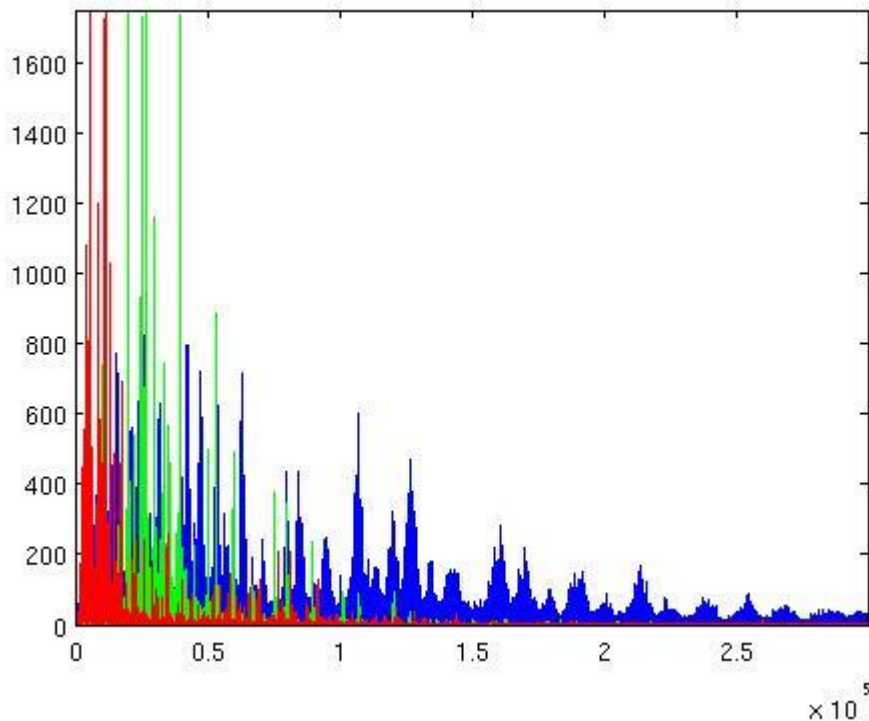
Γ) Εκτέλεση από σόλο κιθάρά με αλλαγή στο ρυθμό και κάποια λάθη  
Πεδίο του χρόνου



Πεδίο της συχνότητας



#### Δ) Παράθεση όλων των DFT σε κοινό διάγραμμα



Από τα προηγούμενα διαγράμματα συμπεραίνουμε ότι οι διαφοροποιήσεις μεταξύ των διαφορετικών εκτελέσεων της ίδιας μουσικής σύνθεσης είναι πολύ μεγάλες τόσο στο πεδίο του χρόνου όσο και στο πεδίο της συχνότητας.

Για παράδειγμα, από το διάγραμμα του DFT της πρότυπης εκτέλεσης (δηλαδή της εκτέλεσης από ορχήστρα) φαίνεται ο πλούτος του ακούσματος, καθώς παρατηρούμε ότι ο DFT του σήματος είναι κατανομημένος σε αρκετά μεγάλο εύρος συχνοτήτων. Επίσης έχει πολύ μεγαλύτερο αριθμό κορυφών και εν γένει διαφορετικού πλάτους, ακόμα και ανηγμένου, από ότι οι κορυφές του πιάνου ή της κιθάρας.

Όσον αφορά την εκτέλεση από σόλο πιάνο, παρατηρούμε ότι ο DFT του σήματος κατανέμεται σε πάρα πολύ μικρό εύρος συχνοτήτων, ενώ ταυτόχρονα περιλαμβάνει πολύ μικρότερο πλήθος αρμονικών. Παρατηρούμε επίσης ότι έχει πολύ λιγότερες κορυφές από την πρότυπη εκτέλεση αλλά κάποιες από αυτές έχουν αξιοσημείωτο πλάτος.

Αντίστοιχα, στην εκτέλεση από κιθάρα παρατηρούμε και πάλι ότι ο Διακριτός Μετασχηματισμός Fourier του σήματος εκτείνεται σε πολύ μικρό εύρος συχνοτήτων και μάλιστα σε ακόμη πιο χαμηλές συχνότητες από το πιάνο.

Γενικά θα θέλαμε να τονίσουμε ότι το εύρος των συχνοτήτων κατά την απόδοση μιας μουσικής σύνθεσης επηρεάζεται αρκετά από τις δυνατότητες του οργάνου.

Ακόμη είναι φανερό ότι οι θέσεις όπου βρίσκονται οι κορυφές και στις τρεις εκτελέσεις διαφέρουν. Παρ' όλες αυτές τις έντονες διαφοροποιήσεις, υπάρχει ένας κοινός πυρήνας ο οποίος αναζητάμε να βρούμε.

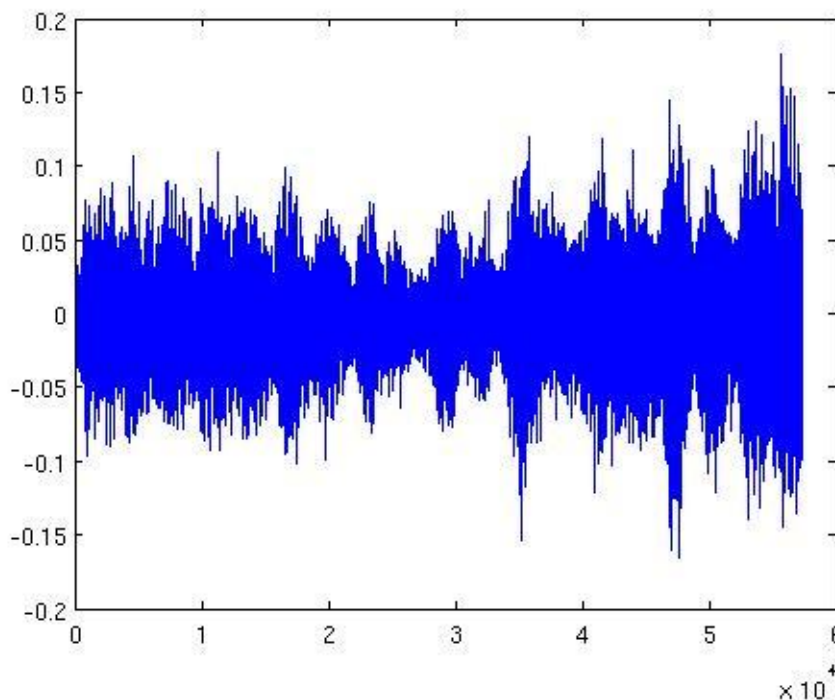
### 3.1.2 Τμήμα των σημάτων που ταιριάζει ακουστικά

Στα παρακάτω διαγράμματα έχουμε εντοπίσει ένα τμήμα 2,5 δευτερολέπτων του σήματος της πρότυπης εκτέλεσης που ταιριάζει ακουστικά επακριβώς με ένα αντίστοιχο τμήμα της εκτέλεσης από σόλο πιάνο και το ίδιο έχουμε κάνει και για την κιθάρα με ένα άλλο τμήμα της ίδια διάρκειας. Αυτά τα διαγράμματα μπορούν να γίνουν οδηγός ώστε να εξάγουμε ακόμα περισσότερες πληροφορίες και συμπεράσματα γιατί πλέον διαθέτουμε δύο τμήματα σημάτων που αντιστοιχούν μουσικά επακριβώς.

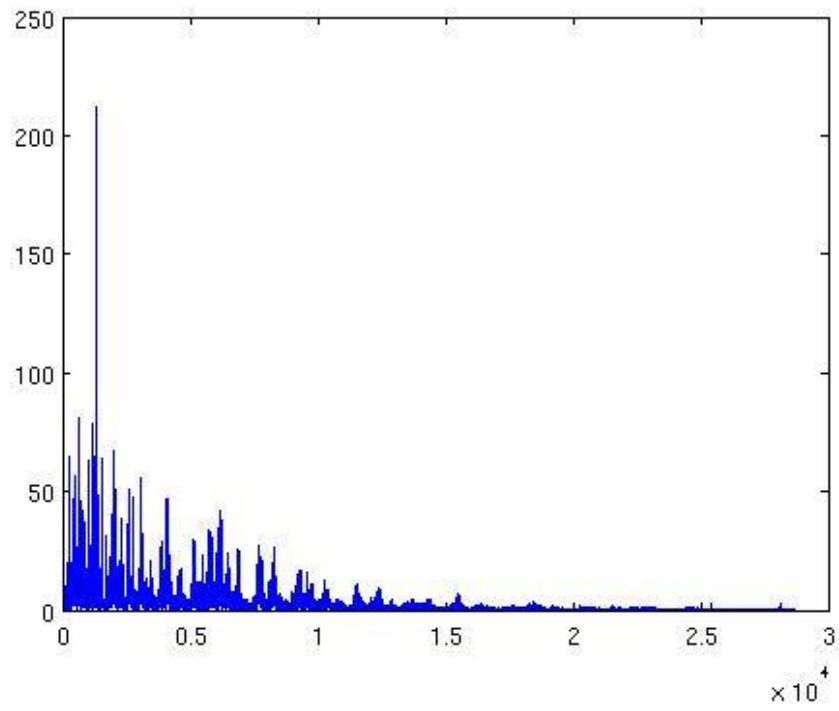
#### Κοινό τμήμα ορχήστρα-πιάνο

A) Εκτέλεση από ορχήστρα

Πεδίο του χρόνου

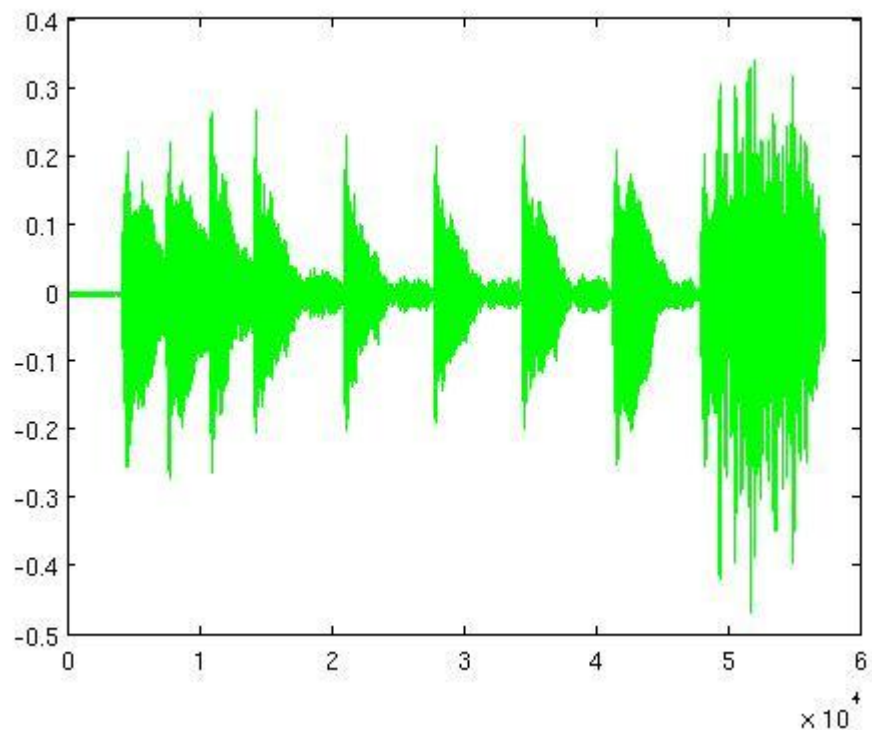


Πεδίο της συχνότητας

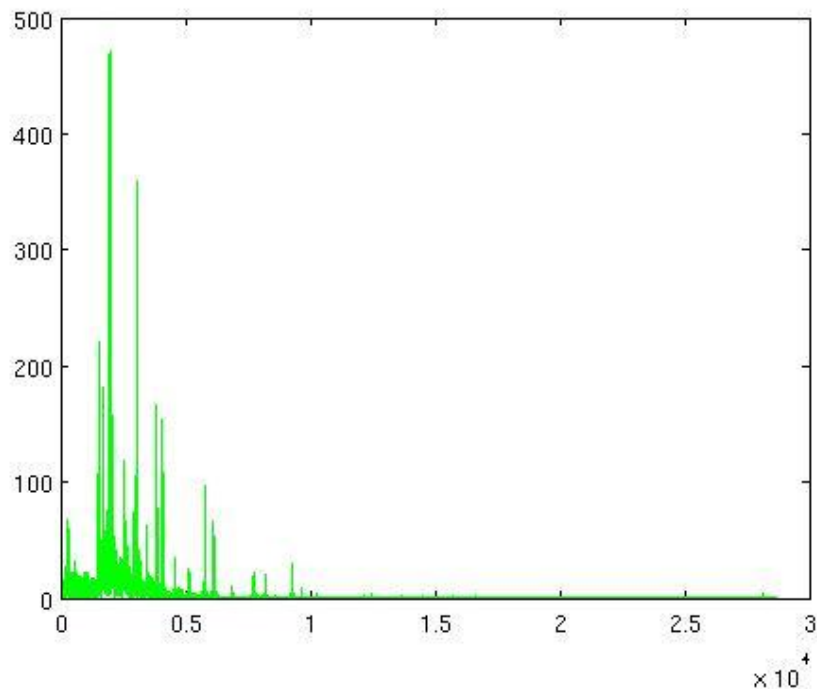


Β) Σόλο πιάνο στην ίδια τονικότητα

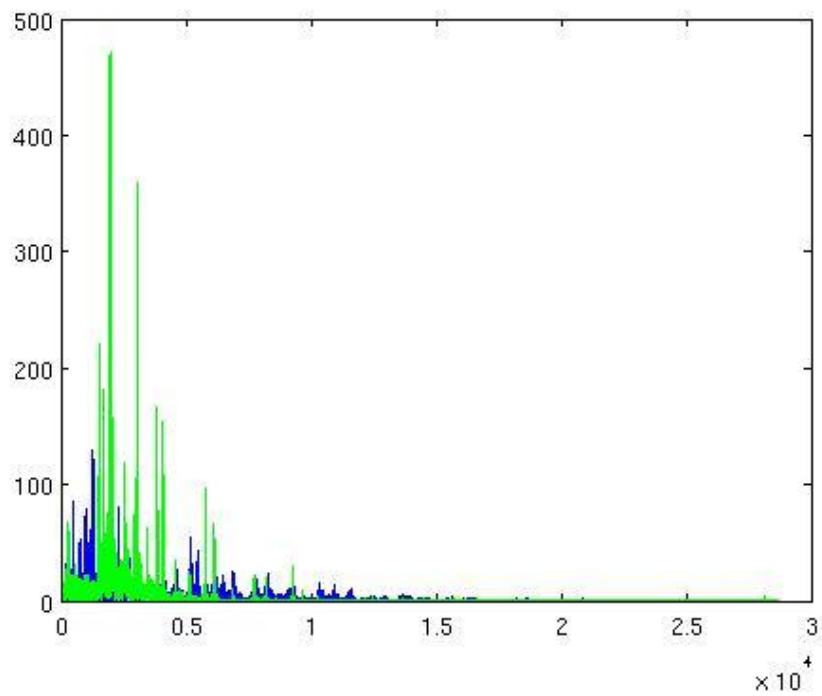
Πεδίο του χρόνου



Πεδίο της συχνότητας



Γ) Παράθεση των DFT ορχήστρας-πιάνο σε κοινό διάγραμμα



#### Δ) Παρατηρήσεις

Παρατηρούμε και σε αυτήν την περίπτωση όπου το τμήμα του σήματος της πρότυπης εκτέλεσης και αυτό του σήματος της εκτέλεσης από σόλο πιάνο αντιστοιχούν στο ίδιο ακουστικά τμήμα της σύνθεσης ότι οι διαφορές τόσο στο πεδίο του χρόνου όσο και στο πεδίο της συχνότητας είναι πολύ μεγάλες.

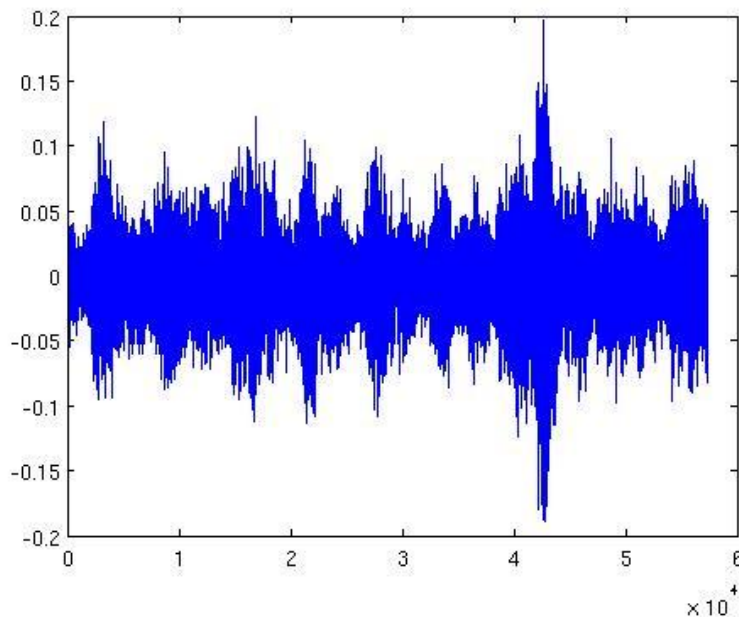
Ο αριθμός δειγμάτων που αντιστοιχούν σε αυτά τα τμήματα των 2,5 δευτερολέπτων είναι  $N=56*1024$  δείγματα. Στο πεδίο του χρόνου βλέπουμε ότι χαρακτηριστικές ποσότητες για κάθε παράθυρο μεγέθους  $N$ , όπως ο αριθμός των μηδενισμών (zero-crossings), η θέση των κορυφών καθώς και το πλάτος τους έχουν τεράστιες διαφορές ανάμεσα στα δύο σήματα.

Αντίστοιχα, στο πεδίο της συχνότητας παρατηρούμε ότι διαφέρει ο αριθμός των κορυφών μεταξύ της πρότυπης εκτέλεσης και της σόλο εκτέλεσης από πιάνο, που είναι σημαντικά περισσότερες στην πρώτη περίπτωση. Επίσης το πλάτος των κορυφών στην περίπτωση της εκτέλεσης από πιάνο είναι εν γένει διαφορετικό από αυτό των κορυφών της πρότυπης εκτέλεσης.

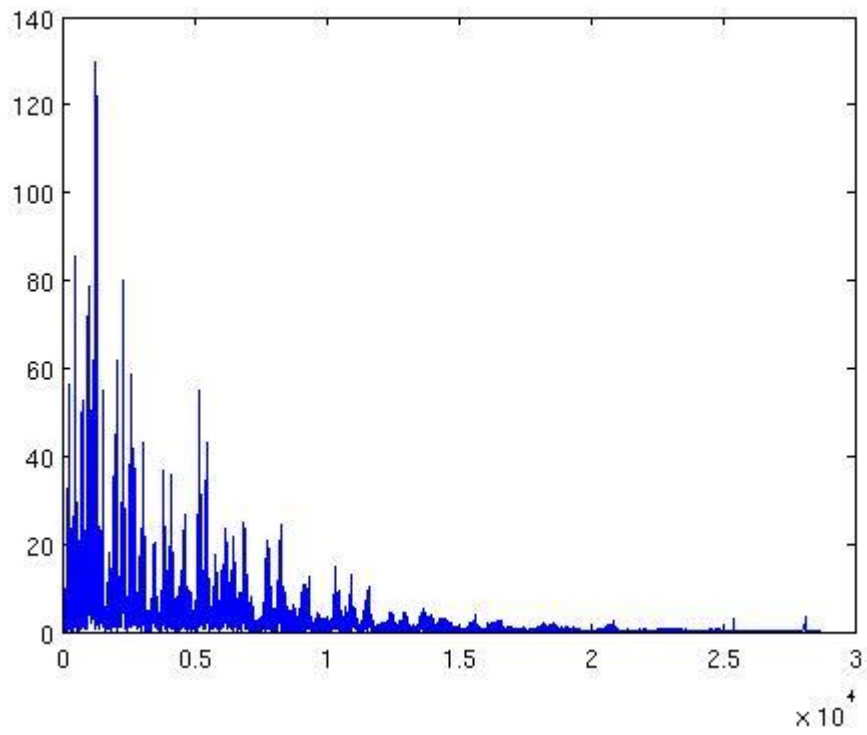
#### Κοινό τμήμα ορχήστρα-κιθάρα

##### Α) Εκτέλεση από ορχήστρα

Πεδίο του χρόνου

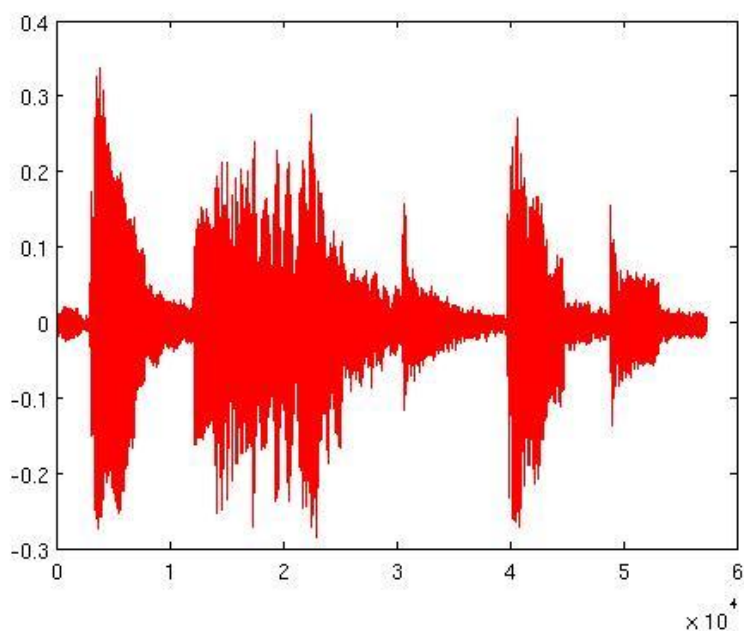


Πεδίο της συχνότητας

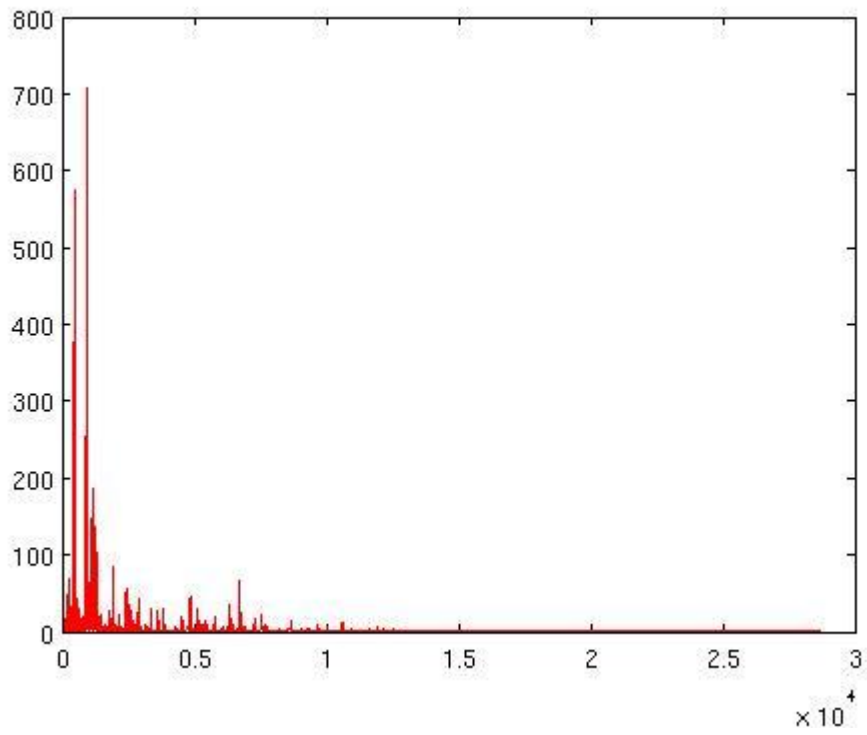


Β) Σόλο κιθάρα στην ίδια τονικότητα με πιο αργό ρυθμό

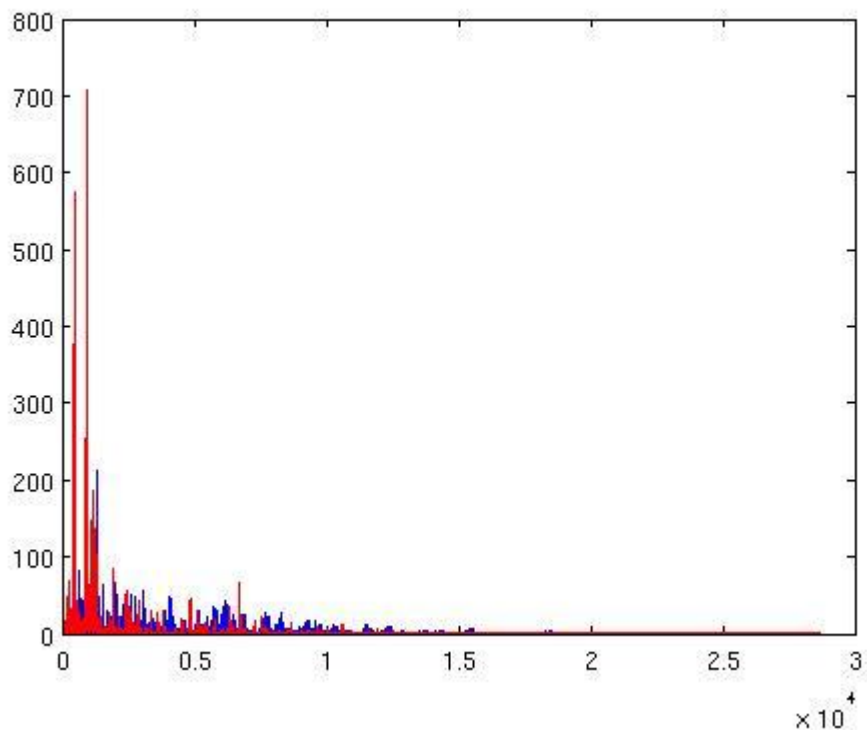
Πεδίο του χρόνου



Πεδίο της συχνότητας



Γ) Παράθεση των DFT ορχήστρας-κιθάρας σε κοινό διάγραμμα





Δ) Παρατηρήσεις

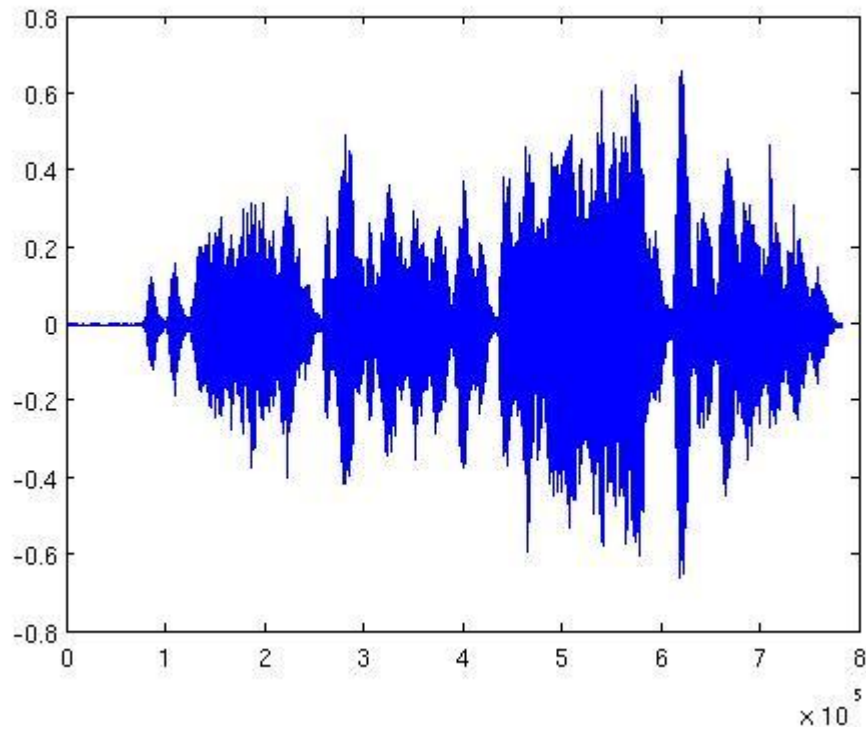
Αντίστοιχα με το κοινό τμήμα μεταξύ ορχήστρας και πιάνο, διαπιστώνουμε και εδώ ότι ισχύουν οι ίδιες παρατηρήσεις.

### 3.2 Μουσική σύνθεση: *Romanze Andante* από το K525

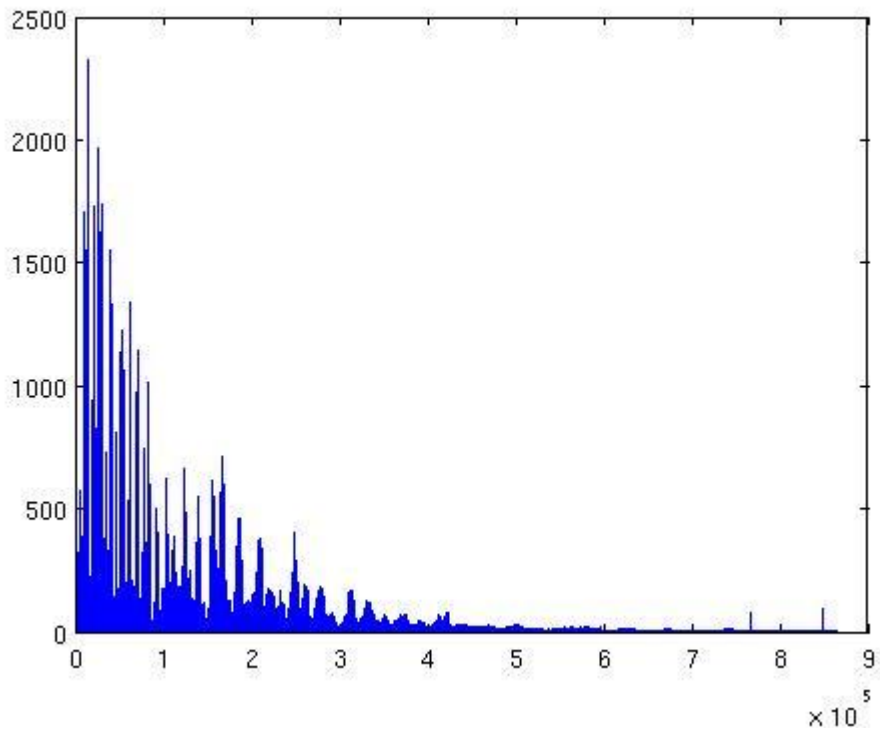
#### 3.2.1 Όλο το σήμα

Α) Εκτέλεση από ορχήστρα

Πεδίο του χρόνου

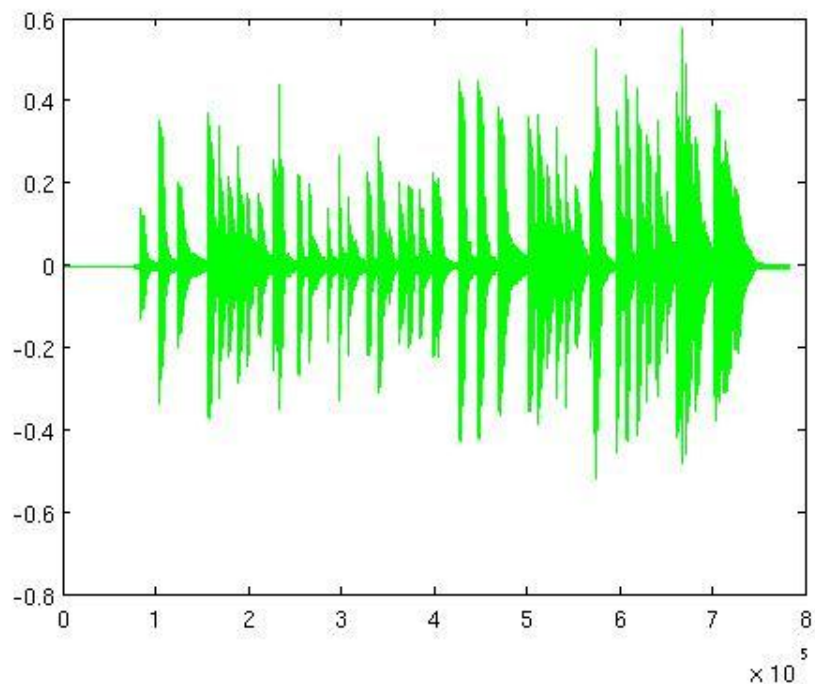


Πεδίο της συχνότητας

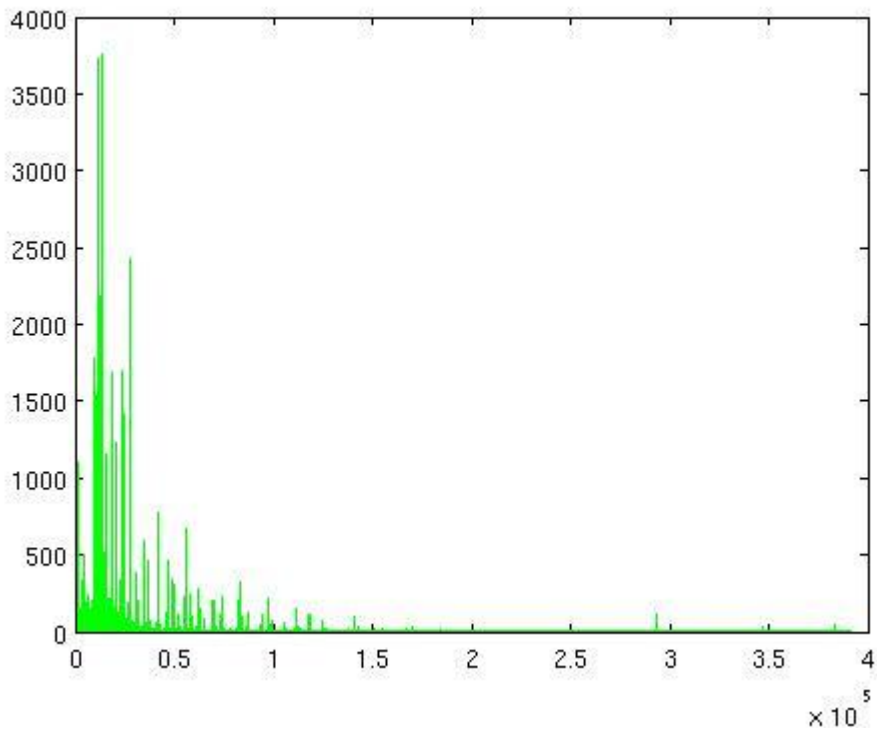


Β) Εκτέλεση από σόλο πιάνο με αλλαγή οκτάβας

Πεδίο του χρόνου

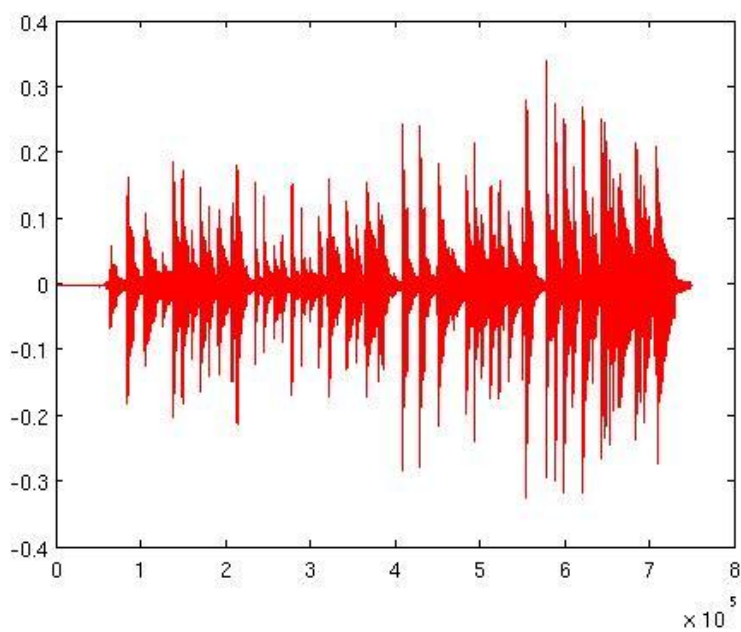


Πεδίο της συχνότητας

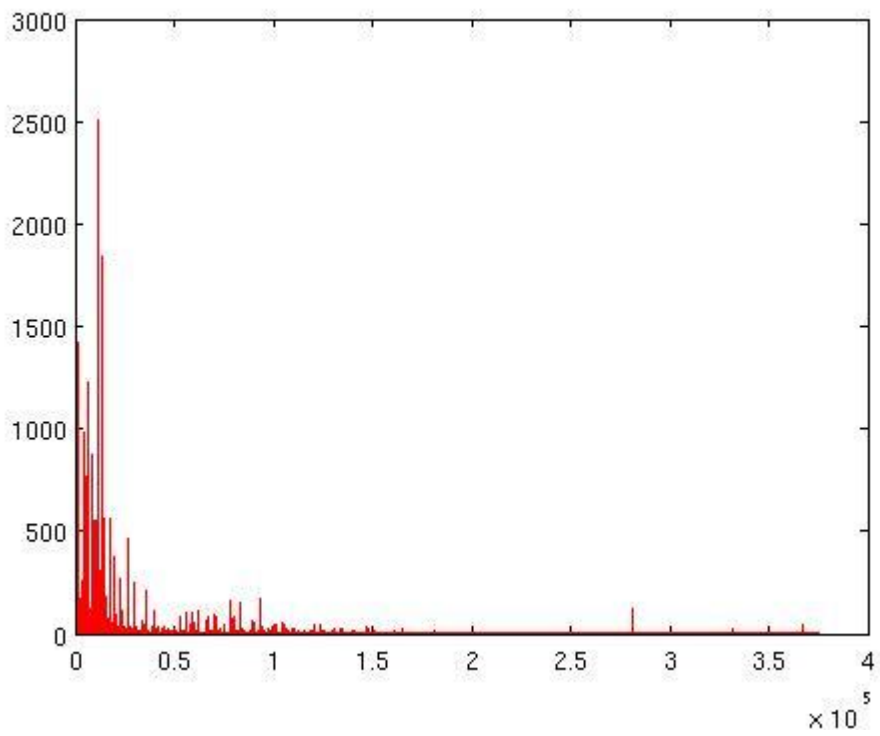


Γ) Εκτέλεση από σόλο κλασική κιθάρα με αλλαγή ρυθμού και σε άλλη τονικότητα

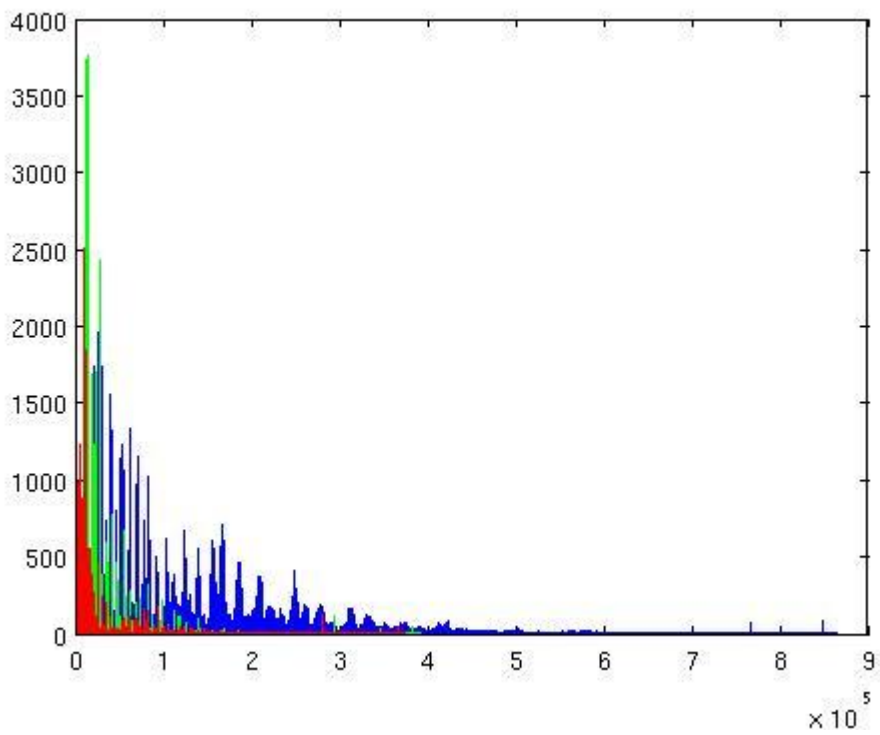
Πεδίο του χρόνου



Πεδίο της συχνότητας



Δ) Παράθεση όλων των DFT σε κοινό διάγραμμα



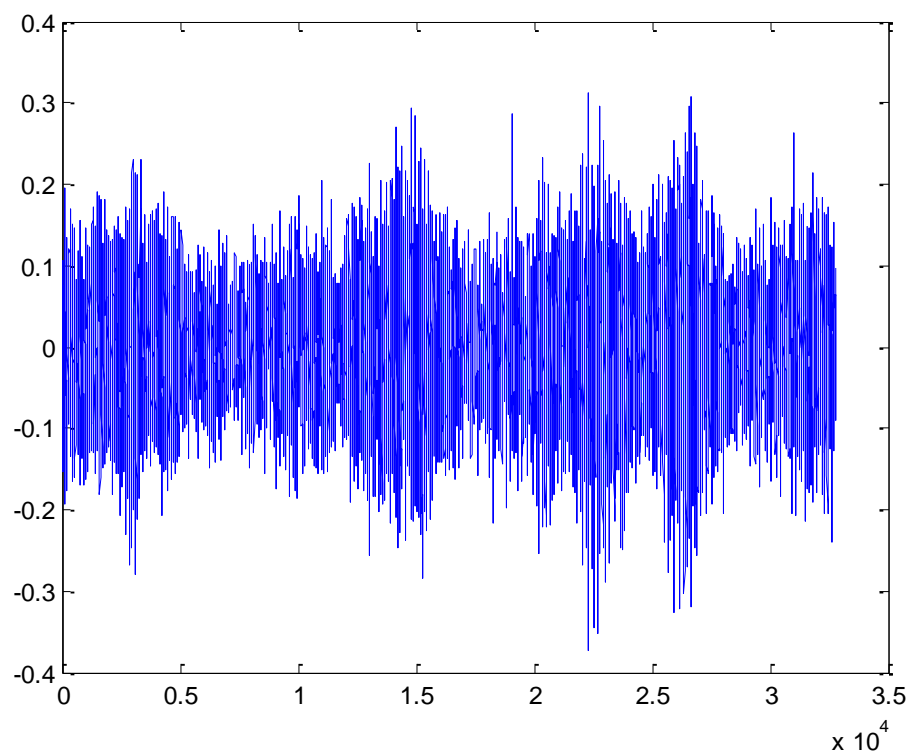
Οι ίδιες παρατηρήσεις ισχύουν και σε αυτή την ενότητα.

### 3.2.2 Τμήμα των σημάτων που ταιριάζει ακουστικά

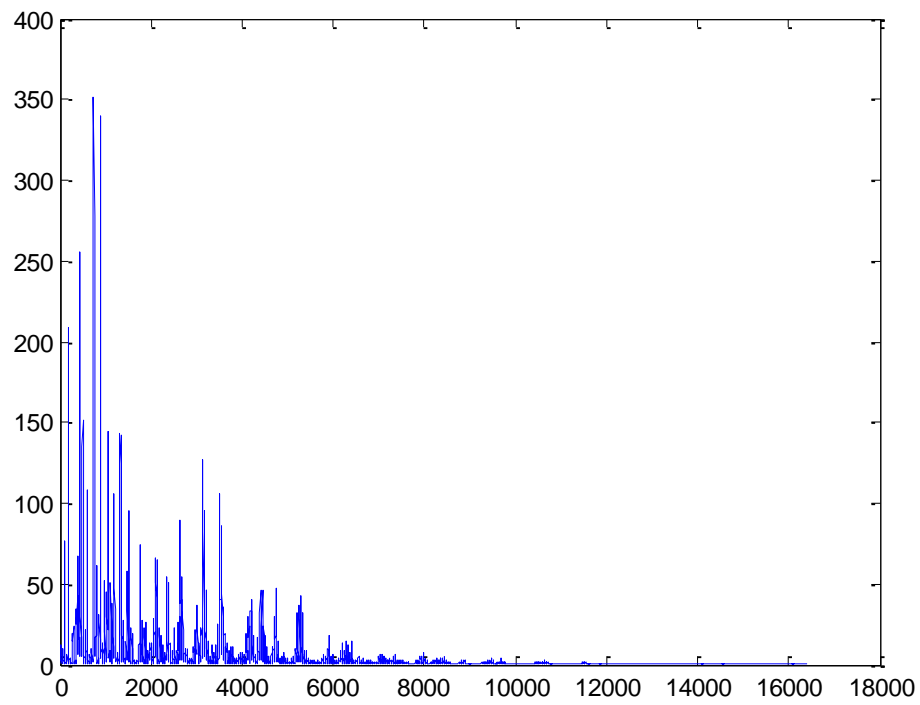
#### Κοινό τμήμα ορχήστρα-πιάνο

Α) Εκτέλεση από ορχήστρα

Πεδίο του χρόνου

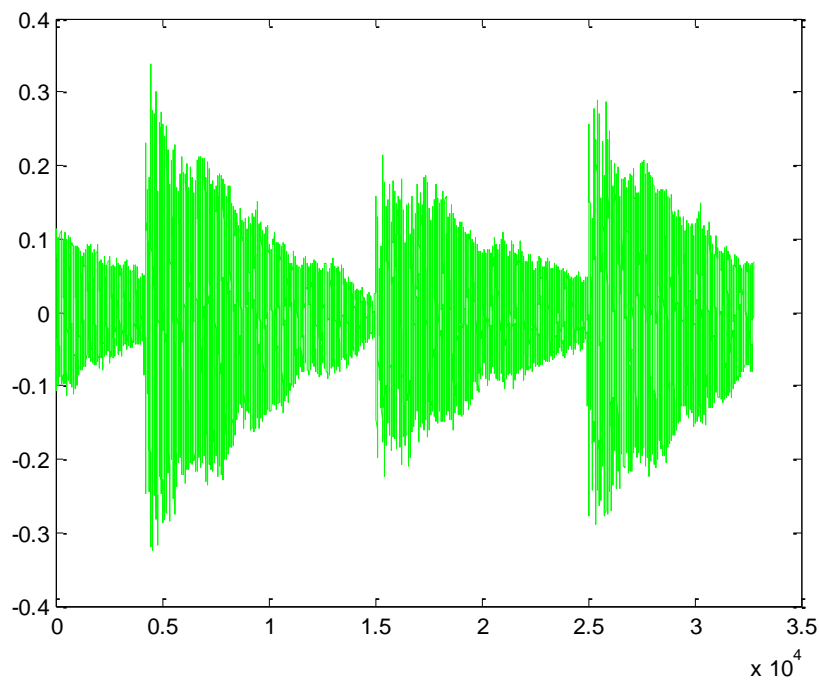


Πεδίο της συχνότητας

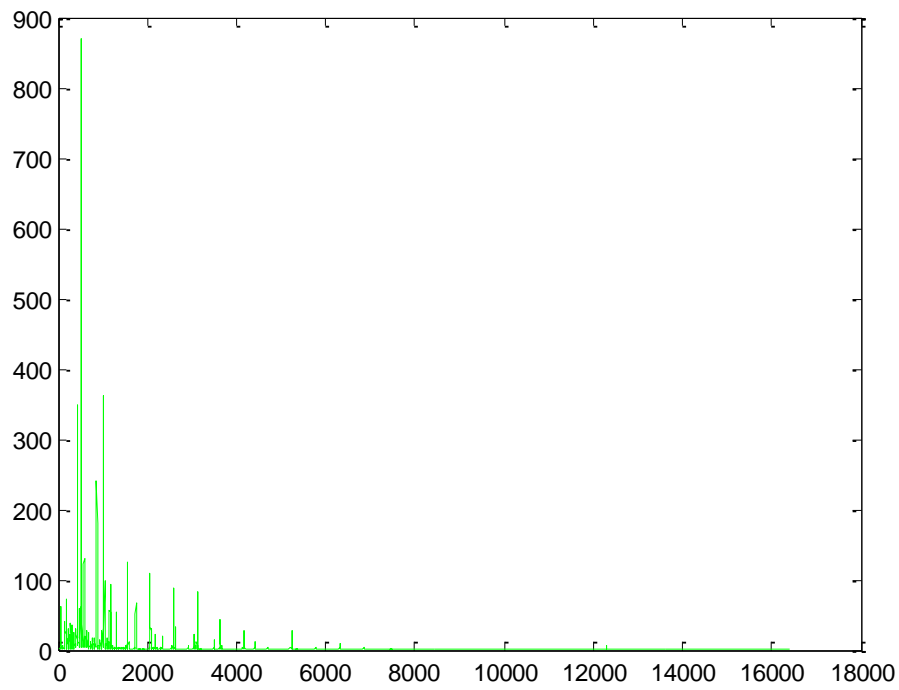


Β) Σόλο πιάνο

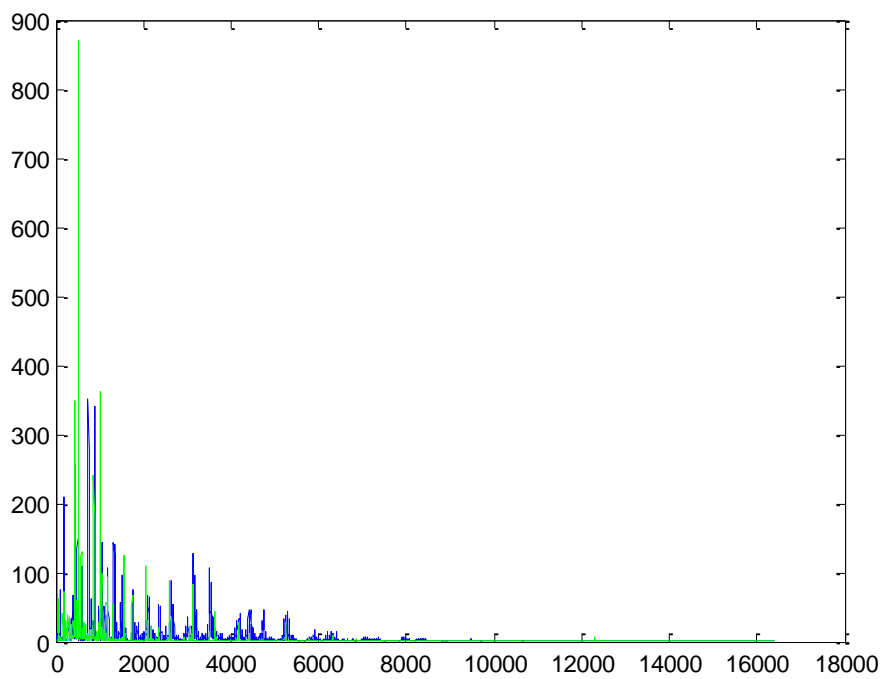
Πεδίο του χρόνου



### Πεδίο της συχνότητας



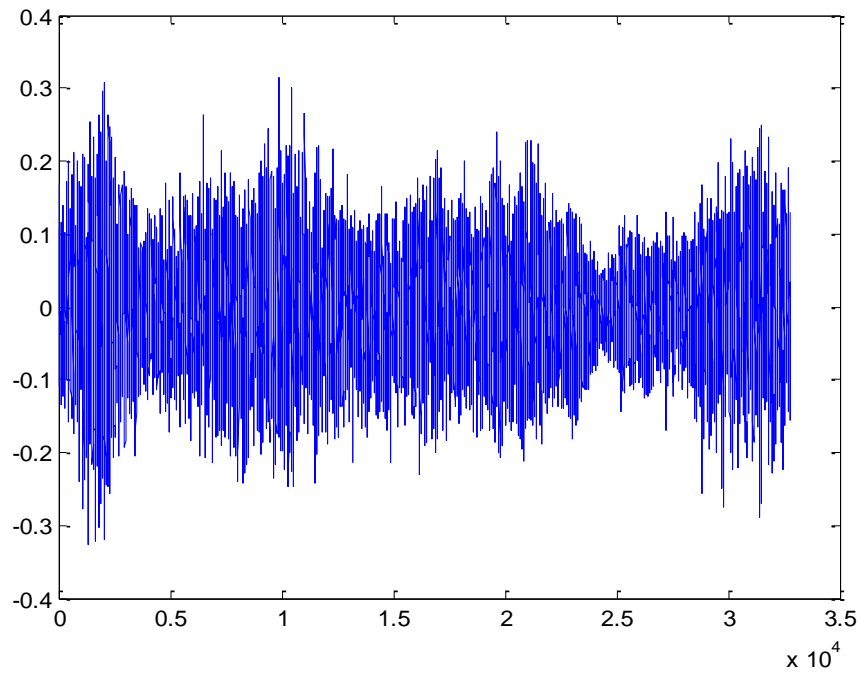
### Γ) Παράθεση των DFT ορχήστρας-πιάνο σε κοινό διάγραμμα



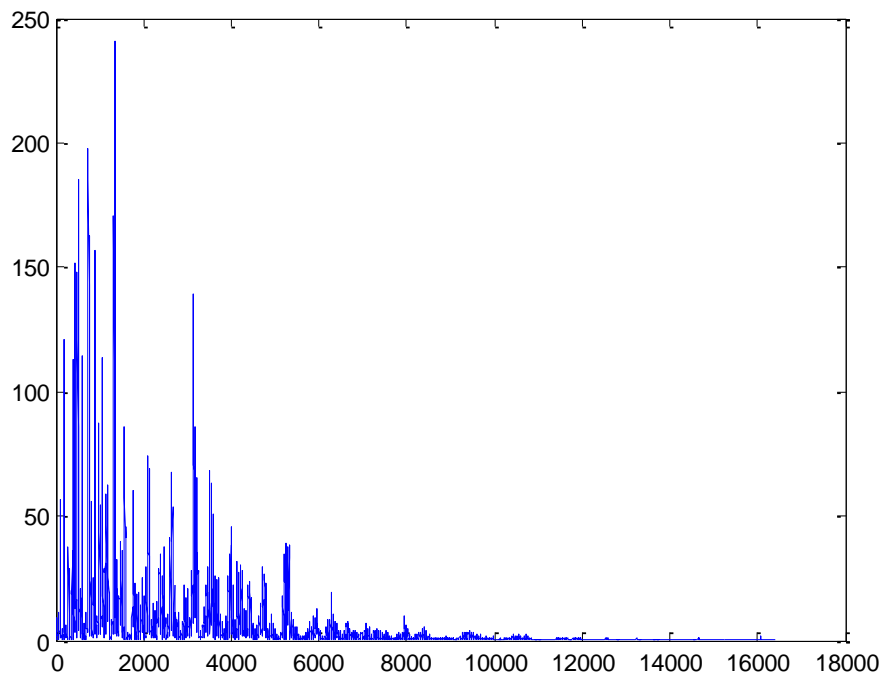
## Κοινό τμήμα ορχήστρα- κιθάρα

A) Εκτέλεση από ορχήστρα

Πεδίο του χρόνου



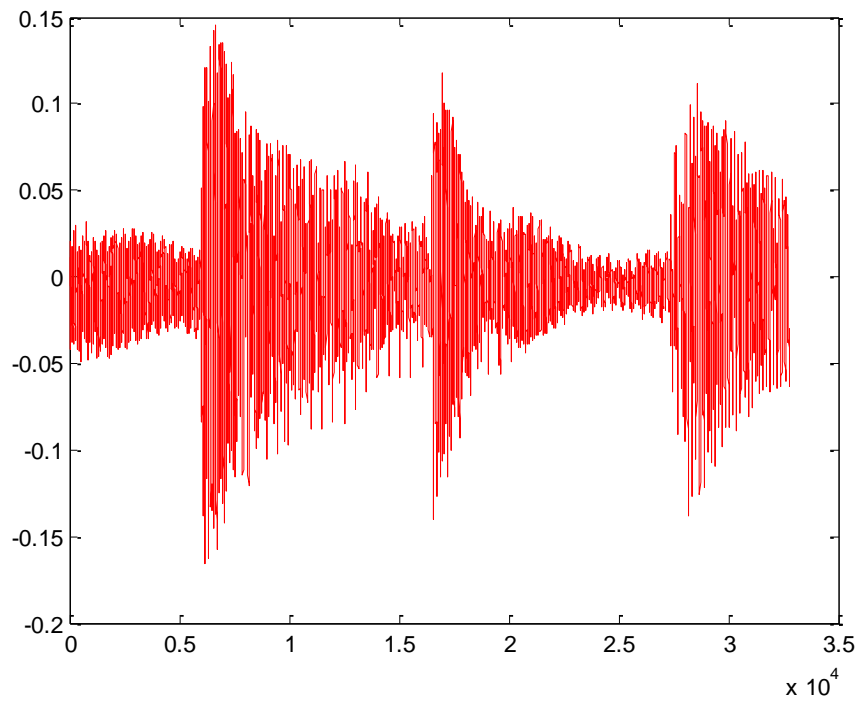
Πεδίο της συχνότητας



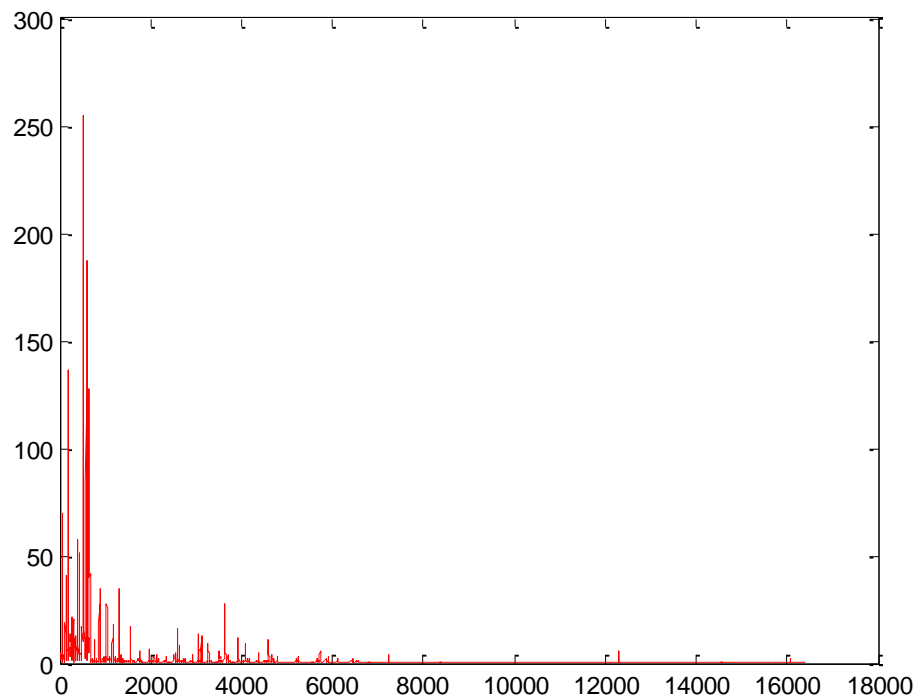


## Β) Σόλο κιθάρα

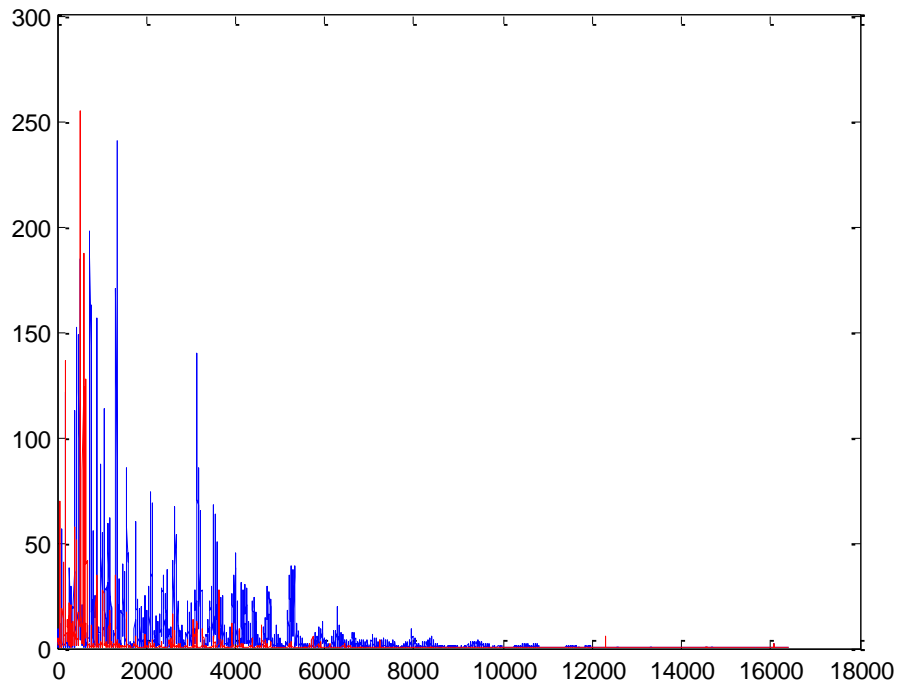
Πεδίο του χρόνου



Πεδίο της συχνότητας



### Γ) Παράθεση των DFT ορχήστρας-κιθάρας σε κοινό διάγραμμα



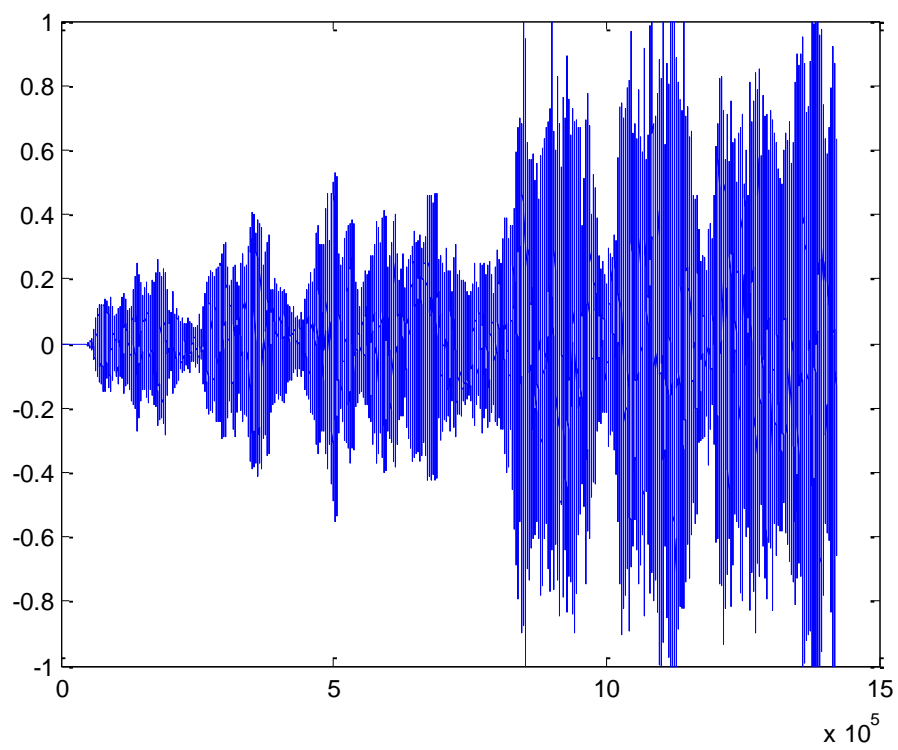
Παρατηρούμε και σε αυτή την ενότητα ότι οι βασικές παρατηρήσεις είναι οι ίδιες, δηλαδή οι διαφορές μεταξύ των ακουστικά ίδιων τμημάτων των σημάτων της ορχήστρας και του πιάνο ή της ορχήστρας και της κιθάρας είναι πολύ μεγάλες. Παρ' όλα αυτά αν παρατηρήσουμε τους Διακριτούς μετασχηματισμούς Fourier των σημάτων αυτών, σε κάποιες συχνότητες βλέπουμε ότι συμπίπτουν κάποιες μεγάλες σε πλάτος κορυφές της ορχήστρας με κάποιες μεγάλες σε πλάτος κορυφές του πιάνο ή της κιθάρας. Αυτή είναι και η βασική ιδέα της προσέγγισης που ακολουθήσαμε.

### 3.3 Μουσική σύνθεση: Adagio από το K622 (Concerto for Clarinet)

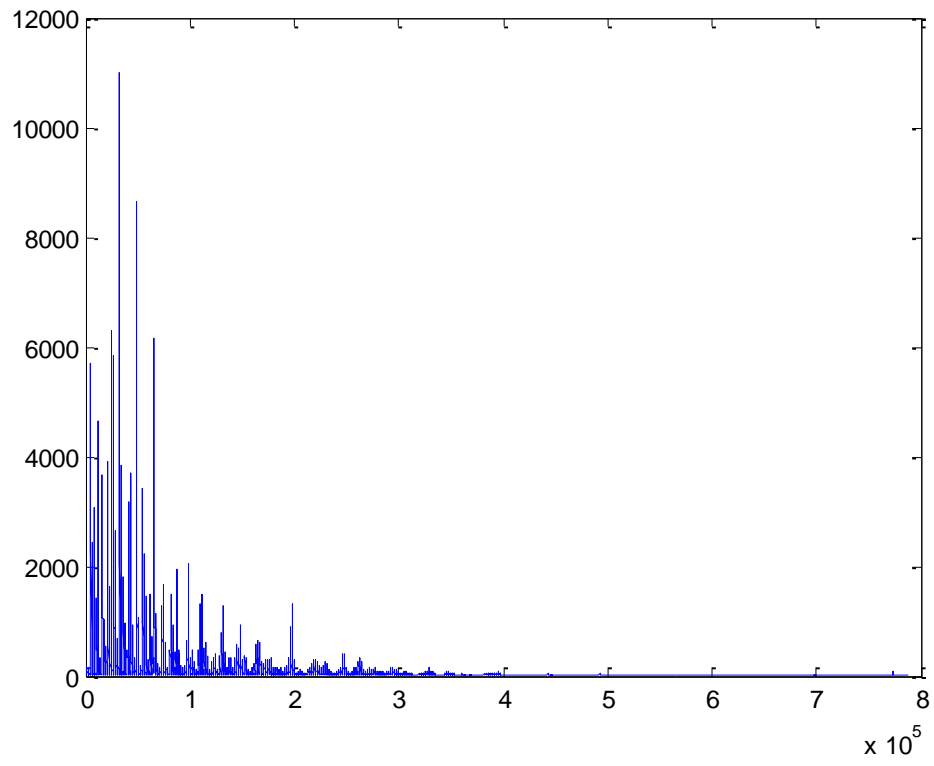
#### 3.3.1 Όλο το σήμα

A) Πρότυπη εκτέλεση

Πεδίο του χρόνου

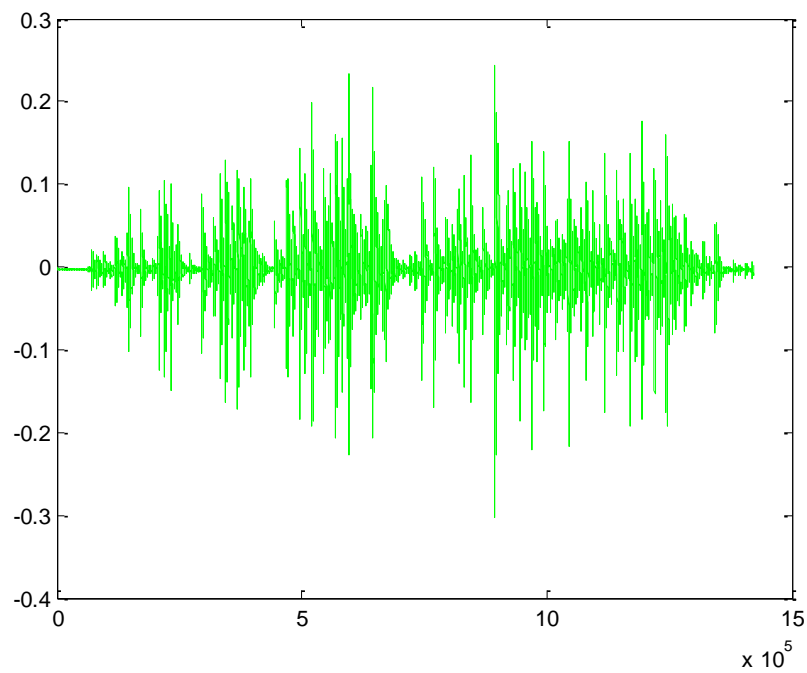


Πεδίο της συχνότητας

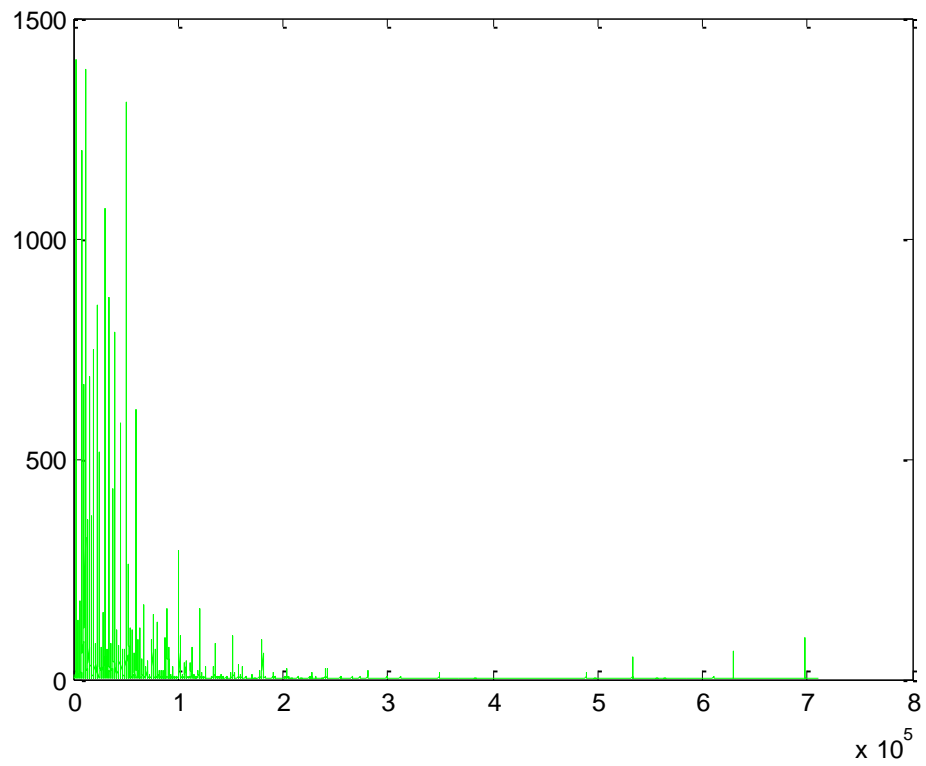


Β) Σόλο πιάνο με αλλαγή του ρυθμού

Πεδίο του χρόνου

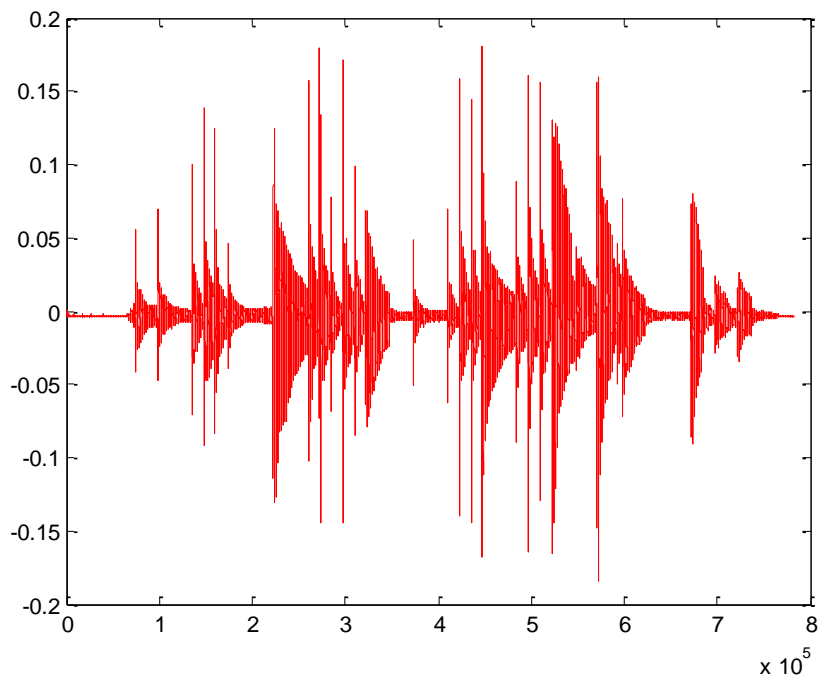


Πεδίο της συχνότητας

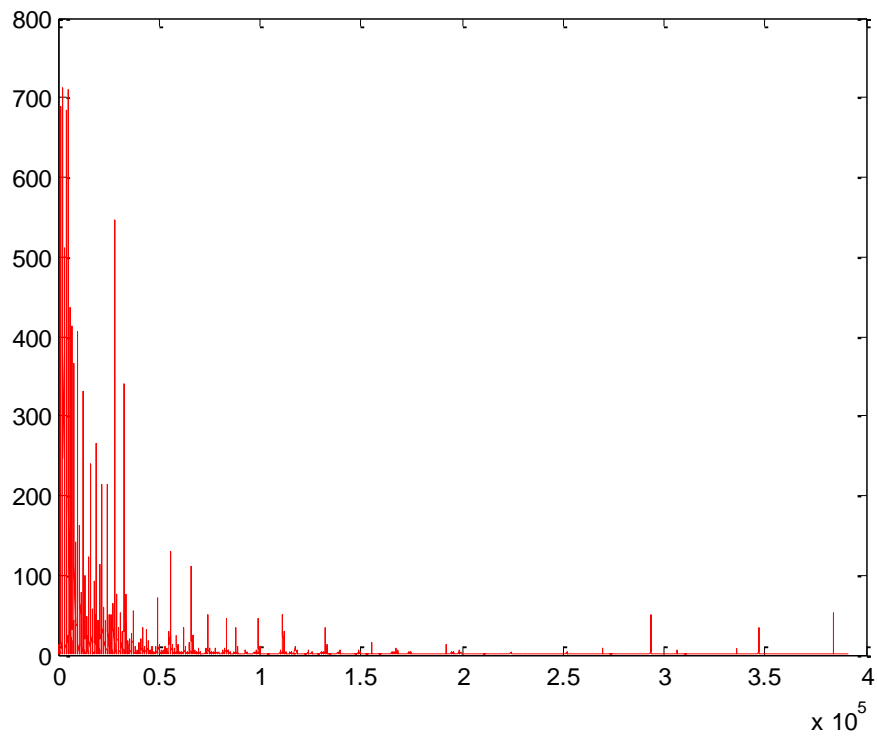


Γ) Σόλο κιθάρα

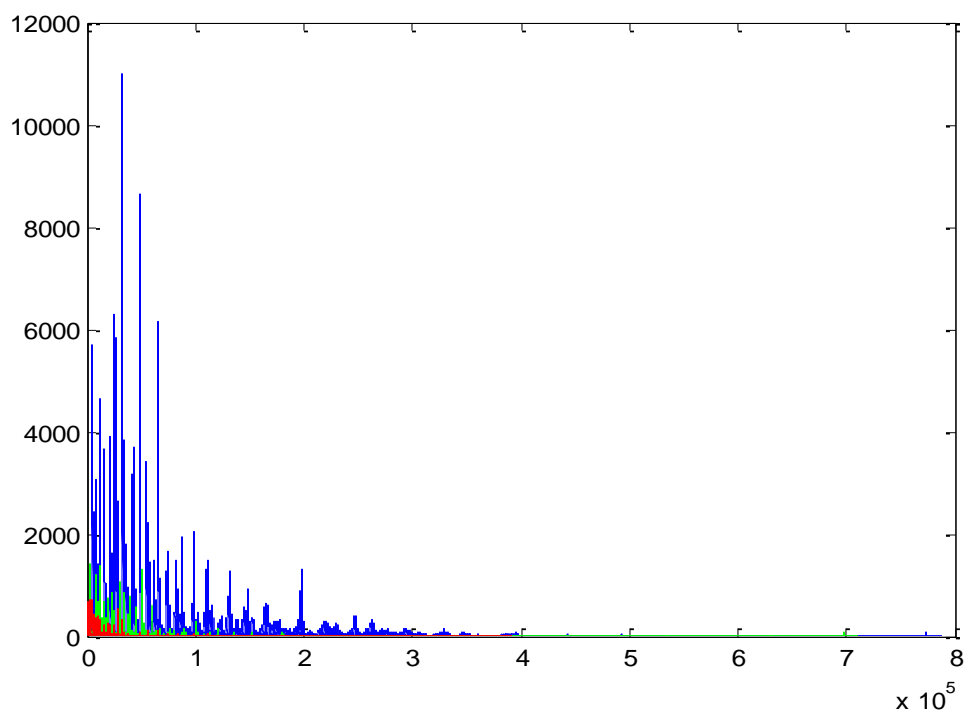
Πεδίο του χρόνου



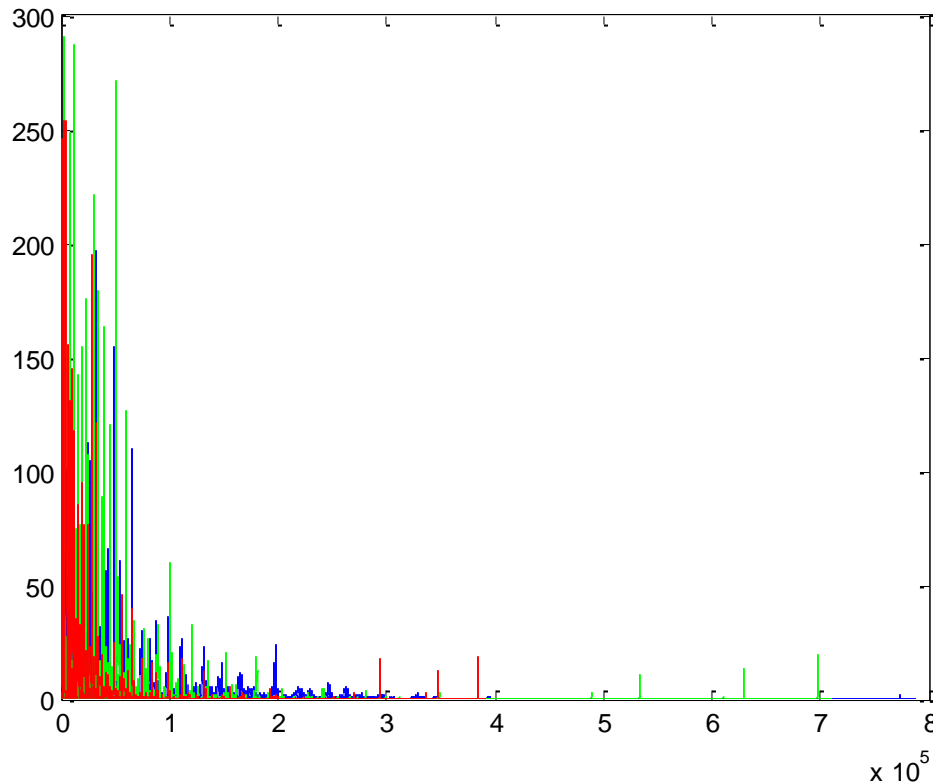
Πεδίο της συχνότητας



Δ) Παράθεση όλων των DFT (μη ανηγμένων) σε κοινό διάγραμμα



Το παρακάτω διάγραμμα είναι οι Διακριτοί Μετασχηματισμοί Fourier των σημάτων όλων των εκτελέσεων αλλά ανηγμένοι στους μέσους όρους των πλατών του κάθε φάσματος:



Από το διάγραμμα αυτό μπορούμε να παρατηρήσουμε ότι αρκετές κορυφές σημαντικού πλάτους τόσο της κιθάρας όσο και του πιάνου συμπίπτουν ή βρίσκονται πολύ κοντά με σημαντικού πλάτους κορυφές της εκτέλεσης από ορχήστρα. Θα δούμε σε παρακάτω κεφάλαιο πως μπορούμε να αντιμετωπίσουμε τις κορυφές που λόγω του διαφορετικού ρυθμού μετατοπίζονται σε άλλες συχνότητες πάνω στο φάσμα ( βλ. η τεχνική του stretch, ενότητα 4.4).

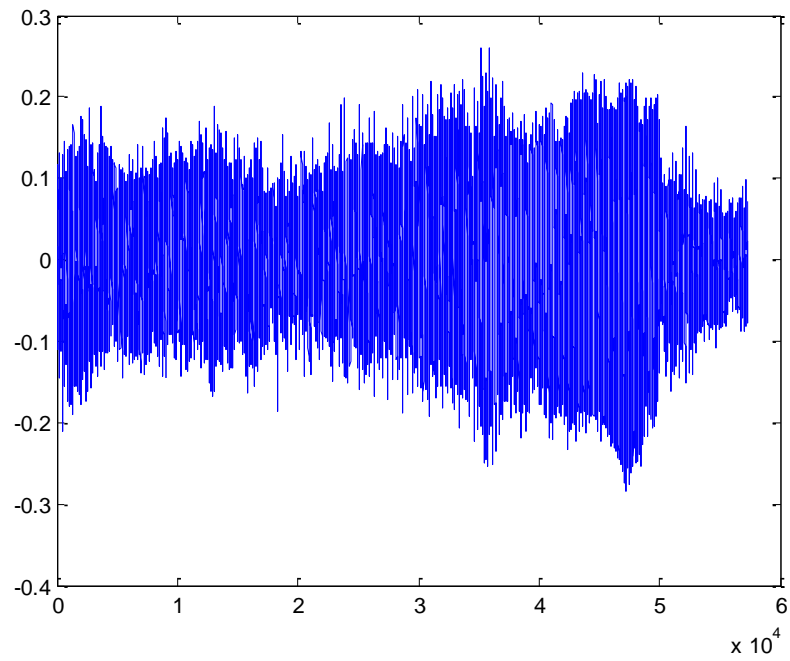
Εδώ παρατηρούμε, επίσης, ότι τόσο η εκτέλεση από πιάνο όσο και η εκτέλεση από κιθάρα έχουν πολύ πιο πλούσιο συχνοτικό περιεχόμενο σε σχέση με τα προηγούμενα δύο κομμάτια που παρουσιάστηκαν στις ενότητες 3.1 και 3.2. Αυτό επιβεβαιώνει την εκτίμησή μας ότι το πρόβλημα προς αντιμετώπιση είναι πολύ σύνθετο και δύσκολο.

### 3.3.1 Τμήμα των σημάτων που ταιριάζει ακουστικά

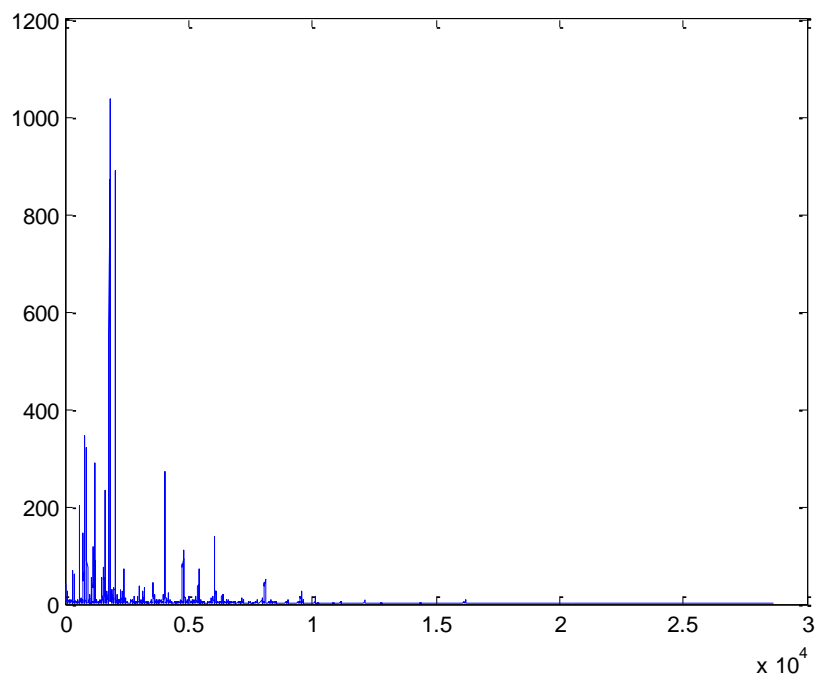
#### Κοινό τμήμα ορχήστρα-πιάνο

Α) Εκτέλεση από ορχήστρα

Πεδίο του χρόνου



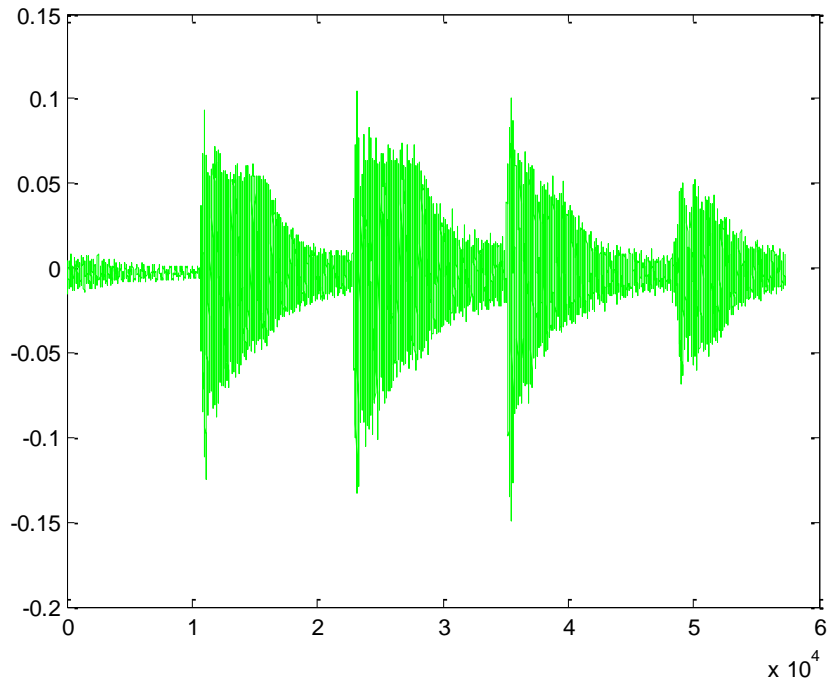
Πεδίο της συχνότητας



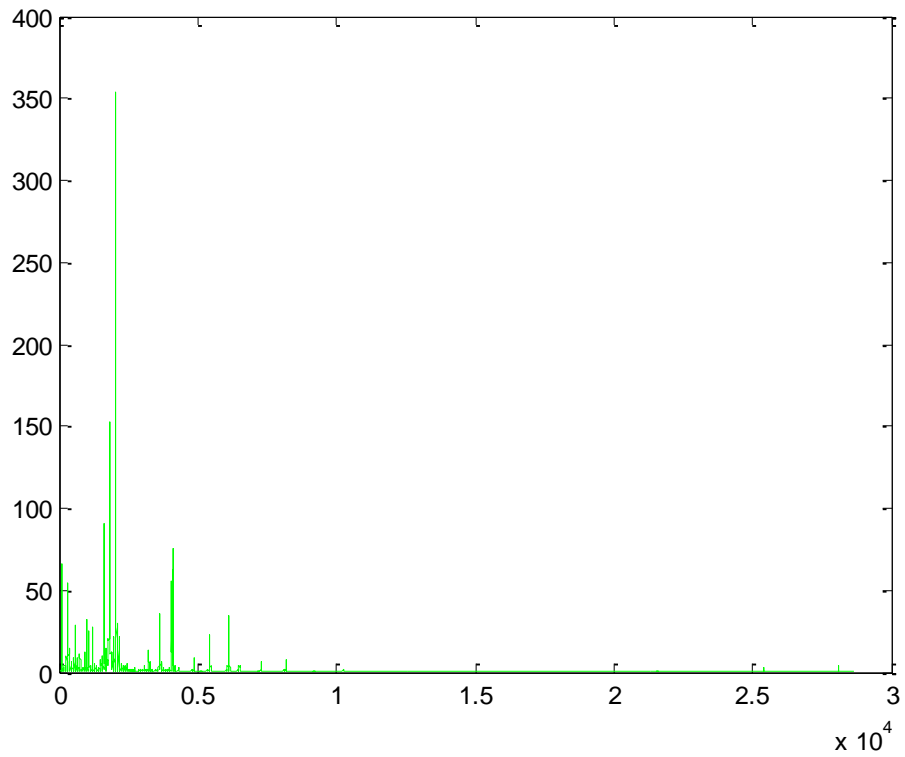


## Β) Εκτέλεση από σόλο πιάνο

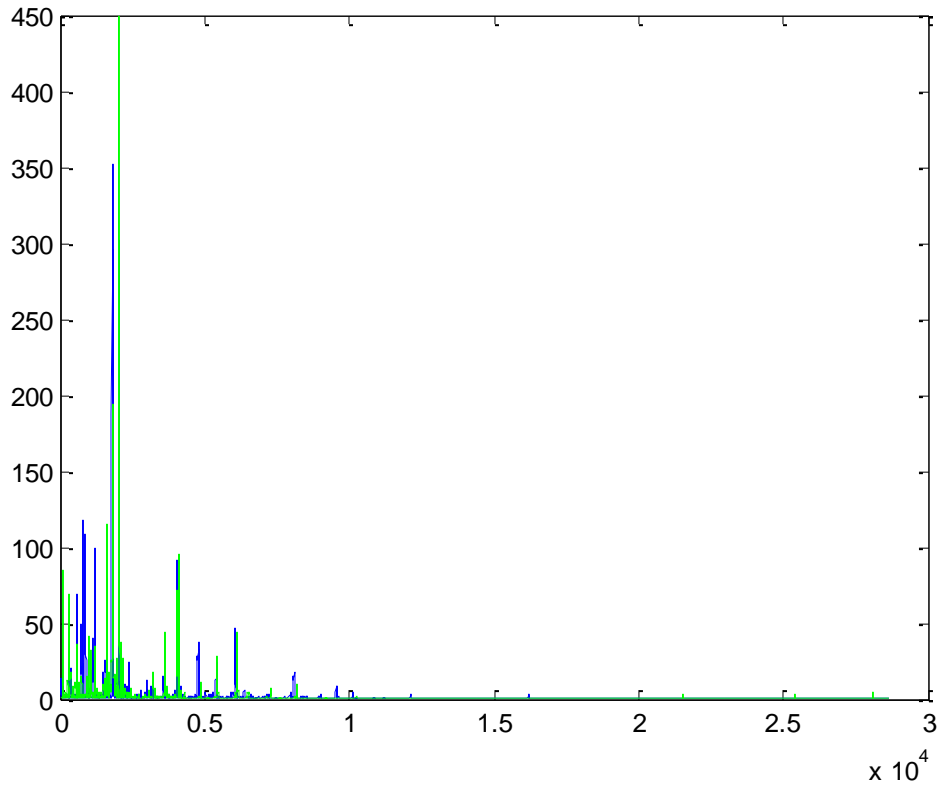
Πεδίο του χρόνου



Πεδίο της συχνότητας



Γ) Παράθεση των DFT (ανηγμένων) ορχήστρας-πιάνο σε κοινό διάγραμμα

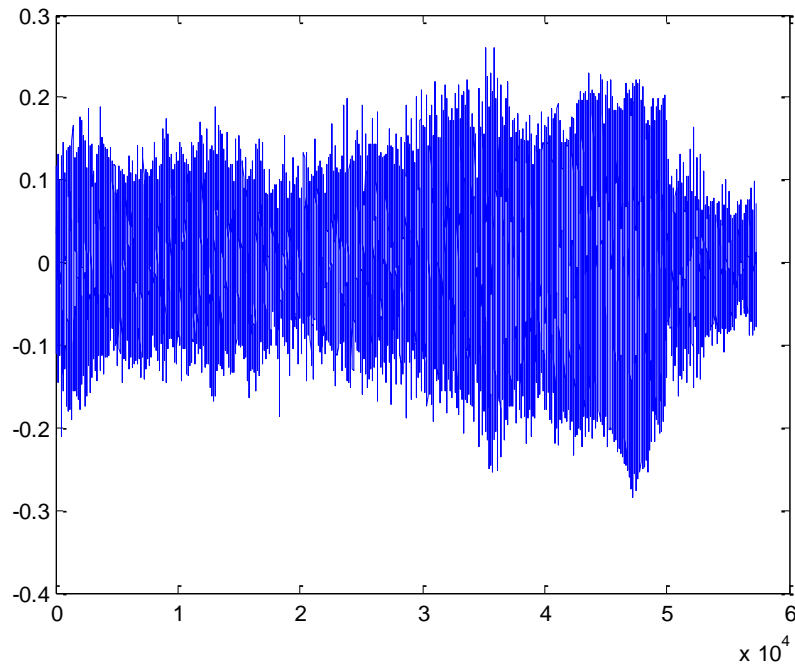


Σε αυτή την ενότητα, όπου πρόκειται για ένα κοινό ακουστικά τμήμα μεταξύ της ορχήστρας και του πιάνο, είναι ακόμα πιο φανερή η σύμπτωση μεγάλου αριθμού κορυφών της ορχήστρας και του πιάνο, που έχουν αξιοσημείωτο σχετικό πλάτος. Αυτό σημαίνει ότι και στις δύο εκτελέσεις κάποιες νότες συμπίπτουν (περισσότερα για το πώς οι κορυφές αντιστοιχούν σε νότες βρίσκονται στην ενότητα 4.2).

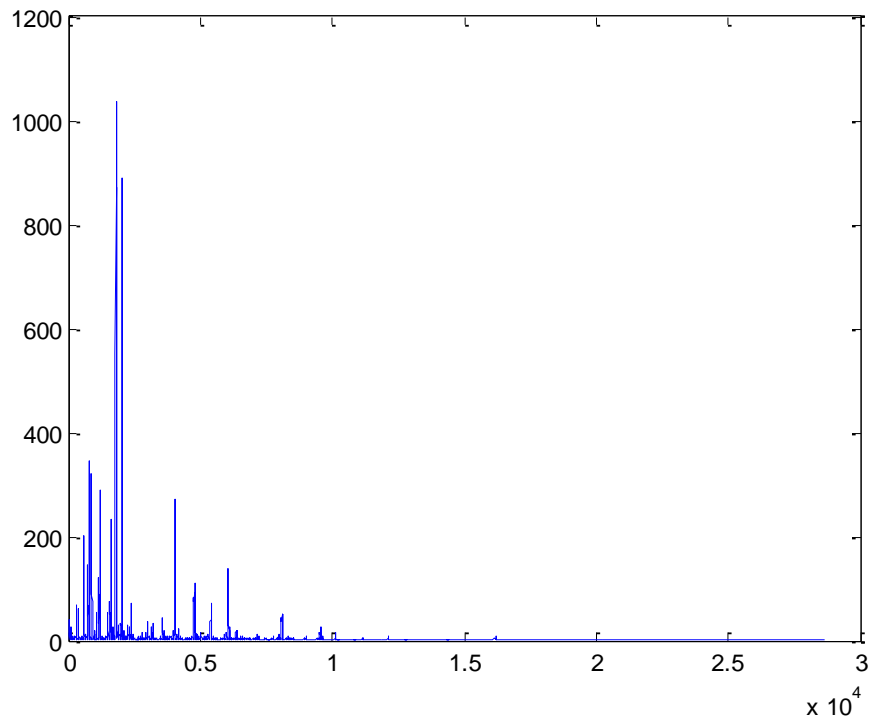
## Κοινό τμήμα ορχήστρα-κιθάρα

A) Εκτέλεση από ορχήστρα

Πεδίο του χρόνου

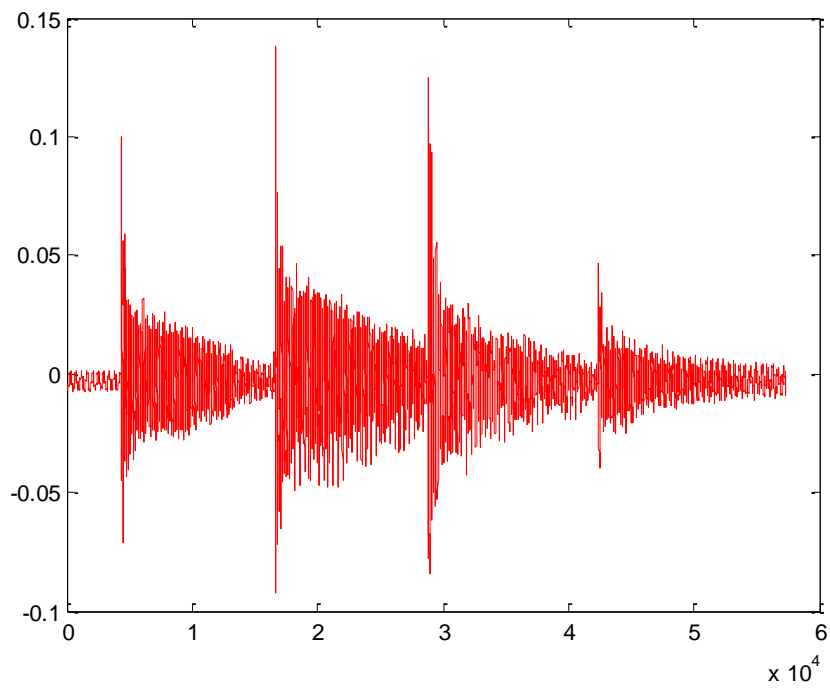


Πεδίο της συχνότητας

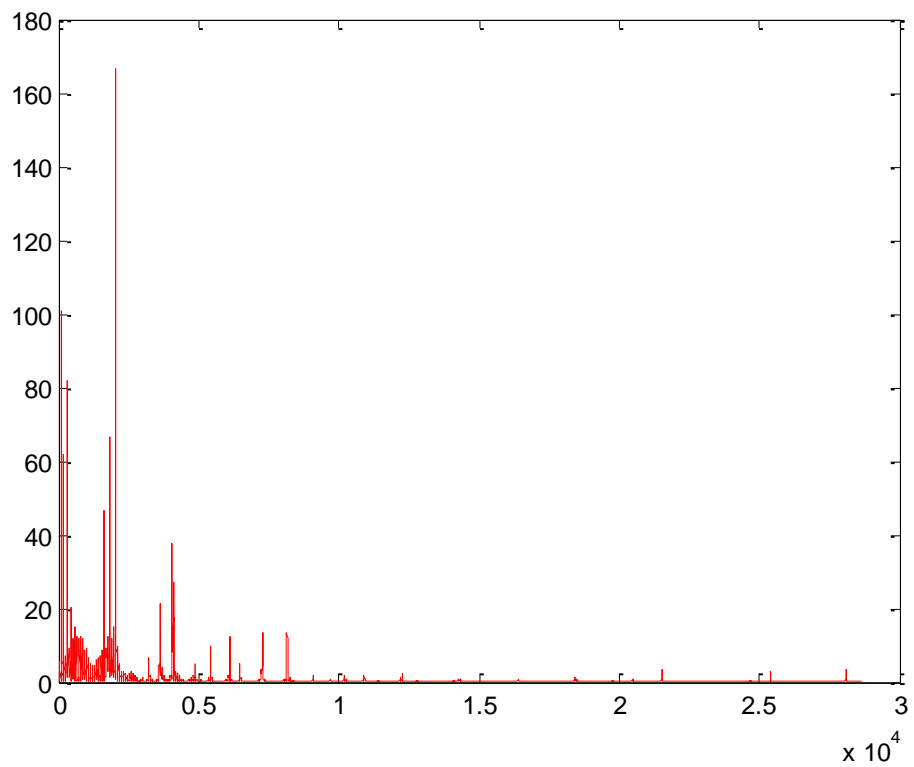


## Β) Εκτέλεση από σόλο κιθάρά

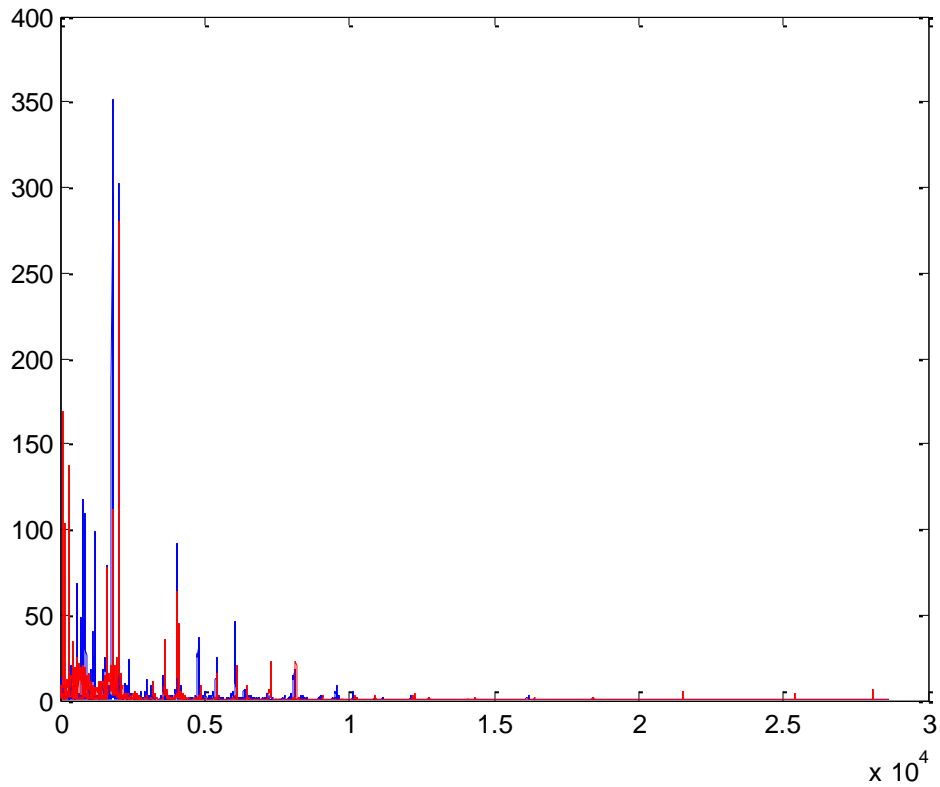
Πεδίο του χρόνου



Πεδίο της συχνότητας



Γ) Παράθεση των DFT (ανηγμένων) ορχήστρας-κιθάρας σε κοινό διάγραμμα



Ισχύουν και εδώ οι ίδιες παρατηρήσεις με αυτές που αφορούν το κοινό τμήμα ορχήστρας και πιάνου αυτής της μουσικής σύνθεσης. Φαίνονται δηλαδή αρκετά καθαρά κάποιες ομοιότητες ανάμεσα στις δύο εκτελέσεις της ίδιας μουσικής σύνθεσης. Βεβαίως, συνεχίζουν να υπάρχουν και οι διαφορές που έχουν αναφερθεί σε προηγούμενες ενότητες, οι οποίες δυσκολεύουν τη διαδικασία της αναγνώρισης.



# 4

## *Θεμελιώδεις τεχνικές και αλγόριθμοι στους οποίους βασίστηκε η μέθοδος που εφαρμόστηκε*

### *4.1 Μαθηματικό υπόβαθρο και αναδρομικός FFT*

Παρά τις τεράστιες διαφορές που παρατηρήσαμε στην προηγούμενη ενότητα τόσο στο πεδίο του χρόνου όσο και στο πεδίο της συχνότητας μεταξύ των δύο σημάτων, εκτελώντας έναν αριθμό από πειράματα παρατηρήσαμε ότι υπάρχουν κάποιες ομοιότητες μεταξύ των απολύτων τιμών των φασμάτων του πρότυπου μουσικού σήματος και του προς αναγνώριση μουσικού σήματος, εφ' όσον αυτές αφορούν την ίδια μουσική σύνθεση. Χρησιμοποιώντας αυτές τις ομοιότητες προκύπτει μία ικανοποιητική και αποδοτική διαδικασία αναγνώρισης μουσικών συνθέσεων. Η προσέγγιση που χρησιμοποιήσαμε βασίζεται στη σύγκριση δύο τμημάτων από τα δύο σήματα (ένα τμήμα από το πρότυπο και ένα από το προς αναγνώριση) που έχουν ακριβώς το ίδιο μήκος, έστω  $N$  δειγμάτων. Η βασική παρατήρηση που οδήγησε στην υλοποίηση της προτεινόμενης μεθοδολογίας ήταν ότι όταν το ηχητικό αποτέλεσμα δύο μουσικών τμημάτων είναι σχεδόν όμοιο, τότε οι κορυφές του σήματος στο μετασχηματισμό

Fourier μπορεί να διαφέρουν αισθητά μεν στο πλάτος, αλλά ελαφρά ως προς τη θέση που εμφανίζονται.

Για τον υπολογισμό του Διακριτού Μετασχηματισμού Fourier των σημάτων χρησιμοποιήθηκε ένας αποδοτικός αλγόριθμος, ο οποίος λέγεται αναδρομικός FFT, που βασίζεται όμως στον DFT. Στο παρακάτω υποκεφάλαιο αναφέρεται ο ορισμός του DFT καθώς και ορισμένες βασικές έννοιες και ιδιότητές του, ώστε στο αμέσως επόμενο υποκεφάλαιο να περιγραφεί η ιδέα του αναδρομικού/προσαρμοστικού FFT.

#### 4.1.1 Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Transform-DFT)

Κατ' αρχήν πρέπει να τονισθεί ότι ο διακριτός μετασχηματισμός Fourier (DFT) αφορά πάντα ακολουθίες πεπερασμένου μήκους (finite duration sequences), δηλαδή ακολουθίες  $x: \mathbf{A} \rightarrow \mathbf{C}$  όπου  $A \subset \mathbf{N}$  και το σύνολο  $\mathbf{A}$  έχει πεπερασμένο πλήθος στοιχείων.

**ΟΡΙΣΜΟΣ 1:** Έστω μία πεπερασμένου μήκους ακολουθία  $x[n]$ , ορισμένη στα  $n$  σημεία ( $n \in \mathbf{v}$ )  $0, 1, 2, \dots, v-1$ . Τότε ο διακριτός μετασχηματισμός Fourier (DFT) αυτής της ακολουθίας είναι το πεπερασμένο άθροισμα:

$$X[k] = \begin{cases} \sum_{n=0}^{N-1} x[n] e^{-j(2\pi/N)kn}, & k \in \mathbf{N}, 0 \leq k \leq (N-1) \\ 0 & \text{σε όλους τους άλλους ακραίους} \end{cases} \quad (\Delta 1)$$

Δηλαδή, σε αντίθεση με τον Ημι-Συνεχή Μετασχηματισμό Fourier της  $x[n]$ ,

$$\left( X(e^{j\omega}) = \sum_{i=-\infty}^{\infty} x[i] e^{-j\omega i} \right), \text{ στο προκείμενο,}$$

α) το άθροισμα λαμβάνεται μόνον από την τιμή 0 έως την τιμή  $v-1$ , και

β) το  $\omega$  στον διακριτό μετασχηματισμό Fourier παίρνει και αυτό μόνο διακριτές τιμές



και συγκεκριμένα τις τιμές  $2\pi \frac{k}{N}, k = 0, 1, \dots, N-1$ .

Ο τύπος (Δ1) αναφέρεται συχνά και σαν “εξίσωση ή τύπος ανάλυσης”.

**ΟΡΙΣΜΟΣ 2:** Ο **αντίστροφος** διακριτός μετασχηματισμός Fourier ή “εξίσωση σύνθεσης” ορίζεται μέσω του αθροίσματος:

$$x[n] = \begin{cases} \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j(2\pi/N)kn}, & k \in \mathbf{N}, 0 \leq k \leq (N-1) \\ 0 & \text{σε όλους τους άλλους ακαραίους} \end{cases} \quad (\Delta.2)$$

Στα επόμενα, θα θεωρούμε πάντα ότι τόσο η ακολουθία  $X[k]$  όσο και η  $x[n]$  θα είναι σε όλα τα άλλα σημεία 0 εκτός των σημείων  $0, 1, 2, \dots, v-1$ , ακόμα και αν δεν το αναφέρουμε σαφώς.

Ισοδύναμα, θα μπορούσε κανείς να υποθέσει ότι τόσο η  $X[k]$  όσο και η  $x[n]$  είναι ακολουθίες πεπερασμένες και περιορισμένες στα σημεία  $0, 1, 2, \dots, v-1$ .

Η δυϊκή σχέση που αποκαθίσταται μεταξύ  $x[n]$  και  $X[k]$  συχνά συμβολίζεται και ως εξής:

$$x[n] \leftrightarrow (\text{DFT})X[k]$$

ή και ως

$$X[k] = \text{DFT}\{x[n]\}.$$

Για παράδειγμα, έστω ότι ζητάμε να βρούμε το μετασχηματισμό Fourier της σταθερής ακολουθίας, δηλαδή της  $x[n] = c, 0, 1, 2, \dots, v-1, N \in \mathbf{N}$ . Πράγματι,

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j(2\pi/N)kn} = \sum_{n=0}^{N-1} c e^{-j(2\pi/N)kn} = c \sum_{n=0}^{N-1} e^{-j(2\pi/N)kn}.$$

Αλλά το τελευταίο άθροισμα είναι άθροισμα όρων γεωμετρικής προόδου με λόγο

$\lambda = e^{-j(2\pi/N)k}$ , άρα ισούται με

$$\alpha_0 \frac{1 - \lambda^v}{1 - \lambda} = \frac{1 - (e^{-j2\pi k})}{1 - e^{-j(2\pi/N)k}}, \quad k = 0, 1, \dots, N-1.$$

Τώρα, επειδή το  $k \in \mathbf{Z} \Rightarrow e^{-j2\pi k} = 1, \forall k$ . Άρα, το άθροισμα  $\sum_{n=0}^{N-1} e^{-j(2\pi/N)kn} = 0$ , σε όλες τις

περιπτώσεις, εκτός από  $k = 0, \pm N, \pm 2N, \pm 3N, \dots$  όπου ο τύπος αθροίσματος όρων

γεωμετρικής προόδου γίνεται απροσδιόριστος. Σ' αυτήν την περίπτωση όμως,

ανατρέχοντας απ' ευθείας στο αρχικό άθροισμα, παρατηρούμε ότι όλοι οι όροι του

αθροίσματος  $\sum_{n=0}^{N-1} e^{-j(2\pi/N)kn}$  είναι ίσοι με 1 (ένα). Άρα, τότε,  $\sum_{n=0}^{N-1} e^{-j(2\pi/N)kn} = N$ .

Επομένως, τελικά,

$$X[k] = c \sum_{n=0}^{N-1} e^{-j(2\pi/N)kn} = \begin{cases} N & k = 0, \pm N, \pm 2N, \pm 3N, \dots \\ 0, & \text{για όλα τα άλλα } k \in \mathbf{Z} \end{cases}$$

### *Ιδιότητες του Διακριτού Μετασχηματισμού Fourier.*

- **Η ιδιότητα της γραμμικότητας .**

**ΙΔΙΟΤΗΤΑ 1:** Έστω  $X(k)$  ο Διακριτός Μετασχηματισμός Fourier της ακολουθίας  $x[n]$

και  $Y(k)$  ο Διακριτός Μετασχηματισμός Fourier της ακολουθίας  $y[n]$ . Τότε για κάθε

αριθμούς μιγαδικούς  $\alpha, \beta$ , ο Διακριτός Μετασχηματισμός Fourier της ακολουθίας

$\alpha x[n] + \beta y[n]$  υπάρχει και ισούται με  $\alpha X(k) + \beta Y(k)$ . Σε συμβολική γραφή:

Εάν  $X[k] = \text{DFT}\{x[n]\}$  και  $Y[k] = \text{DFT}\{y[n]\} \Rightarrow \alpha X(k) + \beta Y(k) = \text{DFT}\{\alpha x[n] + \beta y[n]\}$ .

Γενικότερα, αν  $X_i(k) = \text{DFT}\{x_i[n]\}$ ,  $i \in \mathbf{v}$ ,  $\Rightarrow$

$$\sum_{\square} \alpha_{\square} X_{\square}(k) = \text{DFT}\left\{ \sum_{\square} \alpha_{\square} x_{\square}[n] \right\}, \alpha_{\square} \in \mathbf{C}.$$

*Απόδειξη.* Εξ υποθέσεως και εξ ορισμού,

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j(2\pi/N)kn} \quad \text{και} \quad Y[k] = \sum_{n=0}^{N-1} y[n] e^{-j(2\pi/N)kn}.$$

Άρα,

$$\text{DFT}\{\alpha x[n] + \beta y[n]\} = \sum_{n=0}^{N-1} (\alpha x[n] + \beta y[n]) e^{-j(2\pi/N)kn}$$

$$= \alpha \sum_{n=0}^{N-1} x[n] e^{-j(2\pi/N)kn} + \beta \sum_{n=0}^{N-1} y[n] e^{-j(2\pi/N)kn} = \alpha X(k) + \beta Y(k).$$

Στην αμέσως παραπάνω απόδειξη, υποθέσαμε ότι οι δύο ακολουθίες είχαν **ίδιο πλήθος μη μηδενικών στοιχείων**  $N$ . Εάν κάτι τέτοιο δεν συμβαίνει, πρέπει, για να εφαρμόσουμε την ιδιότητα της γραμμικότητας, να θεωρήσουμε ότι ο κοινός DFT γίνεται σε ένα διάστημα τουλάχιστον ίσο με το μέγιστο του πλήθους μη μηδενικών στοιχείων των δύο ακολουθιών, ενώ η ακολουθία με μικρότερο πλήθος μη μηδενικών στοιχείων “πρέπει να συμπληρωθεί με μηδενικά”. Πιο συγκεκριμένα, έστω ότι το πλήθος μη μηδενικών στοιχείων της  $x[n]$  είναι  $N_1$ , ενώ της  $y[n]$  είναι  $N_2$ , με  $N_1 < N_2$ . Τότε, ο κοινός DFT θα πρέπει να γίνει τουλάχιστον σε ένα διάστημα  $N_2$  σημείων και οι τελευταίοι  $N_2 - N_1$  όροι της ακολουθίας  $x[n]$ , δηλαδή οι όροι  $x[N_1], x[N_1 + 1], \dots, x[N_2 - 1]$  να γίνουν 0.

- **Ο μετασχηματισμός της “μετατοπισμένης στον χρόνο” ακολουθίας (time-shifted sequence)**

**ΣΥΜΒΟΛΙΣΜΟΣ 1:** Για τυχόν ζεύγος αριθμών  $k \in \mathbf{Z}, \ell \in \mathbf{N}$  το σύμβολο  $\text{mod}_\ell(k)$  σημαίνει το υπόλοιπο της διαίρεσης του  $k$  με το  $\ell$ . Δηλαδή εάν  $v$  είναι ο **μικρότερος θετικός ακέραιος**  $0 \leq v < \ell$  τέτοιος ώστε:  $k = m\ell + v, m \in \mathbf{Z}$ ,  $\Rightarrow \text{mod}_\ell(k) = v$ .

Π.χ.  $\text{mod}_3(5) = 2$ ,  $\text{mod}_2(-7) = 1$ ,  $\text{mod}_5(-7) = 3$ ,  $\text{mod}_4(-24) = 0$ , κλπ.

**ΟΡΙΣΜΟΣ 3:** Έστω μία πεπερασμένου μήκους ακολουθία  $x[n]$ , ορισμένη στα  $v$  σημεία  $(v \in \mathbf{v})$   $0, 1, 2, \dots, v-1$ . Η **περιοδική επέκταση της**  $x[n]$  είναι η ακολουθία  $\tilde{x}[n]$ :

$$\tilde{x}[n] = x[\text{mod}_v(n)].$$

Είναι προφανές ότι μ’ αυτόν τον ορισμό  $\tilde{x}[n + bv] = \tilde{x}[n], \forall b \in \mathbf{Z}$ , αφού

$$\text{mod}_v(n + bv) = \text{mod}_v(n), \forall b \in \mathbf{Z}.$$

**ΛΙΘΟΤΗΤΑ 2:** Έστω  $X(k)$  ο Διακριτός Μετασχηματισμός Fourier της ακολουθίας  $x[n]$ , δηλαδή, έστω ότι  $X[k] = \text{DFT}\{x[n]\}$  και έστω ένας οποιοσδήποτε ακέραιος  $\ell \in \mathbf{a}$ . Τότε, εάν ορίσουμε την ακολουθία:

$$x_s[n] = \begin{cases} \tilde{x}[n - \ell] = x[\text{mod}_N(n - \ell)], & 0 \leq n \leq (N - 1) \\ 0 & \text{παντού αλλού} \end{cases} \Rightarrow$$

$$e^{-j(2\pi/N)k\ell} X[k] = \text{DFT}\{x_s[n]\}.$$

Η δημιουργία της  $x_s[n]$  γίνεται με τον εξής τρόπο:

α) Βήμα 1: Δημιουργώ περιοδική επέκταση της  $x[n]$ ,  $\tilde{x}[n]$ .

β) Βήμα 2: Μετατοπίζω κατά  $\ell$  την περιοδική επέκταση της  $x[n]$ .

γ) Βήμα 3: Περιορίζω την  $\tilde{x}[n]$  στο διάστημα  $[0, v - 1]$

Απόδειξη. Εκ τού ορισμού,

$$X[k] = \sum_{n=0}^{N-1} x_s[n] e^{-j(2\pi/N)kn} = \sum_{n=0}^{N-1} \tilde{x}[n] e^{-j(2\pi/N)kn}.$$

Κάνουμε στην αδιάφορη μεταβλητή την αντικατάσταση  $n \rightarrow (m + \ell)$ , οπότε,

$$X[k] = \sum_{m=\ell}^{N+\ell-1} \tilde{x}[m + \ell] e^{-j(2\pi/N)k(m+\ell)}. \text{ Θα υποθέσουμε κατ' αρχήν, χωρίς καμία βλάβη της}$$

γενικότητας, ότι  $\ell < N$ , οπότε, χωρίζουμε το ανωτέρω άθροισμα σε δύο τμήματα:

$$\Sigma_1 = \sum_{m=\ell}^{N-1} \tilde{x}[m] e^{-j(2\pi/N)k(m+\ell)} \quad \text{και} \quad \Sigma_2 = \sum_{m=v}^{N+\ell-1} \tilde{x}[m] e^{-j(2\pi/N)k(m+\ell)}$$

και υπολογίζουμε το κάθε ένα χωριστά.

Πρώτα το  $\Sigma_1$ , το οποίον δεν έχει κανένα πρόβλημα, διότι ο δείκτης άθροισης  $m$  ανήκει στο διάστημα  $[0, v - 1]$ , επομένως  $\tilde{x}[m] = x_s[m] = x[m]$ , άρα,

$$\Sigma_1 = e^{-j(2\pi/N)k\ell} \sum_{m=\ell}^{N-1} x[m] e^{-j(2\pi/N)km}.$$

Όσον αφορά το  $\Sigma_2$ , ισχύει εκ τού ορισμού της ακολουθίας  $\tilde{x}[m]$  ότι

$$\tilde{x}[v] = x[\text{mod}_N(N)] = x[0], \quad \tilde{x}[v+1] = x[\text{mod}_N(N+1)] = x[1], \dots,$$

$$\tilde{x}[v] = x[\text{mod}_N(N + \ell - 1)] = x[\ell - 1]$$

ενώ, επιπλέον,

$$\begin{aligned}
e^{-j(2\pi/N)kv} &= e^{-j(2\pi k)} = e^{-j(2\pi/N)k \cdot 0} \\
e^{-j(2\pi/N)k(N+1)} &= e^{-j(2\pi k)} e^{-j(2\pi/N)k \cdot 1} = e^{-j(2\pi/N)k \cdot 1}, \dots, \\
e^{-j(2\pi/N)k(N+\ell-1)} &= e^{-j(2\pi k)} e^{-j(2\pi/N)k \cdot (\ell-1)} = e^{-j(2\pi/N)k \cdot (\ell-1)}
\end{aligned}$$

και γενικά,

$$\begin{aligned}
e^{-j(2\pi/N)km} &= e^{-j(2\pi km)} e^{-j(2\pi/N) \text{mod}_N(m)}, \quad m \in \mathbf{Z}, \Rightarrow \\
e^{-j(2\pi/N)km} &= e^{-j(2\pi/N) \text{mod}_N(m)} \Rightarrow \\
\Sigma_2 &= \sum_{m=v}^{N+\ell-1} \tilde{x}[m] e^{-j(2\pi/N)k(m+\ell)} = e^{-j(2\pi/N)k\ell} \sum_{m=v}^{N+\ell-1} \tilde{x}[m] e^{-j(2\pi/N)km} = \\
&= e^{-j(2\pi/N)k\ell} \sum_{m=0}^{\ell-1} x[m] e^{-j(2\pi/N)km}.
\end{aligned}$$

Άρα,

$$X[k] =$$

$$\begin{aligned}
\Sigma_1 + \Sigma_2 &= e^{-j(2\pi/N)k\ell} \sum_{m=\ell}^{N-1} x[m] e^{-j(2\pi/N)km} + e^{-j(2\pi/N)k\ell} \sum_{m=0}^{\ell-1} x[m] e^{-j(2\pi/N)km} \Leftrightarrow \\
X_s[k] &= e^{-j(2\pi/N)k\ell} \sum_{m=0}^{N-1} x[m] e^{-j(2\pi/N)km} = e^{-j(2\pi/N)k\ell} X[k].
\end{aligned}$$

Αν, τέλος, το  $\ell$  (η χρονική μετατόπιση της  $\tilde{x}[n]$  είναι εκτός του διαστήματος  $[0, v]$ ), τότε, αφού λάβουμε υπ' όψιν ότι

$$\exists u : \ell = mN + u, 0 \leq u \leq v \Leftrightarrow \exists u : u = \text{mod}_N(\ell),$$

οπότε, επαναλαμβάνουμε εντελώς ανάλογα όλη την ως άνω διαδικασία, μόνο που τώρα τον ρόλο του  $\ell$  τον παίζει το  $u$ . Π.χ., τότε,

$$\Sigma_1 = e^{-j(2\pi/N)k\ell} \sum_{m=u}^{N-1} x[m] e^{-j(2\pi/N)km} \quad \text{και} \quad \Sigma_2 = e^{-j(2\pi/N)k\ell} \sum_{m=0}^{u-1} x[m] e^{-j(2\pi/N)km}.$$

- **Η δυαδικώς αντίστοιχη ιδιότητα.**

**ΙΔΙΟΤΗΤΑ 3:** Έστω  $X(k)$  ο Διακριτός Μετασχηματισμός Fourier της ακολουθίας  $x[n]$ , δηλαδή, έστω ότι  $X[k] = \text{DFT}\{x[n]\}$  και έστω ένας οποιοσδήποτε ακέραιος  $\ell \in \mathbf{a}$ . Τότε, εάν ορίσουμε την ακολουθία:

$$X_s[n] = \begin{cases} \tilde{X}[n - \ell] = X[\text{mod}_N(n - \ell)], & 0 \leq n \leq (N - 1) \\ 0 & \text{παντού αλλού} \end{cases} \Rightarrow$$

$$X_s[k] = \text{DFT}\{e^{j(2\pi/N)n\ell} x[n]\}.$$

Απόδειξη. Εκ τού ορισμού,

$$Y[k] = \sum_{n=0}^{N-1} (e^{j(2\pi/N)n\ell} x[n]) e^{-j(2\pi/N)kn} = \sum_{n=0}^{N-1} x[n] e^{-j(2\pi/N)(k-\ell)n}.$$

Εν συνεχεία, ορίζουμε την περιοδική επέκταση  $\tilde{Y}[k]$  της ακολουθίας  $Y[k]$  κατά τα γνωστά:

$$\tilde{Y}[k] = Y[\text{mod}_N(k)],$$

οπότε, επειδή παραμένουμε στο διάστημα  $[0, (N-1)]$  συνεπάγεται ότι,

$$\tilde{Y}[k] = \sum_{n=0}^{N-1} x[n] e^{-j(2\pi/N)(k-\ell)n}.$$

Κάνουμε στο άθροισμα την αντικατάσταση  $k' = (k - \ell)$ , οπότε φαίνεται αμέσως ότι αυτό δεν είναι τίποτε άλλο από το  $\tilde{X}(k - \ell)$ . Άρα, τελικά, στο διάστημα  $[0, (N-1)]$ :

$$Y[k] = \tilde{Y}[k] = \tilde{X}(k - \ell) = X[\text{mod}_N(n - \ell)].$$

Για παράδειγμα: Έστω μία ακολουθία 7 (επτά) σημείων η  $x[0], x[1], x[2], \dots, x[6]$  (δηλαδή  $N = 7$ ) και έστω ότι ο DFT αυτής είναι η ακολουθία  $X[0], X[1], X[2], \dots, X[6]$ . Τότε, σύμφωνα με την παραπάνω ιδιότητα, ο DFT της

$$e^{4j(2\pi/7)n} x[n] \quad (N = 7, \ell = 4)$$

είναι η

$$X[\text{mod}_7(k - 4)], \quad 0 \leq k \leq 6.$$

Αναλυτικότερα, αν θέσουμε σαν  $Y[n] = \text{DFT}\{e^{4j(2\pi/7)n} x[n]\}$ , τότε:

$$Y[0] = x[\text{mod}_7(-4)] = x[3]$$

$$Y[1] = x[\text{mod}_7(-3)] = x[4]$$

$$Y[2] = x[\text{mod}_7(-2)] = x[5]$$

$$Y[3] = x[\text{mod}_7(-1)] = x[6]$$

$$Y[4] = x[\text{mod}_7(0)] = x[0]$$

$$Y[5] = x[\text{mod}_7(1)] = x[1]$$

$$Y[6] = x[\text{mod}_7(2)] = x[2].$$

- **“Ο μετασχηματισμός τού μετασχηματισμού”**

Εάν παρατηρήσουμε τις εξισώσεις (ορισμούς) σύνθεσης και ανάλυσης του DFT, θα παρατηρήσουμε ότι διαφέρουν μόνο κατά τον παράγοντα και κατά το πρόσημο των μιγαδικών εκθετικών συναρτήσεων-”βαρών”. Επομένως, λογικά αναμένουμε κάποια “αντανεκλαστικότητα” ή κάποιο “δυϊσμό” στις ιδιότητες. Πράγματι, μεταξύ άλλων ισχύει και η εξής:

**ΙΔΙΟΤΗΤΑ 4:** Έστω  $X(k)$  ο Διακριτός Μετασχηματισμός Fourier της ακολουθίας  $x[n]$ , δηλαδή, έστω ότι  $X[k] = \text{DFT}\{x[n]\}$ . Τότε, εάν θεωρήσουμε τον DFT της  $X[k]$ , δηλαδή “τον DFT του DFT”, ισχύει:

$$\text{DFT}\{X[n]\} = Nx[\text{mod}_N(-k)], 0 \leq k \leq (N-1).$$

Για παράδειγμα:

Έστω μία ακολουθία 7 (επτά) σημείων η  $x[0], x[1], x[2], \dots, x[6]$  (δηλαδή  $N = 7$ ) και έστω ότι ο DFT αυτής είναι η ακολουθία  $X[0], X[1], X[2], \dots, X[6]$ . Τότε, σύμφωνα με την παραπάνω ιδιότητα, ο DFT της  $X[n]$  είναι η  $7x[\text{mod}_7(-k)], 0 \leq k \leq 6$ . Αναλυτικότερα, αν θέσουμε σαν  $Y[n] = \text{DFT}\{X[n]\}$ , τότε:

$$Y[0] = 7x[\text{mod}_7(-0)] = 7x[0]$$

$$Y[1] = 7x[\text{mod}_7(-1)] = 7x[6]$$

$$Y[2] = 7x[\text{mod}_7(-2)] = 7x[5]$$

$$Y[3] = 7x[\text{mod}_7(-3)] = 7x[4]$$

$$Y[4] = 7x[\text{mod}_7(-4)] = 7x[3]$$

$$Y[5] = 7x[\text{mod}_7(-5)] = 7x[2]$$

$$Y[6] = 7x[\text{mod}_7(-6)] = 7x[1].$$

Εν κατακλείδι, η  $Y[n] = \text{DFT}\{X[n]\}$  έχει τις τιμές της αρχικής  $x[n]$ , πολλαπλασιασμένες επί  $N = 7$  και αναδιαταγμένες όπως καθορίζει το mod.

Απόδειξη: Πρώτα, αφού  $X[k] = \text{DFT}\{x[n]\}$ , εκ τού ορισμού

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j(2\pi/N)kn}.$$

Στο σημείο αυτό ορίζουμε την περιοδική επέκταση  $\tilde{x}[n]$  της ακολουθίας  $x[n]$  κατά τα γνωστά:

$$\tilde{x}[k] = x[\text{mod}_N(k)],$$

οπότε, επειδή παραμένουμε στο διάστημα  $[0, (v-1)]$  συνεπάγεται ότι,

$$\tilde{x}[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j(2\pi/N)kn}.$$

Στον τελευταίο τύπο πρώτα πολλαπλασιάζουμε και τα δύο μέλη επί  $N$  και εν συνεχεία θέτουμε όπου  $n \rightarrow (-n)$ , οπότε,

$$N\tilde{x}[-n] = \sum_{k=0}^{N-1} \tilde{X}[k] e^{j(2\pi/N)k(-n)} = \sum_{k=0}^{N-1} \tilde{X}[k] e^{-j(2\pi/N)kn}.$$

Αλλά αφού  $Y[n] = \text{DFT}\{X[n]\}$ , τότε, εξ ορισμού:

$$Y[n] = \sum_{k=0}^{N-1} X[k] e^{-j(2\pi/N)kn}.$$

Τα δεύτερα μέλη των δύο ανωτέρω εξισώσεων είναι ίσα, άρα,

$$Y[n] = N\tilde{x}[-n] = Nx[\text{mod}_N(-n)]$$

εκ τού ορισμού της περιοδικής επέκτασης της  $x[n]$ .

- **Η Κυκλική Συνέλιξη και ο διακριτός μετασχηματισμός Fourier αυτής.**

**ΟΡΙΣΜΟΣ 4:** Έστω δύο πεπερασμένου μήκους ακολουθίες  $x_1[n], x_2[n]$ , ορισμένες στα  $v$



σημεία  $0,1,2,\dots,v-1$ ,  $v \in \mathbf{v}$  και ας είναι  $\tilde{x}_1[n], \tilde{x}_2[n]$  οι περιοδικές επεκτάσεις τους, όπως αυτές ορίστηκαν προηγούμενα. Τότε η κυκλική συνέλιξη αυτών των δύο ακολουθιών, ορίζεται σαν η πεπερασμένου μήκους ακολουθία  $\sigma[n]$ :

$$\sigma[n] = \sum_{p=0}^{p=v-1} x_1[p] \tilde{x}_2[n-p] = \sum_{\text{TM}=0}^{\text{TM}=v-1} x_1[p] x_2[\text{mod}_N(n-p)], \quad n \in N, 0 \leq n \leq N-1.$$

Αυτή είναι ίδια με την

$$\sigma[n] = \sum_{p=0}^{p=v-1} \tilde{x}_1[n-p] x_2[p] = \sum_{p=0}^{p=v-1} x_1[\text{mod}_N(n-p)] x_2[p], \quad n \in N, 0 \leq n \leq N-1.$$

**Δηλαδή τελικά ισχύει:**

$$\sum_{p=0}^{p=v-1} x_1[p] \tilde{x}_2[n-p] = \sum_{p=0}^{p=v-1} \tilde{x}_1[n-p] x_2[p]$$

**ΙΔΙΟΤΗΤΑ 5:** Έστω  $X(k)$  ο Διακριτός Μετασχηματισμός Fourier της ακολουθίας  $x[n]$ , και  $Y(k)$  της  $y[n]$ , δηλαδή, έστω ότι  $X[k] = \text{DFT}\{x[n]\}$ ,  $Y[k] = \text{DFT}\{y[n]\}$ . Τότε, ο DFT της κυκλικής συνέλιξης των  $x[n]$  και  $y[n]$ , είναι το γινόμενο των  $X(k)$ ,  $Y(k)$ . Σε συμβολική γραφή:

$$\text{DFT}\left\{\sum_{p=0}^{p=v-1} x[p] y[\text{mod}_N(n-p)]\right\} = X[k] Y[k].$$

*Απόδειξη.*

$$\text{DFT}\left\{\sum_{p=0}^{p=v-1} x[p] y[\text{mod}_N(n-p)]\right\} = \sum_{n=0}^{N-1} \left( \sum_{p=0}^{p=v-1} x[p] y[\text{mod}_N(n-p)] \right) e^{-j(2\pi/N)kn}$$

και αφού εναλλάξουμε την σειρά των πεπερασμένων αθροισμάτων προκύπτει

$$\text{DFT}\{\sigma[n]\} = \sum_{p=0}^{p=v-1} \left\{ x[p] \left( \sum_{n=0}^{N-1} y[\text{mod}_N(n-p)] e^{-j(2\pi/N)kn} \right) \right\}.$$

Αλλά, σύμφωνα με την ιδιότητα 2 ανωτέρω,

$$\sum_{n=0}^{N-1} y[\text{mod}_N(n-p)] e^{-j(2\pi/N)kn} = \text{DFT}\{y[\text{mod}_N(n-p)]\} = e^{-j(2\pi/N)kp} Y[k],$$

οπότε,

$$\text{DFT}\{\sigma[n]\} = \sum_{p=0}^{p=N-1} \{x[p]e^{-j(2\pi/N)kp} Y[k]\} = Y[k] \sum_{p=0}^{p=N-1} \{x[p]e^{-j(2\pi/N)kp}\} = Y[k]X[k]$$

- **Η δυαδικώς αντίστοιχη ιδιότητα.**

**ΙΔΙΟΤΗΤΑ 6:** Έστω  $X(k)$  ο Διακριτός Μετασχηματισμός Fourier της ακολουθίας  $x[n]$ , και  $Y(k)$  της  $y[n]$ , δηλαδή, έστω ότι  $X[k] = \text{DFT}\{x[n]\}$ ,  $Y[k] = \text{DFT}\{y[n]\}$ . Τότε, ο DFT του γινομένου των  $x[n]$  και  $y[n]$ , είναι το  $1/N$  της κυκλικής συνέλιξης των  $X(k)$ ,  $Y(k)$ . Σε συμβολική γραφή:

$$\text{DFT}\{x[n]y[n]\} = \frac{1}{N} \sum_{p=0}^{p=N-1} X[p]Y[\text{mod}_N(k-p)].$$

Απόδειξη.

$$\text{FT}\{x[n]y[n]\} = \sum_{n=0}^{N-1} (x[n]y[n])e^{-j(2\pi/N)kn}$$

οπότε, χρησιμοποιώντας τον τύπο για τον αντίστροφο διακριτό μετασχηματισμό Fourier της  $y[n]$ , προκύπτει,

$$\begin{aligned} \text{DFT}\{x[n]y[n]\} &= \sum_{n=0}^{N-1} \left( x[n] \frac{1}{N} \sum_{p=0}^{N-1} Y[p]e^{j(2\pi/N)pn} \right) e^{-j(2\pi/N)kn} = \\ &= \frac{1}{N} \sum_{p=0}^{N-1} Y[p] \sum_{n=0}^{N-1} x[n]e^{j(2\pi/N)pn} e^{-j(2\pi/N)kn} = \frac{1}{N} \sum_{p=0}^{N-1} Y[p] \sum_{n=0}^{N-1} x[n]e^{-j(2\pi/N)(k-p)n}. \end{aligned}$$

Αλλά εκ τού ορισμού τού DFT και εξ όλης της προηγούμενης ανάλυσης έπεται

$$\sum_{n=0}^{N-1} x[n]e^{-j(2\pi/N)(k-p)n} = X[\text{mod}_N(k-p)],$$

οπότε,

$$\text{DFT}\{x[n]y[n]\} = \frac{1}{N} \sum_{p=0}^{N-1} Y[p]X[\text{mod}_N(k-p)].$$

**ΙΔΙΟΤΗΤΑ 7:** Εάν  $X[k] = \text{DFT}\{x[n]\} \Rightarrow \text{DFT}\{x^*[n]\} = X^*[\text{mod}_N(-k)]$ .

Απόδειξη. Έστω  $Y[k] = \text{DFT}\{x^*[n]\} = \sum_{n=0}^{N-1} x^*[n]e^{-j(2\pi/N)kn} \Rightarrow$

$$Y[k] = \left( \sum_{n=0}^{N-1} x[n]e^{j(2\pi/N)kn} \right)^*$$

Σε αυτό το σημείο, στο τελευταίο άθροισμα, κάνουμε την αντικατάσταση  $(-k) \rightarrow k$ , ενώ ταυτόχρονα αλλάζουμε το συνολικό πρόσημο τού εκθέτη, οπότε, προκύπτει

$$Y[k] = \left( \sum_{n=0}^{N-1} x[n]e^{-j(2\pi/N)(-k)n} \right)^* = \left( \sum_{n=0}^{N-1} x[n]e^{-j(2\pi/N)(\text{mod}_N(-k))n} \right)^*$$

Αλλά το εντός παρενθέσεως τελευταίο άθροισμα, είναι εξ ορισμού ο DFT της ακολουθίας  $x[n]$  ορισμένος στα σημεία  $(\text{mod}_N(-k))$ . Επομένως,

$$Y[k] = (X[\text{mod}_N(-k)])^* \Leftrightarrow \text{DFT}\{x^*[n]\} = X^*[\text{mod}_N(-k)]$$

**ΛΙΘΟΤΗΤΑ 8:** (Η δυαδικώς αντίστοιχη της 7) Εάν  $X[k] = \text{DFT}\{x[n]\} \Rightarrow$

$$\text{DFT}\{x^*[\text{mod}_N(-n)]\} = X^*[k].$$

Απόδειξη. Έστω  $Y[k] = \text{DFT}\{x^*[\text{mod}_N(-n)]\} = \sum_{n=0}^{N-1} x^*[\text{mod}_N(-n)]e^{-j(2\pi/N)kn} \Leftrightarrow$

$$Y[k] = \sum_{n=0}^{N-1} x^*[\text{mod}_N(-n)]e^{j(2\pi/N)k \text{mod}_N(-n)} \Leftrightarrow$$

$$Y[k] = \left( \sum_{n=0}^{N-1} x[\text{mod}_N(-n)]e^{-j(2\pi/N)k \text{mod}_N(-n)} \right)^*$$

Στο άθροισμα στο δεξί μέλος της τελευταίας ισότητας, κάνουμε την αντικατάσταση  $m = \text{mod}_N(-n)$ , και παρατηρούμε ότι όταν το  $n$  διατρέχει τις τιμές  $0, 1, 2, \dots, N-1$ , το  $m$  διατρέχει τις ίδιες ακριβώς τιμές, οπότε προκύπτει:

$$Y[k] = \left( \sum_{m=0}^{N-1} x[m]e^{-j(2\pi/N)km} \right)^*$$

Αλλά το εντός παρενθέσεως τελευταίο άθροισμα, είναι εξ ορισμού ο DFT της ακολουθίας  $x[n]$ . Επομένως,

$$Y[k] = (X[k])^*$$

- **Ο ορισμός της συζυγούς-συμμετρικής και συζυγούς-αντισυμμετρικής ακολουθίας, στην περίπτωση των πεπερασμένων ακολουθιών.**

**ΟΡΙΣΜΟΣ 5:** Μία πεπερασμένη ακολουθία  $x_{\Sigma}[n]:\mathbf{N} \rightarrow \mathbf{C}$  λέγεται *συζυγής-συμμετρική*, όταν η περιοδική επέκτασή της είναι *συζυγής-συμμετρική* δηλαδή όταν

$$\tilde{x}_{\Sigma}[n] = \tilde{x}_{\Sigma}^*[-n], \forall n \in \mathbf{N}.$$

**ΟΡΙΣΜΟΣ 6:** Μία πεπερασμένη ακολουθία  $x_{\eta}[n]:\mathbf{N} \rightarrow \mathbf{C}$  λέγεται *συζυγής-αντισυμμετρική*, όταν η περιοδική επέκτασή της είναι *συζυγής-αντισυμμετρική*, δηλαδή όταν :

$$\tilde{x}_{\Sigma}[n] = -\tilde{x}_{\Sigma}^*[-n], \forall n \in \mathbf{N}.$$

**ΛΗΜΜΑ:** Μία οποιαδήποτε πεπερασμένη ακολουθία  $x[n]:\mathbf{N} \rightarrow \mathbf{C}$ , μπορεί να γραφεί σαν άθροισμα μίας συζυγούς-συμμετρικής και μίας συζυγούς-αντισυμμετρικής πεπερασμένης ακολουθίας.

*Απόδειξη.* Προφανώς για την περιοδική επέκτασή της  $x[n]$ , ισχύει:

$$\tilde{x}[n] = \frac{1}{2}(\tilde{x}[n] - \tilde{x}^*[-n]) + \frac{1}{2}(\tilde{x}[n] + \tilde{x}^*[-n]) = \tilde{x}_{\Sigma}[n] + \tilde{x}_{\eta}[n], \forall n \in \mathbf{N}.$$

όπου είναι προφανές ότι η  $\tilde{x}_{\Sigma}[n]$  είναι συζυγής-συμμετρική, ενώ η  $\tilde{x}_{\eta}[n]$  είναι συζυγής-αντισυμμετρική ακολουθία. Τότε, εάν ορίσουμε

$$x_{\Sigma}[n] = \frac{1}{2}(x[n] + x^*[\text{mod}_N(-n)]), n = 0, 1, 2, \dots, N-1$$

$$x_{\eta}[n] = \frac{1}{2}(x[n] - x^*[\text{mod}_N(-n)]), n = 0, 1, 2, \dots, N-1,$$

προφανώς, η συζυγής-συμμετρική  $\tilde{x}_{\Sigma}[n]$  είναι η περιοδική επέκτασή της  $x_{\Sigma}[n]$  και η συζυγής-συμμετρική  $\tilde{x}_{\eta}[n]$  είναι η περιοδική επέκτασή της  $x_{\eta}[n]$  και

$$x[n] = \frac{1}{2}(x_{\Sigma}[n] + x_{\eta}[n]).$$

**ΠΟΡΙΣΜΑ 1:** Μία οποιαδήποτε πεπερασμένη ακολουθία  $x[n] : \mathbf{N} \rightarrow \mathbf{C}$ , μπορεί να γραφεί σαν άθροισμα μίας συζυγούς-συμμετρικής και μίας συζυγούς-αντισυμμετρικής πεπερασμένης ακολουθίας ως εξής:

$$x[n] = \frac{1}{2}(x_{\Sigma\varsigma}[n] + x_{A\varsigma}[n])$$

$$x_{\Sigma\varsigma}[n] = \frac{1}{2}(x[n] + x^*[N-n]), \quad n = 0, 1, 2, \dots, N-1$$

$$x_{A\varsigma}[n] = \frac{1}{2}(x[n] - x^*[N-n]), \quad n = 0, 1, 2, \dots, N-1,$$

**ΙΔΙΟΤΗΤΑ 9:** Έστω  $X(k)$  ο Διακριτός Μετασχηματισμός Fourier της ακολουθίας  $x[n]$ , δηλαδή, έστω ότι  $X[k] = \text{DFT}\{x[n]\}$ . Τότε, ο DFT τού πραγματικού μέρους του  $x[n]$  είναι  $\frac{1}{2}(X[\text{mod}_N(k)] + X^*[\text{mod}_N(-k)])$ . Σε συμβολική γραφή:

$$\begin{aligned} \text{DFT}\{\text{Re}\{x[n]\}\} &= \frac{1}{2}(X[\text{mod}_N(k)] + X^*[\text{mod}_N(-k)]) = \\ &= \frac{1}{2}(X[k] + X^*[\text{mod}_N(-k)]), \quad k = 0, 1, \dots, N-1 \end{aligned}$$

*Απόδειξη.*

$$\text{Re}\{x[n]\} = \frac{1}{2}(x[n] + x^*[n])$$

οπότε με εφαρμογή των ιδιοτήτων 1 (γραμμικότητας) και 7 έπεται,

$$\text{DFT}\{\text{Re}\{x[n]\}\} = \text{DFT}\left\{\frac{1}{2}(x[n] + x^*[n])\right\} = \frac{1}{2}\text{DFT}\{x[n]\} + \frac{1}{2}\text{DFT}\{x^*[n]\} \Leftrightarrow$$

$$\text{DFT}\{\text{Re}\{x[n]\}\} = \frac{1}{2}(X[k] + X^*[\text{mod}_N(-k)]), \quad k = 0, 1, \dots, N-1.$$

**ΙΔΙΟΤΗΤΑ 10:** Έστω  $X(k)$  ο Διακριτός Μετασχηματισμός Fourier της ακολουθίας  $x[n]$ , δηλαδή, έστω ότι  $X[k] = \text{DFT}\{x[n]\}$ . Τότε, ο DFT τού μιγαδικού τμήματος ( $j\text{Im}\{x[n]\}$ ) τού  $x[n]$  είναι  $\frac{1}{2}(X[\text{mod}_N(k)] - X^*[\text{mod}_N(-k)])$ . Σε συμβολική γραφή:

$$\begin{aligned} \text{DFT}\{j\text{Im}\{x[n]\}\} &= \frac{1}{2}(X[\text{mod}_N(k)] - X^*[\text{mod}_N(-k)]) = \\ &= \frac{1}{2}(X[k] - X^*[\text{mod}_N(-k)]), \quad k = 0, 1, \dots, N-1 \end{aligned}$$

*Απόδειξη.*

$$j\text{Im}\{x[n]\} = \frac{1}{2}(x[n] - x^*[n])$$

οπότε με εφαρμογή των ιδιοτήτων 1 (γραμμικότητας) και 7 έπεται,

$$\begin{aligned} \text{DFT}\{j\text{Im}\{x[n]\}\} &= \text{DFT}\left\{\frac{1}{2}(x[n] - x^*[n])\right\} = \frac{1}{2}\text{DFT}\{x[n]\} - \frac{1}{2}\text{DFT}\{x^*[n]\} \Leftrightarrow \\ \text{DFT}\{j\text{Im}\{x[n]\}\} &= \frac{1}{2}(X[k] + X^*[\text{mod}_N(-k)]), \quad k = 0, 1, \dots, N-1. \end{aligned}$$

**ΙΔΙΟΤΗΤΑ 11:** Έστω ότι  $X[k] = \text{DFT}\{x[n]\}$ . Τότε, ο DFT τού συζυγούς-συμμετρικού μέρους τού  $x[n]$  είναι  $\text{Re}\{X[k]\}$ . Σε συμβολική γραφή:

$$\text{DFT}\{x_{\text{sc}}[n]\} = \text{DFT}\left\{\frac{1}{2}(x[n] + x^*[\text{mod}_N(-n)])\right\} = \text{Re}\{X[k]\} \quad k = 0, 1, \dots, N-1.$$

*Απόδειξη.* Με βάση τις ιδιότητες:

1 (γραμμικότητα) και 8 ( $\text{DFT}\{x^*[\text{mod}_N(-n)]\} = X^*[k]$ ) προκύπτει άμεσα ότι

$$\begin{aligned} \text{DFT}\{x_{\text{sc}}[n]\} &= \frac{1}{2}\text{DFT}\{x[n]\} + \frac{1}{2}\text{DFT}\{x^*[\text{mod}_N(-n)]\} = \\ &= \frac{1}{2}(X[k] + X^*[k]) = \text{Re}\{X[k]\} \quad k = 0, 1, \dots, N-1 \end{aligned}$$

**ΙΔΙΟΤΗΤΑ 12:** Έστω ότι  $X[k] = \text{DFT}\{x[n]\}$ . Τότε, ο DFT τού συζυγούς-αντισυμμετρικού μέρους τού  $x[n]$  είναι  $j\text{Im}\{X[k]\}$ . Σε συμβολική γραφή:

$$\text{DFT}\{x_{\text{ac}}[n]\} = \text{DFT}\left\{\frac{1}{2}(x[n] - x^*[\text{mod}_N(-n)])\right\} = j\text{Im}\{X[k]\} \quad k = 0, 1, \dots, N-1.$$

*Απόδειξη.* Πάλι με βάση τις ιδιότητες:

1 (γραμμικότητα) και 8 ( $\text{DFT}\{x^*[\text{mod}_N(-n)]\} = X^*[k]$ )

προκύπτει άμεσα ότι

$$\begin{aligned} \text{DFT}\{x_{Ac}[n]\} &= \frac{1}{2} \text{DFT}\{x[n]\} - \frac{1}{2} \text{DFT}\{x^*[\text{mod}_N(-n)]\} = \\ &= \frac{1}{2} (X[k] - X^*[k]) = j\text{Im}\{X[k]\} \quad k = 0, 1, \dots, N-1 \end{aligned}$$

- **Ιδιότητες για πεπερασμένες πραγματικές ακολουθίες.**

**ΙΔΙΟΤΗΤΑ 13:** Έστω ότι  $X[k] = \text{DFT}\{x[n]\}$  όπου η ακολουθία  $x[n]$  είναι πεπερασμένη και πραγματική. Τότε, ισχύει:

$$X[k] = X^*[\text{mod}_N(-k)], \quad k = 0, 1, 2, \dots, N-1 \Rightarrow$$

$$X[k] = X^*[N-k], \quad k = 1, 2, \dots, N-1.$$

*Απόδειξη.* Σύμφωνα με την ιδιότητα 10,

$$\text{DFT}\{j\text{Im}\{x[n]\}\} = \frac{1}{2} (X[\text{mod}_N(k)] - X^*[\text{mod}_N(-k)]).$$

Αλλά αφού η  $x[n]$  είναι πραγματική, έπεται ότι  $\text{Im}\{x[n]\} = 0 \Rightarrow$

$$\text{DFT}\{j\text{Im}\{x[n]\}\} = 0 \Leftrightarrow \frac{1}{2} (X[\text{mod}_N(k)] - X^*[\text{mod}_N(-k)]) = 0 \Leftrightarrow$$

$$X[k] = X^*[\text{mod}_N(-k)], \quad k = 0, 1, 2, \dots, N-1 \Rightarrow$$

$$X[k] = X^*[N-k], \quad k = 1, 2, \dots, N-1.$$

**ΙΔΙΟΤΗΤΑ 14:** Έστω ότι  $X[k] = \text{DFT}\{x[n]\}$  όπου η ακολουθία  $x[n]$  είναι πεπερασμένη και πραγματική. Τότε, ισχύει:

$$\text{Re}\{X[k]\} = \text{Re}\{X[\text{mod}_N(-k)]\}, \quad k = 0, 1, 2, \dots, N-1 \Rightarrow$$

$$\text{Re}\{X[k]\} = \text{Re}\{X[N-k]\}, \quad k = 1, 2, \dots, N-1$$

δηλαδή, το πραγματικό μέρος του DFT είναι συμμετρικό ως προς το κέντρο του διαστήματος  $[0, N]$  (που μπορεί να μην είναι ακέραιος).

Παρατήρηση: Το  $\text{Re}\{X[k]\}$  είναι συμμετρικό ως προς το κέντρο του διαστήματος  $[0, N]$ , όχι του  $[0, N-1]$  που είναι το πεδίο ορισμού του DFT. Δηλαδή, αναλυτικά,

$$\text{Re}\{X[1]\} = \text{Re}\{X[N-1]\}, \text{Re}\{X[2]\} = \text{Re}\{X[N-2]\}, \text{Re}\{X[3]\} = \text{Re}\{X[N-3]\},$$

κ.ο.κ, ενώ, στο διάστημα  $[0, N-1]$  το  $x[0]$  “είναι μόνο του”. Στο διάστημα  $[0, N]$ ,

προφανώς,  $X[0] = X^*[\text{mod}_N(v)] = X^*[0]$ , δηλαδή, το  $X[0]$  είναι πραγματικό.

Απόδειξη. Σύμφωνα με την ως άνω ιδιότητα 15,

$$X[k] = X^*[\text{mod}_N(-k)], \quad k = 0, 1, 2, \dots, N-1,$$

άρα, προφανώς,

$$\text{Re}\{X[k]\} = \text{Re}\{X^*[\text{mod}_N(-k)]\}, \quad k = 0, 1, 2, \dots, N-1.$$

Αλλά, βεβαίως, για κάθε μιγαδικό αριθμό ισχύει εξ ορισμού:

$$\text{Re}\{z\} = \text{Re}\{z^*\} \Rightarrow$$

$$\text{Re}\{X[k]\} = \text{Re}\{X[\text{mod}_N(-k)]\}, \quad k = 0, 1, 2, \dots, N-1.$$

**ΙΔΙΟΤΗΤΑ 15:** Έστω ότι  $X[k] = \text{DFT}\{x[n]\}$  όπου η ακολουθία  $x[n]$  είναι πεπερασμένη και πραγματική. Τότε, ισχύει:

$$\text{Im}\{X[k]\} = -\text{Im}\{X[\text{mod}_N(-k)]\}, \quad k = 0, 1, 2, \dots, N-1 \Rightarrow$$

$$\text{Im}\{X[k]\} = -\text{Im}\{X[N-k]\}, \quad k = 1, 2, \dots, N-1, \quad \wedge \quad X[0] \in \mathbf{R}$$

δηλαδή, το φανταστικό μέρος τού DFT είναι αντισυμμετρικό ως προς το κέντρο του διαστήματος  $[0, N]$  (που μπορεί να μην είναι ακέραιος).

Παρατήρηση: Το  $\text{Re}\{X[k]\}$  είναι αντισυμμετρικό ως προς το κέντρο τού διαστήματος  $[0, N]$ , όχι τού  $[0, N-1]$  που είναι το πεδίο ορισμού τού DFT. Δηλαδή, αναλυτικά,

$$\begin{aligned} \text{Im}\{X[1]\} &= -\text{Im}\{X[N-1]\}, \quad \text{Im}\{X[2]\} = -\text{Im}\{X[N-2]\}, \quad \text{Im}\{X[3]\} \\ &= -\text{Im}\{X[N-3]\} \end{aligned}$$

κ.ο.κ, ενώ, στο διάστημα  $[0, N-1]$  το  $x[0]$  “είναι μόνο του”. Στο διάστημα  $[0, N]$ ,

προφανώς,  $\text{Im}\{x[0]\} = -\text{Im}\{x[\text{mod}_N(v)]\} = -\text{Im}\{x[0]\} \Leftrightarrow \text{Im}\{x[0]\} = 0$ .

Απόδειξη. Σύμφωνα με την ως άνω ιδιότητα 15,

$$X[k] = X^*[\text{mod}_N(-k)], \quad k = 0, 1, 2, \dots, N-1,$$

άρα, προφανώς,

$$\text{Im}\{X[k]\} = \text{Im}\{X^*[\text{mod}_N(-k)]\}, \quad k = 0, 1, 2, \dots, N-1.$$

Αλλά, βεβαίως, για κάθε μιγαδικό αριθμό ισχύει εξ ορισμού:

$$\text{Im}\{z\} = -\text{Im}\{z^*\} \Rightarrow$$



$$\text{Im}\{X[k]\} = -\text{Im}\{X[\text{mod}_N(-k)]\}, \quad k = 0, 1, 2, \dots, N-1.$$

**ΙΔΙΟΤΗΤΑ 16:** Έστω ότι  $X[k] = \text{DFT}\{x[n]\}$  όπου η ακολουθία  $x[n]$  είναι πεπερασμένη και πραγματική. Τότε, ισχύει:

α)

$$|X[k]| = |X[\text{mod}_N(-k)]|, \quad k = 0, 1, 2, \dots, N-1 \Rightarrow$$

$$|X[k]| = |X[N-k]|, \quad k = 1, 2, \dots, N-1$$

β)

$$\arg(X[k]) = -\arg(X[\text{mod}_N(-k)]), \quad k = 0, 1, 2, \dots, N-1 \Rightarrow$$

$$\arg(X[k]) = -\arg(X[N-k]), \quad k = 1, 2, \dots, N-1 \Rightarrow$$

δηλαδή, το μέτρο του DFT είναι συμμετρικό ενώ το όρισμα τού DFT είναι αντισυμμετρικό ως προς το κέντρο του διαστήματος  $[0, N]$  (που μπορεί να μην είναι ακέραιος).

Η συμμετρία και η αντισυμμετρία θεωρούνται ως προς το κέντρο τού διαστήματος  $[0, N]$ , όχι του  $[0, N-1]$  που είναι το πεδίο ορισμού τού DFT.

*Απόδειξη.* Είναι άμεση συνέπεια της ιδιότητας 13:

$X[k] = X^*[\text{mod}_N(-k)]$ ,  $k = 0, 1, 2, \dots, N-1 \Rightarrow$  ως και των στοιχειωδών ιδιοτήτων των μιγαδικών αριθμών:

$$|z| = |z^*|, \quad \arg(z) = -\arg(z^*).$$

#### 4.1.2 Αλγόριθμος προσαρμοστικού/αναδρομικού (adaptive) DFT

Ας υποθέσουμε ότι έχουμε υπολογίσει τον Διακριτό Μετασχηματισμό Fourier ενός σήματος  $x[n]$  των  $N$  δειγμάτων, με σημείο εκκίνησης το σημείο  $a$  του πεδίου του χρόνου και σημείο περάτωσης στο σημείο  $(a+N-1)$ . Ας υποθέσουμε επιπλέον ότι επιθυμούμε να υπολογίσουμε τον μετασχηματισμό Fourier  $N$  δειγμάτων του ίδιου

σήματος  $x[n]$ , αλλά τώρα ξεκινώντας από το σημείο  $(a+1)$  και τελειώνοντας στο σημείο  $(a+N)$ . Ανεξάρτητα του τρόπου με τον οποίο έχει υπολογιστεί ο πρώτος μετασχηματισμός που αντιστοιχεί στο σημείο  $a$ , ο υπολογισμός του επόμενου μετασχηματισμού Fourier μπορεί να γίνει με πολύ λιγότερους υπολογισμούς, χρησιμοποιώντας την πληροφορία που διαθέτουμε από τον υπολογισμό του πρώτου DFT. Κατ' αρχήν σημειώνουμε ότι όπως διατυπώθηκε παραπάνω ο μετασχηματισμός Fourier ενός παραθύρου  $N$  δειγμάτων, που ξεκινά από το σημείο  $a$  και τελειώνει στο σημείο  $(a+N-1)$  δίνεται από την παρακάτω σχέση:

$$X[k] = \sum_{n=0}^{N-1} x[n+a]W^{kn} \quad (\text{A.5}), \text{ όπου } W = e^{-j\left(\frac{2\pi}{N}\right)}$$

ενώ ο μετασχηματισμός Fourier ενός παραθύρου  $N$  δειγμάτων, που ξεκινά από το σημείο  $a+1$  και τελειώνει στο σημείο  $(a+N)$  δίνεται από την πιο κάτω σχέση:

$$X_s[k] = \sum_{n=0}^{N-1} x[n+a+1]W^{kn} \quad (\text{A.6})$$

Όμως ευθύγραμμα προκύπτει ότι:

$$X_s[k] = (-x[a] + x[a]) * W^{-k} + \left( \sum_{n=0}^{N-2} x[n+a+1]W^{k(n+1)} \right) * W^{-k} + x[a+N] * W^{k(N-1)} \quad (\text{A.7})$$

Στο τελευταίο άθροισμα αντικαθιστούμε το  $n+1$  με την βοηθητική μεταβλητή  $i$

$$i = n + 1$$

οπότε στη συνέχεια λαμβάνουμε:

$$X_s[k] = -x[a] * W^{-k} + x[a] * W^{-k} + \left( \sum_{i=1}^{N-1} x[i+a]W^{ki} \right) * W^{-k} + x[a+N] * W^{k(N-1)} \quad (\text{A.8})$$

ή ισοδύναμα:

$$X_s[k] = -x[a] * W^{-k} + \left( \sum_{i=0}^{N-1} x[i+a]W^{ki} \right) * W^{-k} + x[a+N] * W^{k(N-1)} \quad (\text{A.9})$$

Αλλά η ποσότητα  $\sum_{i=0}^{N-1} x[i+a]W^{ki}$  είναι ο μετασχηματισμός Fourier του παραθύρου

μήκους  $N$  που ξεκινά από το δείγμα  $a$ . Επομένως, τελικά έχει αποδειχθεί ότι:

$$X_s[k] = -x[a] * W^{-k} + X[k] * W^{-k} + x[a + N] * W^{k(N-1)} \quad (\text{A.10})$$

το οποίο και χρησιμοποιούμε ως βασικό τύπο για τον αναδρομικό υπολογισμό του Διακριτού Μετασχηματισμού Fourier, με ιδιαίτερος σημαντική μείωση της υπολογιστικής πολυπλοκότητας.

Η υλοποίηση αυτού του αλγορίθμου γίνεται με τα εξής βήματα: Κατ' αρχήν παρατηρούμε ότι με χρήση του τύπου (A.10) κάθε δείγμα του μετασχηματισμού Fourier υπολογίζεται πολύ απλά. Πράγματι, οι ποσότητες  $W^{-k}$  και  $W^{k(N-1)}$  υπολογίζονται μόνο μια φορά στην αρχή της διαδικασίας. Στη συνέχεια, για την πραγματοποίηση του αναδρομικού υπολογισμού, για κάθε δείγμα απαιτούνται 6 πολλαπλασιασμοί πραγματικών αριθμών και 5 προσθαιρέσεις. Ο αλγόριθμος επομένως έχει συνολικά πολυπλοκότητα  $O(N)$  ή πιο συγκεκριμένα χρειάζεται  $6N$  πολλαπλασιασμούς. Αντιθέτως, ο απλός αλγόριθμος γρήγορου μετασχηματισμού Fourier έχει πολυπλοκότητα

$$O\left(\frac{N}{2} * \log_2(N)\right) \text{ αλλά σε μιγαδικές πράξεις, επομένως χρειάζεται } 2N * \log_2(N)$$

πολλαπλασιασμούς πραγματικών αριθμών. Προφανώς ο παρουσιαζόμενος αλγόριθμος είναι πολύ γρηγορότερος από τον κλασικό DFT. Για μέγεθος παραθύρου μεγαλύτερο από 16 δείγματα, ο προσαρμοστικός μετασχηματισμός Fourier είναι πάντα πιο γρήγορος, ενώ με το χρησιμοποιούμενο παράθυρο των  $56 * 1024$  δειγμάτων, είναι 4,5 περίπου φορές γρηγορότερος, αν κανείς λάβει υπόψη τον αριθμό των πλέον πολύπλοκων πράξεων, των πολλαπλασιασμών. Σύγκριση του απαιτούμενου αριθμού πράξεων για τον υπολογισμό του φάσματος, πρώτα μέσω του απλού DFT και εν συνεχεία με τη χρήση του προσαρμοστικού (Adaptive) DFT, για διάφορα μεγέθη παραθύρων, με τα οποία η εφαρμογή μπορεί να λειτουργήσει με πολύ καλά αποτελέσματα, φαίνεται στον πιο κάτω πίνακα.

	Μέγεθος παραθύρου			
	N=8*1024	N=16*1024	N=32*1024	N=56*1024
Αριθμός πράξεων για απλό DFT	212992	458752	983040	1812913
Αριθμός πράξεων για Adaptive DFT	49152	98304	196608	344064

Η πράξη επιβεβαίωσε ότι ο αλγόριθμος έδινε πραγματικά αποτελέσματα τάξεις φορές ταχύτερα από τον απλό DFT σύμφωνα με τα στοιχεία που παρουσιάζονται στον πίνακα. Σε αυτόν το χρόνο παίζουν ρόλο τόσο οι προσθέσεις και οι αφαιρέσεις, όσο και προβλήματα κατανομής και εκχώρησης μνήμης

Η τελική εφαρμογή του προσαρμοστικού αλγορίθμου ξεκινά με την εκτέλεση ενός μετασχηματισμού Fourier. Στη συνέχεια εφαρμόζεται ο προσαρμοστικός μετασχηματισμός Fourier, ενώ κάθε 30000 περίπου δείγματα, το σύστημα επαναρχικοποιείται, εφαρμόζοντας ξανά ένα κλασσικό DFT, για να αποφευχθεί η περίπτωση του αριθμητικού λάθους. Να σημειωθεί πάντως, ότι ο προσαρμοστικός DFT δοκιμάστηκε και σε πλήρεις μουσικές συνθέσεις, χωρίς καμία επανεκκίνηση, χωρίς να παρουσιάσει κανένα πρόβλημα αριθμητικής ευστάθειας και δίνοντας ακριβώς τα ίδια αποτελέσματα με τον κλασσικό DFT, όταν αυτός λειτουργεί χωρίς λάθος πεπερασμένης ακρίβειας. Διότι αξίζει να σημειωθεί ότι ο κλασσικός DFT μπορεί να αποτύχει τελείως για διάφορα σήματα εισόδου, λόγω του γεννώμενου λάθους πεπερασμένης ακρίβειας.

## 4.2 Διαχωρισμός σε νότες

Είναι γνωστό ότι το ανθρώπινο αυτί ακούει περιοχές συχνοτήτων και όχι συγκεκριμένες συχνότητες. Για να μιμηθούμε αυτή την ιδιότητα, χωρίζουμε όλο το φάσμα των συχνοτήτων σε ζώνες, όπου κάθε ζώνη αντιστοιχεί σε μία νότα. Εξάλλου είναι γνωστό ότι η μουσική κλίμακα είναι εκθετική.

Η εκθετική κατανομή των περιοχών συχνοτήτων, που δίνει το καλύτερο αποτέλεσμα, μπορεί να εξηγηθεί εύκολα αν σκεφθούμε το γεγονός ότι το ανθρώπινο αυτί ακούει με παρόμοιο τρόπο, καταλαβαίνει δηλαδή συχνοτικές περιοχές σχεδόν εκθετικά. Κλασσικό παράδειγμα είναι οι νότες των μουσικών οργάνων που βρίσκονται εκθετικά κατανεμημένες, με την έννοια ότι μια συγκεκριμένη νότα μιας οκτάβας αντιστοιχεί ακριβώς στη διπλάσια συχνότητα από την ίδια νότα της προηγούμενης οκτάβας. Για παράδειγμα, το ΛΑ της πρώτης οκτάβας, αυτό που γράφεται δηλαδή στο κλειδί του ΣΟΛ στη δεύτερη γραμμή του πενταγράμμου αντιστοιχεί στα 440 Hz, ενώ το ΛΑ της δεύτερης οκτάβας αντιστοιχεί στα 880 Hz. Αντίστοιχα, το ΛΑ κάτω από την πρώτη οκτάβα γράφεται στις βοηθητικές γραμμές κάτω από το πεντάγραμμο αντιστοιχεί στα 220 Hz.

Από την εκθετική μορφή της κατανομής των συχνοτικών περιοχών ξεφεύγουν μόνο οι πρώτες συχνοτικές περιοχές. Και αυτό το γεγονός μπορεί να εξηγηθεί. Αρχικά, είναι διαπιστωμένο, ότι σε χαμηλές συχνότητες το ανθρώπινο αυτί τείνει να ακούει περισσότερο γραμμικά παρά εκθετικά. Επιπλέον, το γεγονός της χρήσης του διακριτού μετασχηματισμού Fourier θα δημιουργούσε προβλήματα στις χαμηλές συχνότητες. Μια περιοχή χαμηλών συχνοτήτων ίδιου εύρους με μία περιοχή υψηλότερων συχνοτήτων αντιστοιχεί σε πολύ λιγότερα δείγματα του DFT. Για παράδειγμα, εάν χρησιμοποιήσουμε ένα παράθυρο 8192 δειγμάτων, μία περιοχή εύρους 32 Hz με κέντρο στα 440 Hz αντιστοιχεί σε 12 περίπου δείγματα, ενώ με κέντρο στα 55 Hz αντιστοιχεί μόνο σε 1,5 δείγματα.

Έτσι προκύπτει ο παρακάτω πίνακας ο οποίος δείχνει αναλυτικά τις διαμερίσεις σε νότες που επιλέχθηκαν. Η κατασκευή των διαμερίσεων - νοτών αυτών έγινε ξεκινώντας από την θεμελιώδη συχνότητα ΛΑ των 440 Hz, ανεβαίνοντας με βήμα  $2^{\frac{1}{4}}\sqrt{2}$  και κατεβαίνοντας με βήμα  $2^{\frac{1}{4}}\sqrt{2}$ .

Η πιο χαμηλή νότα βρίσκεται στη συχνότητα 53,43 Hz και η υψηλότερη στη

συχνότητα 7246,3Hz. Ως αναγνωριστικό της συχνοτικής περιοχής έχουμε λάβει το κάτω όριο του διαστήματος, ενώ αναθέτουμε σε κάθε περιοχή και έναν αριθμό από 1 έως 85.

Επιπλέον για κάθε νότα που συναντάμε και δεν ανήκει σε κάποια από τις συχνοτικές περιοχές κατατάσσεται ως περιοχή 0 αν είναι μικρότερη από 53,43 Hz και ως περιοχή 86 αν είναι μεγαλύτερη από 7246,3 Hz.

Αύξων αριθμός διαμέρισης	Διάστημα συχνοτήτων(Hz)	Αύξων αριθμός διαμέρισης	Διάστημα συχνοτήτων (Hz)
0	< 53,43	44	640,48-678,57
1	53,43-56,61	45	678,57-718,92
2	56,61-59,97	46	718,92-761,67
3	59,97-63,54	47	761,67-806,96
4	63,54-67,32	48	806,96-854,94
5	67,32-71,32	49	854,94-905,78
6	71,32-75,56	50	905,78-959,64
7	75,56-80,06	51	959,64-1016,71
8	80,06-84,82	52	1016,71-1077,17
9	84,82-89,86	53	1077,17-1141,23
10	89,86-95,20	54	1141,23-1209,17
11	95,20-100,87	55	1209,17-1280,97
12	100,87-106,86	56	1280,97-1357,14
13	106,86-113,22	57	1357,14-1437,81
14	113,22-119,95	58	1437,81-1523,34
15	119,95-127,08	59	1523,34-1613,93
16	127,08-134,64	60	1613,93-1709,92
17	134,64-142,65	61	1709,92-1811,57
18	142,65-151,13	62	1811,57-1919,33
19	151,13-160,12	63	1919,33-2033,48
20	160,12-169,64	64	2033,48-2154,33
21	169,64-179,73	65	2154,33-2282,47

22	179,73-190,41	66	2282,47-2418,21
23	190,41-201,74	67	2418,21-2561,94
24	201,74-213,73	68	2561,94-2714,37
25	213,73-226,44	69	2714,37-2875,69
26	226,44-239,91	70	2875,69-3046,68
27	239,91-254,17	71	3046,68-3227,91
28	254,17-269,29	72	3227,91-3419,83
29	269,29-285,30	73	3419,83-3623,14
30	285,30-302,27	74	3623,14-3838,58
31	302,27-320,24	75	3838,58-4066,84
32	320,24-339,28	76	4066,84-4308,77
33	339,28-359,46	77	4308,77-4564,87
34	359,46-380,83	78	4564,87-4836,31
35	380,83-403,48	79	4836,31-5123,89
36	403,48-427,47	80	5123,89-5428,58
37	427,47-452,89	81	5428,58-5751,38
38	452,89-479,82	82	5751,38-6093,37
39	479,82-508,35	83	6093,37-6455,70
40	508,35-538,58	84	6455,70-6839,58
41	538,58-570,60	85	6839,58-7246,31
42	570,60-604,53	86	> 7246,3
43	604,53-640,48		

### 4.3 Η έννοια του *masking*, για τον εντοπισμό των ακουστικά σημαντικών κορυφών και αποθορυβοποίηση

Το ακουστικό σύστημα του ανθρώπου έχει μερικές πολύ ενδιαφέρουσες ιδιότητες τις οποίες κανείς εκμεταλλεύεται στην κωδικοποίηση και συμπίεση των ηχητικών σημάτων, όπως γίνεται στο πρότυπο MPEG – 3 (MP3). Πιο συγκεκριμένα:

Το ανθρώπινο αυτί ακούει ήχους που ανήκουν σε δυναμικό εύρος συχνοτήτων από 20 μέχρι 20000 Hz και με ένταση που ποικίλλει σε μια πολύ μεγάλη κλίμακα. Επιπλέον, το ανθρώπινο ακουστικό σύστημα έχει μεγάλες δυνατότητες προσαρμογής, σε βαθμό που αυτό που ο άνθρωπος ακούει επηρεάζεται άμεσα από το ηχητικό περιβάλλον. Για παράδειγμα, στην παρουσία ενός πολύ ισχυρού λευκού θορύβου, πιο αδύναμοι ήχοι «καλύπτονται» και έτσι δεν μπορούμε να τους ακούσουμε καθόλου. Αυτό το ακουστικό φαινόμενο, κατά το οποίο ισχυρής εντάσεως ήχοι καλύπτουν άλλους χαμηλότερης εντάσεως αναφέρεται στην διεθνή βιβλιογραφία ως *masking*.

Συγκεκριμένα, αναφέρονται στη διεθνή βιβλιογραφία διάφορα πειράματα στα οποία» κατάλληλος θόρυβος προστέθηκε σε ηχητικά δεδομένα, με SNR μέχρι 25 dB, χωρίς να υπάρξει καμιά επίδραση στο ηχητικό αποτέλεσμα. Απ' την άλλη μεριά, θόρυβος με SNR 90 dB, δηλ. θόρυβος με υψηλή ενέργεια σε ευαίσθητες περιοχές, μπορεί να γίνει ακουστός.

Πολλά από τα χαρακτηριστικά του *masking* οφείλονται στην ύπαρξη κρίσιμων ζωνών συχνοτήτων που χαρακτηρίζουν την ανθρώπινη ακοή, σύμφωνα με όσα προαναφέραμε. Αυτές οι συχνοτικές ζώνες δεν λειτουργούν σαν διακόπτες αποκοπής αλλά αντιθέτως αντιστοιχούν σε φίλτρα που το πλάτος τους είναι συνάρτηση της συχνότητας.

Η συνάρτηση μεταβολής του πλάτους των κρίσιμων συχνοτικών ζωνών προέκυψε από την εφαρμογή του *masking* σε λευκό θόρυβο (βλέπε [13]). Δεδομένου ότι παρατηρήθηκε πως μία συχνότητα «καλύπτει» το θόρυβο όταν η ισχύς του θορύβου στη συγκεκριμένη κρίσιμη συχνοτική ζώνη είναι ίση με το 1/4 της ισχύος της υπό εξέταση συχνότητας, τα πλάτη των κρίσιμων συχνοτικών ζωνών ορίστηκαν ως εξής:

$$BW(f) = \begin{cases} 100 \text{ Hz} & f < 500 \text{ Hz} \\ 0.2f \text{ Hz} & f \geq 500 \text{ Hz} \end{cases} \quad (\text{A.1})$$



Αυτά τα πλάτη όρισαν μία κλίμακα συχνοτικών περιοχών αρκετά παρόμοια με την εκθετική μουσική κλίμακα. Απευθείας μετατροπή μεταξύ αυτής της κλίμακας συχνοτήτων (Bark) και Hz μπορεί να γίνει με χρήση της εξίσωσης [13] [14]

$$z(f) = 13 * \arctan(0.00076f) + 3.5 * \arctan\left(\left(\frac{f}{7500}\right)^2\right) \quad (\text{A.2})$$

Το  $z$  θα δηλώνει από εδώ και πέρα την συχνότητα σε Bark ενώ το  $f$  την συχνότητα σε Hz.

Με τη χρήση των παραπάνω, στα [13] και [14] αναφέρεται ένα σημαντικό πείραμα masking σύμφωνα με το οποίο ένας masker «καλύπτει» όλες τις συχνότητες που ανήκουν στην ίδια κρίσιμη ζώνη σχεδόν με τον ίδιο τρόπο.

Το masking βέβαια από έναν masker δεν περιορίζεται μόνο στη κρίσιμη ζώνη στην οποία αυτός ανήκει, αλλά επεκτείνεται και στις γειτονικές ζώνες. Ορίστηκε, λοιπόν, μία συνάρτηση διασποράς  $SF(z, a)$ , με  $z$  την συχνότητα και  $a$  το πλάτος του masker. Με αυτή τη συνάρτηση μπορεί να υπολογιστεί ένα κατώφλι masking από έναν και μόνο masker για τις γειτονικές του συχνότητες.

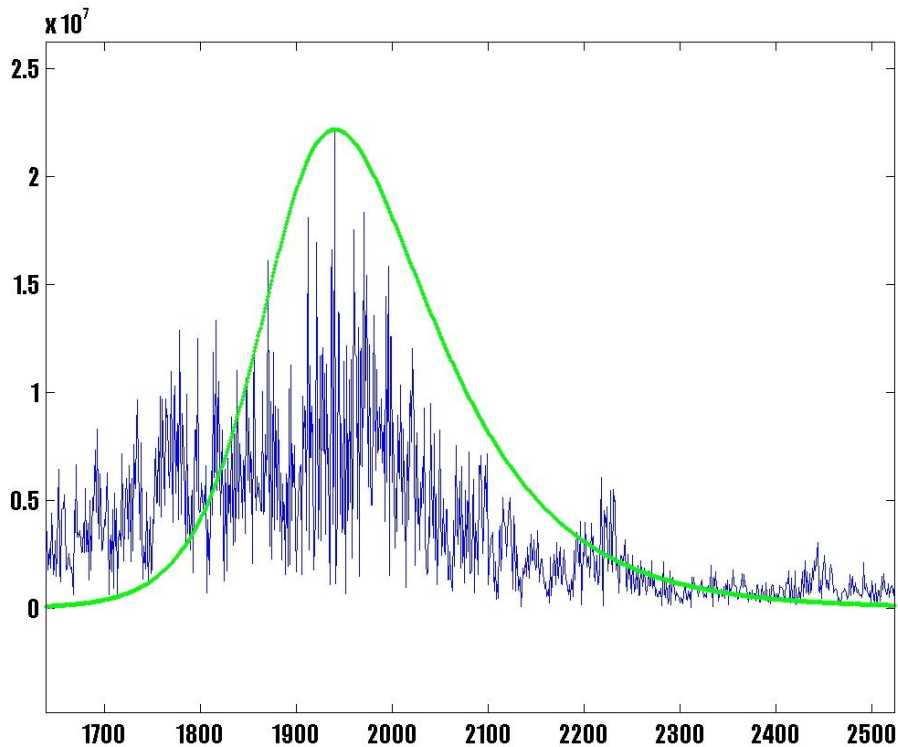
Μία συνάρτηση διασποράς η οποία λαμβάνει υπ' όψη τόσο το masking στην ίδια κρίσιμη ζώνη όσο και έξω από αυτή (βλέπε [14]) είναι η

$$F(z) = 15.81 + 7.5 * (z + 0.474) - 17.5 * \sqrt{1 + (z + 0.474)^2} \quad (\text{A.3})$$

με

$$z = 13 * \arctan(0.00076(f - f_0)) + 3.5 * \arctan\left(\left(\frac{f - f_0}{7500}\right)^2\right) \quad (\text{A.4})$$

η συχνότητα σε Bark,  $f_0$  η συχνότητα γύρω από την οποία εφαρμόζεται η συνάρτηση διασποράς του masking σε Hz και  $f$  η συχνότητα στις οποίες εφαρμόζεται το masking σε Hz (βλέπε το παρακάτω σχήμα).



Η διαδικασία εφαρμογής του masking που εφαρμόστηκε περιγράφεται ακολούθως:

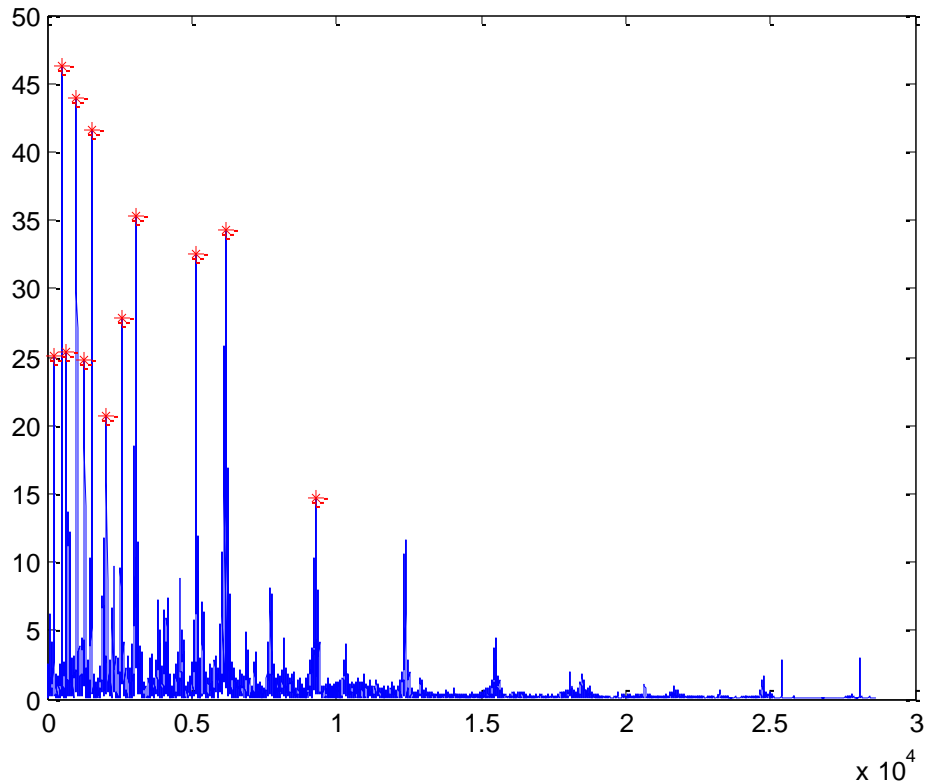
Σε κάθε παράθυρο του μουσικού σήματος, είτε της εκτέλεσης από ορχήστρα είτε της σόλο εκτέλεσης υπολογίζουμε το Διακριτό Μετασχηματισμό Fourier (DFT). Μετά τον υπολογισμό της απόλυτης τιμής του DFT εντοπίζουμε τις κορυφές (μέγιστα) αυτού και αποθηκεύουμε τις θέσεις τους στο φάσμα καθώς και τα αντίστοιχα πλάτη τους. Ταξινομούμε αυτές τις κορυφές ανάλογα με το πλάτος τους, κρατάμε αυτές που βρίσκονται πάνω από ένα ορισμένο κατώφλι και στη συνέχεια χρησιμοποιούμε τη συνάρτηση διασποράς του masking όπως περιγράφεται παρακάτω: Χρησιμοποιούμε τις σχέσεις (A.3) και (A.4) ξεκινώντας από τη συχνότητα-κορυφή  $f_0$  με το μεγαλύτερο πλάτος. Εάν το πλάτος μιας κορυφής στη συχνότητα  $f$  είναι μικρότερο από την τιμή της συνάρτησης διασποράς  $F(f)$  στη συγκεκριμένη θέση, τότε διαγράφουμε την κορυφή  $f$  από τη λίστα με τις ταξινομημένες κορυφές. Επαναλαμβάνουμε την ίδια διαδικασία επιλέγοντας ως  $f_0$  τη συχνότητα της αμέσως μικρότερης κορυφής που έχει εναπομένει στη λίστα μέχρι να εξαντλήσουμε όλες τις κορυφές. Η ανανεωμένη πλέον λίστα των κορυφών για κάθε μουσικό τμήμα φυλάσσεται προκειμένου να εφαρμοσθεί η διαδικασία που θα περιγραφεί σε επόμενη ενότητα.

Μία σημαντική διαφορά της χρησιμοποιούμενης τύπου masking (“masking-like”) διαδικασίας από την αυθεντική, που παρουσιάστηκε εκτενώς προηγουμένως, είναι το γεγονός ότι περιορίζουμε κάθε φορά τη δράση της συνάρτησης διασποράς του masking μεταξύ των συχνοτήτων  $f_0 - W$  και  $f_0 + W$ , με  $f_0$  την συχνότητα masker και  $W$  κατάλληλο επιλεγμένο εύρος. Τα πειράματα που πραγματοποιήσαμε δείχνουν ότι μια καλή τιμή του  $W$  είναι εκείνη που αντιστοιχεί σε 400 δείγματα σε παράθυρο παρατήρησης  $56 \cdot 1024$  δειγμάτων. Σύμφωνα με τα πειράματα που πραγματοποιήσαμε ως τώρα, παρατηρήσαμε ότι όσο μικρότερη είναι η τιμή του  $W$  τόσο μεγαλύτερη είναι η διακριτική ικανότητα του συστήματος, τόσο μεγαλύτερη είναι όμως παρ’ όλα αυτά και η δυσκολία στην αναγνώριση μουσικών σημάτων που οι εκτελέσεις τους διαφέρουν σημαντικά μεταξύ τους.

Ουσιαστικά η διαδικασία του masking μας βοηθάει αν εξαλείψουμε σημαντικά το θόρυβο, δεδομένου ότι αυτή διαγράφει όλο το «περιττό» συχνοτικό περιεχόμενο μιας σύνθεσης που βρίσκεται εκατέρωθεν μιας σημαντικής σε πλάτος κορυφής.

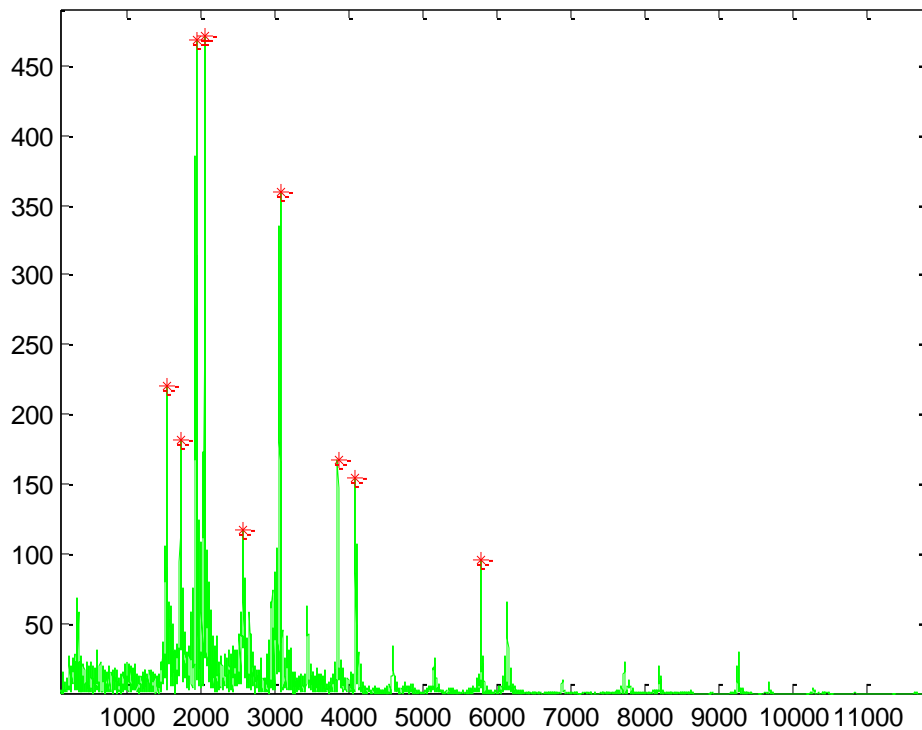
Μια σημαντική συμβολή του masking στη διαδικασία αναγνώρισης έγκειται στο γεγονός ότι βοηθάει στον εντοπισμό των ακουστικά πιο σημαντικών κορυφών και ταυτόχρονα ελαχιστοποιεί την ποσότητα πληροφορίας που πρέπει να κρατήσουμε για κάθε σήμα. Ας αναλογιστεί κανείς πόσο μεγάλο υπολογιστικό φόρτο μπορούμε να αποφύγουμε όταν αντί για όλα τα δείγματα ενός σήματος (που είναι εκατοντάδες χιλιάδες όταν μιλάμε για μουσικό σήμα), μπορούμε να κρατήσουμε μόνο 10 ή 20 ανά παράθυρο (δηλαδή σε σύνολο  $56 \cdot 1024$ ), τα οποία να προσδιορίζουν το σήμα.

Στο παρακάτω σχήμα φαίνεται η διαδικασία του masking για ένα παράθυρο  $56 \cdot 1024 \dots$ . Απεικονίζεται δηλαδή ο DFT του σήματος και οι κορυφές που επιστρέφει η διαδικασία του masking με αστερίσκο. Το εύρος της «καμπάνας» μετά από πειραματισμούς θεωρήθηκε ότι δίνει βέλτιστα αποτελέσματα όταν είναι 400.



Επίσης, η εφαρμογή των συναρτήσεων του masking που αναφέρθηκαν και παραπάνω δεν εφαρμόζονται σε όλο το σήμα, αλλά μόνο για τις κορυφές που βρίσκονται πάνω από ένα κατώφλι πλάτους. Αυτό συμβαίνει γιατί κορυφές κάτω από ένα κατώφλι θεωρούνται θόρυβος. Στη συγκεκριμένη διπλωματική, το κατώφλι που χρησιμοποιήσαμε μετά από δοκιμές ήταν 1.5 φορές το μέσο όρο του πλάτους όλου του φάσματος συν 3.5 φορές την τυπική απόκλιση αυτού.

Άλλο ένα παράδειγμα εφαρμογής του masking είναι αυτό που εμφανίζεται στο κάτωθι σχήμα:



Το συγκεκριμένο αφορά ένα παράθυρο του σήματος της εκτέλεσης από σόλο όργανο. Παρατηρούμε και εδώ ότι οι κορυφές που βρίσκονται κάτω από ένα κατώφλι ουσιαστικά διαγράφονται και δεν ελέγχονται από τη διαδικασία του masking. Για τη σόλο σύνθεση μία καλή τιμή για το κατώφλι θεωρήθηκε το ο μέσος όρος του πλάτους ολόκληρου του φάσματος συν 4.5 φορές την τυπική απόκλιση αυτού.

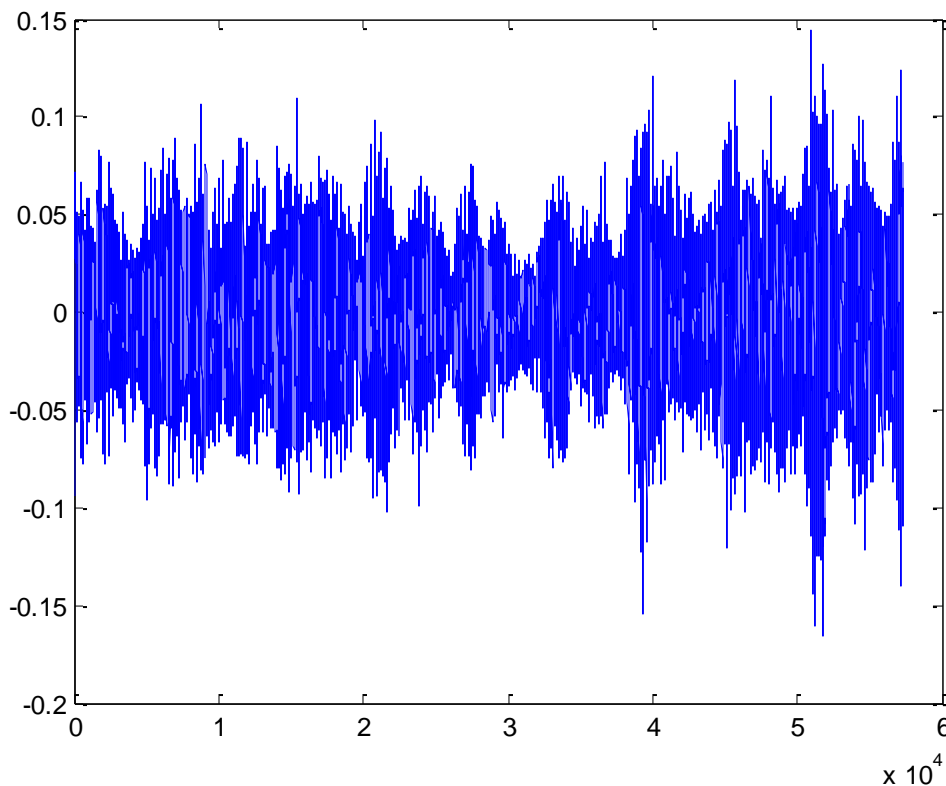
#### ***4.4 Η έννοια του stretch (ομοιοθεσία) για την αντιστάθμιση της διαφοράς ρυθμού και τονικότητας***

Όπως έχει αναφερθεί και σε προηγούμενη ενότητα, η μουσική σύνθεση εκτελεσμένη από ορχήστρα διαφέρει σε πάρα πολλά σημεία από τη σόλο εκτέλεση. Κάποια από τα προβλήματα που δημιουργούνται λόγω των διαφορετικών εκτελέσεων είναι επιλύσιμα,

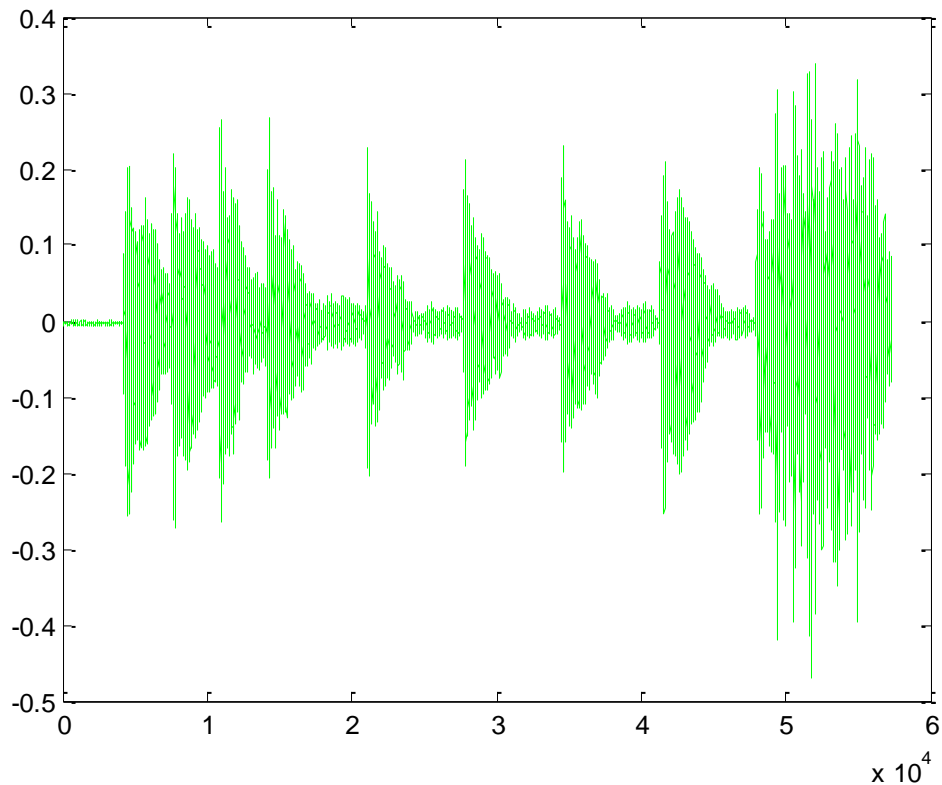
όπως για παράδειγμα η διαφορά στο ρυθμό εκτέλεσης ή η τονικότητα. Αυτά τα προβλήματα έρχεται να επιλύσει η τεχνική του stretching (στα ελληνικά ομοιοθεσία).

Πολλές φορές εμφανίζεται αλλαγή ρυθμού από εκτέλεση σε εκτέλεση, η οποία μπορεί να εντοπίζεται σε μερικά σημεία της σύνθεσης ή και να απλώνεται σε όλη της την έκταση. Σε όλες τις εκτελέσεις που μελετήθηκαν στην παρούσα εργασία, υπήρχαν έως και σημαντικές διαφορές στο ρυθμό σε διάφορα σημεία της εκτέλεσης, αλλά και συνολικά. Γεννάται τώρα το ερώτημα: Αυτή η υποκειμενική αίσθηση αλλαγής ρυθμού, εν πολλοίς απρόβλεπτη αντανακλά σε μαθηματικά χαρακτηριστικά του σήματος μουσικής; Κι αν πράγματι αυτό συμβαίνει, τότε διερωτάται κανείς η αλλαγή ρυθμού αντανακλά περισσότερο στην παράσταση του μουσικού σήματος στο πεδίο του χρόνου ή στην παράσταση του στο πεδίο της συχνότητας;

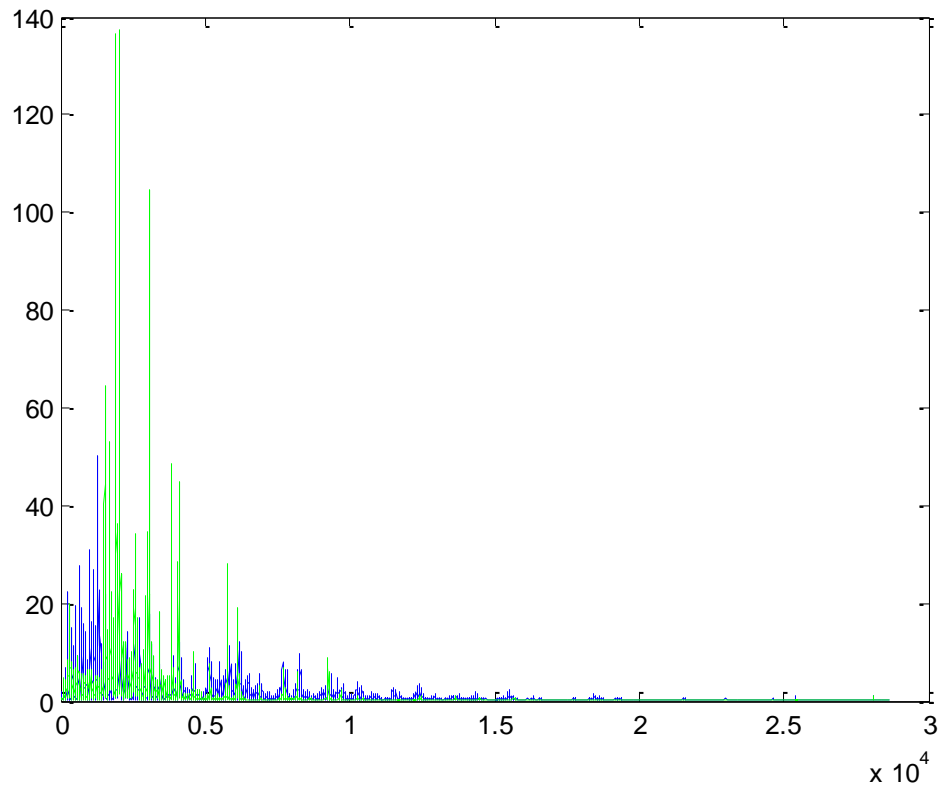
Οι παρατηρήσεις και η ανάλυση που έγιναν ως τώρα δείχνουν τα εξής: Έστω για παράδειγμα ένα τμήμα μουσικής σύνθεσης διάρκειας περίπου 2.5 δευτερολέπτων εκτελεσμένο από συμφωνική ορχήστρα, σαν αυτό που παρουσιάζεται στο σχήμα:



Ας θεωρήσουμε παράλληλα το ίδιο περίπου τμήμα της μουσικής σύνθεσης παιγμένο από ένα σόλο όργανο π.χ. κιθάρα ή πιάνο. Στο παρακάτω σχήμα φαίνεται η εκτέλεση από πιάνο.

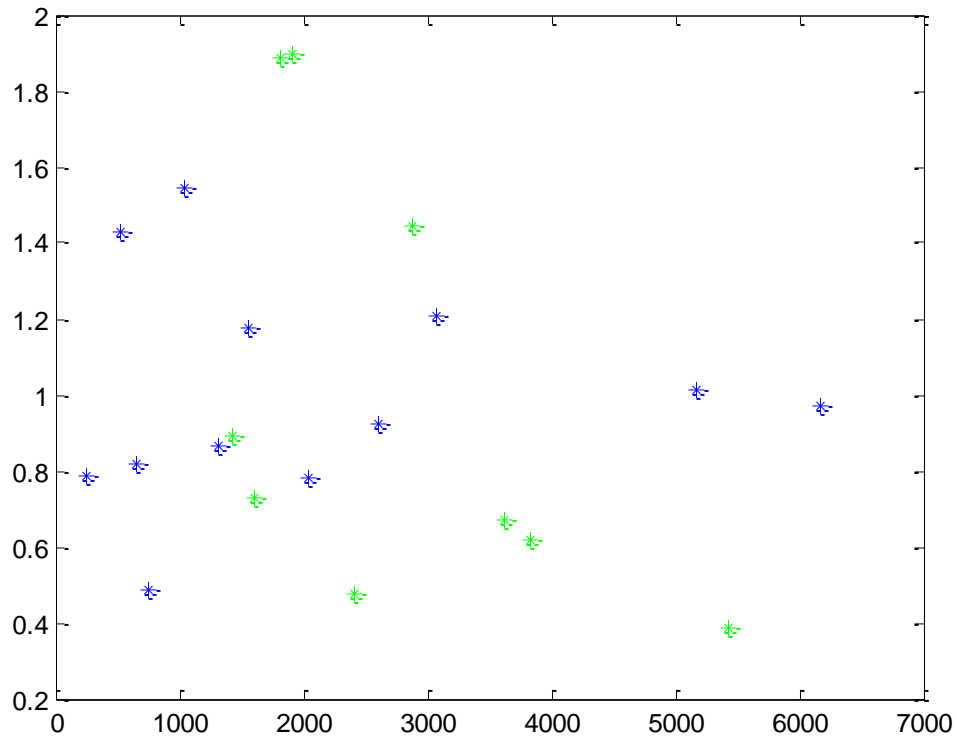


Θα θεωρήσουμε ότι το τμήμα αυτής της σύνθεσης που έχει εκτελεστεί από τη συμφωνική ορχήστρα, είναι χωρίς καμιά βλάβη της γενικότητας η πρότυπη σύνθεση, την οποία θα συμβολίσουμε ως ΠΕ (πρότυπη εκτέλεση). Αντιστρόφως, το ίδιο τμήμα εκτελεσμένο από το σόλο όργανο θα το συμβολίσουμε ΣΕ (σόλο εκτέλεση). Αν εφαρμόσουμε το διακριτό μετασχηματισμό Fourier και στο ΠΕ και στο ΣΕ λαμβάνουμε τα αποτελέσματα που φαίνονται στο παρακάτω σχήμα:

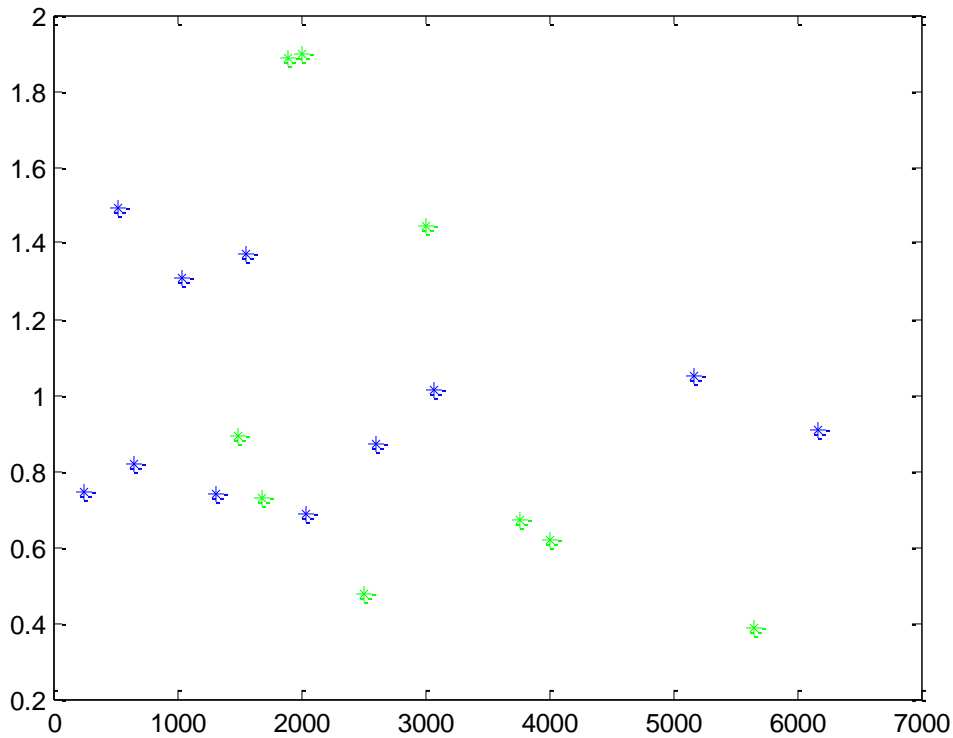


Στο παρακάτω σχήμα παρουσιάζονται οι κορυφές της ΠΕ και της ΣΕ που επιβιώνουν μετά η διαδικασία του masking. Με μπλε χρώμα είναι οι κορυφές της ΠΕ και με πράσινο οι κορυφές της ΣΕ ( στην προκειμένη περίπτωση είναι σόλο πιάνο).





Από αυτό το σχήμα φαίνεται ότι οι κορυφές των ΠΕ και ΣΕ που θα έπρεπε να συμπίπτουν, έστω κάποιες από αυτές, εάν δεν υπήρχε αλλαγή ρυθμού, εδώ διαφοροποιούνται. Επιπλέον παρατηρήσαμε ότι αν εφαρμόσουμε ομοιοθεσία στις κορυφές της ΣΕ με κέντρο ομοιοθεσίας την αρχή των αξόνων και κατάλληλο συντελεστή ομοιοθεσίας, τότε ένα σημαντικό πλήθος ζευγών αντιστοίχων κορυφών θα συμπέσουν στην ίδια νότα. Αυτό μπορούμε να το διαπιστώσουμε στο αμέσως επόμενο σχήμα:



Αυτό υποβάλλει την ιδέα ότι η αλλαγή του ρυθμού κατά την εκτέλεση του ΣΕ προκαλεί μία ομοιοθεσία στις τιμές των κορυφών της ΣΕ με κέντρο ομοιοθεσίας την αρχή των αξόνων και κατάλληλο συντελεστή ομοιοθεσίας. Επιπλέον για την επιβεβαίωση των ανωτέρω επιτελέσαμε το κάτωθι πείραμα: Πολλαπλασιάσαμε όλες τις μιγαδικές τιμές του  $DFT\{ΠΕ\}$  με έναν συντελεστή  $f$  μεταξύ  $1+-0.01$  έως  $1+- 0.09$  και για κάθε τέτοια τιμή του  $f$ , εφαρμόσαμε τον αντίστροφο μετασχηματισμό Fourier  $IDFT(f*DFT(ΣΕ))$  και ακούσαμε το πραγματικό μέρος αυτού. Το ακουστικό αποτέλεσμα ήταν πολύ κοντά στην αρχική σύνθεση ΠΕ, σαν να ήταν όμως εκτελεσμένη πιο γρήγορα ή πιο αργά. Περαιτέρω απόκλιση του  $f$  από τη μονάδα και ειδικότερα  $\geq 15\%$  αλλοίωσε πολύ ενοχλητικά το ακουστικό αποτέλεσμα.

Επομένως μπορούμε να χρησιμοποιήσουμε τις ανωτέρω παρατηρήσεις και ανάλυση, για να αντισταθμίσουμε πιθανές αλλαγές του ρυθμού μεταξύ μίας τυχαίας εκτέλεσης όπως η ΠΕ και μίας αντίστοιχής της ΣΕ που είναι παιγμένη διαφορετικά.

Πράγματι, αυτό γίνεται επιτελώντας την εξής σειρά βημάτων:

- **Βήμα 1:** Επιτελούμε το διακριτό μετασχηματισμό Fourier της ΣΕ, πολλαπλασιάζουμε όλα τα στοιχεία αυτού με  $f$  και υπολογίζουμε το μέτρο αυτού. Το  $f$  είναι ο συντελεστής ομοιοθεσίας που αντιπροσωπεύει την ενδεχόμενη αλλαγή ρυθμού.
- **Βήμα 2:** Εντοπίζουμε όλες τις κορυφές αυτού του ομοιοθετημένου απόλυτου DFT στο ήμισυ του πεδίου ορισμού του, λόγω συμμετρίας, οι οποίες έχουν τιμή πάνω από ένα κατώφλι. Το κατώφλι αυτό επελέγη να είναι η μέση τιμή του  $\text{abs}(f * \text{DFT}\{\Sigma E\}) + 3.5 * \text{την τυπική απόκλιση αυτού}$ , δεδομένου ότι το 99,9% ενός κανονικού πληθυσμού βρίσκεται κάτω από την τιμή αυτή. Επομένως εάν εύλογα αποδεχτούμε ότι ο θόρυβος υποβάθρου (background noise) ακολουθεί κανονική κατανομή, η χρήση αυτού του κατωφλίου πρακτικά αποκλείει κορυφές που οφείλονται σε θόρυβο.
- **Βήμα 3:** Με εφαρμογή της προαναφερθείσας διαδικασίας masking εντοπίζουμε τις  $N$  μεγαλύτερες κορυφές του  $\text{abs}(f * \text{DFT}\{\Sigma E\})$  στο ήμισυ του πεδίου ορισμού του. Τα πειράματα και οι παρατηρήσεις που πραγματοποιήθηκαν έδειξαν ότι κατ' αρχήν μια καλή τιμή για το  $N$  είναι 26.
- **Βήμα 4:** Αντιστοιχούμε αυτές τις  $N$  μεγαλύτερες κορυφές σε νότες όπως περιγράφηκε στο κεφάλαιο 4.2
- **Βήμα 5:** Αίρουμε τις πολλαπλές εμφανίσεις των ίδιων νοτών, δημιουργώντας μία μόνο ακολουθία διαφορετικών νοτών, πλήθους μικρότερου ίσου ενός άλλου κατάλληλα επιλεγμένου αριθμού  $NN$ . Τα πειράματα έδειξαν ότι μια καλή τιμή για το  $NN$  είναι 16.
- **Βήμα 6:** Εφαρμόζουμε τα προηγούμενα για μια ακολουθία τιμών του συντελεστή ομοιοθεσίας, έστω  $f_i, i=1,2,\dots$
- **Βήμα 7:** Για κάθε  $f_i$  υπολογίζουμε τον κοινό αριθμό νοτών της ΠΕ αφενός και της ομοιοθετημένης ΣΕ αφετέρου. Εκείνο το  $f_i$  που δίνει το μεγαλύτερο πλήθος κοινών νοτών, θεωρούμε ότι αντιπροσωπεύει την τροποποίηση του ρυθμού η οποία βέλτιστα κάνει την ΣΕ να ακούγεται σε πολύ παραπλήσιο ρυθμό με την ΠΕ.



# 5

## *Πρωτότυπες επεκτάσεις των τεχνικών και αλγορίθμων*

### *5.1 Αλλαγή οκτάβας*

Είναι πιθανό η μουσική σύνθεση εκτελεσμένη από ορχήστρα, έστω πάλι ΠΕ, και η μουσική σύνθεση εκτελεσμένη από σόλο όργανο, έστω ΣΕ, να βρίσκονται σε διαφορετικές οκτάβες. Αυτό συμβαίνει είτε διότι η ίδια η έκταση του οργάνου δεν περιλαμβάνει την οκτάβα στην οποία είναι εκτελεσμένη η ΠΕ, είτε διότι ο εκτελεστής επιλέγει κάποια άλλη οκτάβα επειδή το όργανο εκεί ηχεί καλύτερα. Είναι απαραίτητο να επεκτείνουμε την προηγούμενη μεθοδολογία ώστε να συμπεριλάβει και την αλλαγή οκτάβας για τους εξής λόγους: Ενώ και στην ΠΕ και στην ΣΕ πρόκειται για την ίδια νότα απλά παιγμένη σε διαφορετική οκτάβα, οι δύο νότες-κορυφές θεωρούνται διαφορετικές σύμφωνα με την παράγραφο 4.2 όπου περιγράφηκε ο διαχωρισμός σε νότες.

Συγκεκριμένα οι δύο αύξοντες αριθμοί που υποδηλώνουν τη νότα θα έχουν διαφορά μεταξύ τους πολλαπλάσιο του 12 (φυσικά αν πρόκειται για την ίδια νότα σε

άλλη οκτάβα). Για παράδειγμα, αν το ΛΑ μιας οκτάβας έχει τον αριθμό 15, το ΛΑ# θα έχει τον αριθμό 16, κοκ, το ΣΟΛ θα έχει τον αριθμό 26 και άρα το ΛΑ της επόμενης οκτάβας θα έχει τον αριθμό  $15+12=27$ .

Το ίδιο συμβαίνει και για τις παρακάτω οκτάβες. Δηλαδή, σύμφωνα με το προηγούμενο παράδειγμα, αν θέλουμε να βρούμε τον αύξοντα αριθμό του ΛΑ της χαμηλότερης οκτάβας θα είναι  $15-12=3$ .

Με αυτόν τον τρόπο εξασφαλίζουμε ότι η αναγνώριση μεταξύ δύο εκτελέσεων της ίδιας μουσικής σύνθεσης θα είναι επιτυχημένη ανεξάρτητα από τις οκτάβες στις οποίες εκτελούνται.

## ***5.2 Αριθμός κορυφών $N$ ανά παράθυρο***

Μία πολύ σημαντική παράμετρος στη διαδικασία ταυτοποίησης μουσικής σύνθεσης είναι ο αριθμός κορυφών  $N$  ανά παράθυρο. Όπως αναφέρθηκε και στις παρατηρήσεις που προέκυψαν από τα διαγράμματα στο 3<sup>ο</sup> κεφάλαιο, η εκτέλεση από συμφωνική ορχήστρα είναι πολύ πιο πλούσια σε ηχόχρωμα και σε βάθος από την εκτέλεση από σόλο όργανο, γεγονός που αντικατοπτρίζεται και στην παράσταση στο πεδίο της συχνότητας. Όπως φάνηκε από τα διαγράμματα, ο DFT της ΠΕ εκτεινόταν σε πολύ μεγαλύτερο εύρος συχνοτήτων από ότι η ΣΕ.

Ως εκ τούτου έχει πολύ μεγάλη σημασία ο αριθμός  $N$  των κορυφών του DFT που θα κρατάμε σε κάθε παράθυρο τόσο για την ΠΕ όσο και για τη ΣΕ αφού προηγηθεί η διαδικασία του masking, η οποία αναφέρθηκε αναλυτικά στο κεφάλαιο 4. Κατ' αρχάς επισημαίνεται ότι εξαιτίας των προηγούμενων παρατηρήσεων, ο αριθμός των κορυφών που επιστρέφει η διαδικασία του masking (που εφαρμόζεται τόσο στην πρότυπη εκτέλεση όσο και στη σόλο εκτέλεση) δεν είναι ο ίδιος στην πρότυπη εκτέλεση και στη σόλο εκτέλεση. Συνήθως στην περίπτωση της ΠΕ είναι μεγαλύτερος για όλους τους παραπάνω λόγους.

Έγιναν αρκετοί πειραματισμοί σχετικά με τον αριθμό  $N$  των κορυφών που θα χρησιμοποιούνται κατά τη διαδικασία της αναγνώρισης. Αρχικά είχαμε επιλέξει να

κρατάμε όλες τις κορυφές που επιστρέφει η διαδικασία του *masking*, οι οποίες ήταν πάνω από ένα κατώφλι τόσο για την ΠΕ όσο και για τη ΣΕ. Αυτή η επιλογή δεν απέδωσε πολύ καλά αποτελέσματα διότι η ΠΕ, όπως αναφέρθηκε και παραπάνω είχε πολύ μεγαλύτερο αριθμό κορυφών από τη ΣΕ και ως εκ τούτου σημειώνονταν μεγάλα ποσοστά αναγνώρισης, σε μέρη όμως που δεν αντιστοιχούσαν ακουστικά.

Ας είναι  $N1$  το πλήθος των κορυφών της ΠΕ, οι οποίες επιβίωσαν μετά τη διαδικασία του *masking* και που έχουν πλάτος μεγαλύτερο από το προαναφερθέν κατώφλι. Αποφασίσαμε να κρατήσουμε τις πιο σημαντικές κορυφές, τις οποίες επιλέξαμε ως εξής: Υπολογίσαμε το μέσο όρο του πλάτους αυτών των  $N1$  κορυφών, έστω  $\mu_1$  και την τυπική απόκλισή τους, έστω  $\sigma_1$ . Θεωρούμε ότι η πλέον σημαντικές από αυτές τις κορυφές είναι αυτές που έχουν πλάτος μεγαλύτερο από  $\mu_1 + \alpha \cdot \sigma_1$ , όπου το  $\alpha$  κατάλληλα επιλεγμένη σταθερά, π.χ.  $\alpha=1$ . Έστω ότι το πλήθος αυτών των κορυφών είναι  $N_{ΠΕ}$  για ένα παράθυρο παρατήρησης.

Όσον αφορά το πλήθος των διατηρητέων κορυφών στη ΣΕ, διακρίναμε δύο περιπτώσεις: Πρώτα, την περίπτωση που το πλήθος των κορυφών της ΣΕ ήταν μεγαλύτερο από το  $N_{ΠΕ}$ , οπότε κρατήσαμε ακριβώς τις  $N_{ΠΕ}$  μεγαλύτερες κορυφές της σόλο εκτέλεσης. Εν συνεχεία, την περίπτωση όπου το πλήθος των κορυφών της ΣΕ ήταν μικρότερο από το  $N_{ΠΕ}$ , οπότε κρατήσαμε όλες τις κορυφές αυτές. Σε κάθε περίπτωση, το πλήθος των κορυφών της σόλο εκτέλεσης σε ένα παράθυρο παρατήρησης το ονομάσαμε  $N_{ΣΕ}$ .

### ***5.3 Εύρος του *masking****

Διαπιστώσαμε ότι ο βαθμός σύμπτωσης των κορυφών, πλήθους  $N_{ΠΕ}$ , της πρότυπης σύνθεσης και αυτών, πλήθους  $N_{ΣΕ}$  της σόλο εκτέλεσης σε ίδια ακουστικά παράθυρα, εξαρτιόταν από το εύρος συχνοτήτων στο οποίο εφαρμόζαμε τη διαδικασία του *masking*. Για να διαλέξουμε το βέλτιστο αυτό εύρος συχνοτήτων πραγματοποιήσαμε ένα μεγάλο πλήθος πειραμάτων. Ειδικότερα, διαλέξαμε ένα ικανό πλήθος παραθύρων τόσο στην πρότυπη όσο και στη σόλο εκτέλεση που τα αναγνωρίσαμε ακουστικά σαν ίδια. Εν συνεχεία, σε κάθε τέτοιο ζεύγος ακουστικά ταυτόσημων παραθύρων

υπολογίσαμε το ποσοστό των κοινών κορυφών ως προς το ΝΠΕ για ένα μεγάλο διάστημα συχνοτήτων masking με πυκνή διαμέριση. Συγκεκριμένα εάν  $f_0$  ήταν η συχνότητα της κυρίαρχης κορυφής γύρω από την οποία εφαρμόστηκε η διαδικασία masking, θεωρήσαμε το εύρος συχνοτήτων  $[f_0-500, f_0+500]$ , το οποίο διαμερίσαμε σε ισαπέχουσες ανά 20 Hz συχνότητες. Σε καθεμία συχνότητα αυτής της διαμέρισης υπολογίσαμε το ποσοστό  $\Pi =$  κοινές κορυφές στα παράθυρα ΠΕ και ΣΕ/ΝΠΕ. Τελικά επιλέξαμε εκείνη τη συχνότητα η οποία έδινε το μεγαλύτερο μέσο ποσοστό  $\Pi$  για όλα τα  $f_0$  σε όλα τα παράθυρα παρατήρησης με κοινό ακουστικό περιεχόμενο. Το εύρος αυτό ήταν  $400 \cdot f_s / \text{winlen}$  η συχνότητες που αποκόπτεται εκατέρωθεν της κορυφής, οπότε εφαρμόσαμε τη διαδικασία masking στην περιοχή  $[f_0 - \Delta f, f_0 + \Delta f]$  για όλες τις κεντρικές συχνότητες  $f_0$  που προέκυψαν, σε όλα τα παράθυρα παρατήρησης τόσο στην ΠΕ όσο και στη ΣΕ.



# 6

## *Η μεθοδολογία που εφαρμόστηκε και μερικά πρώτα αποτελέσματα*

### *6.1 Συνοπτική περιγραφή της διαδικασίας ταυτοποίησης μουσικών συνθέσεων που αναπτύχθηκε*

- *Βήμα 1:* Διαβάζουμε τα δύο σήματα και αποκόπτουμε από το καθένα ένα τμήμα ίδιου μεγέθους, 56x1024 δειγμάτων, που αποτελεί το μέγεθος παραθύρου. Έστω ότι το τμήμα αυτό της πρότυπης εκτέλεσης είναι το ΠΕ και το τμήμα της σόλο εκτέλεσης το ΣΕ.
- *Βήμα 2:* Χωρίζουμε τη μουσική εκθετική κλίμακα σε νότες-διαμερίσεις, σύμφωνα με τη διαδικασία που περιγράφηκε στην ενότητα 4.2.

- **Βήμα 3:** Υπολογίζουμε το Διακριτό Μετασχηματισμό Fourier τόσο για την ΠΕ όσο και για τη ΣΕ.
- **Βήμα 4:** Εφαρμόζουμε τη διαδικασία του masking και για την ΠΕ και για τη ΣΕ, η οποία μας επιστρέφει ένα σύνολο από κορυφές πλήθους  $N_{ΠΕ}$  και  $N_{ΣΕ}$  αντίστοιχα.
- **Βήμα 5:** Πολλαπλασιάζουμε όλα τα στοιχεία του συνόλου  $N_{ΣΕ}$  με  $f$ . Το  $f$  είναι ο συντελεστής ομοιοθεσίας που αντιπροσωπεύει την ενδεχόμενη αλλαγή ρυθμού.
- **Βήμα 6:** Αντιστοιχίζουμε τις κορυφές που προέκυψαν μετά από το masking και την ομοιοθεσία σε νότες, όπως περιγράφηκε στο κεφάλαιο 4.2, τόσο για τη ΠΕ όσο και για τη ΣΕ
- **Βήμα 7:** Αίρουμε τις πολλαπλές εμφανίσεις των ίδιων νοτών και για την ΠΕ και για τη ΣΕ, δημιουργώντας μία μόνο ακολουθία διαφορετικών νοτών, πλήθους μικρότερου ίσου ενός άλλου κατάλληλα επιλεγμένου αριθμού  $NN$ . Τα πειράματα έδειξαν ότι μια καλή τιμή για το  $NN$  για την ΠΕ είναι 16.
- **Βήμα 8:** Εφαρμόζουμε τα προηγούμενα για μια ακολουθία τιμών του συντελεστή ομοιοθεσίας, έστω  $f_i, i=1,2,\dots$
- **Βήμα 9:** Για κάθε  $f_i$  υπολογίζουμε τον κοινό αριθμό νοτών της ΠΕ αφενός και της ομοιοθετημένης ΣΕ αφετέρου. Εκείνο το  $f_i$  που δίνει το μεγαλύτερο πλήθος κοινών νοτών, θεωρούμε ότι αντιπροσωπεύει την τροποποίηση του ρυθμού η οποία βέλτιστα κάνει την ΣΕ να ακούγεται σε πολύ παραπλήσιο ρυθμό με την ΠΕ.
- **Βήμα 10:** Υπολογίζουμε το ποσοστό επιτυχίας της σύγκρισης ως το πλήθος των κοινών νοτών ΠΕ και ΣΕ προς το πλήθος των νοτών της ΠΕ. Εάν αυτό το ποσοστό βρίσκεται πάνω από ένα κατάλληλα επιλεγμένο κατώφλι, τότε προχωράμε κατά έναν αριθμό δειγμάτων (Window Hop) και στις δύο εκτελέσεις και παίρνουμε άλλο ένα παράθυρο όπου εφαρμόζουμε πάλι όλα τα προηγούμενα βήματα, εν συνεχεία ένα τρίτο παράθυρο, ένα τέταρτο, ..., ένα  $v$ -οστό, που απέχουν το καθένα από το προηγούμενο Window Hop. Αν το ποσοστό δε βρίσκεται πάνω από το κατώφλι που έχουμε επιλέξει, τότε προχωράμε στη σύγκριση του ίδιου παραθύρου της ΣΕ με το επόμενο παράθυρο της ΠΕ, εφαρμόζοντας όλα τα προηγούμενα βήματα, κ.ο.κ.

- *Βήμα 11:* Αν διαδοχικά παράθυρα δώσουν ποσοστό επιτυχίας πάνω από το επιλεγμένο κατώφλι, και συνολικά το μέσο ποσοστό αναγνώρισης βρίσκεται πάνω από μία συγκεκριμένη τιμή, τότε θεωρούμε ότι έχουμε επιτυχή αναγνώριση. Διαφορετικά, θεωρούμε ότι οι δύο εκτελέσεις αντιστοιχούν σε διαφορετικές μουσικές συνθέσεις.

## **6.2 Μερικά πρώτα αποτελέσματα**

Τα αποτελέσματα που διαθέτουμε προέκυψαν από την εφαρμογή της μεθοδολογίας ταυτοποίησης που αναφέρθηκε στην προηγούμενη ενότητα σε τρεις διαφορετικές μουσικές συνθέσεις με διαφορετικές εκτελέσεις από σόλο όργανο. Οι τρεις αυτές μουσικές συνθέσεις είναι οι ίδιες που αναφέρθηκαν και στο κεφάλαιο 3, δηλαδή α) το Rondo Allegro από το K525 (Eine Kleine Nachtmusik), β) το Romanze Andante από το K525 (Eine Kleine Nachtmusik) και γ) το Adagio από το K622 (Concerto for Clarinet). Και οι τρεις μουσικές συνθέσεις είναι έργα του Mozart. Στις τρεις αυτές μουσικές συνθέσεις είχαμε την πρότυπη εκτέλεση, δηλαδή την εκτέλεση από συμφωνική ορχήστρα.

Για την πρώτη μουσική σύνθεση, χρησιμοποιήσαμε για τη σύγκριση εκτέλεση από σόλο πιάνο στην ίδια τονικότητα, οκτάβα και ρυθμό και σόλο κιθάρα στην ίδια τονικότητα και οκτάβα αλλά με σημαντικά πιο αργό ρυθμό. Για τη δεύτερη μουσική σύνθεση, χρησιμοποιήσαμε πάλι μία εκτέλεση από σόλο πιάνο στην ίδια τονικότητα αλλά σε μεγαλύτερη οκτάβα και με παραπλήσιο ρυθμό και μία εκτέλεση από σόλο κιθάρα στην ίδια τονικότητα, οκτάβα και ρυθμό με την ορχηστρική. Τέλος, για την τρίτη μουσική σύνθεση, χρησιμοποιήσαμε μια εκτέλεση από σόλο πιάνο και μία άλλη από σόλο αρμόνιο σε ίδιες τονικότητες και ρυθμούς αλλά σε άλλες οκτάβες.

Εφαρμόσαμε τη μεθοδολογία της ενότητας 6.1 και ταυτόχρονα ακούγαμε παράλληλα τις δύο εκτελέσεις της ίδιας μουσικής σύνθεσης, εντοπίζοντας παράθυρα που αντιστοιχούν ακουστικά επακριβώς και στις δύο εκτελέσεις. Μετά την εφαρμογή της διαδικασίας αναγνώρισης μετρήσαμε το μέσο ποσοστό επιτυχίας που έδιναν όλα τα

παράθυρα της ΠΕ με ένα παράθυρο της ΣΕ και ύστερα το συγκρίναμε με το μέσο όρο των ποσοστών που έδινε το ίδιο παράθυρο της ΣΕ με τα παράθυρα της ΠΕ που αντιστοιχούσαν ακουστικά επακριβώς. Διαπιστώσαμε ότι το δεύτερο ποσοστό ήταν σε όλες τις περιπτώσεις στατιστικά ικανά μεγαλύτερο από το μέσο ποσοστό. Με βάση αυτό το γεγονός, μπορεί να θεωρηθεί ότι το σύστημα αναγνώρισε τα τμήματα της σύνθεσης που αντιστοιχούσαν μουσικά επακριβώς στην ΠΕ και στην ΣΕ.

# 7

## *Συμπεράσματα και μελλοντικές επεκτάσεις*

Στην παρούσα διπλωματική μελετήθηκε η αυτόματη ταυτοποίηση μουσικών συνθέσεων που έχουν εκτελεστεί από διαφορετικούς εκτελεστές και διαφορετικά όργανα. Οι τεχνικές που εφαρμόστηκαν σχετίζονται με: α) εντοπισμό των πιο σημαντικών κορυφών του φάσματος κάθε μουσικής σύνθεσης, β) κατάλληλη διαλογή των κορυφών με χρήση της διαδικασίας masking, γ) αντιστοίχιση των επιλεγμένων κορυφών σε νότες-διαμερίσεις που προσομοιάζουν τη μουσική εκθετική κλίμακα, δ) πολλαπλή εφαρμογή της διαδικασίας μετά από ομοιοθεσία στο χρόνο, ε) κάποιες νέες πρωτοποριακές τεχνικές. Η μεθοδολογία που αναπτύχθηκε εφαρμόστηκε σε τρεις κλασικές μουσικές συνθέσεις, η καθεμία από τις οποίες είχε εκτελεστεί τόσο από συμφωνική ορχήστρα, όσο και από δύο σόλο όργανα. Τα σχετικά πειράματα έδωσαν στατιστικά ιδιαίτερος ικανοποιητικό ποσοστό επιτυχούς ταυτοποίησης.

Πρέπει να αναφερθεί ότι κατά την πραγματοποίηση της παρούσας εργασίας δοκιμάστηκαν και νέες πρωτότυπες επεκτάσεις, όσον αφορά τις μεθόδους και τεχνικές ταυτοποίησης μουσικών συνθέσεων, επεκτάσεις οι οποίες φαίνεται να είναι πολλά υποσχόμενες. Επομένως η παρούσα διπλωματική εργασία μπορεί να αποτελέσει μία

βάση για περαιτέρω μελέτη και έρευνα που οδηγούν στην αυτόματη ταυτοποίηση μουσικών συνθέσεων ανεξαρτήτως του εκτελεστή αυτών και του μουσικού οργάνου.

Σχετικά αντικείμενα μελλοντικής έρευνας μπορεί επίσης να είναι τα ακόλουθα: Θα πρέπει να μελετηθεί και να προβλεφθεί από ένα αντίστοιχο σύστημα αναγνώρισης στο μέλλον η δυνατότητα ταυτοποίησης δύο μουσικών εκτελέσεων μιας μουσικής σύνθεσης ακόμα και αν είναι εκτελεσμένες σε διαφορετικές τονικότητες (και όχι μόνο σε διαφορετικές οκτάβες). Θα πρέπει να βρεθούν τα κατάλληλα κριτήρια τα οποία θα υποδεικνύουν σύμπτωση νοτών σε αυτή την περίπτωση.

Άλλη μελλοντική επέκταση θα μπορούσε να είναι ο έλεγχος του κατάλληλου Window Hop, που αναφέρθηκε στην ενότητα 6.1, ώστε να μην έχει πάντα την ίδια τιμή, αλλά αυτή να μπορεί να είναι μεταβλητή και να εξαρτάται από το βέλτιστο συντελεστή ομοιοθεσίας που προκύπτει για κάθε παράθυρο. Επίσης, με την ίδια έννοια πρέπει να μελετηθεί εάν το Window Hop πρέπει να είναι ίδιο μεταξύ ΠΕ και ΣΕ (αυτή η προσέγγιση ακολουθείται σε αυτή τη διπλωματική) ή αν θα μπορούσε να είναι διαφορετικό, ανάλογα με το ρυθμό εκτέλεσης των ΠΕ και ΣΕ.

Επιπλέον, είναι αναγκαίος ο καλύτερος εντοπισμός και προσδιορισμός του κατώφλιου του ελάχιστου ποσοστού σύμπτωσης των κορυφών στα παράθυρα των ΠΕ και ΣΕ, που ουσιαστικά καθορίζει αν θα ελεγχθούν και τα επόμενα παράθυρα των ΠΕ, ΣΕ, παράλληλα. Το κατώφλι αυτό είναι αναγκαίο να ελεγχθεί με πολλές διαφορετικές μουσικές συνθέσεις, ώστε να μπορεί να εξαχθεί ένα πιο ασφαλές συμπέρασμα σχετικά με την τιμή που θα πρέπει να έχει.

Αντίστοιχη μελέτη πρέπει να γίνει και με το κριτήριο που καθορίζει εάν έχουμε επιτυχή αναγνώριση ή όχι, το κατώφλι, δηλαδή, του μέσου όρου ποσοστού σύμπτωσης των κορυφών πάνω σε όλα τα παράθυρα των ΠΕ και ΣΕ. Πρέπει και σε αυτή την περίπτωση να γίνει έλεγχος του κριτηρίου αυτού σε ένα ευρύ φάσμα μουσικών συνθέσεων και διαφόρων εκτελέσεών τους.

Τέλος, μία ακόμη επέκταση θα μπορούσε να είναι ο εντοπισμός των θεμελιωδών συχνοτήτων σε κάθε παράθυρο (pitch) και να απαιτούμε είτε αυτές να είναι παραπλήσιες για τις ΠΕ, ΣΕ εφόσον και οι δύο εκτελέσεις αντιστοιχούν στο ίδιο ακουστικό περιεχόμενο και είναι στην ίδια τονικότητα, είτε να βρίσκονται σε πολλαπλάσιες συχνότητες εφόσον οι δύο εκτελέσεις βρίσκονται σε διαφορετικές οκτάβες.

# 8

## *Βιβλιογραφία*

- [1] Fiss, X.; Kwasinski, A. Automatic real-time electric guitar audio transcription, 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Page(s): 373 - 376
- [2] Faruqe, M.O.; Hasan, M.A.-M.; Ahmad, S.; Bhuiyan, F.H. Template music transcription for different types of musical instruments, 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)
- [3] Pikrakis, A.; Theodoridis, S.; Kamarotos, D. Recognition of Isolated Musical Patterns Using Context Dependent Dynamic Time Warping, IEEE Transactions on Speech and Audio Processing, Volume: 11 , Issue: 3, Publication Year: 2003 , Page(s): 175 – 183
- [4] de Leon, P.P.; Inesta, J. Pattern Recognition Approach for Music Style Identification Using Shallow Statistical Descriptors, IEEE Transactions on Systems, Man, and

Cybernetics, Part C: Applications and Reviews, Volume: 37 , Issue: 2, Publication Year: 2007 , Page(s): 248 – 257

[5] Tzanetakis, G.; Cook, P. Musical Genre Classification of Audio Signals, IEEE Transactions on Speech and Audio Processing, Volume: 10 , Issue: 5, Publication Year: 2002 , Page(s): 293 – 302

[6] Clausen, M.; Kurth, F. A Unified Approach to Content-Based and Fault-Tolerant Music Recognition, IEEE Transactions on Multimedia, Volume: 6 , Issue: 5, Publication Year: 2004 , Page(s): 717 – 731

[7] Pikrakis, A.; Theodoridis, S.; Kamarotos, D. Classification of musical patterns using variable duration hidden Markov models, IEEE Transactions on Audio, Speech, and Language Processing, Volume: 14 , Issue: 5, Publication Year: 2006 , Page(s): 1795 – 1807

[8] Marolt, M. A Connectionist Approach to Automatic Transcription of Polyphonic Piano Music, IEEE Transactions on Multimedia, Volume: 6 , Issue: 3, Publication Year: 2004 , Page(s): 439 – 449

[9] Sethares, W.A.; Morris, R.D.; Sethares, J.C. Beat Tracking of Musical Performances Using Low-Level Audio Features, IEEE Transactions on Speech and Audio Processing, Volume: 13 , Issue: 2, Publication Year: 2005 , Page(s): 275 – 285

[10] Barbancho, A.M.; Klapuri, A.; Tardon, L.J.; Barbancho, I. Automatic Transcription of Guitar Chords and Fingering From Audio, IEEE Transactions on Audio, Speech, and Language Processing, Volume: 20 , Issue: 3, Publication Year: 2012 , Page(s): 915 – 921

[11] Bharathi, V.; Asaph, A.A.; Ramya, R. Vocal pitch detection for musical transcription, 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), Publication Year: 2011 , Page(s):



724 - 726

[12] Fragoulis, D.; Rousopoulos, G.; Panagopoulos, T.; Alexiou, C.; Papaodysseus, C. On the automated recognition of seriously distorted musical recordings, IEEE Transactions on Signal Processing, Volume: 49 , Issue: 4, Publication Year: 2001 , Page(s): 898 – 908

[13] E. Zwicker, H. Fastl, "Psychoacoustics", Springer - Verlag, Berlin, Germany, 1990.

[14] Painter T., Spanias A., A Review of Algorithms for Perceptual Coding of Digital Audio Signals