



ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΑΜΦΙΔΡΟΜΗ ΠΡΑΓΜΑΤΙΚΟΥ ΧΡΟΝΟΥ  
ΜΕΤΑΤΡΟΠΗ ΦΩΝΗΤΙΚΩΝ ΜΟΥΣΙΚΩΝ  
ΦΘΟΓΓΩΝ ΣΕ ΨΗΦΙΑΚΗ ΠΑΡΤΙΤΟΥΡΑ**

**ΜΙΧΑΛΗΣ ΑΝΘΗΣ**

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ

ΦΛΩΡΟΣ ΑΝΔΡΕΑΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΑΜΦΙΔΡΟΜΗ ΠΡΑΓΜΑΤΙΚΟΥ ΧΡΟΝΟΥ  
ΜΕΤΑΤΡΟΠΗ ΦΩΝΗΤΙΚΩΝ ΜΟΥΣΙΚΩΝ  
ΦΘΟΓΓΩΝ ΣΕ ΨΗΦΙΑΚΗ ΠΑΡΤΙΤΟΥΡΑ**

**ΜΙΧΑΛΗΣ ΑΝΘΗΣ**

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ

ΚΕΡΚΥΡΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2011

Ευχαριστίες

Ιδιαίτερες ευχαριστίες θα ήθελα να απευθύνω στον Δρ. Ανδρέα Φλώρο για την καθοδήγηση και την υποστήριξή του ως προς την διεκπεραίωση της παρούσας πτυχιακής. Χωρίς τη βοήθεια του η ολοκλήρωση αυτής της μελέτης θα ήταν αδύνατη.

Ευχαριστώ από καρδιάς τον Δρ. Ανδρέα Μνιέστρη για την εμπιστοσύνη που έδειξε δίνοντας μου την ευκαιρία να συμμετέχω σε αυτό το μεταπτυχιακό πρόγραμμα.

Επίσης, θα ήθελα να ευχαριστήσω επιπλέον τους καθηγητές Θεόδωρο Λώτη, Διονύσιο Μπατζάκη, Ιωάννη Ζάννο και Tim Ward για την ανυπολόγιστη συμβολή τους σε όλη τη διάρκεια του προγράμματος.

Τέλος, ευχαριστώ θερμά την οικογένεια μου για τη συνεχή συμπαράσταση, την αγάπη και ηθική υποστήριξη τους καθ' όλη τη διάρκεια των σπουδών μου.

## ΠΕΡΙΛΗΨΗ

Η εργασία αυτή έχει σαν θέμα την έρευνα γύρω από τις τεχνικές και μεθόδους του εντοπισμού της θεμελιώδης συχνότητας με σκοπό την σχεδίαση ενός συστήματος που θα εντοπίζει τη συχνότητα αυτή από ένα εισαγόμενο σήμα και θα το απεικονίζει πάνω σε μία εικονική παρτιτούρα υπό μορφή μουσικών φθόγγων. Το σύστημα αυτό έχει σαν στόχο τα σήματα φωνής και λειτουργεί σε πραγματικό χρόνο ώστε η αλληλεπίδραση με τον χρήστη να έχει μεγαλύτερη ρεαλιστικότητα. Στόχος του είναι η εξάσκηση του χρήστη πάνω σε επίπεδα μουσικής μελέτης, όπως ορθοφωνία και σολφέζ, μέσω οπτικής και ακουστικής ανάδρασης. Στην εργασία γίνεται μια αρκετά πλούσια αναφορά ενός μεγάλου μέρους των μεθόδων σχετικών με τον εντοπισμό του τονικού ύψους, οι οποίες περιγράφονται, συγκρίνονται και αξιολογούνται ενώ μία από αυτές τις μεθόδους χρησιμοποιείται και στο σύστημα που θα υλοποιηθεί στην εργασία. Το σύστημα, όπως θα αναφερθεί, βρίσκεται σε πειραματικό στάδιο καθώς παρέχει στο χρήστη τις βασικές ακόμα πληροφορίες για το σήμα που εισάγεται, όπως ονομασία φθόγγου και θέση στο πεντάγραμμο. Παρόλα αυτά, γίνεται πλήρη περιγραφή και αξιολόγηση του συστήματος γύρω από τη συμπεριφορά του, τα σφάλματα και τα αποτελέσματα που αυτό παρουσιάζει. Επίσης, παρέχει στον χρήστη άλλες χρήσιμες πληροφορίες όπως το φασματογράφημα και το ιστόγραμμα.

## Πίνακας Περιεχομένων

<b>ΕΙΣΑΓΩΓΗ</b> .....	<b>6</b>
<b>1 ΘΕΩΡΙΑ</b> .....	<b>8</b>
1.1 ΧΡΗΣΙΜΗ ΟΡΟΛΟΓΙΑ .....	8
1.1.1 Τονικό ύψος.....	8
1.1.2 Θεμελιώδης συχνότητα.....	9
1.1.3 Απλοί τόνοι.....	9
1.1.4 Μέτρηση συχνότητας .....	10
1.1.5 Ψηφιακή επεξεργασία σήματος.....	12
1.2 ΜΕΘΟΔΟΙ ΕΝΤΟΠΙΣΜΟΥ ΘΕΜΕΛΙΩΔΟΥΣ ΣΥΧΝΟΤΗΤΑΣ – ΙΣΤΟΡΙΚΑ ΣΤΟΙΧΕΙΑ .....	13
1.2.1 Μέθοδοι στο πεδίο του χρόνου (Time domain).....	13
1.2.2 Μέθοδοι στο πεδίο της συχνότητας (Frequency domain) .....	17
1.3 ΑΝΑΛΥΣΗ ΣΥΧΝΟΤΗΤΑΣ.....	22
1.3.1 Ανάλυση Fourier.....	23
1.3.2 Μετασχηματισμός Fourier (Fourier Transform) – Βασικές Αρχές.....	24
1.3.3 Διακριτός Σχηματισμός Fourier.....	25
1.3.4 Γρήγορος Μετασχηματισμός Fourier.....	26
<b>2 ΥΛΟΠΟΙΗΣΗ</b> .....	<b>28</b>
2.1 ΑΝΑΛΥΣΗ ΣΗΜΑΤΟΣ ΕΙΣΟΔΟΥ.....	28
2.1.1 Τεχνική συμπλήρωσης μηδενικών ( Zero Padding) .....	28
2.1.2 Παραθυροποίηση (Windowing).....	29
2.1.3 Μετασχηματισμός FFT.....	32
2.1.4 Κλάση Επεξεργασίας και ανάλυσης τονικού ύψους .....	34
2.1.5 Εντοπισμός θεμελιώδους συχνότητας .....	38
2.1.6 Μετατροπή συχνότητας σε MIDI πληροφορία.....	40
2.2 ΓΡΑΦΙΚΟ ΠΕΡΙΒΑΛΛΟΝ (USER INTERFACE) .....	41
2.3 ΠΕΡΙΒΑΛΛΟΝ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ – PROCESSING 1.2.1 .....	43
<b>3 ΑΠΟΤΕΛΕΣΜΑΤΑ</b> .....	<b>44</b>
<b>4 ΣΥΜΠΕΡΑΣΜΑΤΑ</b> .....	<b>51</b>

# ΕΙΣΑΓΩΓΗ

Ο εντοπισμός του τονικού ύψους έχει απασχολήσει τις τελευταίες πέντε δεκαετίες ένα ευρύ πεδίο του επιστημονικού, και όχι μόνο, τομέα με έναν μεγάλο αριθμό ερευνών στο ενεργητικό του. Οι έρευνες αυτές ποικίλουν ως προς τις μεθόδους και τεχνικές που χρησιμοποιούν αλλά και ως προς τον τρόπο που προσεγγίζουν το θέμα αυτό. Η εργασία αυτή ξεκινάει με το θεωρητικό μέρος στο οποίο εισάγονται οι βασικότερες έννοιες που σχετίζονται με τον εντοπισμό του τονικού ύψους, με σκοπό την καλύτερη κατανόηση του αναγνώστη στα επόμενα κεφάλαια. Στη συνέχεια, γίνεται μια αναφορά στις πιο γνωστές μεθόδους και τεχνικές που έχουν χρησιμοποιηθεί σε ερευνητικό επίπεδο για τον εντοπισμό του τονικού ύψους. Η κάθε μία αξιολογείται σύμφωνα με τα αποτελέσματα που δίνει και το περιεχόμενο στο οποίο αναφέρεται (πχ φωνή) ενώ για λόγους απλότητας περιγράφεται μέσω απλών μαθηματικών εξισώσεων και συνοδεύεται με σχήματα που αναπαριστούν το τρόπο που λειτουργίας τους ή τις αρχές πάνω στις οποίες βασίζεται. Το θεωρητικό μέρος τελειώνει με την ανάλυση συχνότητας, του λόγους ύπαρξης της αλλά και τους βασικότερους τρόπους εφαρμογής της. Πιο συγκεκριμένα, χρησιμοποιώντας και εδώ απλά μαθηματικά, θα περιγράψει την ανάλυση Fourier και τον μετασχηματισμό Fourier (DFT - FFT), αναλύοντας περισσότερο θέματα γύρω από τον γενικό τρόπο που λειτουργίας της τεχνικής FFT και την σχέση της με την DFT. Στο δεύτερο κεφάλαιο αναλύεται η όλη διαδικασία υλοποίησης του συστήματος, περιγράφοντας όλες τις τεχνικές και μεθόδους που χρησιμοποιήθηκαν αλλά και τον λόγο που επιλέχθηκαν. Εδώ, περιγράφεται και το γραφικό περιβάλλον που βλέπει ο χρήστης αλλά και οι τρόποι που μπορεί να αλληλεπιδράσει με το σύστημα. Το σύστημα υλοποιήθηκε αποκλειστικά σε προγραμματιστικό περιβάλλον γι'αυτό το λόγο το κεφάλαιο αυτό εμπεριέχει και μια αναφορά σε αυτό. Τέλος, θα διαιρεθεί σε υπομονάδες, αναλύοντας και περιγράφοντας την συμβολή της κάθε μίας ξεχωριστά στο συνολικό ρόλο του συστήματος. Συνεχίζοντας στο επόμενο κεφάλαιο, γίνεται η αξιολόγηση του ίδιου του

συστήματος μέσω των αποτελεσμάτων που απορρέουν από αυτό. Περιγράφεται η συμπεριφορά του μέσω σχημάτων για τις διάφορες αναλογίες δεδομένων, ορισμάτων και παραμέτρων που το αποτελούν αλλά και πως αυτά μπορούν να επηρεάσουν την έξοδο του. Τέλος, η εργασία καταλήγει στα συμπεράσματα μέσω των οποίων γίνεται μια γενική κριτική του συνολικού συστήματος σχετικά την εγκυρότητα και την αξιοπιστία του. Επίσης, αναφέρονται και οι διάφορες βελτιώσεις και προσθήκες που θεωρούνται αναγκαίες για την ορθότερη συμπεριφορά του συστήματος.

# 1 ΘΕΩΡΙΑ

Το κεφάλαιο αυτό αρχίζει με μια αναφορά πάνω στις βασικές έννοιες που σχετίζονται άμεσα με το θέμα της εργασίας, όπως τονικό ύψος, θεμελιώδης συχνότητα, απλοί τόνοι κτλ, οι οποίες αναφέρονται σε όλο το μέρος της. Στη συνέχεια, γίνεται μια θεωρητική περιγραφή ενός συνόλου αλγορίθμων εντοπισμού της θεμελιώδης συχνότητας και η κατηγοριοποίηση τους σύμφωνα με τις αρχές λειτουργίας τους. Τέλος, περιγράφεται η βασική θεωρία με όλες τις σχετικές έννοιες που χρησιμοποιήθηκαν στην εργασία αυτή, όπως επεξεργασία σήματος, ανάλυση Fourier κτλ.. Εδώ, θα πρέπει να επισημανθεί ότι η μαθηματική ανάλυση των αλγορίθμων που θα αναφερθούν στη παρακάτω είναι πέραν από το θέμα αυτής της εργασίας. Παρόλα αυτά, γίνεται μια περιορισμένη χρήση εξισώσεων με βιβλιογραφικό χαρακτήρα που έχει σαν σκοπό την καλύτερη κατανόηση των αλγορίθμων αυτών.

## 1.1 Χρήσιμη ορολογία

### 1.1.1 Τονικό ύψος

Σύμφωνα με ένα πρότυπο του American National Standards το 1994, “το τονικό ύψος είναι μια ιδιότητα της αίσθησης της ακοής με το οποίο, σαν όρο, οι ήχοι κατανέμονται σε μία κλίμακα που εκτείνεται από χαμηλές σε υψηλές τιμές. Το τονικό ύψος εξαρτάται κυρίως από το συχνотικό περιεχόμενο του ηχητικού ερεθίσματος αλλά επίσης και από την ηχητική πίεση και κυματομορφή του ερεθίσματος”. Οι Dodge και Jerse (1997) ορίζουν το τονικό ύψος σαν μια υποκειμενική απόκριση στην συχνότητα, όπου σαν κύρια αίσθηση, δημιουργείται από το ερέθισμα των νευρικών καταλήξεων πάνω στη βασική μεμβράνη. Ο Wizard(1994) αναφέρει ότι το τονικό ύψος προκύπτει όταν οι αρμονικές του φάσματος του ήχου έχουν μια συγκεκριμένη και απλή σχέση μεταξύ τους, πχ όταν όλες είναι πολλαπλάσια μια σταθερής τιμής γνωστής ως θεμελιώδης συχνότητα.



Μια χονδρική εκτίμηση του εύρους των συχνοτήτων που μπορεί να αντιληφθεί το ανθρώπινο αυτί ξεκινάει από τα 20Hz και φτάνει μέχρι τα 20000 Hz. Πιο συγκεκριμένα, η περιοχή ανάμεσα στα 200 Hz – 2000 Hz, απασχολεί το μεγαλύτερο μέρος της βασικής μεμβράνης καθώς έχει την μεγαλύτερη αντιλαμβανόμενη οξύτητα και ευαισθησία στην αλλαγή της συχνότητας.

Η αντίληψη του τονικού ύψους έχει παίξει μεγάλο ρόλο στην ανάπτυξη της θεωρίας της ακοής. Αυτό οφείλεται στην άριστη πειραματική ακρίβεια που μπορεί κάποιος να πάρει από μετρήσεις πάνω στο τονικό ύψος. Η ακρίβεια αυτή, έχει επιτρέψει στον τομέα της ψυχολογιστικής στο να εξερευνήσει μία μεγάλη ποικιλία από φαινόμενα τονικής μετατόπισης και μεταβλητότητας.

### 1.1.2 Θεμελιώδης συχνότητα

Η θεμελιώδης συχνότητα είναι το βασικότερο συστατικό στοιχείο του φάσματος κάθε ήχου. Σε ένα περιοδικό σήμα η συχνότητα αυτή είναι αντιστρόφως ανάλογη με τη περίοδο, και ορίζεται ως το χαμηλότερο θετικό μέλος ενός άπειρου συνόλου χρονικών μετατοπίσεων που αφήνουν το σήμα αναλλοίωτο[10]. Στα μαθηματικά, η θεμελιώδης συχνότητα αποτελεί τον υψηλότερο κοινό παράγοντα (HCF<sup>1</sup>) στους αρμονικούς. Για την ομιλία ορίζεται ως το ποσοστό των δονήσεων στις φωνητικές χορδές. Παρόλα αυτά, όπως ο αναφέρει ο Wizard(1994), το αντιλαμβανόμενο τονικό ύψος είναι μια νοητική κατασκευή ενός αρμονικού φάσματος και όχι απλά η άμεση η κάθετη αντίληψη της θεμελιώδους συχνότητας. Η σχέση μεταξύ συχνότητας και αντίληψης του τονικού ύψους είναι κάθε άλλο παρά γραμμική καθώς έχει βρεθεί μέσω πειραμάτων ότι το ερέθισμα της βασικής μεμβράνης συμβαίνει σε σημεία που είναι σχεδόν ανάλογα με τον λογάριθμο της συχνότητας.

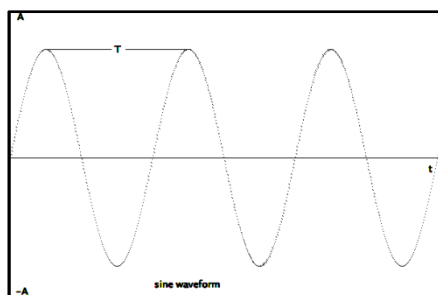
### 1.1.3 Απλοί τόνοι

Ένας απλός τόνος καταλαμβάνει μια μοναδική θέση στην ακουστική αλλά και στην επεξεργασία σήματος, καθώς είναι το πιο στοιχειώδες από όλα τα σήματα. Στο πεδίο των μαθηματικών ο απλός τόνος μεταφράζεται σαν μια

---

<sup>1</sup> Highest Common Factor

ημιτονοειδής κυματομορφή, η οποία είναι συνάρτηση του χρόνου, όπως φαίνεται στην εικ. 1.1.



*Εικόνα 1.1 Ημιτονοειδής κυματομορφή στο πεδίο του χρόνου.*

Η εξίσωση που μας δίνει την παραπάνω κυματομορφή δίνεται παρακάτω:

$$x(t) = A \sin(2\pi/T)t + \Phi \quad (1.1)$$

η αλλιώς

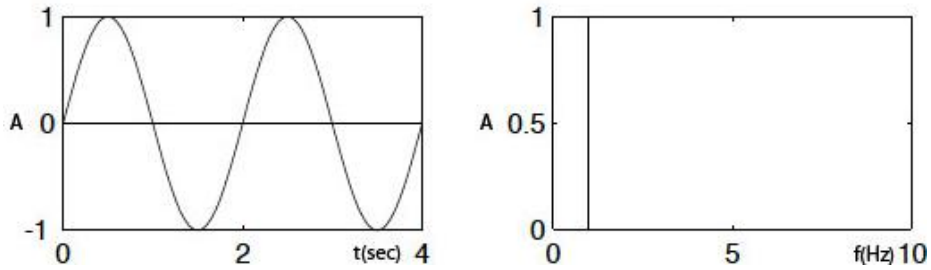
$$x(t) = A \sin(\omega t + \Phi) \quad (1.2)$$

όπου  $A$  το πλάτος(κάθετη απομάκρυνση από το μηδέν) της κυματομορφής,  $T$  η περίοδος σε δευτερόλεπτα,  $\omega$  η γωνιακή συχνότητα και  $\Phi$  η φάση σε ακτίνια.

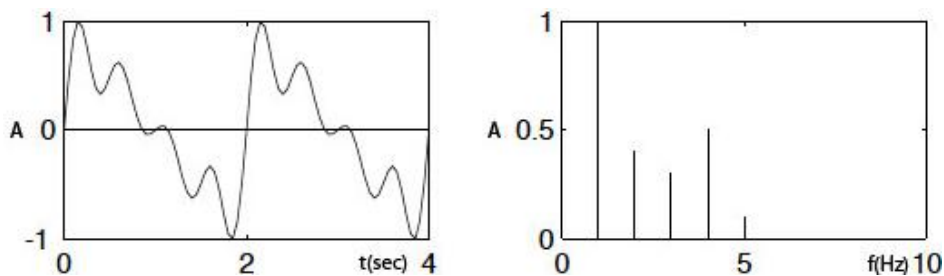
#### 1.1.4 Μέτρηση συχνότητας

Εφόσον το τονικό ύψος αποτελεί μια υποκειμενική ποσότητα που σχετίζεται με την θεμελιώδη συχνότητα ενός περιοδικού ή ψευδό-περιοδικού ήχου, τότε βρίσκοντας την περίοδο της κυματομορφής του μπορεί πολύ απλά να υπολογιστεί και η συχνότητα του. Όταν όμως η κυματομορφή δεν είναι απλή αλλά εμπεριέχει και αρμονικούς, η εύρεση την θεμελιώδους συχνότητας αλλά και η ύπαρξη του τονικού ύψους δεν είναι τόσο απλό ζήτημα. Έτσι λοιπόν, η δυσκολία εύρεσης της θεμελιώδους εξαρτάται από την ίδια την κυματομορφή. Λιγότερες υψηλότερες αρμονικές ή ψηλότερες αρμονικές με χαμηλότερη ενέργεια από την θεμελιώδη μας βοηθάει στο να μπορεί να καθοριστεί ευκολότερα η βασική συχνότητα. Τα παρακάτω σχήματα περιγράφουν το

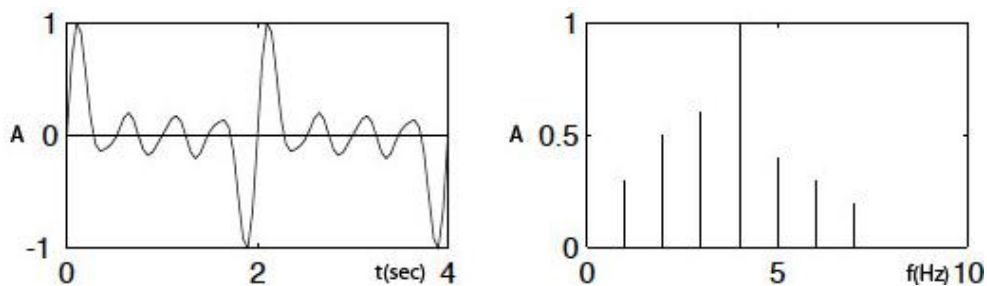
παραπάνω φαινόμενο σε διάφορες πιθανές περιπτώσεις. Για κάθε εικόνα, στα αριστερά απεικονίζεται η κυματομορφή στο πεδίο του χρόνου ενώ στα δεξιά στο πεδίο της συχνότητας.



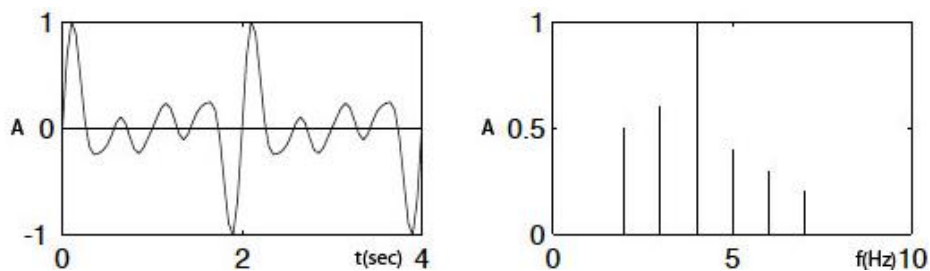
**Εικόνα 1.2** Απλή κυματομορφή χωρίς αρμονικούς.



**Εικόνα 1.3** Κυματομορφή όπου η θεμελιώδης συχνότητα έχει μεγαλύτερη ενέργεια από τους αρμονικούς



**Εικόνα 1.4** Κυματομορφή με τη θεμελιώδη συχνότητα να έχει χαμηλότερη ενέργεια από τους υψηλούς αρμονικούς.



**Εικόνα 1.5** Κυματομορφή με αρμονικούς χωρίς θεμελιώδη συχνότητα.

### 1.1.5 Ψηφιακή επεξεργασία σήματος

Γενικά, ένα σήμα μπορεί να πάρει πολλές μορφές με διαφορετικά χαρακτηριστικά για το καθένα ή και συνδυασμός αυτών. Έτσι, ένα σήμα μπορεί να είναι μία τάση σαν συνάρτηση του χρόνου ή μία δύναμη σαν συνάρτηση του χρόνου και της θέσης. Επίσης, μπορεί να είναι ένα δυναμικό σαν συνάρτηση της θέσης σε τρισδιάστατο πεδίο ή η ένταση σαν συνάρτηση συντεταγμένων και του χρόνου. Ο Ludeman (1986) αναφέρεται συγκεκριμένα στο κάθε όρο (1.1.5) λέγοντας ότι:

*Σήμα*, είναι μια συνάρτηση από ένα σύνολο ανεξάρτητων μεταβλητών, με τον χρόνο να αποτελεί την πιο κυρίαρχη μεταβλητή. Επίσης, εμπεριέχει πληροφορίες προς παρατήρηση.

Με τον όρο *επεξεργασία*, ορίζει την ασκούμενη μέθοδο πάνω σε ένα σήμα που έχει σαν σκοπό να εξάγει πληροφορίες από το σήμα.

Τέλος, ο όρος *ψηφιακή* αναφέρει απλά ότι η επεξεργασία γίνεται σε ένα ψηφιακό υπολογιστή ή ειδικού τύπου ψηφιακού λογισμικού.

Η ψηφιακή επεξεργασία σήματος ασχολείται με ένα σύνολο από διαχειρίσεις σήματος και δεδομένων σε μορφή δειγμάτων. Χρησιμοποιώντας τέτοιες στοιχειώδεις διεργασίες όπως ψηφιακή αποθήκευση και καθυστερήσεις, πρόσθεση, αφαίρεση και πολλαπλασιασμό μπορούν να παραχθούν πολλές χρήσιμες λειτουργίες όπως για παράδειγμα η εξαγωγή συγκεκριμένων συχνοτήτων από ένα θορυβώδες σήμα ή μεταχείριση αποκλειστικά αυτών των συχνοτήτων στο σήμα.

## **1.2 Μέθοδοι εντοπισμού θεμελιώδους συχνότητας – Ιστορικά στοιχεία.**

Ο εντοπισμός της θεμελιώδους συχνότητας ενός ήχου, τα τελευταία χρόνια αποτελεί ένα αρκετά δημοφιλές αντικείμενο έρευνας σε πολλά πεδία και για διαφορετικούς σκοπούς, όπως εκπαιδευτικούς αλλά και εμπορικούς. Πολύ συχνά αναφέρεται και ως εντοπισμός του τονικού ύψους αλλά όπως είδαμε και παραπάνω, οι δύο έννοιες σχετίζονται άμεσα χωρίς όμως να είναι ταυτόσημες. Για τον εντοπισμό της θεμελιώδους συχνότητας έχουν αναπτυχθεί κατά καιρούς διάφορων ειδών αλγόριθμοι που διαφέρουν ως προς αρχές λειτουργίας αλλά και ως προς το περιεχόμενο για το οποίο αναφέρονται, πχ μουσική ή ομιλία. Οι βασικοί αυτοί μέθοδοι είναι δύο. Οι μέθοδοι της πρώτης κατηγορίας, λειτουργούν πάνω στο πεδίο του χρόνου και γενικά αναζητούν επαναλαμβανόμενα μέρη μέσα στο εισερχόμενο σήμα, τα οποία αντιστοιχούν στη θεμελιώδη συχνότητα. Η δεύτερη κατηγορία λειτουργεί πάνω στο πεδίο της συχνότητας και προσπαθεί να αναλύσει τα σημεία την απόσταση ανάμεσα από υποθετικούς αρμονικούς του φάσματος.

Επίσης, υπάρχουν και οι στατιστικές μέθοδοι οι οποίοι χρησιμοποιούν τη θεωρία των πιθανοτήτων ώστε να καταλήξουν στο επιθυμητό αποτέλεσμα. Παρακάτω, περιγράφονται οι δύο αυτοί μέθοδοι όπως επίσης και οι διάφοροι αλγόριθμοι που στηρίζουν την λειτουργία τους πάνω αυτές.

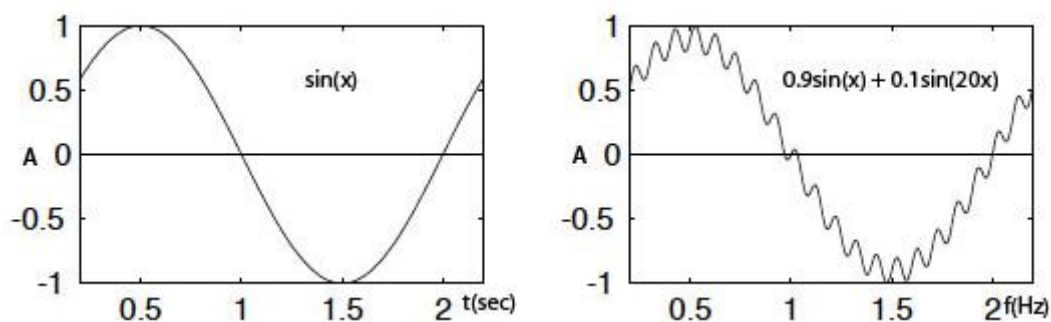
### **1.2.1 Μέθοδοι στο πεδίο του χρόνου (Time domain)**

#### **1.2.1.1 Εντοπισμός ποσοστιαίου χρονικού γεγονότος (Time event Rate detection)**

Σε αυτή την υποκατηγορία όλοι οι μέθοδοι του εντοπισμού της βασικής συχνότητας μετρούν το πόσο συχνά η ίδια η κυματομορφή του ηχητικού περιεχομένου επαναλαμβάνεται στο χρόνο. Κάθε μια από αυτές είναι χρήσιμες για συγκεκριμένο είδος κυματομορφών. Γενικά από τη θεωρία γνωρίζουμε ότι εάν μια κυματομορφή είναι περιοδική, τότε από αυτήν μπορούν να εξαχθούν διάφορα επαναλαμβανόμενα στο χρόνο γεγονότα, ο αριθμός των οποίων ανά δευτερόλεπτο σχετίζεται αντίστροφα με τη συχνότητα. Παρακάτω αναφέρονται οι μέθοδοι αυτοί:

### 1.2.1.2 Εντοπισμός Μηδενικών Σημείων (Zero Crossing Rate)

Η μέθοδος αυτή μετράει το πόσο συχνά μια κυματομορφή διανύει το μηδέν στη μονάδα του χρόνου ώστε να μας δώσει πληροφορίες για το φασματικό της περιεχόμενο. Ο βασικότερος λόγος που χρησιμοποιήθηκε η Ε.Μ.Σ ήταν για τον εντοπισμό της βασικής συχνότητας. Σύντομα όμως εμφανίστηκαν σημαντικά προβλήματα όσον αφορά την εγκυρότητα των αποτελεσμάτων και αυτό γιατί, εάν για παράδειγμα η ενέργεια του φασματικού περιεχομένου της κυματομορφής συγκεντρώνονταν γύρω από την θεμελιώδη συχνότητα τότε η κυματομορφή διανύει το μηδέν δύο φορές ανά κύκλο. Εάν όμως το φάσμα της κυματομορφής εμπεριέχει και υψηλότερες αρμονικές τότε η διάβαση από μηδενικό σημείο μπορεί να είναι και παραπάνω από δύο φορές κάνοντας έτσι τον εντοπισμό της βασικής συχνότητας δυσκολότερο έως και αδύνατο. Στο παρακάτω σχήμα βλέπουμε το παραπάνω φαινόμενο:



Εικόνα 1.6 Gerhard[2003] Επιρροή των υψηλών συχνοτήτων στον εντοπισμό μηδενικών σημείων.

### 1.2.1.3 Εντοπισμός Μεγίστων

Η μέθοδος αυτή μετράει τις μέγιστες τιμές (θετικές ή αρνητικές) μιας κυματομορφής ανά δευτερόλεπτο με σκοπό τον υπολογισμό της βασικής συχνότητας. Και εδώ, αν στη περίοδο της κυματομορφής υπάρχει μόνο ένα μέγιστο, τότε ο αριθμός αυτών στο χρόνο θα μας δώσει τη βασική συχνότητα. Σε αντίθετη περίπτωση η μέθοδος δεν είναι αποτελεσματική. Μια εναλλαγή της είναι ο υπολογισμός της απόστασης μεταξύ των μεγίστων ώστε να βρεθεί το μήκος κύματος το οποίο είναι αντιστρόφως ανάλογο με την συχνότητα.

Οι παραπάνω μέθοδοι είναι απλοί ως προς τον τρόπο λειτουργίας τους αλλά και την εφαρμογή τους, ενώ επίσης καταναλώνουν πολύ μικρή υπολογιστική ισχύ. Επίσης, λειτουργούν θετικά και χωρίς λάθη όταν η φύση του ηχητικού περιεχομένου είναι γνωστή. Για πολλά χρόνια μάλιστα χρησιμοποιούνται σε συσκευές ανίχνευσης της συχνότητας καθώς η σχεδίαση του κυκλώματος τους είναι απλή. Από την άλλη μεριά όμως, σπάνια στη φύση συναντάμε απλές κυματομορφές όπου χρονικά γεγονότα συμβαίνουν μία φορά ανά κύκλο. Πλούσιες σε φασματικό περιεχόμενο κυματομορφές μπορούν να διανύουν το μηδέν αρκετές φορές αν κύκλο ή να παρουσιάζουν περισσότερα από ένα μέγιστα. Σε αυτή τη περίπτωση οι μέθοδοι αυτοί παρουσιάζουν προβλήματα.

#### 1.2.1.4 Εντοπισμός της βασικής συχνότητας μέσω της συνάρτησης της αυτοσυσχέτισης (AutocorrelationFunction)

Η συνάρτηση της αυτοσυσχέτισης ενός σήματος είναι μια μέση ποσότητα των ιδιοτήτων του πάνω στο πεδίο του χρόνου και είναι αρκετά συναφής όταν ένα τέτοιο σήμα είναι τυχαίο(Lynn,1989). Ο τύπος που γενικά δίνει την συνάρτηση αυτή είναι:

$$r_{xx}(\tau) = \lim_{T_0 \rightarrow \infty} \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} f(t)f(t+\tau)dt \quad (1.3)$$

Η παραπάνω εξίσωση ισχύει στην περίπτωση ενός συνεχόμενου σήματος άπειρης διάρκειας και ισούται με το μέσο παράγωγο του αρχικού σήματος  $f(t)$  και του ίδιου σήματος όπου υστερεί χρονικά κατά χρόνο  $t$ . Στην περίπτωση τώρα που το σήμα έχει πεπερασμένη χρονική διάρκεια, το προϊόν της συνάρτησης για ένα μεγάλο χρονικό διάστημα  $T$  θα τείνει στο μηδέν για όλες τις τιμές του  $t$ . Αυτό οδηγεί σε μια πιο απλοποιημένη και πεπερασμένη μορφή της (1.3), η οποία ορίζεται ως:

$$r_{xx}(\tau) = \int_{-\infty}^{\infty} f_1(t)f_1(t+\tau)dt \quad (1.4)$$

Τα μέγιστα που εμφανίζονται σε αυτήν τη συνάρτηση αντιστοιχούν σε πολλαπλάσια της θεμελιώδους συχνότητας. Συχνά, τέτοιου είδους μέθοδοι επιλέγουν το υψηλότερο μη μηδενικό υστερημένο μέγιστο, πάνω από ένα συγκεκριμένο κατώφλι και για ένα συγκεκριμένο εύρος χρονικών καθυστερήσεων. Παρόλα αυτά, τέτοιες τεχνικές είναι αρκετά ευαίσθητες σε δομές των *formants*<sup>2</sup> με αποτέλεσμα να δημιουργούν σφάλματα οκτάβας.

#### 1.2.1.5 ΜΕΘΟΔΟΣ YIN

Το όνομα της μεθόδου, de Cheveigne και Kawahara (2002), προέρχεται από το yin-yang, μια φιλοσοφική αρχή των Ανατολικών χωρών για την ισορροπία, υποδηλώνοντας έτσι την προσπάθεια των δημιουργών να ισορροπήσουν μέσα στο αλγόριθμο τα φαινόμενα αυτοσυσχέτισης και ματαίωσης (Gerhard, 2003). Η μέθοδος αυτή χρησιμοποιεί την συνάρτηση διαφοράς τετραγώνου, η οποία και αυτή χρησιμοποιείται με σκοπό την προσέγγιση ανομοιότητας μεταξύ των συγκρινόμενων κυματομορφών (Ibanez, 2010). Η εξίσωση που χρησιμοποιεί ο αλγόριθμος αυτό είναι η εξής:

$$SDF_x[r] = \sum_{k=t}^{t+w-1} (x[k] + x[k-r])^2 \quad (1.5)$$

Επίσης, υπολογίζει τον αθροιστικό κανονικοποιημένο μέσο όρο της συνάρτησης  $SDF'$ , η οποία προέρχεται από την διαίρεση κάθε τιμής της  $SDF$  με τον μέσο όρο των χαμηλότερων τιμών υστέρησης.

$$SDF'_x[r] = \begin{cases} \frac{1}{(1/r) \sum_{j=1}^r SDF_x[r]} SDF_x[r] & , r = C \end{cases} \quad (1.6)$$

Γενικά, οι διάφοροι αλγόριθμοι αυτοσυσχέτισης παρουσιάζουν πρόβλημα όταν τα μέγιστα παρουσιάζονται και στις υπό-αρμονικές, κάνοντας έτσι δύσκολο τον

<sup>2</sup> Το formant ορίζεται ως μια συσσώρευση ενέργειας στο φάσμα ενός ήχου, το οποίο μπορεί να εμπεριέχει ταυτόχρονα αρμονικούς και μη αρμονικούς ή και θόρυβο. (Roads, 2001)



διαχωρισμό των μεγίστων που αντιπροσωπεύουν την θεμελιώδη συχνότητα με αυτών που αντιπροσωπεύουν τις αρμονικές ή τις μερικές αρμονικές.

### 1.2.2 Μέθοδοι στο πεδίο της συχνότητας (Frequency domain)

Στο πεδίο την συχνότητας υπάρχει αρκετή πληροφορία η οποία σχετίζεται με την θεμελιώδη. Τα ηχητικά σήματα που έχουν κάποια τονικότητα τείνουν να συνθέτονται από αρμονικά σχετιζόμενες συχνότητες(μερικούς αρμονικούς) με την θεμελιώδη και οι οποίες μπορούν να εντοπιστούν και να χρησιμοποιηθούν για τον καθορισμό την θεμελιώδους. Παρακάτω αναφέρονται οι πιο γνωστοί μέθοδοι που λειτουργούν στο πεδίο της συχνότητας:

#### 1.2.2.1 Μέθοδος συχνοτικής αναλογίας (Component Frequency Ratios)

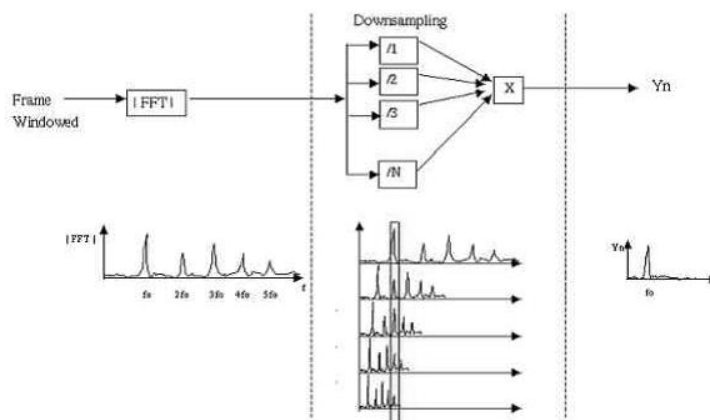
Το 1979 ο Martin Piszczalski (Πανεπιστήμιο του Michigan) εργάζονταν πάνω στην υλοποίηση ενός συστήματος αυτόματης μουσικής καταγραφής. Το αρχικό στάδιο του συστήματος αυτού ήταν εντοπισμός του τονικού ύψους ενώ στη συνέχεια (με την προϋπόθεση ότι στο σύστημα θα εισέρχονταν μία νότα για κάθε χρονικό σημείο) υπολόγιζε τα χρονικά όρια κάθε νότας, όριζε το μουσικό κλειδί και το παρουσίαζε την παρτιτούρα.

Η υλοποίηση του συστήματος ξεκινούσε με έναν φασματικό μετασχηματισμό και την αναγνώριση των μερικών αρμονικών του σήματος χρησιμοποιώντας την μέθοδο εντοπισμού μεγίστων. Για κάθε ζεύγος των αρμονικών αυτών, ο αλγόριθμος προσπαθούσε να βρίσκει τους δύο μικρότερους αριθμούς οι οποίοι θα αντιστοιχούσαν στις αρμονικές εκείνες σειρές που θα εμπεριείχαν το συγκεκριμένο ζεύγος. Αν για παράδειγμα δύο μερικοί αρμονικοί βρίσκονταν στα 435 Hz και 488 Hz, τότε οι δύο αυτοί αριθμοί που αναφέρονται πιο πάνω (και υπό ένα συγκεκριμένο κατώφλι) θα ήταν οι 6 και 7, αντίστοιχα. Στην συνέχεια οι δύο αυτοί αριθμοί θα χρησιμοποιούνταν σαν υπόθεση για τον εντοπισμό της θεμελιώδης συχνότητας που στο προηγούμενο παράδειγμα και με τις μερικές που αναφέρθηκαν η συχνότητα του σήματος θα ήταν περίπου 70 Hz. Η παραπάνω διαδικασία πραγματοποιείται για όλες τις μερικές αρμονικές που

εντοπίζονται από το σύστημα ενώ τελικά επιλέγεται το ζεύγος όπου οι αρμονικές αυτές έχουν το μεγαλύτερο πλάτος. Η μέθοδος αυτή δεν προϋποθέτει την παρουσία της ίδιας της αρμονικής (βλέπε εικ. 4) ενώ λειτουργεί ικανοποιητικά και σε μη αρμονικά σήματα αλλά και σε σήματα με ελλείψεις αρμονικές.

### 1.2.2.2 Μέθοδος αρμονικού φασματικού παράγωγου (Harmonic Product Spectrum)

Όπως έχει αναφερθεί, το φάσμα μιας μουσικής νότας αποτελείται από την θεμελιώδη και τους αρμονικούς, σαν ακέραια πολλαπλάσια της. Η μέθοδος αυτή, συμπιέζει το φάσμα και το συγκρίνει με το αρχικό. Έτσι, το πρώτο μέγιστο που εμφανίζεται στο αρχικό φάσμα ταυτίζεται με το δεύτερο μέγιστο του φάσματος που συμπιέστηκε με συντελεστή δύο, το οποίο συμπίπτει με το τρίτο μέγιστο του φάσματος που συμπιέστηκε με συντελεστή τρία κ.ο.κ. Η παραπάνω διαδικασία φαίνεται στο παρακάτω σχήμα.



**Εικόνα 1.7 Middleton [2003] Εφαρμογή της HPS. Από αριστερά προς τα δεξιά, το αρχικό φάσμα συμπιέζεται κατά ακέραια πολλαπλάσια και καταλήγει στην εύρεση της θεμελιώδους.**

Στο παραπάνω σχήμα, από αριστερά βλέπουμε το αρχικό φάσμα το σήματος στο πεδίο της συχνότητας. Στο κέντρο απεικονίζεται η διαδικασία συμπίεσης και τα σχετικά φάσματα τα οποία πολλαπλασιάζονται στο τέλος για να μας δώσουν την αντίστοιχη θεμελιώδη συχνότητα.

Η μέθοδος αυτή είναι υπολογιστικά μη δαπανηρή και ανθεκτική στο θόρυβο. Παρόλα αυτά, παρουσιάζει κάποιους περιορισμούς πάνω στις αρχές λειτουργίας του αλγορίθμου πάνω στον οποίο βασίζεται. Αρχικά, το γεγονός ότι η αντίληψη του τονικού ύψους περιγράφεται λογαριθμικά, σημαίνει ότι με την παραπάνω μέθοδο οι χαμηλές συχνότητες θα εντοπίζονται δυσκολότερα από τις υψηλές. Επίσης, η ανάλυση της διαδικασίας εντοπισμού εξαρτάται από την ποιότητα του μετασχηματισμού Fourier. Πχ εφαρμόζοντας τον αλγόριθμο για μικρά τμήματα περιοριζόμαστε στον αριθμό των διακριτών συχνοτήτων που μπορούν να εντοπιστούν. Για μεγαλύτερη ανάλυση, χρειάζονται μεγαλύτερα τμήματα τα οποία όμως απαιτούν και περισσότερο χρόνο κάνοντας την μέθοδο πιο αργή.

### 1.2.2.3 Μέθοδοι βασισμένοι σε φίλτρα

Οι μέθοδοι αυτοί προσπαθούν να εντοπίσουν την θεμελιώδη συχνότητα ενός σήματος χρησιμοποιώντας φίλτρα ζώνης με διαφορετικές κεντρικές συχνότητες το καθένα τα οποία συγκρίνονται στην έξοδο. Όταν κάποιο μέγιστο που εντοπίζεται στο φάσμα αντιστοιχεί σε μια ζώνη διέλευσης ενός φίλτρου τότε η τιμή της εξόδου του φίλτρου θα είναι μεγαλύτερη από ότι πριν.

- Μέθοδος βέλτιστων φίλτρων τύπου κτένας. (Comb filter)

Γενικά, τα φίλτρα τύπου κτένας αποτελούνται από πολλά ισαπέχοντα φίλτρα διέλευσης ζώνης. Στην περίπτωση ενός αλγορίθμου που βασίζεται σε αυτή τη μέθοδο, η κεντρική συχνότητα του πρώτου φίλτρου διέλευσης ζώνης είναι και αυτή που καθορίζει και την θέση των υπολοίπων. Όταν το σήμα εισέρχεται στον αλγόριθμο φιλτράρεται ανάλογα και στην περίπτωση που οι αρμονικές του σήματος ταιριάζουν με τις κεντρικές συχνότητες των φίλτρων διέλευσης ζώνης τότε στην έξοδο του ο αλγόριθμος θα δίνει μέγιστες τιμές εντοπίζοντας έτσι και την θεμελιώδη συχνότητα του σήματος. Αν όμως για παράδειγμα το εισερχόμενο στο αλγόριθμο σήμα είναι μια απλή ημιτονοειδής κυματομορφή τότε η μέθοδος είναι εσφαλμένη καθώς, όποτε η θεμελιώδη συμπίπτει με ένα φίλτρο, τα περισσότερα φίλτρα διέλευσης ζώνης θα δίνουν την ίδια τιμή στην έξοδο με αυτή της εισόδου.

- Μέθοδος συντονισμένων φίλτρων IIR

Η μέθοδος αυτή είναι μια πιο σύγχρονη μέθοδος, η οποία αποτελείται από ένα φίλτρο διέλευσης ζώνης του οποία η κεντρική συχνότητα επιλέγεται από τον χρήστη και μπορεί να σαρώνει όλο το φάσμα. Όταν το φίλτρο συμπίπτει με κάποια αρμονική του σήματος τότε η έξοδος του θα μεγιστοποιείται και από την κεντρική συχνότητα θα καθορίζεται και η θεμελιώδης. Ο δημιουργός (Lane) της μεθόδου αυτής αναφέρει ότι αν ο χρήστης είναι έμπειρος τότε κατά την σάρωση του φάσματος με σκοπό των εντοπισμό της θεμελιώδους, θα μπορεί να διαχωρίσει αν το φάσμα είναι αρμονικό με όλες τις αρμονικές να είναι πολλαπλάσια της θεμελιώδους ή αν το φάσμα του σήματος έχει παραπάνω από ένα τονικά ύψη. Επίσης, αναφέρει και τρόπους σύμφωνα με τους οποίους θα μπορούσε ένας υπολογιστής να χρησιμοποιηθεί ώστε η διαδικασία διαχωρισμού διάφορων ειδών φασμάτων να είναι πιο γρήγορες και ακριβής.

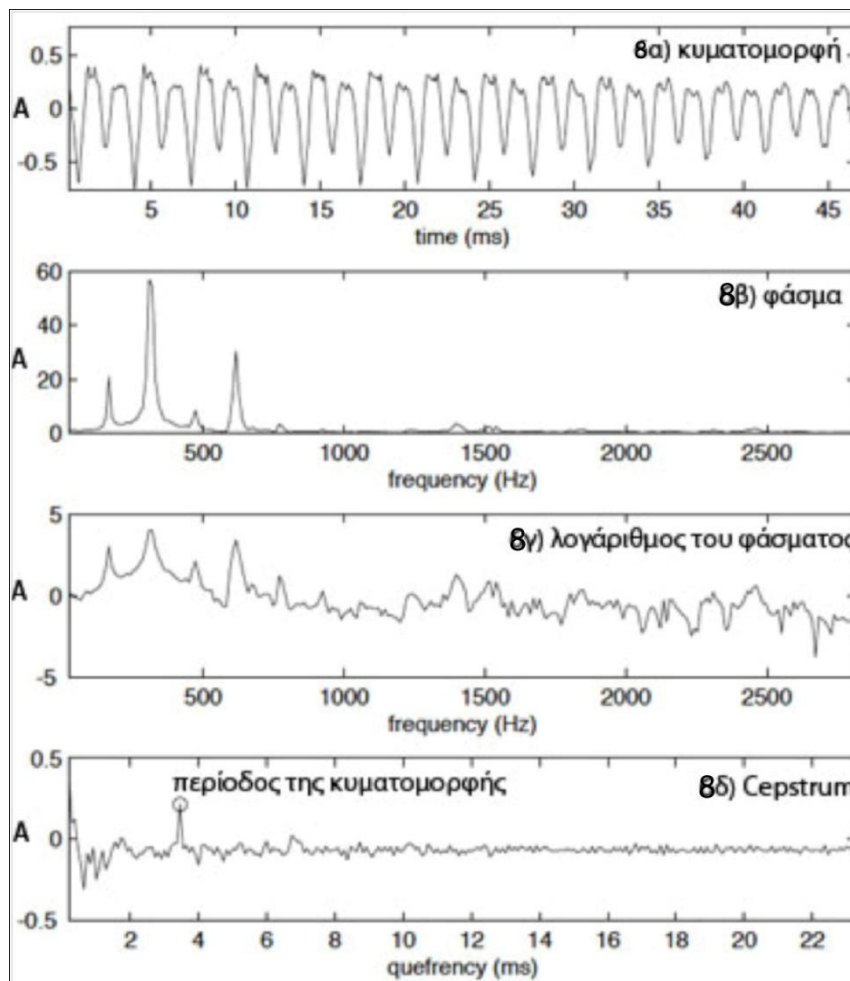
- Ανάλυση Cepstrum

Η ανάλυση Cepstrum είναι μια τεχνική που έχει ως σκοπό την μετατροπή μη γραμμικών συστημάτων σε γραμμικά και αποτελεί ένα είδος φασματικής ανάλυσης, στην έξοδο της οποίας λαμβάνουμε το αποτέλεσμα από τον μετασχηματισμό Fourier του λογαρίθμου του φάσματος του εισερχόμενου σήματος. Το όνομα Cepstrum είναι αναγραμματισμός της αγγλικής λέξης Spectrum και υποδηλώνει έτσι μια παραλλαγή του φάσματος.

Η θεωρία πίσω από την μέθοδο αυτή βασίζεται στο γεγονός ότι κατά τον μετασχηματισμό Fourier ενός τονικού σήματος, συχνά παρουσιάζεται ένας αριθμός από μέγιστα τα οποία αναπαριστούν το αρμονικό φάσμα του ήχου αυτού. Αν αντί του κανονικού φάσματος χρησιμοποιηθεί ο λογάριθμος του τότε τα μέγιστα αυτά μειώνονται αισθητά, το πλάτος τους μετατρέπεται σε πιο χρήσιμη κλίμακα και το αποτέλεσμα είναι ένα περιοδικό σήμα (κυματομορφή) στο πεδίο της συχνότητας, η περίοδος του οποίου σχετίζεται άμεσα με την θεμελιώδη συχνότητα καθώς αποτελεί την απόσταση μεταξύ των μεγίστων που αναφέρθηκαν παραπάνω.

Τέλος, η μέθοδος αυτή προϋποθέτει το εισερχόμενο στον αλγόριθμο σήμα να είναι πλούσιο σε αρμονικές που ισαπέρχουν μεταξύ τους. Σε άλλη περίπτωση

(πχ μη αρμονικούς ήχου(καμπάνα, κρουστά) ή ημιτονοειδής κυματομορφές) η μέθοδος δίνει εσφαλμένα αποτελέσματα. Συγκριτικά με άλλες μεθόδους εντοπισμού της θεμελιώδης συχνότητας, η συγκεκριμένη μέθοδος λειτουργεί πιο αποδοτικά σε σήματα φωνής, τα οποία είναι αρμονικά αλλά και οι αρμονικές του απέχουν κατά ίσες αποστάσεις πάνω στο φάσμα. Στο παρακάτω σχήμα φαίνεται διαδικασία της παραπάνω μεθόδου. Στο σχήμα 8α φαίνεται η κυματομορφή του σήματος εισόδου στο πεδίου του χρόνου ενώ στο 8β, το ίδιο σήμα απεικονίζεται στο πεδίο της συχνότητας. Στο επόμενο σχήμα 8γ απεικονίζεται ο λογάριθμος του προηγούμενου και τέλος, στο 8δ το αποτέλεσμα της ανάλυσης Cepstrum.



Εικόνα 1.7Gerhard[2003]. Τα τέσσερα στάδια της διαδικασίας Cepstrum.

### 1.3 Ανάλυση συχνότητας

Το 1807, ο Jean Baptiste Joseph, Baron de Fourier κατέθεσε στο Institute de France μια έρευνα του πάνω στο πρόβλημα για την ροή της θερμότητας. Οι ιδέες που εμπεριέχονταν σε αυτό το έργο του χαρακτηρίστηκαν ως αμφιλεγόμενες και για αυτό το λόγο η έρευνα του δημοσιοποιήθηκε 15 χρόνια αργότερα (Lynn, 1994). Μέσα στο έργο του δείχνει ότι τα περιοδικά σήματα μπορούν αναπαρασταθούν σαν σταθμισμένα αθροίσματα αρμονικά σχετιζόμενων ημιτόνων. Επίσης, έδειξε ότι μη επαναλαμβανόμενα ή μη περιοδικά σήματα μπορούν να θεωρηθούν σαν το ολοκλήρωμα των ημιτόνων τα οποία δεν είναι σχετίζονται αρμονικά. Αυτές οι δύο ιδέες αποτέλεσαν αντίστοιχα την βάση για τους όρους Σειρές Fourier και μετασχηματισμό Fourier.

Η ανάλυση συχνότητας οποιουδήποτε σήματος προϋποθέτει τον μετασχηματισμό του από το πεδίο του χρόνου, στο πεδίο της συχνότητας το οποίο προσθέτει μια άλλη διάσταση στην ανάλυση δεδομένων, που στο πεδίο του χρόνου είναι αρκετά δύσκολο έως και ακατόρθωτο να επιτευχθεί (Khan, 2005). Οι ανάγκες ενός τέτοιου μετασχηματισμού, οφείλεται στο γεγονός ότι η ψηφιακή επεξεργασία σήματος γενικά επιτυγχάνεται χρησιμοποιώντας συστήματα που περιγράφονται μέσω όρων απόκρισης συχνότητας, τα οποία συμπεριφέρονται διαφορετικά σε διαφορετικές συχνότητες. Μετατρέποντας σήματα αλλά και συστήματα από το πεδίο του χρόνου στο πεδίο της συχνότητας βοηθάει στη καλύτερη κατανόηση των χαρακτηριστικών και των δύο. Στο 1.3.1 γίνεται μια εισαγωγή πάνω στις σειρές Fourier και στον μετασχηματισμό τους. Οι σειρές Fourier είναι μια τεχνική η οποία διαχειρίζεται αποδοτικά περιοδικές συναρτήσεις. Παρέχει μεθόδους σύμφωνα με τους οποίους μια περιοδική συνάρτηση μπορεί να εκφραστεί σε ένα γραμμικό συνδυασμό ημιτονοειδών συναρτήσεων. Γνωρίζοντας τα συχνотικά συστατικά μιας συνάρτησης ή ενός σήματος, η συνολική επεξεργασία του συστήματος μπορεί να εκτιμηθεί σαν το άθροισμα των αποκρίσεων των συχνοτήτων αυτών σαν να ήταν ξεχωριστά σήματα εισόδου του συστήματος (Khan, 2005). Ο μετασχηματισμός Fourier από

την άλλη, είναι απαραίτητος στο να δημιουργηθεί η έννοια του πεδίου της συχνότητας καθώς και η επεξεργασία του (Kuo, 2001). Στο 1.3.2 γίνεται μια πιο εκτεταμένη αναφορά στην τεχνική FFT, καθώς είναι αυτή που χρησιμοποιείται στην εργασία αυτή.

### 1.3.1 Ανάλυση Fourier

Ο μετασχηματισμός Fourier είναι μια ιδέα με τεράστια θεωρητική και πρακτική σπουδαιότητα σύμφωνα με την οποία, όπως αναφέρθηκε και παραπάνω, μια κυματομορφή, πχ ένα ακουστικό σήμα, μπορεί να αναπαρασταθεί σαν ένα σύνολο απλών τόνων με ξεχωριστή συχνότητα, πλάτος και φάση. Εάν η κυματομορφή είναι περιοδική τότε οι τόνοι αυτοί θα είναι αρμονικοί, δηλαδή ακέραια πολλαπλάσια της θεμελιώδους συχνότητας. Η ανάλυση Fourier αποτελεί ένα τρόπο ώστε να γίνονται προφανής τα συχνοτικά στοιχεία τέτοιων τόνων. Επίσης, μπορεί να πάρει και μια καθαρά υλική υπόσταση. Για παράδειγμα, πολλά αντικείμενα στον υλικό περιβάλλον συμπεριφέρονται σαν αντηχεία, το οποίο σημαίνει ότι αποκρίνονται βέλτιστα σε συγκεκριμένες συχνότητες. Αυτό που κάνουν τα αντικείμενα αυτά ουσιαστικά είναι μια ανάλυση Fourier με την έννοια ότι προβάλλουν συγκεκριμένες συχνότητες όποτε αυτές εμπεριέχονται στο σήμα που αλληλεπιδρά με το αντικείμενο. Παρακάτω γίνεται μία απλή μαθηματική αναφορά στις σειρές Fourier.

Αρχικά θεωρούμε μια χρονική συνάρτηση  $x(t)$  σαν περιοδική συνάρτηση με περίοδο  $T$ . Η περιοδικότητα ενός σήματος συνεπάγεται στο ότι μπορούμε να προσθέσουμε οποιοδήποτε άρτιο πολλαπλάσιο  $mT$  στην μεταβλητή του χρόνου  $t$  και η συνάρτηση θα έχει την ίδια ακριβώς τιμή,

$$x(t + mT) = x(t), \quad \text{για όλα τα άρτια } m \quad (1.6)$$

Επειδή ο αριθμός  $m$  μπορεί να πάρει θετικές και αρνητικές τιμές χωρίς κάποιο όριο, τότε το  $x$  μπορεί να είναι περιοδικό σε ένα άπειρο παρελθόν και μέλλον. Το θεώρημα του Fourier λέει τότε ότι το  $x$  μπορεί να αναπαρασταθεί σαν ένα σύνολο από σειρές σύμφωνα με τον παρακάτω τύπο:

$$x(t) = A_0 + \sum_{n=1}^{\infty} [A_n \cos(\omega_n t) + B_n \sin(\omega_n t)] \quad (1.7)$$

Με τις παραπάνω σειρές το  $x(t)$  εκφράζεται σαν ένα άθροισμα ημιτόνων και συνημίτονων. Εκτός του όρου  $A_0$ , όπου είναι μια σταθερά, όλοι οι άλλοι όροι είναι μια συνάρτηση ημιτόνου ή συνημίτονου. Οι σειρές εδώ αποτελούν μια συνάρτηση του χρόνου και όχι της συχνότητας. Ο χρόνος είναι η μόνη μεταβλητή στην εξίσωση και όλοι οι άλλοι παράμετροι θεωρούνται σταθεροί. Όλα τα ημίτονα και συνημίτονα εμπεριέχουν συχνότητες  $\omega_n$  που είναι ακέραια πολλαπλάσια (αρμονικές) της θεμελιώδους γωνιακής συχνότητας  $\omega_0$ ,

$$\omega_n = n\omega_0 = 2\pi n/T \quad (1.8)$$

όπου  $n$  είναι ένας ακέραιος αριθμός.

### 1.3.2 Μετασχηματισμός Fourier (Fourier Transform) – Βασικές Αρχές

Ένα από τα στάδια της υλοποίησης του συστήματος που θα αναλυθεί στο επόμενο κεφάλαιο είναι η μετατροπή του εισαγόμενου στο σύστημα σήματος σύμφωνα με την παραπάνω τεχνική. Τα τελευταία 50 χρόνια, (Ludeman, 1986) έχουν μελετηθεί πάρα πολύ μέθοδοι πάνω σε αυτόν τον αλγόριθμο, που έχουν σαν σκοπό την μείωση των πολλαπλασιασμών, που είναι συχνά πιο σημαντικοί από τις προσθέσεις. Η τεχνική αυτή αναπτύχθηκε από τους Cooley και Turkey και κατάφερε να προσφέρει ακόμα περισσότερες δυνατότητες πάνω στην ανάλυση του φάσματος σημάτων. Σε σχέση με την μέθοδο DFT<sup>3</sup>, η τεχνική FFT μειώνει τον υπολογιστικό χρόνο επεξεργασίας, ανακατανέμοντας έξυπνα της λειτουργίες της DFT και εκμεταλλεύονταν συγκεκριμένες ιδιότητες του, προς όφελος της ταχύτητας των αποτελεσμάτων. Η τεχνική αυτή βασίζεται στην αποσύνθεση ή διάσπαση του μετασχηματισμού σε ακόμα μικρότερα τμήματα τα

---

<sup>3</sup> Discrete Fourier Transform



οποία συνδυάζει στη συνέχεια ώστε να δώσει τον τελικό μετασχηματισμό. Όπως έχουν δείξει οι δημιουργοί της τεχνικής αυτής, ο αποδεκατισμός μπορεί να γίνει στο πεδίο του χρόνου αλλά και της συχνότητας. Παρακάτω, γίνεται μια βασική αναφορά στην τεχνική DFT πάνω στην οποία βασίζονται οι FFT αλγόριθμοι.

### 1.3.3 Διακριτός Σχηματισμός Fourier

Στην ψηφιακή επεξεργασία σήματος τα συνήθως διαχειριζόμενα σήματα είναι μη περιοδικά με πεπερασμένο αριθμό μη μηδενικών τιμών δειγμάτων. Ο μετασχηματισμός ενός σήματος  $X[n]$  κατά την παραπάνω τεχνική για το διάστημα  $0 \leq n \leq (N-1)$  δίνεται από τον παρακάτω τύπο:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N} = \sum_{n=0}^{N-1} x[n] W_N^{kn} \quad (1.9)$$

όπου  $W = e^{-j2\pi/N}$ , ενώ οι φασματικές συντεταγμένες  $X[k]$  υπολογίζονται για το διάστημα  $0 \leq k \leq (N-1)$ . Η αντίστροφη διαδικασία της παραπάνω τεχνικής ώστε να επιστρέψουμε στο αρχικό μας σήμα περιγράφεται με την παρακάτω εξίσωση:

$$X[n] = \frac{1}{N} \sum_{k=0}^{N-1} x[k] X[k] W_N^{-kn} \quad (1.10)$$

Από τα παραπάνω προκύπτει ότι οι ίδιες τιμές του  $x[n] W_N^{kn}$  υπολογίζονται αρκετές φορές κατά την διαδικασία της. Αυτό συμβαίνει γιατί ο όρος  $W_N^{kn}$  είναι μια περιοδική συνάρτηση με περιορισμένο αριθμό ευδιάκριτων τιμών. Αυτόν τον πλεονασμό προσπαθεί να απαλείψει η τεχνική του FFT. Η γενική αρχή του βασίζεται στην αποδόμησή του DFT σε έναν αριθμό από απλούστερα και μικρότερα συνεχόμενα τμήματα (DFTs).

### 1.3.4 Γρήγορος Μετασχηματισμός Fourier

Εδώ θα περιγραφεί η διαδικασία αποδόμησής του DFT σε μικρότερες ακολουθίες μέσω του FFT. Έστω ένα σήμα  $N$  δειγμάτων, όπου  $N$  ακέραια πολλαπλάσια της δύναμης του δύο. Αρχικά διαχωρίζεται η  $x[n]$  σε δύο ακολουθίες με  $N/2$  δείγματα η καθεμία. Η πρώτη ακολουθία αποτελείται από ζυγό αριθμό σημείων ενώ η δεύτερη από μονό αριθμό. Γράφοντας  $n = 2r$  για την πρώτη και  $n = 2r + 1$  για τη δεύτερη ο DFT μπορεί να αναδιαμορφωθεί ως:

$$\begin{aligned}
 X[k] &= \sum_{n=0}^{N-1} x[n] W_N^{kn}, 0 \leq k \leq (N-1) \\
 &= \sum_{r=0}^{N/2-1} x[2r] (W_N^{2rk}) + \sum_{r=0}^{N/2-1} x[2r+1] (W_N^{2r+1})^k \\
 &= \sum_{r=0}^{N/2-1} x[2r] (W_N^2)^{rk} + W_N^k \sum_{r=0}^{N/2-1} x[2r+1] (W_N^2)^{rk}
 \end{aligned} \tag{1.11}$$

Αν τώρα  $W_N^2 = e^{(-2j2\pi/N)} = W_{N/2}$  καταλήγουμε ότι:

$$\begin{aligned}
 X[k] &= \sum_{r=0}^{N/2-1} x[2r] W_{N/2}^{rk} + W_N^k \sum_{r=0}^{N/2-1} x[2r+1] W_{N/2}^{rk} \\
 &= G[k] + W_N^k H[k]
 \end{aligned} \tag{1.12}$$

Από την παραπάνω σχέση, το αρχικό DFT σήμα  $N$ -σημείων εκφράζεται μέσω δύο όρων  $N/2$ -σημείων, του  $G[k]$  και του  $H[k]$  όπου αποτελούν τους μετασχηματισμούς ζυγών και περιττών σημείων αντίστοιχα. Η παραπάνω μέθοδος ονομάζεται αποδεκατισμός στον χρόνο και αποτελεί έναν κοινό FFT αλγόριθμο. Αν για παράδειγμα έχουμε ένα σήμα  $x[n]$  όπου  $N=8$  τότε το σήμα αυτό καθορίζεται για:

$$n = \{0, 1, 2, 3, 4, 5, 6, 7\}$$

Σύμφωνα με το διαχωρισμό ζυγών και περιττών σημείων παίρνουμε τις εξής ακολουθίες:

$$n = \{0,2,4,6\} \text{ και } n = \{1,3,5,7\}$$

Και συνεχίζοντας ακόμα περισσότερο τον αποδεκατισμό έχουμε:

$$n = \{0,4\} \text{ και } n = \{2,6\} \text{ και } n = \{1,5\} \text{ και } n = \{3,7\}$$

Όπου φαίνεται πως ένα αρχικό σήμα  $N=8$  σημείων εκφράζεται σαν ζεύγος δειγμάτων χωρισμένων κατά το μισό του μήκους του μετασχηματισμού.

## 2 Υλοποίηση

Το κεφάλαιο αυτό θα περιγράψει όλη τα στάδια σχεδίασης και υλοποίησης του συστήματος, το οποίο γενικά μπορεί να διαχωριστεί σε δύο βασικές μονάδες. Η πρώτη μονάδα αφορά την ανάλυση και επεξεργασία του σήματος εισόδου με σκοπό τον εντοπισμό της θεμελιώδους συχνότητας, την εξαγωγή διάφορων πληροφοριών σχετικών με το σήμα και διάφορες άλλες επεξεργασίες ενώ η δεύτερη στοχεύει στην απεικόνιση του αποτελέσματος της πρώτης πάνω σε ένα εικονικό πεντάγραμμο σε μορφή μουσικών φθόγγων. Η όλη διαδικασία γίνεται σε πραγματικό χρόνο και περιγράφεται παρακάτω.

### 2.1 Ανάλυση σήματος εισόδου

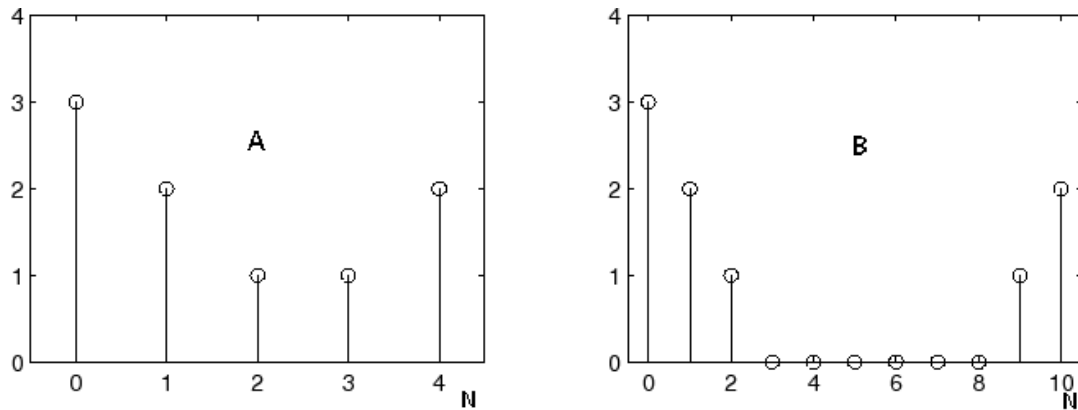
Η μονάδα αυτή, όπως αναφέρθηκε παραπάνω, αναλύει το σήμα που εισάγεται στο σύστημα με σκοπό την αναγνώριση της θεμελιώδους συχνότητας. Να τονιστεί εδώ ότι δεν γίνεται αναφορά στο τονικό ύψος καθώς, όπως περιγράφηκε στο κεφάλαιο 1, κάτι τέτοιο προϋποθέτει αρκετά υποκειμενικά κριτήρια για την ορθότητα των αποτελεσμάτων. Από την άλλη, ο εντοπισμός της θεμελιώδους συχνότητας μιας κυματομορφής μπορεί να χαρακτηριστεί αρκετά πιο απλή διαδικασία καθώς γενικά εξαρτάται από την φύση και το είδος της ίδιας της κυματομορφής και όχι τόσο από την ανθρώπινη αντίληψη.

#### 2.1.1 Τεχνική συμπλήρωσης μηδενικών ( Zero Padding)

Η τεχνική αυτή αποτελεί το πρώτο στάδιο πριν από οποιαδήποτε επεξεργασία. Σύμφωνα με αυτήν, το δειγματοληπτημένο σήμα μεγέθους  $N$  γεμίζει με μηδενικά και εκτείνεται σε ένα νέο μήκος  $M > N$  όπου το πρώτο δεν διαιρεί το τελευταίο. Ο ορισμός του παραπάνω έχει ως εξής:

$$ZeroPad_{M,N}(W) = \begin{cases} W[n], & |n| < N/2 \\ 0 & \text{αλλιώς} \end{cases} \quad (2.1)$$

όπου  $n = 0, \pm 1, \pm 2, \dots, M_h$  για  $M_h = \begin{cases} (M-1)/2, & M = \text{θετικο} \\ (M/2)-1, & M = \text{αρνητικο} \end{cases}$



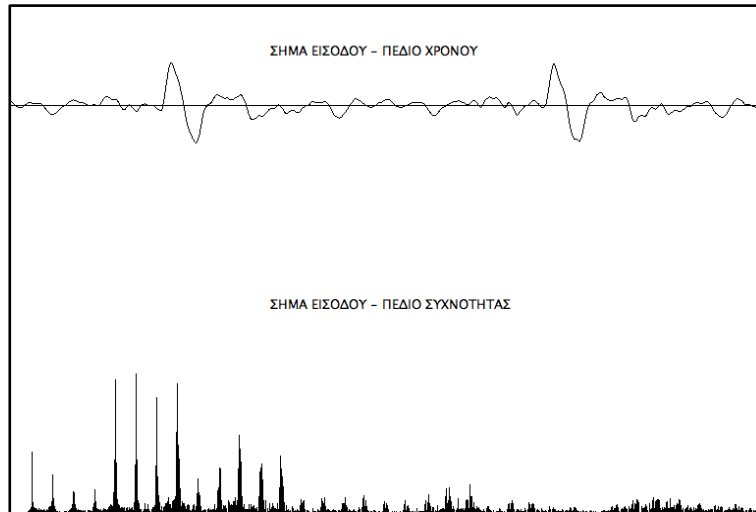
Εικόνα 2.1 Η τεχνική zero padding σε ένα σήμα όπου  $N=5$  και  $A) n = [3, 2, 1, 1, 2] B) ZeroPad_{11}(n)$

Στο παρακάτω σχήμα βλέπουμε την παραπάνω τεχνική για ένα σήμα όπου  $N=5$ .

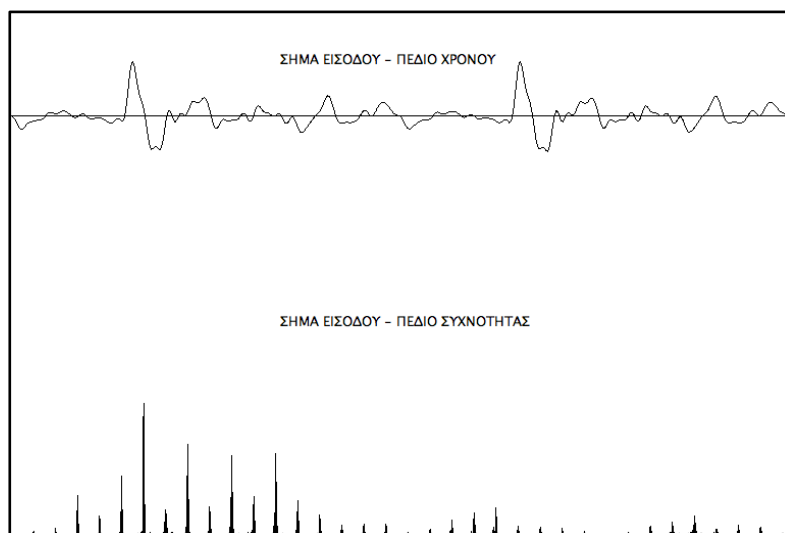
### 2.1.2 Παραθυροποίηση (Windowing)

Κατά την φασματική ανάλυση ακουστικών σημάτων, όπως αναφέρθηκε παραπάνω, το σήμα αναλύεται κατά τμήματα παρά ολόκληρο. Ο κυριότερος ίσως λόγος γι' αυτό είναι ότι, παρόμοια με την ανάλυση Fourier, το ανθρώπινο αυτί αναλύει τα ακουστικά σήματα κατά τμήματα κατά χρονικά διαστήματα της τάξης των 10-20 ms[10]. Έτσι λοιπόν, το σήμα που εισάγει ο χρήστης στο σύστημα, δειγματοληπτείται κατά τμήματα (buffers), με μια συχνότητα δειγματοληψίας  $f_s$  και εισέρχεται στο στάδιο της παραθυροποίησης (windowing), στο οποίο τα δείγματα του σήματος διαμορφώνονται σύμφωνα με μία συνάρτηση, με σκοπό την μείωση των ασυνεχειών ανάμεσα από τα τμήματα αυτά. Οι ασυνέχειες αυτές δημιουργούνται στη περίπτωση που το σήμα δεν επαναλαμβάνεται περιοδικά με αποτέλεσμα να προκαλούνται φασματικές διαρροές στο συχνοτικό περιεχόμενο που θα εισέλθει στην FFT. Εφαρμόζοντας σε κάθε τμήμα του σήματος την συνάρτηση της παραθυροποίησης, μειώνονται

οι τιμές των δειγμάτων στα άκρα του κάθε τμήματος και απαλείφονται οι ασυνέχειες ανάμεσα από τα τμήματα. Επίσης, η επιλογή της συνάρτησης της παραθυροποίησης καθορίζει την αντίστροφη αναλογία του χρόνου και την ανάλυση συχνότητας το οποίο με τη σειρά του επηρεάζει την ομαλότητα του φάσματος. Στα παρακάτω σχήματα, φαίνονται δύο διαφορετικές μορφές της συνάρτησης της παραθυροποίησης όπως υλοποιήθηκαν στο σύστημα:



**Εικόνα 2.2** Σήμα εισόδου μήκους  $L = 8192$  δειγμάτων, πάνω στο πεδίο του χρόνου και της συχνότητας, χωρίς την εφαρμογή της συνάρτησης παραθυροποίησης.



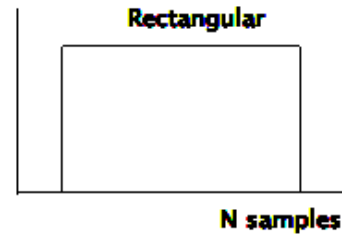
**Εικόνα 2.3** Σήμα εισόδου μήκους  $L = 8192$  δειγμάτων, πάνω στο πεδίο του χρόνου και της συχνότητας, με εφαρμογή της συνάρτησης παραθυροποίησης σχήματος Hamming.

Συνολικά, στο σύστημα εμπεριέχονται 7 διαφορετικές συναρτήσεις παραθυροποίησης οι οποίες περιγράφονται παρακάτω :

- **Rectangular**

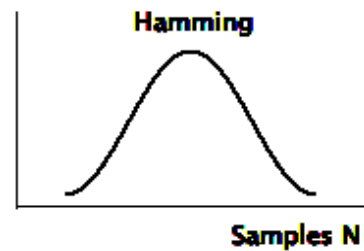
Η πιο απλή μορφή παραθυροποίησης κατά την τα δείγματα παραμένουν ανεπηρέαστα.

$$W_n = \begin{cases} 1, & -\frac{N-1}{2} \leq n \leq \frac{N-1}{2} \\ 0, & \text{αλλιώς} \end{cases}$$



- **Hamming**

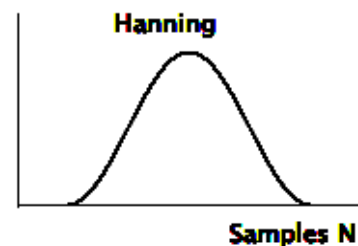
Αυτή η κατηγορία μπορεί να περιγραφεί ως η παραπάνω με την προσθήκη ενός ημιτόνου το οποίο έχει μια περίοδο κατά μήκος του παραθύρου.



$$W_n = 0.54 + 0.46 \cos n\pi/N, \quad -N < n < N$$

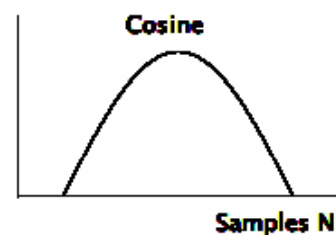
- **Hanning**

$$W_n = 0.5 + 0.5 \cos n\pi/N, \quad -N < n < N$$



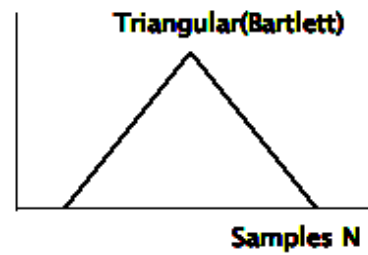
- **Cosine**

$$W_n = \cos\left(\frac{\pi n}{N-1} - \frac{\pi}{2}\right)$$



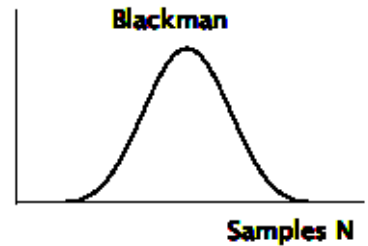
- **Triangular (Bartlett)**

$$W_n = \begin{cases} \frac{2n}{N}, & 0 \leq n \leq \frac{N}{2} \\ 2 - \frac{2n}{N}, & \frac{N}{2} \leq n \leq N \end{cases}$$



- **Blackman**

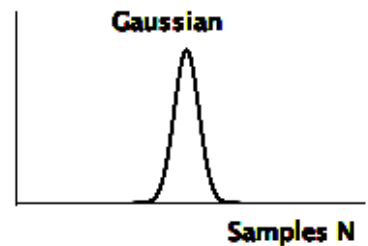
$$W_n = 0.42 - 0.5 \cos\left(2\pi \frac{n}{N}\right) + 0.08 \cos\left(4\pi \frac{n}{N}\right), \quad 0 \leq n \leq N$$



- **Gaussian**

$$W_n = e^{-\frac{1}{2}\left(\alpha \frac{n}{N/2}\right)^2}, \quad \frac{-N}{2} \leq n \leq \frac{N}{2}$$

$\alpha = 2.5$



### 2.1.3 Μετασχηματισμός FFT

Μετά το στάδιο της παραθυροποίησης το σήμα κατακεραματίζεται μέσω της τεχνικής FFT σε ίσες συχνοτικές μπάντες (bins), των οποίων το μέγεθος και ο αριθμός εξαρτάται από το μέγεθος του σήματος που εισέρχεται στην FFT και την συχνότητα δειγματοληψίας. Η κάθε μία μπάντα έχει μία κεντρική συχνότητα, η οποία εκφράζεται ως ένα κλάσμα της συχνότητας δειγματοληψίας του αρχικού σήματος και είναι ίση με το αριθμητικό δείκτη της μπάντας προς τον συνολικό αριθμό των συχνοτικών μπαντών. Ο αριθμός των μπαντών αυτών ισούται με το συνολικό μήκος (αριθμός δειγμάτων  $N$ ) του εισερχόμενου σήματος (πεδίο του χρόνου), όμως χρησιμοποιούνται μόνο αυτές με δείκτες



μικρότερες του μισού του μήκους του σήματος οι οποίες αντιστοιχούν σε συχνότητες μικρότερες της συχνότητας Nyquist<sup>4</sup>. Έτσι, αν ένα σήμα έχει  $N$  δείγματα τότε στο τελικό φάσμα μετά την FFT θα έχει  $N/2$  μπάντες. Για παράδειγμα, σε μια εφαρμογή όπου το εισερχόμενο σήμα θα αναλύεται από την FFT ανά  $L = 2048$  δείγματα και με δειγματοληψία  $f_s = 44100\text{Hz}$ , η συχνοτική μπάντα με δείκτη  $i = 5$  θα αντιστοιχεί σε εκείνη την μπάντα όπου η κεντρική της συχνότητα  $f_c$  θα ισούται με:

$$f_c = (i/L) * f_s = (5/2048) * 44100 = 107\text{Hz} \quad (2.2)$$

Το εύρος της παραπάνω μπάντας ισούται με  $2/L$  και εκφράζεται ως κλάσμα του συνολικού φάσματος, το οποίο όπως αναφέρθηκε ισούται με την συχνότητα Nyquist,

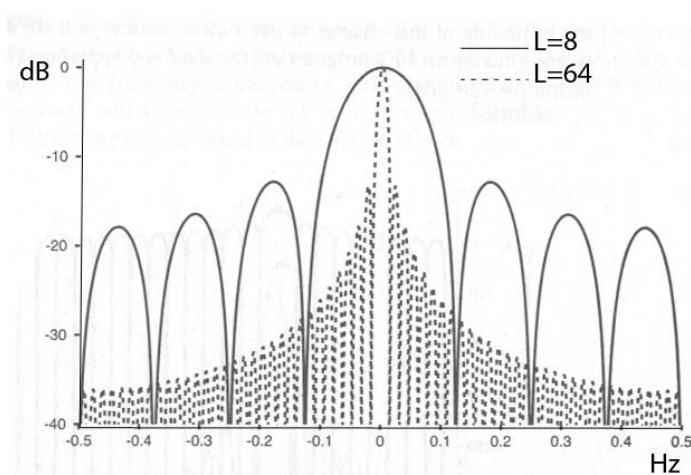
$$f_{bw} = (2/L) * f_{\text{Nyquist}} = (2/L) * 22050 = 21\text{Hz} \quad (2.3)$$

Από τα παραπάνω, φαίνεται ότι όσο μεγαλύτερος είναι ο αριθμός των δειγμάτων  $L$  τόσο μικρότερο είναι το εύρος της κάθε μπάντας. Πρακτικά, στο δυτικό μουσικό σύστημα αυτό συνεπάγεται στο ότι όσο μικρότερο το μήκος του σήματος που εφαρμόζεται η FFT, τόσο δυσκολότερος θα γίνεται ο εντοπισμός ή διαχωρισμός στις χαμηλότερες μουσικές νότες λόγω του ότι το ανθρώπινο αυτί ακούει σχεδόν λογαριθμικά. Για παράδειγμα, το συχνοτικό διάστημα ανάμεσα από την Λα(A) και ΣΙ(B) της τρίτης οκτάβας είναι περίπου 13 Hz, ενώ για τις ίδιες νότες στη πέμπτη οκτάβα το διάστημα εκτείνεται στα 215 Hz. Άρα με τα δεδομένα τις (2) η νότα Σι στην τρίτη οκτάβα δεν θα γίνονταν αντιληπτή από το σύστημα. Από την άλλη όμως, όσο μεγαλύτερο είναι το μήκος του σήματος, τόσο μεγαλύτερη είναι και η επεξεργαστική ισχύς που απαιτεί η διαδικασία

---

<sup>4</sup> Αν  $f_s$  συχνότητα δειγματοληψίας τότε η συχνότητα Nyquist ισούται με  $f_s/2$

κατακερματισμού καθώς είναι περισσότερα τα δείγματα που πρέπει να υπολογιστούν. Άρα εδώ συνήθως πρέπει να γίνεται ένας συμβιβασμός ανάμεσα στην χαμηλότερη χρήσιμη συχνότητα που θα εντοπίζεται από το σύστημα αλλά και στις επεξεργαστικές του δυνατότητες. Στο παρακάτω σχήμα, φαίνεται απλοποιημένο το φαινόμενο αυτό, για ένα δειγματοληπτημένο σήμα, για δύο διαφορετικά μήκη, χωρίς την εφαρμογή της συνάρτησης παραθυροποίησης. Είναι ξεκάθαρο ότι για μεγαλύτερο μήκος δειγμάτων η ανάλυση αυξάνεται με αποτέλεσμα να γίνεται ευκολότερος ο εντοπισμός σε χαμηλότερες συχνότητες.



**Εικόνα 2.4 Steiglitz[1996], Σύγκριση ενός απλοποιημένου φάσματος για δύο διαφορετικά μήκη δειγμάτων. Στο σήμα δεν έχει εφαρμοστεί παραθυροποίηση.**

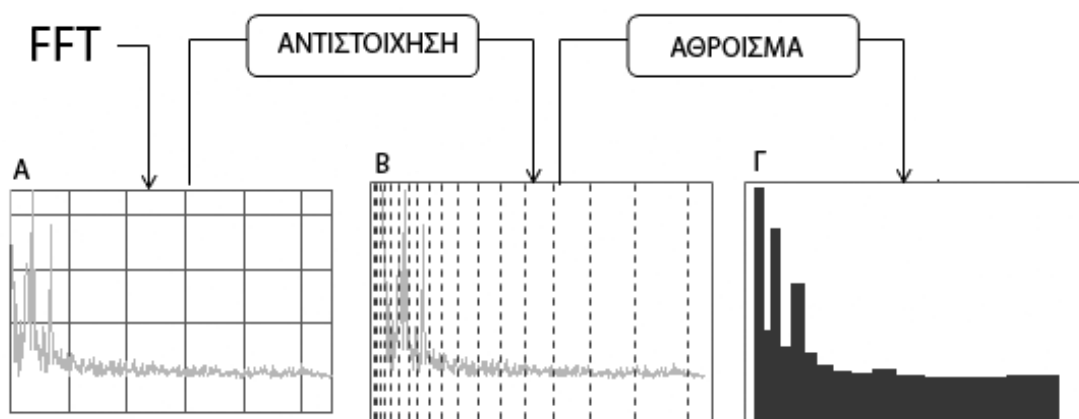
#### 2.1.4 Κλάση Επεξεργασίας και ανάλυσης τονικού ύψους (Pitch Class Profile)

Ο όρος αυτός γενικά εκφράζει ένα διάνυσμα το οποίο αναπαριστά την ένταση κάθε ενός από τα 12 ημιτόνια στην δυτική τονική κλίμακα ενώ χρησιμοποιούνται ευρέως σε εφαρμογές που σχετίζονται με αρμονικό περιεχόμενο, όπως εκτίμηση τονικού ύψους και εντοπισμός μουσικού κλειδιού. Μερικά από τα κύρια πλεονεκτήματα των PCP είναι η απλότητα των υπολογισμών και η σύνοψη της αρμονικής πληροφορίας.

Στην εργασία αυτή υπολογίζοντας τα διανύσματα αυτά, κάθε συχνοτική μπάντα του φάσματος αντιστοιχίζεται με το κοντινότερο από τα 12 τονικά ύψη της μουσικής κλίμακας. Η αντιστοίχιση αυτή γίνεται με τον παρακάτω μαθηματικό τύπο:

$$p_k = [12 \log_2 \left( \frac{k}{N} \frac{f_r}{f_{ref}} \right)] \bmod 12 \quad (2.4)$$

όπου,  $N$  ο αριθμός των δειγμάτων του σήματος,  $k$  ο αριθμητικός δείκτης της συχνοτικής μπάντας του FFT ( $0 \leq k \leq N - 1$ ),  $f_{ref}$  είναι η συχνότητα αναφοράς που αντιστοιχεί στο PCP[0], δηλαδή στο πρώτο ημίτονο της οκτάβας, και  $f_r$  η συχνότητα δειγματοληψίας. Το mod προέρχεται από το modulus και υπολογίζει το υπόλοιπο της διαίρεσης, (πχ  $20 \bmod 6 = 2$ ). Όμως αρχικά, για κάθε μία από τις δώδεκα μουσικές νότες συμπίπτει ένας αριθμός από μπάντες και όχι μόνο μία. Έτσι, για κάθε στοιχείο PCP, τα πλάτη των συχνοτικών μπαντών της κάθε περιοχής (εικ. 2.5β) προστίθενται μεταξύ τους και δίνουν το ιστόγραμμα που φαίνεται στην εικ. 2.5γ και το οποίο είναι παρόμοιο με αυτό που βλέπει ο χρήστης στο σύστημα στην καρτέλα "Analyser".



**Εικόνα 2.5** Cabral[2008]. Τα 3 στάδια υπολογισμού της PCP. Κάθε συχνοτική μπάντα της FFT αντιστοιχεί σε μια περιοχή(B). Οι μπάντες που αντιστοιχούν στην ίδια περιοχή προστίθενται και δίνουν το ιστόγραμμα (Γ).

Τα πλάτη του φάσματος αθροίζονται με τον παρακάτω τύπο:

$$PCP_{[p]} = \sum_{k:p(k)=p} |X[k]|^2 \quad (2.5)$$

όπου  $p$  ο αριθμός κάθε νότας και  $X[k]$  τα πλάτη των μπαντών.

Τις περισσότερες φορές όμως οι δείκτες των συχνοτικών μπαντών δεν συμπίπτουν ακριβώς στη συχνότητα η οποία με τη σειρά της αντιστοιχεί σε κάποια μουσική νότα. Για αυτό το λόγο στη (2.5) υπολογίζεται και η απόσταση  $bin_{distance}$  μεταξύ των συχνοτικών μπαντών η οποία μπορεί να εκφραστεί ως εξής:

$$bin_{distance} = 2 \left( \frac{12 \log\left(\frac{f_{bin}}{440}\right)}{\log 2} - \frac{12 \log\left(\frac{f_{note}}{440}\right)}{\log 2} \right) \quad (2.6)$$

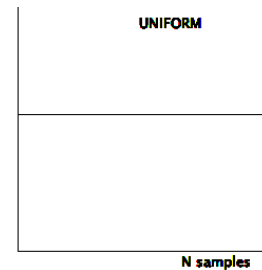
όπου  $f_{bin}$  η συχνότητα κάθε μπάντας και  $f_{note}$  η συχνότητα της κοντινότερης προς τη μπάντα μουσική νότα. Η παραπάνω εξίσωση είναι ανάλογη με τις διαστάσεις κάθε περιοχής στην εικ. 4γ ενώ οι τιμές που δίνει έχουν εύρος από 0 (η συχνοτική μπάντα της FFT συμπίπτει ακριβώς με την  $f_{note}$ ) έως 1 ( $f_{bin}$  βρίσκεται στο ενδιάμεσο δύο νοτών). Προσθέτοντας την παραπάνω εξίσωση στη (2.6), η τελευταία γίνεται ως εξής:

$$PCP_{[p]} = \frac{\sum_{k:p(k)=p} |X[k]|^2 f(bin_{distance})}{\sum_{k:p(k)=p} f(bin_{distance})} \quad (2.7)$$

Οι Cabral, Briot και François Pachet [2008] όρισαν έξι διαφορετικές συναρτήσεις που θα μπορούν να επεξεργάζονται την απόσταση μεταξύ μεταξύ των μπαντών, θα είναι εφαρμόσιμες σε τιμές από 0 έως 1 ενώ θα επιστρέφουν τιμές του ίδιου εύρους. Πέντε από αυτές τις συναρτήσεις εμπεριέχονται στο στάδιο της ανάλυσης του σήματος εισόδου του συστήματος και είναι οι εξής:

- **Uniform**

Η συνάρτηση αυτή δεν επηρεάζει την απόσταση των συχνοτικών μπαντών.



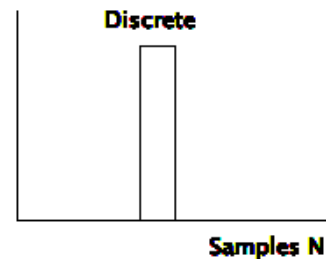
$$W_u = 1$$

- **Discrete**

Η συνάρτηση αυτή λαμβάνει υπόψη της αποστάσεις μικρότερες του 0.2 και αποκόπτει όλες τις υπόλοιπες.

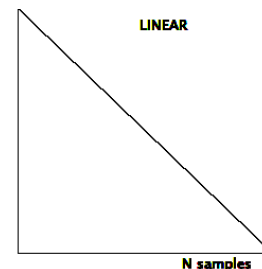
$$W_u = 1, \quad n \leq 0.2$$

$$W_u = 0, \quad n \geq 0.2$$



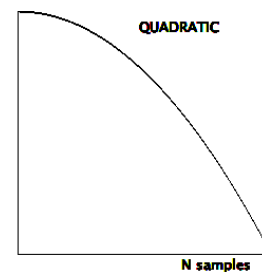
- **Linear**

$$W_u = 1 - n$$



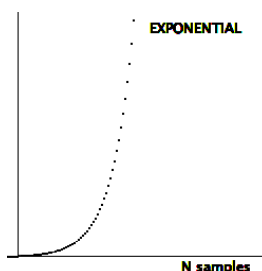
- **Quadratic**

$$W_u = 1 - n^2$$



- **Exponential**

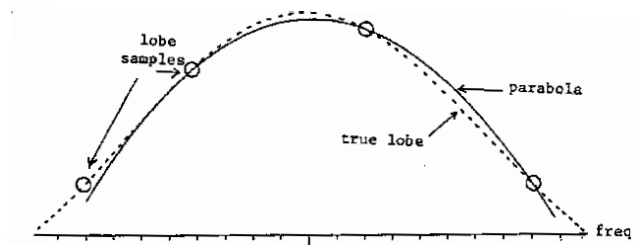
$$W_u = e^{(1-n)/\tau}$$



### 2.1.5 Εντοπισμός θεμελιώδους συχνότητας

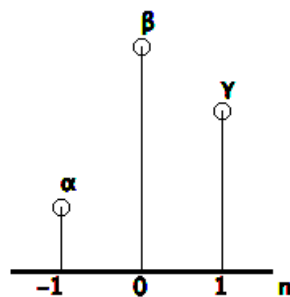
Για την εύρεση της θεμελίου μέσα στο φάσμα χρησιμοποιείται η μέθοδος εντοπισμού των μεγίστων με την τεχνική της παραβολικής παρεμβολής<sup>5</sup> (Interpolation), η οποία περιγράφεται παρακάτω.

Στο παρακάτω σχήμα φαίνεται η μέθοδος του προγράμματος PARSHL<sup>6</sup> η οποία εφαρμόζει την καμπύλη παραβολής για τρία συνεχόμενα και υψηλότερα δείγματα ενός μεγίστου ώστε να εκτιμήσει την πραγματική θέση και πλάτος του.



Εικόνα 2.6 Smith[2010]. Παραβολική καμπύλη των υψηλότερων δειγμάτων ενός μεγίστου.

Αρχικά, ορίζεται ένα σύστημα συντεταγμένων με κέντρο το  $(n_p, 0)$ , όπου  $n_p$  ο αριθμός της συχνотικής μπάντας ενός μεγίστου όπως φαίνεται στο σχήμα:



<sup>5</sup> Τεχνική μοντελοποίησης καμπύλων κατα τρόπο ώστε η παραγόμενη καμπύλη να περνά από όλα τα σημεία ελέγχου. Λεξικό MAZENTA).

<sup>6</sup> Πρόγραμμα ανάλυσης και σύνθεσης μη αρμονιών ήχων (<https://ccrma.stanford.edu/~jos/parshl/>)

Για τις τρεις μεταβλητές  $\alpha, \beta$  και  $\gamma$ , οι οποίες αντιπροσωπεύουν τα τρία υψηλότερα δείγματα, ισχύει:

$$\alpha = W[n-1] \quad \beta = W[n] \quad \gamma = W[n+1] \quad (2.8)$$

Εφαρμόζοντας την συνθήκη  $\alpha < \beta > \gamma$ , η παραβολική θέση  $p$  του μεγίστου μπορεί να δοθεί από τη σχέση:

$$p = \frac{(\gamma - \alpha)}{2(2\beta - \gamma - \alpha)} \quad (2.9)$$

ενώ το πλάτος του μεγίστου υπολογίζεται ως:

$$f(p) = \beta - 0.25 (\alpha - \gamma) p \quad (2.10)$$

Τέλος η συχνότητα του μεγίστου υπολογίζεται εύκολα από τη σχέση:

$$f_{Hz} = (n + p) * f_s / N \quad (2.11)$$

όπου  $f_s$  η συχνότητα δειγματοληψίας και  $N$  ο αριθμός των δειγμάτων της FFT.

### 2.1.6 Μετατροπή συχνότητας σε MIDI πληροφορία

Στο στάδιο αυτό, κάθε συχνότητα  $f_{Hz}$  που εντοπίζεται από το σύστημα μετατρέπεται σε MIDI πληροφορία ώστε να αναπαραχθεί από μια αντίστοιχη γεννήτρια ήχου με σκοπό να δώσει στο χρήστη και ακουστικά πληροφορίες σχετικά με το τονικό ύψος και ένταση της νότας. Ο μαθηματικός τύπος με τον οποίο γίνεται η παραπάνω μετατροπή έχει ως εξής:

$$f_{midi} = 69 + 12 * \frac{(\log(f_{Hz}/440))}{\log(2)} \quad (2.12)$$

ενώ η αντίθετη διαδικασία γίνεται μέσω της σχέσης:

$$f_{hz} = 440 * \left( \frac{f_{midi} - 69}{12} \right)^2 \quad (3.13)$$

---

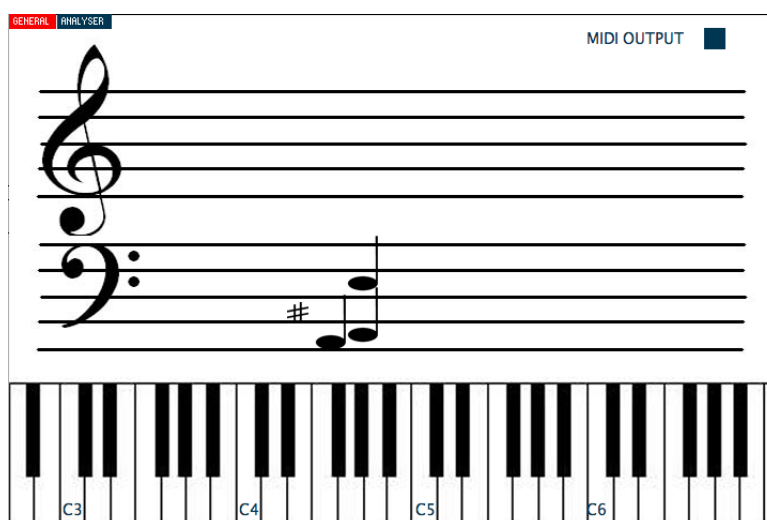
<sup>7</sup> Ο αριθμός 69 αντιστοιχεί στην νότα A4 με συχνότητα 440 Hz και χρησιμοποιείται συνήθως ως αναφορά.



## 2.2 Γραφικό περιβάλλον (User Interface)

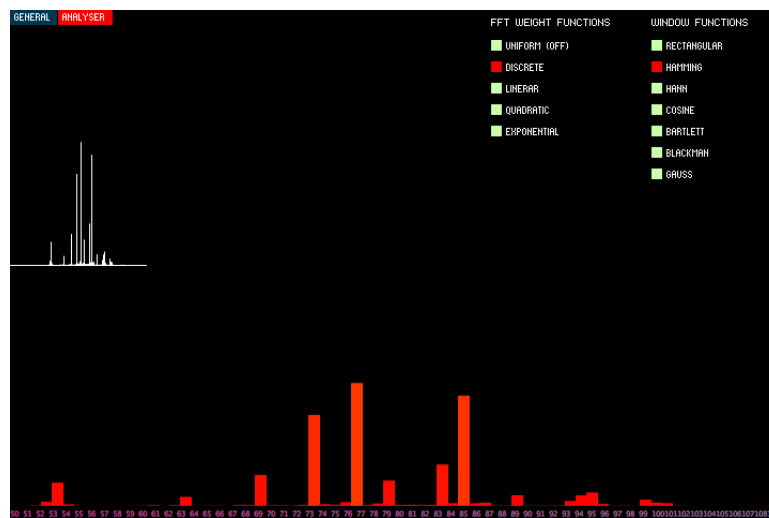
Το γραφικό περιβάλλον του συστήματος χωρίζεται σε δύο βασικές σελίδες τις οποίες ο χρήστης μπορεί να εναλλάσσει οποιαδήποτε στιγμή θέλει. Στη πρώτη σελίδα (αυτή που εμφανίζεται όταν ανοίγει το πρόγραμμα - GENERAL) γίνεται η απεικόνιση της παρτιτούρας σε πραγματικό χρόνο. Εδώ, υπάρχουν δύο πεντάγραμμα, ένα στο κλειδί του σόλ και ένα στο φα με συνολική έκταση σχεδόν 4 οκτάβες (A2 - C5). Ο χρήστης ανάλογα με το σήμα που δίνει στην είσοδο του συστήματος βλέπει την ανάλογη νότα στο πεντάγραμμα. Προς το παρόν, ο εντοπισμός της διάρκειας δεν είναι εφικτός γι' αυτό το λόγο όλες οι νότες έχουν το σχήμα της νότας ενός τετάρτου χωρίς όμως αυτό να σημαίνει ότι διαρκούν τόσο. Για το λόγο αυτό έχει προστεθεί στο σύστημα το φαινόμενο κύλισης προς τα δεξιά.

Στο κάτω μέρος της σελίδας έχει προστεθεί ένα πληκτρολόγιο ενός πιάνου με αντίστοιχη έκταση 5 οκτάβων (A2-C6) στο οποίο απεικονίζονται οι νότες που εντοπίζει το σύστημα και στις οποίες έχει προστεθεί το εφέ του παιχνίματος. Πάνω από κάθε πλήκτρο φαίνεται και η ονομασία κάθε νότας στην αγγλική. Τέλος, στο δεξί πάνω άκρο της σελίδας υπάρχει ένα διακόπτης με τον οποίο μπορεί ο χρήστης να ενεργοποιεί/απενεργοποιεί την MIDI έξοδο από τη γεννήτρια ήχου.



*Εικόνα 2.7 Η πρώτη σελίδα - ετικέτα GENERAL - του προγράμματος όπου ο χρήστης μπορεί να βλέπει την απεικόνιση της παρτιτούρας*

Στην επόμενη σελίδα -ετικέτα ANALYSER- ο χρήστης μπορεί να βλέπει το ιστόγραμμα και το φασματογράφημα, τα οποία κατασκευάστηκαν με τους μαθηματικούς τύπους που αναφέρθηκαν παραπάνω. Ο λόγος που συμπεριλήφθησαν είναι για να δώσουν επιπλέον πληροφορίες στο χρήστη για την φύση του σήματος που εισάγεται στο σύστημα. Στο κάτω μέρος (άξονας χ), φαίνεται η κλίμακα PCP, όπου ο χρήστης μπορεί να βλέπει που αντιστοιχεί κάθε μέγιστο. Επίσης, ο χρήστης έχει την επιλογή να εναλλάσσει το τρόπο, δηλαδή την μαθηματική συνάρτηση, με τον οποίο γίνεται η παραθυροποίηση στο σήμα εισόδου όπως επίσης και την συνάρτηση που ορίζει την απόσταση που θα έχουν οι συχνοτικές μπάντες μεταξύ τους. Και για τις δύο παραπάνω επιλογές ο χρήστης μπορεί να βλέπει και γραφικά την καμπύλη της συνάρτησης.



**Εικόνα 2.8** Η δεύτερη σελίδα του προγράμματος -ετικέτα ANALYSER- στην οποία ο χρήστης παίρνει πληροφορίες για το φάσμα του σήματος εισόδου.

## 2.3 Περιβάλλον προγραμματισμού – Processing 1.2.1

Ο κώδικας υλοποίησης στις εργασίες αυτής έγινε αποκλειστικά στο Processing. Το **Processing**, το οποίο για πολλούς είναι ο απόγονος των **Logo** και **Design by Numbers** [11], δημιουργήθηκε αρχικά από τους Casey Reas και Benjamin Fry το 2001 κατά την διάρκεια του ερευνητικού προγράμματος “*Aesthetics and Computation*” στο MIT<sup>8</sup>. Το Processing βασίζεται πάνω στην γλώσσα προγραμματισμού Java, κάτι το οποίο δημιουργεί μια διαφωνία για το αν αποτελεί από μόνο του μια γλώσσα προγραμματισμού ή ένα προγραμματιστικό περιβάλλον[12]. Στην εργασία αυτή η κατασκευή του κώδικα προσεγγίζεται και με τη μέθοδο συναρτήσεων και με τον αντικειμενοστραφή προγραμματισμό. Η πρώτη προσέγγιση βασίζεται στη χρήση συναρτήσεων, οι οποίες αποτελούν τμήματα κώδικα τα οποία το πρόγραμμα μπορεί να χρησιμοποιεί όποτε τα καλεί ο χρήστης. Η δεύτερη αποτελεί μία πιο αφηρημένη και φιλόδοξη προσέγγιση η οποία μοντελοποιεί ένα προγραμματιστικό πρόβλημα χρησιμοποιώντας όρους από τον φυσικό κόσμο. Στο αντικειμενοστραφή προγραμματισμό, ο κώδικας διαιρείται σε μικρότερα και ανεξάρτητα τμήματα τα οποία ονομάζονται αντικείμενα.

Η εργασία συνοδεύεται από οπτικό δίσκο ο οποίος εμπεριέχει ολόκληρο το κώδικα υλοποίησης μαζί με τα σχετικά σχόλια και διευκρινήσεις.

---

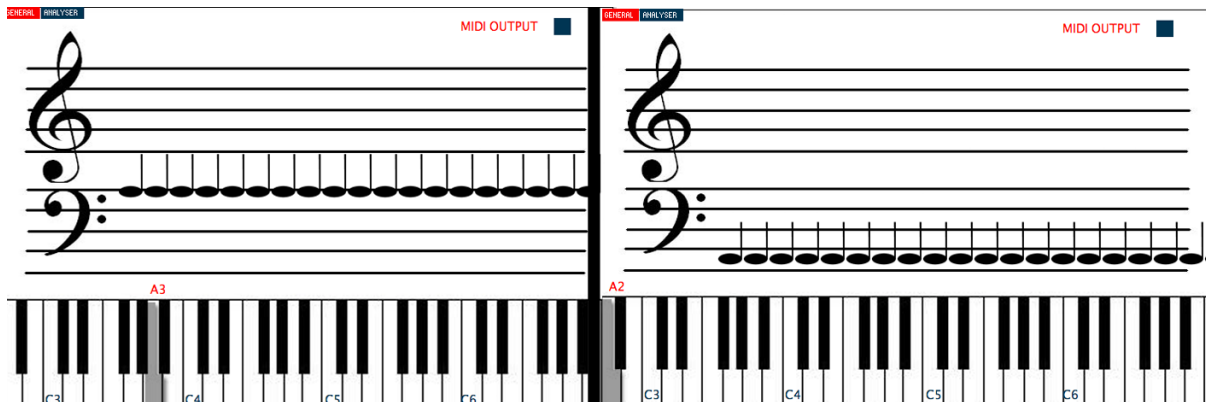
<sup>8</sup> Massachusetts Institute of Technology Media Lab

### 3 Αποτελέσματα

Στο κεφάλαιο αυτό περιγράφεται το πειραματικό μέρος της εργασίας με σκοπό την αξιολόγηση της εξόδου του συστήματος. Στο στάδιο αυτό, στο σύστημα εισήλθαν διάφορα σήματα γνωστής συχνότητας αλλά με διαφορετικό αρμονικό περιεχόμενο ώστε να μπορεί να εκτιμηθεί με ακρίβεια η έξοδος του. Τα σήματα αυτά ήταν, απλή ημιτονοειδής κυματομορφή, πιάνο από ψηφιακό συνθετητή και σήμα φωνής. Για κάθε ένα από αυτά, η συμπεριφορά του συστήματος απεικονίζεται στα σχήματα που φαίνονται παρακάτω.

- **Απλή ημιτονοειδής κυματομορφή**

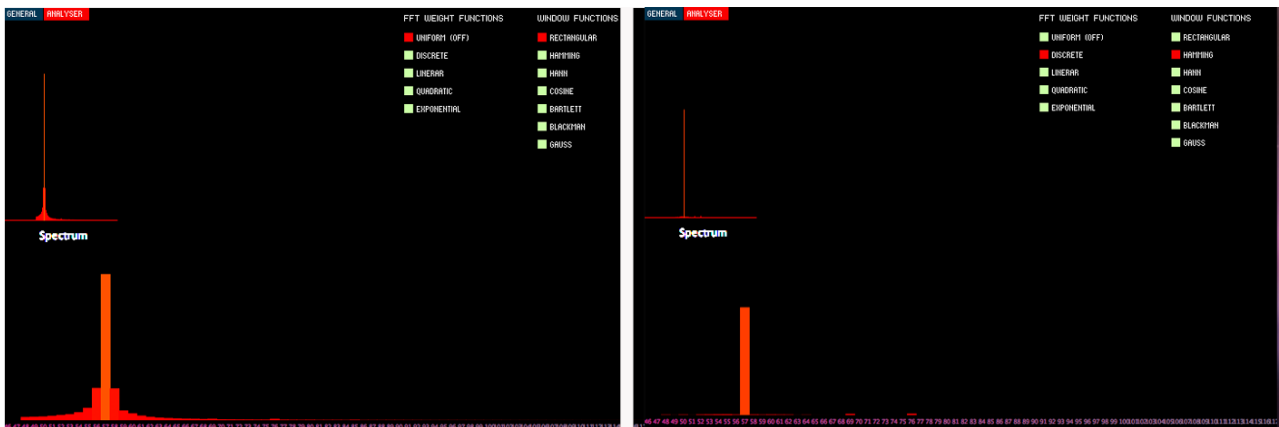
Στο παρακάτω σχήμα φαίνεται η συμπεριφορά του συστήματος για δύο ημιτονοειδές κυματομορφές 440 Hz και 220 Hz αντίστοιχα για την ετικέτα GENERAL. Η διάκριση και απεικόνιση για ένα τέτοιου είδους σήμα είναι επιτυχής.



*Εικόνα 3.1 Απεικόνιση απλής ημιτονοειδής κυματομορφής στην ετικέτα GENERAL για συχνότητες 440 Hz και 220 Hz αντίστοιχα.*

Συνεχίζοντας, για το ίδιο σήμα, το σύστημα δοκιμάζεται για τον τρόπο που πραγματοποιεί την φασματική ανάλυση όπως επίσης και το πώς επιδρούν στο σήμα οι συναρτήσεις της παραθυροποίησης και απόστασης συχνοτικών μπαντών. Παρακάτω, στην αριστερή εικόνα, το σήμα δεν επηρεάζεται από καμία συνάρτηση αντίθετα με την δεξιά στην οποία εφαρμόζεται παραθυροποίηση

HAMMING ενώ η απόσταση μεταξύ των μπαντών εκφράζεται μέσω της συνάρτησης DISCRETE .

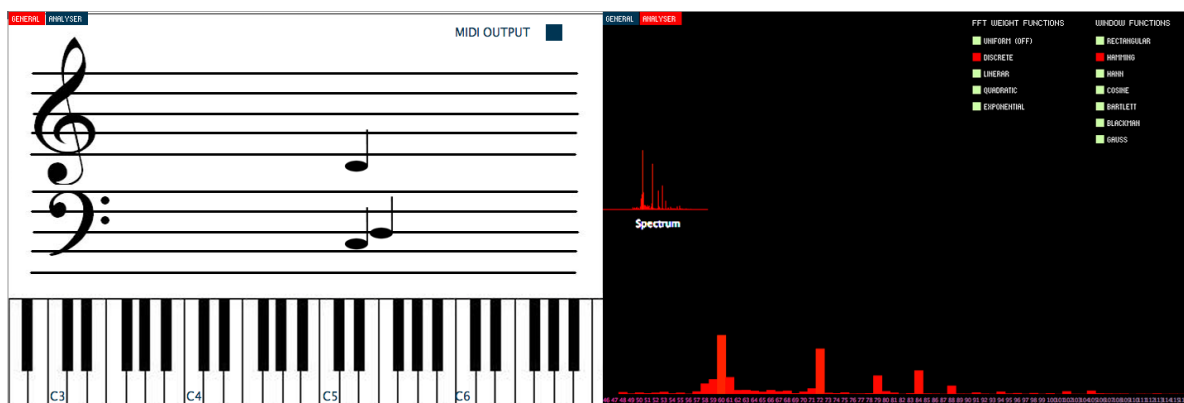


Εικόνα 3.2 Στάδιο της ανάλυσης μιας ημιτονοειδούς κυματομορφής συχνότητας 220 Hz, με διαφορετικούς συνδυασμούς παραμέτρων.

Από τα παραπάνω, φαίνεται πως η εφαρμογή των συναρτήσεων παρουσιάζει κάποια επίδραση στο φάσμα του σήματος εισόδου. Παρόλα αυτά όμως και χωρίς καμία εφαρμογή συνάρτησης(αριστερά) βλέπουμε ότι η συχνότητα εντοπίζεται εύκολα, πράγμα το οποίο οφείλεται στην μη παρουσία αρμονικών.

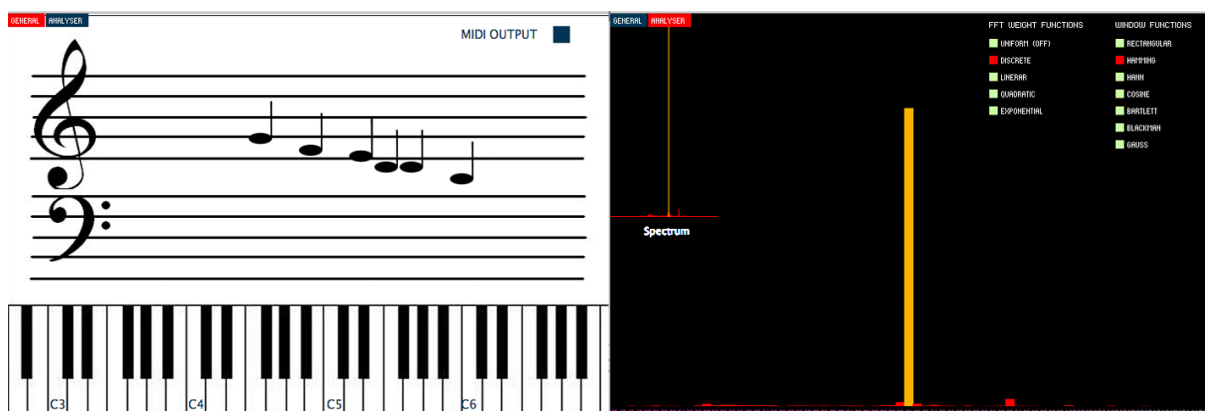
- **Νότα ακουστικού πιάνου**

Το δεύτερο σήμα που εφαρμόστηκε στο σύστημα προήλθε από ένα ψηφιακό συνθετητή (Logic - EFM 1 ) ο οποίος προσομοιώνει ένα ακουστικό πιάνο. Ο λόγος που επιλέχθηκε τέτοιου είδους ήχος είναι το πλούσιο αρμονικό περιεχόμενο του καθώς και ο γρήγορος χρόνος ανόδου της περιβάλλουσας του. Έτσι λοιπόν, αρχικά στο σύστημα εισέρχεται μία νότα ΝΤΟ της τρίτης οκτάβας (C3) ενώ οι συναρτήσεις παραθυροποίησης και απόσταση μπαντών εφαρμόστηκαν ως HAMMING και DISCRETE αντίστοιχα. Όπως όμως φαίνεται στην εικόνα3.3 (αριστερά), εμφανίζεται μαζί με την C3 και η C4. Αυτό, μπορεί εξηγηθεί μελετώντας την ετικέτα Analyser (δεξιά) στην οποία φαίνεται η παρουσία υψηλότερων αρμονικών στο σήμα εισόδου. Να επισημανθεί εδώ, ότι ο λόγος εμφανίζεται μόνο η C4 στην παρτιτούρα ενώ στο ιστόγραμμα βλέπουμε και άλλους αρμονικούς είναι ότι στο σύστημα έχει προστεθεί ένα κατώφλι πλάτους κάτω από το οποίο οι αρμονικοί απορρίπτονται στην έξοδο.



Εικόνα 3.3 Ακουστικό πιάνο νότας C3 όπως απεικονίζεται από το σύστημα. Είναι φανερή η επίδραση των αρμονικών στην απεικόνιση της παρτιτούρας.

Για το ίδιο είδος σήματος τώρα, παίζεται μια ακολουθία από 5 νότες (Ντο-Ρε-Μι-Φα-Σολ) στην τέταρτη οκτάβα. Ο λόγος για αυτό ήταν να δοκιμαστεί το σύστημα σε μεγαλύτερη οκτάβα όπου το αρμονικό περιεχόμενο μεταβάλλεται. Παρακάτω, αριστερά φαίνεται η ακολουθία ενώ στα δεξιά φαίνεται το στιγμιότυπο από τη νότα σολ. Να σημειωθεί εδώ οι συναρτήσεις παραθυροποίησης και απόστασης μεταξύ των συχνοτικών μπαντών διατηρήθηκαν ίδιες με την παραπάνω περίπτωση για λόγους σύγκρισης.

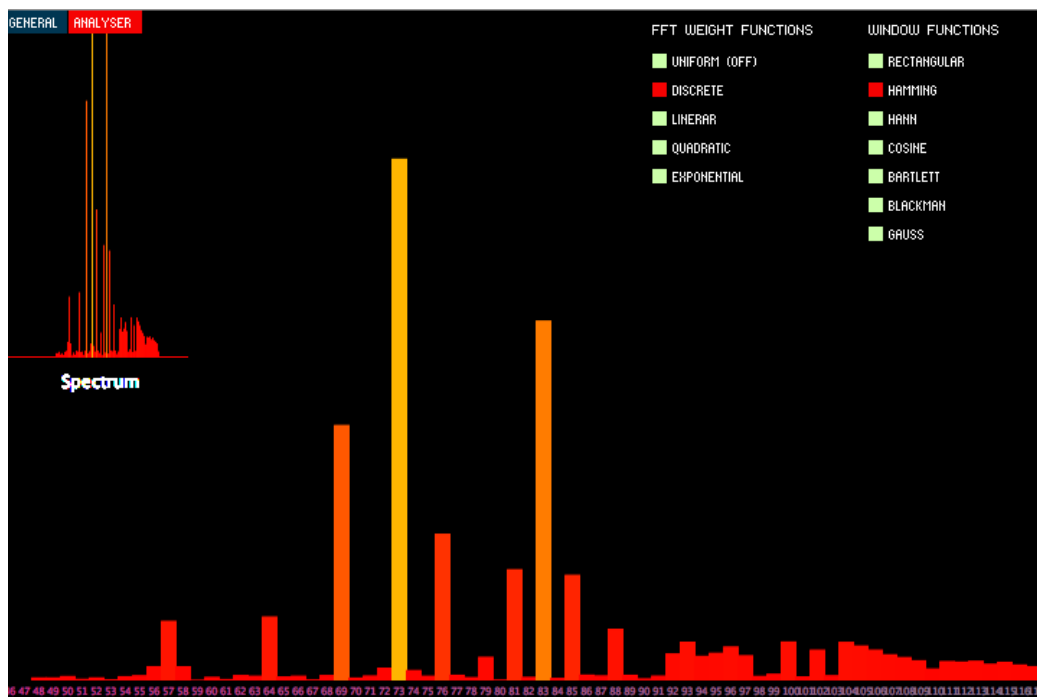


Εικόνα 3.4 Ακολουθία νοτών παιγμένη από ακουστικό πιάνο (αριστερά) και το στιγμιότυπο μίας από τις νότες αυτές(δεξιά).

Στη παραπάνω εικόνα είναι εμφανές ότι ο εντοπισμός της συχνότητας βρίσκεται σε πολύ καλύτερα επίπεδα από ότι στην χαμηλότερη οκτάβα της παραπάνω περίπτωσης. Αυτό όπως φαίνεται οφείλεται στην απουσία αρμονικών του οργάνου στις υψηλότερες νότες.

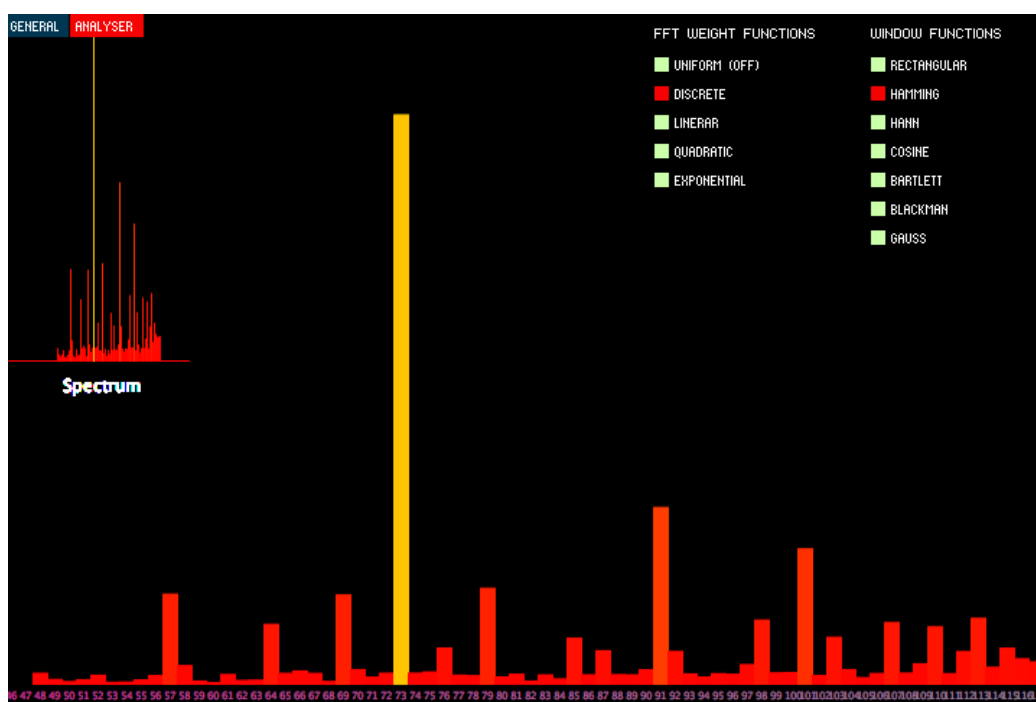
- **Ανθρώπινη φωνή**

Τρίτο και τελευταίο σήμα που δοκιμάζεται το σύστημα είναι η ανθρώπινη φωνή. Για ένα τέτοιου είδους σήμα, ο εντοπισμός της θεμελιώδους συχνότητας είναι αρκετά πιο πολύπλοκος από ότι στις παραπάνω δύο περιπτώσεις. Κάποιοι από τους σημαντικότερους λόγους για αυτό είναι το πλούσιο αρμονικό περιεχόμενο το οποίο μεταβάλλεται δραματικά ανάλογα με τις αλλαγές στην στοματική κοιλότητα και είναι χαρακτηριστικό για κάθε άνθρωπο όπως επίσης και οι απότομες μεταβάσεις της περιόδου στο πεδίο του χρόνου, πχ ανάμεσα από ένα φωνήεν και ένα ρινικό σύμφωνο[13]. Στο παρακάτω πείραμα, χρησιμοποιήθηκε μια ανδρική φωνή η οποία αρχικά εισάγει στο σύστημα διάφορα φωνήεντα ίδια συχνότητας ενώ στη συνέχεια πραγματοποιείται το αντίθετο, δηλαδή ίδιο φωνήεν διαφορετικής συχνότητας.



Εικόνα 3.5 Φασματική ανάλυση του φωνήεντος «ο» για τη νότα A2

Όπως φαίνεται στο παραπάνω σχήμα το αρμονικό περιεχόμενο του σήματος είναι αρκετά πιο πλούσιο από των δύο προηγούμενων περιπτώσεων. Και εδώ, οι επιλογές στις συναρτήσεις παραθυροποίησης και απόστασης των συχνοτικών μπαντών παρέμεινε ίδια με παραπάνω για λόγους σύγκρισης. Το σημαντικότερο όμως χαρακτηριστικό του σχήματος είναι ο τρόπος με τον οποίο διατάσσονται οι αρμονικοί. Αν και η συχνότητα της φωνής είναι A2 (η οποία αντιστοιχεί στον αριθμό 69 του άξονα x), οι C#3 και C#4 έχουν μεγαλύτερο πλάτος το οποίο βέβαια δημιουργεί προβλήματα στον εντοπισμό της θεμελιώδους. Στο παρακάτω σχήμα βλέπουμε το ίδιο πείραμα αλλά για το φωνήεν «Ε». Εδώ, ο εντοπισμός γίνεται ακόμα πιο δύσκολος καθώς όπως φαίνεται η αρμονική C#3 έχει πολύ μεγαλύτερη ενέργεια από ότι η θεμελιώδης.



*Εικόνα 3.6 Φασματική ανάλυση του φωνήεντος «ε» για τη νότα A2. Μπορεί εύκολα να διακριθεί η διαφορά ενέργειας ανάμεσα στον θεμελιώδη και τον πρώτο αρμονικό.*

Συνεχίζοντας, η απεικόνιση των μουσικών φθόγγων από τα παραπάνω σήματα πάνω σε μορφή παρτιτούρας θα επηρεάζεται αρνητικά από αυτήν την προσθήκη των αρμονικών καθώς και άλλες νότες θα εμφανίζονται μαζί με την θεμελιώδη. Σε αυτό το σημείο μπορεί να εξηγηθεί και ο λόγος ο οποίος το κατώφλι που αναφέρθηκε παραπάνω (βλέπε σελ. 42) δεν μπορεί να πάρει μεγάλες τιμές καθώς τα πλάτη των αρμονικών ποικίλουν πάνω και κάτω από



αυτό της θεμελιώδους. Στα δύο επόμενα σχήματα φαίνονται δύο νότες διαφορετικής οκτάβας για το φωνήεν Α.

The image shows a software interface for musical analysis. At the top left, there are two tabs: 'GENERAL' (highlighted in red) and 'ANALYSER'. At the top right, there is a 'MIDI OUTPUT' label next to a blue square. The main part of the image consists of two staves of musical notation. The upper staff is in treble clef and contains a single note on the second line (G4). The lower staff is in bass clef and contains a sequence of notes: C3, D3, E3, F3, G3, A3, B3, C4, D4, E4, F4, G4, A4, B4, C5, D5, E5, F5, G5, A5, B5, C6. Below the staves is a piano keyboard diagram with keys labeled C3, C4, C5, and C6. Three keys are highlighted in grey: A2 (the A below C3), E3 (the E below C3), and A3 (the A below C4).

Εικόνα 3.7 Απεικόνιση νότας A2 του φωνήεντος α.

The image shows a software interface for musical analysis, similar to the previous one. At the top left, there are two tabs: 'GENERAL' (highlighted in red) and 'ANALYSER'. At the top right, there is a 'MIDI OUTPUT' label next to a blue square. The main part of the image consists of two staves of musical notation. The upper staff is in treble clef and contains a sequence of notes: C4, D4, E4, F4, G4, A4, B4, C5, D5, E5, F5, G5, A5, B5, C6. The lower staff is in bass clef and contains a sequence of notes: C3, D3, E3, F3, G3, A3, B3, C4, D4, E4, F4, G4, A4, B4, C5, D5, E5, F5, G5, A5, B5, C6. Below the staves is a piano keyboard diagram with keys labeled C3, C4, C5, and C6. Two keys are highlighted in grey: A3 (the A below C4) and E4 (the E above C4).

Εικόνα 3.8 Απεικόνιση νότας A3 του φωνήεντος α.

Τέλος, να σημειωθεί εδώ ότι, για την περίπτωση της εισόδου της φωνής, τα αποτελέσματα δεν έχουν σαν σκοπό να γενικεύσουν την περίπτωση, καθώς για διαφορετικές χροιές το αρμονικό περιεχόμενο μπορεί αλλάξει δραματικά επηρεάζοντας έτσι και τον εντοπισμό της θεμελιώδους. Στις εικόνες 3.7 και 3.8 φαίνεται ότι και για τη φωνή η συχνότητα επηρεάζει την διάκριση των συχνοτήτων καθώς στη χαμηλότερη νότα παρουσιάζονται περισσότεροι αρμονικοί από ότι στη A3.

## 4 Συμπεράσματα

Στην εργασία αυτή μελετήθηκαν όλα τα βασικά στάδια που αφορούν την υλοποίηση ενός απλού συστήματος εντοπισμού της θεμελιώδους συχνότητας και απεικόνισης σε εικονική παρτιτούρα. Το σύστημα δοκιμάστηκε για διάφορα είδη σημάτων εισόδου με σκοπό την αξιολόγηση της εξόδου του. Όσον αφορά τον εντοπισμό της βασικής συχνότητας για μια απλή κυματομορφή και ένα ακουστικό πιάνο το σύστημα λειτούργησε αποδοτικά με αποτέλεσμα οι μουσικοί φθόγγοι να απεικονίζονται σωστά στην παρτιτούρα. Σφάλματα παρουσιάζονται για το τρίτο σήμα που είναι η φωνή. Όπως περιγράφηκε, ένα τέτοιο σήμα έχει πλούσιο αρμονικό περιεχόμενο όπου οι αρμονικοί μπορεί να έχουν μικρότερη ή και μεγαλύτερη ενέργεια από την θεμελιώδη. Αυτό έχει σαν αποτέλεσμα κάποιοι αρμονικοί να θεωρούνται ως θεμελιώδης και να εισάγονται στην παρτιτούρα. Όσον αφορά το διάστημα κάθε νότας δεν υπολογίζεται προς το παρόν από το σύστημα άρα όλες οι νότες έχουν ένα προκαθορισμένο σχήμα νότας ενός τετάρτου.

Όσον αφορά τις σημαντικότερες βελτιώσεις που πρέπει να εφαρμοστούν στο σύστημα ώστε να βοηθήσουν τον λειτουργικό του χαρακτήρα, αυτές είναι η καλύτερη διάκριση μεταξύ αρμονικών και μεγίστων του σήματος εισόδου, ο υπολογισμός της διάρκειας κάθε νότας και η ορθή απεικόνιση στο πεντάγραμμο. Το τελευταίο προϋποθέτει και την εισαγωγή της έννοιας του μέτρου όπου οι νότες ανάλογα με την διάρκεια τους θα έχουν την ανάλογη σχηματική αξία (ολόκληρο, μισό, κτλ) αλλά και την ανάλογη απόσταση μεταξύ τους.

Περαιτέρω έρευνα στην εργασία θα μπορούσε να γίνει γύρω από τον αριθμό και είδος των σημάτων εισόδου. Περισσότερα σήματα με διαφορετική χροιά θα βοηθήσουν στην αντιμετώπιση των σφαλμάτων. Επίσης, θα μπορούσε να μελετηθεί και ο συνδυασμός τεχνικών εντοπισμού θεμελιώδους συχνότητας, από αυτές που περιγράφηκαν στην θεωρία της εργασίας, με σκοπό την βέλτιστη αναγνώριση της ταυτότητας του φάσματος και την ορθότερη απεικόνιση των μουσικών φθόγγων στην εικονική παρτιτούρα.

## **ΒΙΒΛΙΟΓΡΑΦΙΑ**

BIBΛΙΑ

[1] Dodge, C., Jerse, Th. A., [1997]. **Computer Music – *Synthesis, Composition and Performance***. New York, Schirmer Books

[2] Hartmann, W. M., [2000]. **Signals, Sound, and Sensation**. USA, Edward Brothers, Inc., Ann Arbor, MI.

[3] Ibanez, A. P., [2010]. **Computationally efficient methods for polyphonic music transcription**. PhD Thesis, University of Alicante .

[4] Kuo, S. M, and others, [2001]. **Real Time Digital Processing – Implementations, Applications and Experiments with the TMS320C55x**. West Essex, John Wiley & Sons LTD.

[5] Khan, A. A., [2005]. **Digital Signal Processing Fundamentals**, Massachusetts, Da Vinci Engineering Press.

[6] Lynn, A. P., [1989]. **An introduction to the Analysis and Processing of signals**. 3<sup>rd</sup> ed. London, Macmillan Press LTD.

[7] Lynn, A. P., [1994]. **Introductory Digital Signal Processing with Computer Applications**. 2<sup>nd</sup> Edition, London, John Wiley & Sons Ltd.

[8] Ludeman, L. C., [1976]. **Fundamental of Digital Signal Processing**. Canada, John Wiley & Sons, Inc.

[9] Roads, C., [2001]. **Microsound**, USA, MIT Press

[10] Smith III, J., [2010]. **SPECTRAL AUDIO SIGNAL PROCESSING**, USA, Center for Computer Research in Music and Acoustics

[11] Shiffman, D., [2008]. **Learning Processing - A Beginner's Guide to Programming Images, Animation, and Interaction**, USA, Morgan Kaufmann Publishers

[12] Greenberg, I.,[2007]. **Processing: Creative Coding and Computational Art**, USA, Apress

[13] Gold, B., Morgan, N.,[2000]. **Speech and Audio Processing**, USA, John Wiley & Sons, Inc

APOPA

[14] David Gerhard, [2003]. **Pitch Extraction and Fundamental Frequency: History and Current Techniques**, Technical Report: TR-CS 2003-06.

[15] Alain de Cheveigne' Hideki Kawahara,[2002]. **YIN, A fundamental Frequency Estimator for Speech and Music**, The Journal of The Acoustical Society of America, Vol. 111, No. 4, April 2002.

[16] Cabral, G., Briot, J.P., Pachet F., [2008]. **Impact of Distance in Pitch Class Profile Computation**, Scientific Literature Digital Library and Search Engine [Internet][<http://citeseerx.ist.psu.edu/oai2>]

[17] Harris, F.J.,[ 1978]. **On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform**. *Proceedings of the IEEE*. Vol. 66, No. 1 ,pp 179-185, January 1978.