



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής  
Πρόγραμμα Μεταπτυχιακών Σπουδών  
«Προηγμένα Συστήματα Πληροφορικής»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	<b>Σχεδίαση Ενσωματωμένου Συστήματος σε FPGA για την Ανίχνευση και Διόρθωση των Θεμελιωδών Συχνοτήτων Μουσικών Σημάτων</b>
Όνοματεπώνυμο Φοιτητή	<b>Κωστόπουλος Φώτιος του Χρήστου</b>
Αριθμός Μητρώου	<b>ΜΠΣΠ/07032</b>
Κατεύθυνση	<b>Τεχνολογία Ενσωματωμένων Υπολογιστικών Συστημάτων</b>
Επιβλέπων	<b>Μιχάλης Ψαράκης, Επίκουρος Καθηγητής</b>

Πανεπιστήμιο Πειραιώς-Τμήμα Πληροφορικής  
Πρόγραμμα Μεταπτυχιακών Σπουδών στα  
Προηγμένα Συστήματα Πληροφορικής

Ημερομηνία Παράδοσης **5 Μαΐου 2011**



**Τριμελής Εξεταστική Επιτροπή**

(υπογραφή)

(υπογραφή)

(υπογραφή)

Μιχάλης Ψαράκης  
Επίκουρος Καθηγητής

Άγγελος Πικράκης  
Λέκτορας

Δημήτριος Γκιζόπουλος  
Αναπληρωτής  
Καθηγητής

## Περίληψη

Στόχος της συγκεκριμένης μεταπτυχιακής διατριβής είναι η ανάπτυξη ενός ενσωματωμένου συστήματος για την ανίχνευση και διόρθωση της θεμελιώδους συχνότητας μουσικών σημάτων. Οι ιδιαιτερότητες που παρουσιάζει η συγκεκριμένη εφαρμογή δίνουν την ευκαιρία για μια ενδελεχή μελέτη και βαθύτερη κατανόηση της έννοιας της θεμελιώδους συχνότητας ενός ηχητικού σήματος, του τρόπου που ο άνθρωπος αντιλαμβάνεται την έννοια αυτή και των αλγορίθμων και μεθόδων που έχουν αναπτυχθεί για την αναγνώρισή και χειραγώγηση της από ένα υπολογιστικό σύστημα.

Ταυτόχρονα οι απαιτήσεις της ανάπτυξης ενός ενσωματωμένου συστήματος για την εφαρμογή είναι τέτοιες που οδηγούν στην χρήση νέων εργαλείων σχεδίασης και αποσφαλμάτωσης αυτού, στην βαθύτερη κατανόηση των μεθόδων ταυτόχρονης ανάπτυξης υλικού / λογισμικού και συντελούν στην δημιουργία βαθύτερης αντίληψης για τη λεπτομερή λειτουργία και την γενικότερη σχεδίαση των ενσωματωμένων συστημάτων.

## Abstract

This postgraduate thesis' objective is the design and development of an embedded system that detects and corrects the pitch of musical signals. The peculiar nature of the application presents a unique chance to study and deeply understand the concept of pitch of an audio signal, the way that people interpret it and the different algorithms and methods that have been proposed over the years for its recognition and manipulation by a computer system.

Furthermore, the challenge of development of such an embedded system, leads to the use of new and more advanced design and debug tools, to the deeper comprehension of methods for hardware / software co-development and create a unique perception of the operation details and the design secrets of embedded systems in general.

## Πίνακας Περιεχομένων

Κεφάλαιο 1: Συστήματα επεξεργασίας ψηφιακού ήχου και αλγόριθμοι Αναγνώρισης / Διόρθωσης Θεμελιώδους Συχνότητας.....	7
1.1: Ψηφιακή αναπαράσταση ήχου.....	8
1.1.1: Ρυθμός δειγματοληψίας.....	9
1.1.2 : Ανάλυση.....	11
1.1.3: Χαρακτηριστικά αναπαράστασης.....	12
1.1.4: Αριθμός Καναλιών.....	12
1.1.5 : Συμπύεση.....	13
1.2: Θεμελιώδης συχνότητα και Pitch.....	13
1.2.1: Ορισμός του Pitch.....	13
1.2.2: Αλγόριθμοι ανίχνευσης του Pitch.....	15
1.2.3: Αλγόριθμοι διόρθωσης του Pitch.....	28
1.2.4: Εφαρμογές της ανίχνευσης και διόρθωσης του Pitch.....	34
Κεφάλαιο 2: Περιγραφή υλικού.....	35
2.1: Αναπτυξιακή Πλακέτα Xilinx ML403.....	36
2.2: AC' 97 Audio Codec.....	37
2.2.1: Χαρακτηριστικά και δομή του LM4550 Audio Codec.....	37
2.2.2: Σειριακό πρωτόκολλο διασύνδεσης AC-link.....	42
2.2.3: Ελεγκτής OPB_AC97_controller.....	46
2.3: Ενσωματωμένος Επεξεργαστής MicroBlaze.....	47
Κεφάλαιο 3: Ανάπτυξη εφαρμογής Ανίχνευσης και Διόρθωσης Θεμελιώδους συχνότητας Μουσικών Σημάτων.....	49
3.1: Χαρακτηριστικά του ήχου της ηλεκτρικής κιθάρας.....	50
3.2: Δομή μουσικών φράσεων και διάταξη νοτών σε κλίμακες.....	53
3.3: Ηχογράφηση Σήματος Εισόδου.....	54
3.4: Αλγόριθμος Ανίχνευσης Θεμελιώδους Συχνότητας.....	56
3.5: Ανίχνευση Νοτών Μουσικής Φράσης.....	61
3.6: Αλγόριθμος διόρθωσης της θεμελιώδους συχνότητας και αναπαραγωγή ηχητικού σήματος.....	62
Κεφάλαιο 4: Σχεδιαστική Ροή, Εργαλεία Υλοποίησης, Συμπεράσματα.....	64
4.1: Ολοκληρωμένο περιβάλλον σχεδίασης EDK.....	64
4.2: Συμπεράσματα – Βελτιώσεις.....	67
Βιβλιογραφία.....	68
Παράρτημα Α: Κώδικας Εφαρμογής.....	69

## Κατάλογος Εικόνων

- Εικόνα 1: Βασική δομή συστήματος ψηφιακής επεξεργασίας σημάτων  
Εικόνα 2: Μετατροπή αναλογικού σήματος σε PCM δεδομένα  
Εικόνα 3: Ψηφιοποίηση ενός σήματος σε διαφορετικούς ρυθμούς δειγματοληψίας  
Εικόνα 4: Αναδίπλωση Συχνοτήτων  
Εικόνα 5: Σφάλμα Κβάντισης  
Εικόνα 6: Clipping Distortion  
Εικόνα 7: Θεμελιώδης συχνότητα και αρμονικές σε παλλόμενη χορδή  
Εικόνα 8: Περιοδική Κυματομορφή Χαμηλής Πολυπλοκότητας  
Εικόνα 9: Περιοδική Κυματομορφή Υψηλής Πολυπλοκότητας  
Εικόνα 10: Κυματομορφή και ACF  
Εικόνα 11: Κυματομορφή και AMDF  
Εικόνα 12: Σύγκριση Time – Domain Pitch Detection Αλγορίθμων ως προς την απόδοση  
Εικόνα 13: Αλγόριθμος Time – Domain Maximum Likelihood  
Εικόνα 14: Αλγόριθμος Adaptive Filtering Με Χρήση Bandpass φίλτρου  
Εικόνα 15: Αλγόριθμος Adaptive Filtering Με Χρήση Comb φίλτρου  
Εικόνα 16: Αλγόριθμος Harmonic Product Spectrum  
Εικόνα 17: Ανάλυση Cepstrum  
Εικόνα 18: Frequency Domain Maximum Likelihood  
Εικόνα 19: Διάκριση Pitch / Frequency Shifting  
Εικόνα 20: Time Expansion  
Εικόνα 21: Time Compression  
Εικόνα 22: Αλγόριθμος Synchronous OverLap Add  
Εικόνα 23: Αλγόριθμος Phase Vocoder  
Εικόνα 24: Δομή του υλικού του ενσωματωμένου συστήματος  
Εικόνα 25: Δομή της πλακέτας ML403  
Εικόνα 26: Δομή του ολοκληρωμένου LM4550  
Εικόνα 27: Δομή των πλαισίων εξόδου / εισόδου  
Εικόνα 28: Χρονισμός του πλαισίου εξόδου  
Εικόνα 29: Χρονισμός του πλαισίου εισόδου  
Εικόνα 30: Τυπική διασύνδεση του OPB\_AC97\_controller σε ενσωματωμένο σύστημα  
Εικόνα 31: Δομή του MicroBlaze  
Εικόνα 32: Συχνοτική έκταση της ηλεκτρικής κιθάρας  
Εικόνα 33: Συχνοτικό φάσμα πρόσφατα διεγερμένης χορδής ηλεκτρικής κιθάρας  
Εικόνα 34: Κυματομορφή πρόσφατα διεγερμένης χορδής ηλεκτρικής κιθάρας  
Εικόνα 35: Συχνοτικό φάσμα εξασθενημένης χορδής ηλεκτρικής κιθάρας  
Εικόνα 36: Κυματομορφή εξασθενημένης χορδής ηλεκτρικής κιθάρας  
Εικόνα 37: Κυματομορφή ατάκας χορδής ηλεκτρικής κιθάρας  
Εικόνα 38: Αναλογικό μονοπάτι ηχογράφησης αναπαραγωγής στο LM4550  
Εικόνα 39: Πρώτο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο  
Εικόνα 40: Δεύτερο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο  
Εικόνα 41: Τρίτο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο  
Εικόνα 42: Τέταρτο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο  
Εικόνα 43: Πέμπτο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο

Εικόνα 44: Έκτο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο

Εικόνα 45: Έβδομο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο

Εικόνα 46: Όγδοο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο

Εικόνα 47: Τελικό στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο

Εικόνα 48: Τυπική μορφή AMDF συνάρτησης σε ένα πλαίσιο

Εικόνα 49: Δομή του Embedded Development Kit

Εικόνα 50: Στιγμιότυπο Χρήσης του XPS

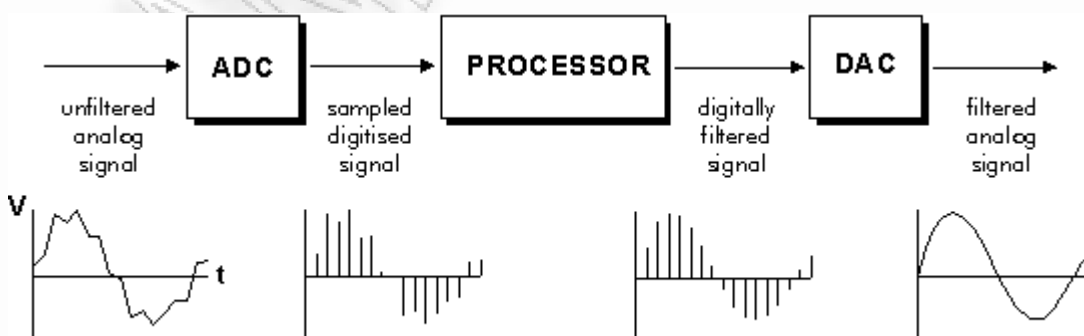
Εικόνα 50: Στιγμιότυπο Χρήσης του XPS

## Κεφάλαιο 1: Συστήματα Επεξεργασίας Ψηφιακού Ήχου και Αλγόριθμοι Αναγνώρισης/Διόρθωσης Θεμελιώδους Συχνότητας

Η ραγδαία ανάπτυξη της τεχνολογίας των ψηφιακών συστημάτων, η συνεχής βελτίωση των επεξεργαστών και η λειτουργία τους σε όλο και μεγαλύτερες συχνότητες έχουν καταστήσει εφικτή τα τελευταία χρόνια την ανάπτυξη και τη γενικευμένη χρήση συστημάτων ψηφιακής επεξεργασίας σήματος σε πλήθος εφαρμογών (τηλεπικοινωνίες, επεξεργασία ομιλίας και ήχου, επεξεργασία εικόνας και βίντεο). Αυτό έχει οδηγήσει σε συστήματα πιο αξιόπιστα από τα αντίστοιχα αναλογικά, η απόδοση των οποίων εξαρτάται από αστάθμητους παράγοντες (θερμοκρασία λειτουργίας, γήρανση των εξαρτημάτων, στατιστικές διακυμάνσεις στις τιμές των ηλεκτρικών στοιχείων), το οποίο οδηγεί σε ασταθή πολλές φορές λειτουργία. Επιπλέον η χρήση μιας ψηφιακής πλατφόρμας για την υλοποίηση τέτοιων συστημάτων προσφέρει ευελιξία στην εφαρμογή και μεγαλύτερη φιλικότητα προς το χρήστη.

Για την υλοποίηση συστημάτων ψηφιακής επεξεργασίας σήματος χρησιμοποιούνται μικροεπεξεργαστές, μικροελεγκτές, επεξεργαστές ειδικού σκοπού (DSP's) καθώς και custom υλικό σε σπανιότερες περιπτώσεις. Η ανάπτυξη της τεχνολογίας των FPGA's όμως τα τελευταία χρόνια τα καθιστά ιδανική πλατφόρμα για την υλοποίηση τέτοιων συστημάτων. Τα FPGA's πλέον είναι κυκλώματα με αρκετά μεγάλο βαθμό ολοκλήρωσης και περιλαμβάνουν εκτός από custom υλικό και επεξεργαστές, block μνήμης, hardware ειδικού σκοπού καθώς και πλήθος επιλογών διασύνδεσης με άλλα μέρη ενός ολοκληρωμένου συστήματος. Έτσι προωθούν την έννοια της προγραμματισιμότητας, από το λογισμικό (κάτι που ισχύει στους επεξεργαστές) στο υλικό. Με τον τρόπο αυτόν ο σχεδιαστής μπορεί να ξεφύγει από τους περιορισμούς του σταθερού υλικού και να αναπτύξει το σύστημα ακριβώς σύμφωνα με τις ανάγκες του. Επιπλέον η εύκολη και οικονομική σχεδίαση και υλοποίηση υλικού δίνει την επιλογή στον σχεδιαστή να ξεφύγει από το σειριακό (ή περιορισμένα παράλληλο) μοντέλο εκτέλεσης κώδικα ενός επεξεργαστή και να εκμεταλλευτεί την παραλληλία που συναντάται σε πολλούς αλγόριθμους επεξεργασίας σημάτων.

Για την υλοποίηση ενός συστήματος ψηφιακής επεξεργασίας σήματος απαιτείται η μετατροπή του σήματος από την αναλογική στην ψηφιακή του μορφή κατά την είσοδο στο σύστημα (με χρήση ενός μετατροπέα αναλογικού-σε-ψηφιακό - ADC) και η αντίστροφη διαδικασία κατά την έξοδο (με χρήση ενός μετατροπέα ψηφιακού-σε-αναλογικό - DAC). Το αναλογικό σήμα δίνεται ως είσοδος στον μετατροπέα αναλογικού-σε-ψηφιακό ως συνεχής μεταβολή μιας τάσης. Ο μετατροπέας δειγματοληπτεί το αναλογικό σήμα σε προκαθορισμένα χρονικά σημεία και εν συνεχεία κβαντίζει την τιμή της τάσης του σε μια ακολουθία αριθμών. Με τον τρόπο αυτόν επιτυγχάνεται μετατροπή από τον συνεχή χρόνο ( $x(t)$ ) στον διακριτό ( $x(n)$ ). Κατά την έξοδο από το σύστημα ακολουθεί η αντίστροφη διαδικασία κατά την οποία μέσω ενός φίλτρου ανακατασκευής το σήμα μετατρέπεται σε αναλογικό. Η βασική δομή ενός τέτοιου συστήματος φαίνεται στην εικόνα 1



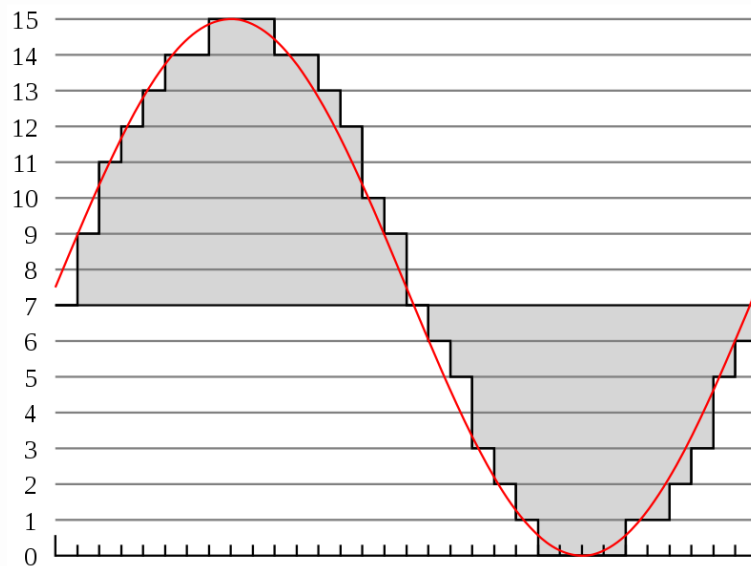
**Εικόνα 1: Βασική δομή συστήματος ψηφιακής επεξεργασίας σημάτων**

Με την μεταφορά των μεθόδων επεξεργασίας σημάτων στον ψηφιακό κόσμο δίνεται η δυνατότητα επεξεργασίας και χειρισμού των σημάτων σε βαθμούς που θα ήταν αδύνατοι στον αναλογικό κόσμο. Είναι εφικτός ο σχεδιασμός πολύπλοκων αλγορίθμων με δραματικά μεγαλύτερη ακρίβεια επεξεργασίας και με ευκολότερο και πιο αποδοτικό έλεγχο. Διευκολύνεται επίσης η δυνατότητα αποθήκευσης των σημάτων δίχως υποβάθμιση της ποιότητας τους λόγω αστοχίας υλικών. Η χρήση μεθόδων ψηφιακής επεξεργασίας σημάτων όμως παρουσιάζει ορισμένα προβλήματα που χαρακτηρίζουν τον ψηφιακό κόσμο. Πρώτον εισάγεται η έννοια του latency. Ως latency ορίζεται η χρονική καθυστέρηση που παρατηρείται μεταξύ εισόδου – εξόδου του συστήματος. Σε ένα αναλογικό σύστημα η καθυστέρηση αυτή είναι ίση με την καθυστέρηση διάδοσης του ηλεκτρικού ρεύματος μέσα από τα ηλεκτρικά στοιχεία του κυκλώματος και είναι πρακτικά αμελητέα. Σε ένα ψηφιακό σύστημα όμως ο χρόνος επεξεργασίας των δεδομένων είναι υπαρκτός και έχει να κάνει πρωτίστως με την ταχύτητα του χρησιμοποιούμενου επεξεργαστή. Αν σε αυτό συνυπολογίσουμε και το γεγονός ότι συχνά οι αλγόριθμοι ψηφιακής επεξεργασίας σημάτων απαιτούν ένα σύνολο δειγμάτων για να εκτελεστούν, συμπεραίνουμε ότι στο υπάρχον latency προστίθεται και ο χρόνος αποθήκευσης του συνόλου των δεδομένων σε ένα buffer εισόδου. Σημαντικό μειονέκτημα της χρήσης ψηφιακών συστημάτων επεξεργασίας σημάτων συνιστά και η περιορισμένη ακρίβεια αναπαράστασης των αναλογικών σημάτων στο ψηφιακό πεδίο. Σε ένα αναλογικό σήμα παρατηρείται συνεχής μεταβολή μιας φυσικής ποσότητας στον χρόνο. Για την απόλυτα ακριβή αναπαράσταση της ποσότητας αυτής σε ένα ψηφιακό σήμα και σύμφωνα με την φύση της αρχιτεκτονικής των σύγχρονων επεξεργαστών θα απαιτούνταν άπειρη μνήμη και άπειρη επεξεργαστική ισχύς, κάτι το οποίο προφανώς είναι αδύνατον. Ευτυχώς όσο αναπτύσσεται η τεχνολογία και η υπολογιστική ισχύς των νέων συστημάτων βρίσκονται τρόποι και μέθοδοι αναπαράστασης των αναλογικών σημάτων σε ψηφιακά με αρκετά καλά αποτελέσματα πιστότητας. Τέλος ένα ακόμα θεμελιώδες πρόβλημα των ψηφιακών συστημάτων σε σύγκριση με τα αναλογικά είναι τα λάθη στην ανακατασκευή του σήματος εξόδου σε περιπτώσεις υλοποίησης συστημάτων πραγματικού χρόνου, καθώς μια ακριβής ανακατασκευή στο αναλογικό πεδίο θα απαιτούσε σύμφωνα με τα μαθηματικά μοντέλα άπειρο χρόνο επεξεργασίας.

## 1.1 Ψηφιακή Αναπαράσταση Ήχου

Η μετατροπή του ήχου από τον αναλογικό στον ψηφιακό κόσμο γίνεται με ποικίλους τρόπους ανάλογα με την εφαρμογή στην οποία θα χρησιμοποιηθούν τα δεδομένα. Ο ήχος γίνεται αντιληπτός στον άνθρωπο ως μεταβολή στην πίεση του αέρα. Για την είσοδό του σε ένα σύστημα ένας αισθητήρας (μικρόφωνο, πιεζοηλεκτρικά) μετατρέπει τις μεταβολές στην πίεση σε αντίστοιχες μεταβολές στην τάση ενός αναλογικού σήματος. Για την μετατροπή του σε ψηφιακά δεδομένα ο μετατροπέας αναλογικού-σε-ψηφιακό παράγει μια σειρά αριθμών ανάλογα με την μεταβαλλόμενη τάση του αναλογικού σήματος εισόδου. Η διαδικασία επιλογής μιας συγκεκριμένης στιγμής στον χρόνο για την μετατροπή της τάσης που αντιστοιχεί στο αναλογικό σήμα τη συγκεκριμένη στιγμή ονομάζεται δειγματοληψία. Η επιλογή του αριθμού που θα αναπαραστήσει την συγκεκριμένη τάση ονομάζεται κβαντοποίηση του σήματος. Η μορφή αυτή αναπαράστασης του ήχου ονομάζεται παλμοκωδική διαμόρφωση (pulse code modulation - PCM). Η παλμοκωδική διαμόρφωση εφευρέθηκε από τον Alec Reeves το 1937 και αποτελεί την πρότυπη πλέον μορφή αποθήκευσης ψηφιακών ηχητικών σημάτων στα σύγχρονα συστήματα. Η μετατροπή ενός αναλογικού σήματος σε PCM δεδομένα απεικονίζεται στην εικόνα 2.

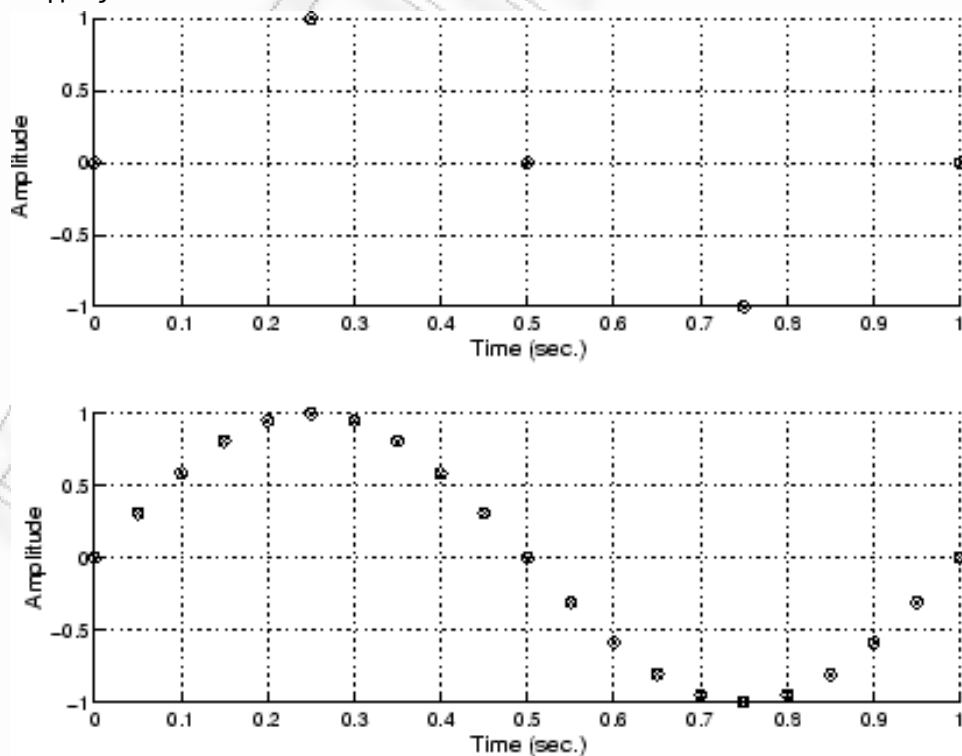




Εικόνα 2: Μετατροπή αναλογικού σήματος σε PCM δεδομένα

### 1.1.1 Ρυθμός Δειγματοληψίας

Ο ρυθμός δειγματοληψίας είναι ο αριθμός των μετατροπών ανά δευτερόλεπτο του αναλογικού σήματος σε ψηφιακό. Όσο μεγαλύτερος είναι ο ρυθμός δειγματοληψίας τόσο πιο πιστή είναι η αναπαράσταση του σήματος στο ψηφιακό σύστημα καθώς και τόσο πιο ακριβής η ανακατασκευή του αναλογικού σήματος στην έξοδο του συστήματος. Ταυτόχρονα όμως, αυξάνεται ο όγκος των προς επεξεργασία δεδομένων καθώς και ο υπολογιστικός φόρτος του συστήματος. Στην εικόνα 3 φαίνεται η ψηφιοποίηση ενός σήματος σε διαφορετικούς ρυθμούς δειγματοληψίας.

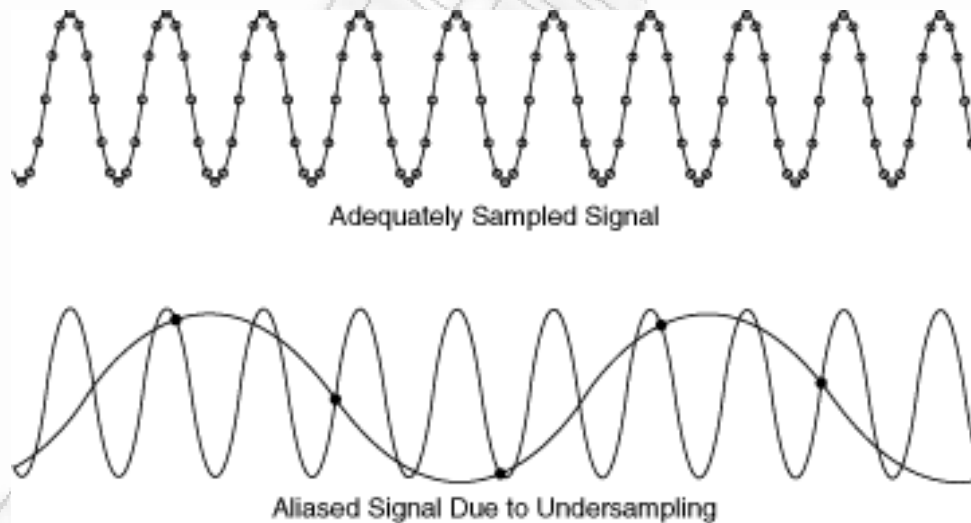


Εικόνα 3: Ψηφιοποίηση ενός σήματος σε διαφορετικούς ρυθμούς δειγματοληψίας

Η επιλογή του ρυθμού δειγματοληψίας για μια εφαρμογή καθορίζεται πρωτίστως από το θεώρημα Nyquist – Shannon. Το θεώρημα ορίζει ότι αν η μέγιστη συχνότητα που περιέχει το αρχικό σήμα είναι  $f$ , τότε αν μετατραπεί σε ψηφιακό με ρυθμό δειγματοληψίας μεγαλύτερο του  $2f$  μπορεί να ανακατασκευαστεί τέλεια από αυτό. Έτσι για παράδειγμα εφόσον τα όρια της ανθρώπινης ακοής είναι 20Hz – 20 kHz αρκεί θεωρητικά ρυθμός δειγματοληψίας 40kHz για τέλεια αναπαραγωγή του ψηφιακού ηχητικού σήματος.

Το θεώρημα Nyquist – Shannon βασίζεται σε μαθηματικά μοντέλα που ορίζουν ότι το σήμα εκτείνεται χρονικά στο άπειρο. Αυτό σε ένα πραγματικό σύστημα δεν ισχύει καθώς το σήμα έχει μια προκαθορισμένη χρονικά αρχή. Η απόκλιση αυτή από το θεωρητικό μαθηματικό μοντέλο οδηγεί σε μεταβατικά φαινόμενα που υποβαθμίζουν την πιστότητα του ανακατασκευασμένου σήματος. Επιπλέον λόγω περιορισμών στη μνήμη και την επεξεργαστική ισχύ ενός πραγματικού ψηφιακού συστήματος η επεξεργασία του σήματος γίνεται όχι σε όλη του την έκταση αλλά σε διακριτά και περιορισμένα τμήματά του. Το ανακατασκευασμένο αναλογικό σήμα έτσι προκύπτει από ένα περιορισμένο παράθυρο του αρχικού σήματος. Τα παραπάνω οδηγούν στο συμπέρασμα ότι ακόμα και αν το σήμα εισόδου είναι αυστηρά συχνοτικά περιορισμένο, μπορούν να υπάρξουν πλεονεκτήματα στην χρήση ρυθμών δειγματοληψίας μεγαλύτερων από αυτόν που υποδεικνύει το θεώρημα Nyquist – Shannon για την συγκεκριμένη εφαρμογή. Για τον λόγο αυτόν για ηχητικές εφαρμογές γενικού σκοπού, αν και ο καθιερωμένος ρυθμός δειγματοληψίας είναι κοντά στο όριο Nyquist (44,1kHz), πολύ συχνά όταν χρειαζόμαστε βέλτιστη ποιότητα επεξεργασίας χρησιμοποιούμε μεγαλύτερους ρυθμούς (48kHz, 96kHz, 192kHz).

Αν το αναλογικό προς ψηφιοποίηση σήμα περιλαμβάνει συχνότητες υψηλότερες του ημιορίου Nyquist, μετά την ψηφιοποίησή του οι συχνότητες αυτές δεν χάνονται αλλά εμφανίζονται ως χαμηλές. Το φαινόμενο αυτό ονομάζεται αναδίπλωση συχνοτήτων (frequency aliasing) και οδηγεί στην εμφάνιση παραμόρφωσης στο σήμα.



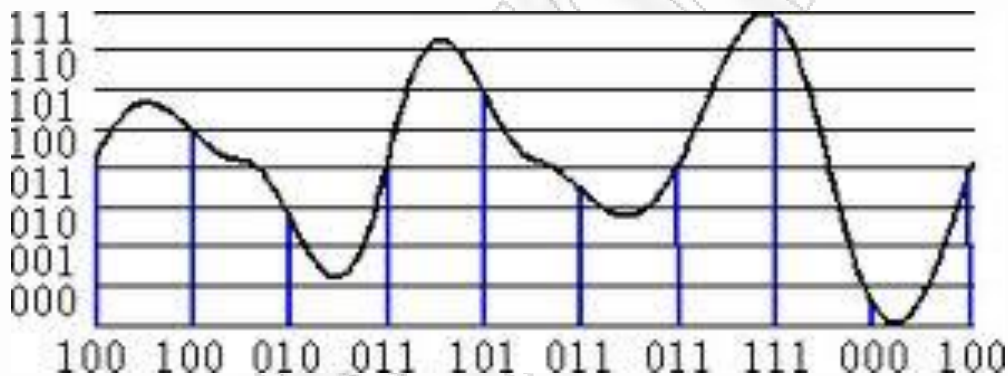
**Εικόνα 4: Αναδίπλωση Συχνοτήτων**

Για την αποφυγή των φαινομένων της αναδίπλωσης συχνοτήτων το σήμα πριν υποστεί δειγματοληψία περνά μέσα από ένα κατωπερατό (low-pass) φίλτρο το οποίο αφαιρεί τις συχνότητες που είναι μεγαλύτερες του ημιορίου Nyquist. Ιδανικά η ενέργεια αυτή θα εξάλειφε το φαινόμενο της αναδίπλωσης συχνοτήτων. Στην πραγματικότητα κανένα φίλτρο δεν έχει απόλυτη συχνότητα αποκοπής και έτσι μέρος της ενέργειας των υψηλότερων συχνοτήτων εισέρχεται στο προς δειγματοληψία σήμα. Αυτό συνιστά έναν ακόμα λόγο χρήσης μεγαλύτερων ρυθμών δειγματοληψίας. Η χρήση των anti-aliasing φίλτρων γίνεται στον μετατροπέα αναλογικού-σε-ψηφιακό. Οι παλαιότεροι μετατροπείς μεταβλητού ρυθμού δειγματοληψίας

επέτρεπαν την επιλογή του κατάλληλου φίλτρου πριν την διαδικασία ψηφιοποίησης. Με την πρόοδο της τεχνολογίας το υλικό έγινε τόσο γρήγορο που πλέον είναι εφικτοί ρυθμοί δειγματοληψίας πολλαπλάσιοι του επιθυμητού για κάθε εφαρμογή. Έτσι χρησιμοποιείται σταθερό anti-aliasing φίλτρο μεγάλης συχνότητας αποκοπής και δειγματοληψία σε υψηλό ρυθμό. Τέλος το σήμα μετατρέπεται στον επιθυμητό ρυθμό με ψηφιακές τεχνικές (downsampling).

### 1.1.2 Ανάλυση

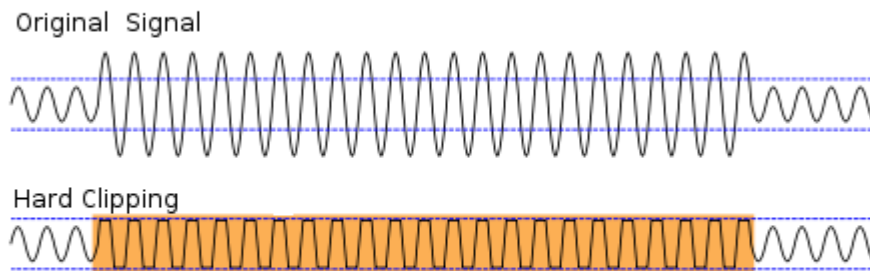
Βασικό χαρακτηριστικό του ψηφιακού ήχου είναι η ανάλυσή του. Ως ανάλυση ορίζουμε την πιστότητα της αναπαράστασης του αναλογικού σήματος κατά την ψηφιοποίησή του. Η κάθε τιμή του αναλογικού σήματος που διαβάζεται κατά την δειγματοληψία κβαντίζεται στην πλησιέστερη ψηφιακή τιμή που ορίζεται από την ανάλυση του σήματος. Μια ανάλυση 16bit ανά δείγμα για παράδειγμα αντιστοιχεί την τάση του αναλογικού σήματος σε μία από 65.536 διακριτές τιμές. Γίνεται έτσι προφανές ότι η αναπαράσταση του αρχικού σήματος δεν είναι απόλυτα ακριβής. Το σφάλμα που προκαλείται ονομάζεται θόρυβος κβάντισης.



Εικόνα 5: Σφάλμα Κβάντισης

Η ποιότητα του κβαντισμένου σήματος μετρείται μέσω του λόγου σήματος προς θόρυβο κβάντισης (signal-to-noise ratio - SNR). Κάθε bit που προσθέτουμε στην ανάλυση του σήματος αυξάνει το SNR κατά 6db. Η συχνότερα χρησιμοποιούμενη ανάλυση για επεξεργασία μουσικής και ομιλίας είναι 16bit και δίνει SNR 96db.

Οι αριθμητικές τιμές που θα δώσει η διαδικασία κβάντισης του σήματος κατά την ψηφιοποίηση εξαρτώνται από τα ηλεκτρικά χαρακτηριστικά του μετατροπέα, καθώς και του αναλογικού σήματος. Για την βέλτιστη μετατροπή σε ψηφιακή μορφή θα πρέπει η τάση του αναλογικού σήματος να βρίσκεται εντός των ορίων της τάσης εισόδου του μετατροπέα. Αυτό επιτυγχάνεται συνήθως με κατάλληλη ρύθμιση της τάσης του εισερχόμενου σήματος μέσω προενίσχυσης. Στην ιδανική περίπτωση χρησιμοποιούμε όλο το δυναμικό εύρος του μετατροπέα. Αν το σήμα εισόδου είναι ασθενές λόγω σταθερού SNR πιθανή ενίσχυσή του στο ψηφιακό πλέον σύστημα θα ενισχύσει και τον θόρυβο. Αν το σήμα είναι πολύ δυνατό και τα ηλεκτρικά του χαρακτηριστικά ξεπερνούν τις αντοχές του μετατροπέα, οι ακραίες τιμές του θα κβαντιστούν στην μέγιστη (και ελάχιστη) τιμή που προσφέρει η δεδομένη ανάλυση. Με τον τρόπο αυτόν παρατηρείται απώλεια πληροφορίας του αρχικού σήματος. Το φαινόμενο ονομάζεται παραμόρφωση ψαλιδισμού (clipping distortion).



Εικόνα 6: Clipping Distortion

### 1.1.3 Χαρακτηριστικά Αναπαράστασης

Η διαδικασία της κβάντισης του κάθε δείγματος του αναλογικού σήματος οδηγεί στην αντιστοίχηση του σε μια ακέραια τιμή. Για την υλοποίησή όμως ενός ψηφιακού συστήματος και για την εξασφάλιση της συμβατότητας μεταξύ συστημάτων ανεπτυγμένων από πληθώρα σχεδιαστών πρέπει να οριστούν τα χαρακτηριστικά της κωδικοποίησης στον ψηφιακό κόσμο. Αυτά είναι η χρήση ή όχι προσήμου στην αναπαράσταση και η σχετική θέση των bytes σε κάθε δείγμα (endianness).

Οι ακέραιοι αριθμοί που χρησιμοποιούνται για να αναπαραστήσουν τις τιμές των δειγμάτων του σήματος μπορεί να είναι προσημασμένοι ή μη προσημασμένοι. Στον αναλογικό κόσμο ένα ηχητικό σήμα αποτελείται και από θετικές και από αρνητικές τιμές τάσης (εναλλασσόμενο ρεύμα). Κατά τη μετατροπή του σε ψηφιακό σήμα μπορούμε να κωδικοποιήσουμε την τιμή της τάσης με οποιονδήποτε τρόπο. Τα διακριτά επίπεδα κβάντισης δεν θα μεταβληθούν όπως επίσης και η σχετική διαφορά μεταξύ ακραίων τιμών. Αυτό που θα αλλάξει όμως θα είναι η αναπαράσταση του σημείου ηρεμίας του σήματος το οποίο στην περίπτωση χρήσης προσημασμένων τιμών είναι μηδέν, ενώ στην περίπτωση χρήσης μη προσημασμένων τιμών βρίσκεται στο μέσο του δυναμικού εύρους της μετατροπής.

Το δεύτερο χαρακτηριστικό της κωδικοποίησης είναι η σχετική θέση των bytes σε κάθε δείγμα. Εφόσον η ανάλυση της μετατροπής είναι μεγαλύτερη από 8 bit πρέπει να γνωρίζει ο σχεδιαστής εφόσον η θέση των bytes σε κάθε δείγμα ακολουθεί το big-endian ή το little-endian μοντέλο. Τα χαρακτηριστικά της ψηφιακής αναπαράστασης του ήχου δεν επηρεάζουν την ποιότητα του αποτελέσματος της ψηφιοποίησης όπως ο ρυθμός δειγματοληψίας ή η ανάλυση αλλά η γνώση τους είναι απαραίτητη για τον σωστό σχεδιασμό συστημάτων ψηφιακής επεξεργασίας ήχου

### 1.1.4 Αριθμός καναλιών

Συχνά σε εφαρμογές ψηφιακής επεξεργασίας ήχου χειριζόμαστε ταυτόχρονα περισσότερα από ένα σήματα. Οι εφαρμογές επεξεργασίας μουσικής είναι χαρακτηριστικό παράδειγμα. Το καθιερωμένο πρότυπο παραγωγής μουσικής τα τελευταία χρόνια (stereo) χρησιμοποιεί δύο κανάλια, δύο δηλαδή ξεχωριστά σήματα για το αριστερό και το δεξί ηχείο, ενώ διαδίδεται όλο και περισσότερο ο εξακάναλος ήχος (συστήματα 5.1). Κατά την ψηφιοποίηση ενός τέτοιου σήματος ο μετατροπέας αναλογικού-σε-ψηφιακό πολυπλέκει χρονικά τα εισερχόμενα σήματα σε μία ροή δεδομένων τοποθετώντας σειριακά τα ψηφιοποιημένα δείγματα. Η αποπλεξία των καναλιών και η επεξεργασία του καθενός ξεχωριστά γίνεται από το σχεδιαστή της εφαρμογής.

### 1.1.5 Συμπίεση

Η αναπαράσταση του ήχου σε ψηφιακή μορφή και οι ολοένα αυξανόμενες απαιτήσεις για ποιότητα επεξεργασίας δημιουργούν μεγάλες ανάγκες αποθήκευσης των δεδομένων καθώς και προβλήματα στην αποδοτική μετάδοσή τους σε κάποιο δίκτυο. Οι ανάγκες αυτές γίνονται μεγαλύτερες όσο αυξάνει ο ρυθμός δειγματοληψίας και η ανάλυση του ψηφιοποιημένου ήχου. Έτσι πολύ συχνά χρησιμοποιούνται τεχνικές συμπίεσης του ψηφιακού ήχου με σκοπό την ελάττωση του όγκου των δεδομένων. Οι τεχνικές που χρησιμοποιούνται διακρίνονται σε lossless, στις οποίες μετά την αποσυμπίεση τα αρχικά δεδομένα ανακατασκευάζονται τέλεια, και lossy, όπου η ποιότητα του συμπιεζόμενου σήματος υποβαθμίζεται.

Οι lossless τεχνικές μειώνουν το μέγεθος των δεδομένων εξαλείφοντας άχρηστη πληροφορία και αναγνωρίζοντας και κωδικοποιώντας μοτίβα στα ηχητικά δεδομένα. Ο λόγος συμπίεσης που επιτυγχάνεται είναι συνήθως μικρός και εξαρτάται από το ηχητικό περιεχόμενο. Οι αλγόριθμοι lossless συμπίεσης επιτυγχάνουν τα καλύτερα αποτελέσματα σε επεξεργασία φωνής και ήχου με περιορισμένο συχνοτικό εύρος. Τέλος απαιτείται και σχετικά μεγάλη επεξεργαστική ισχύς για την συμπίεση / αποσυμπίεση του ήχου. Τα πρότυπα lossless συμπίεσης που έχουν επικρατήσει είναι τα Shorten και FLAC (Free Lossless Audio Codec).

Οι lossy τεχνικές συμπιέζουν τα ηχητικά δεδομένα αφαιρώντας πληροφορία από το σήμα, με τρόπο όμως που αυτό γίνεται όσο το δυνατόν λιγότερο αντιληπτό από τον ακροατή. Συνήθως επιτυγχάνεται καλύτερος λόγος συμπίεσης από τις lossless τεχνικές, ενώ δίνεται και η δυνατότητα επιλογής της σχέσης ποιότητας / μεγέθους των δεδομένων. Σε πολλούς lossy αλγορίθμους συμπίεσης χρησιμοποιούνται ψυχοακουστικές μέθοδοι για να αναγνωρισθεί ηχητική πληροφορία που δεν γίνεται αντιληπτή από τον άνθρωπο και στη συνέχεια να αφαιρεθεί (συνήθως στο υψηλότερο φάσμα συχνοτήτων). Οι κυριότερες lossy μέθοδοι συμπίεσης είναι οι μ-law audio compression, το Adaptive Differential Pulse Code Modulation (ADPCM) και το MPEG compression (υποκατηγορία του οποίου είναι και το MP3 format).

## 1.2 Θεμελιώδης Συχνότητα και Pitch

### 1.2.1 Ορισμός του Pitch

Σύμφωνα με τον ορισμό του American National Standard Institute (ANSI) το pitch είναι «το χαρακτηριστικό της αίσθησης της ακοής σύμφωνα με το οποίο οι ήχοι μπορούν να διαταχθούν σε μια κλίμακα που εκτείνεται από “χαμηλά” έως “ψηλά”». Ο συγκεκριμένος ορισμός είναι ένας από τους πολλούς που έχουν προταθεί για να περιγράψουν το pitch. Κοινό χαρακτηριστικό όλων είναι ότι δεν περιέχουν αναφορά σε φυσικά μετρήσιμα μεγέθη.

Ο τρόπος που ο άνθρωπος αντιλαμβάνεται το pitch είναι δυαδικός. Σε πρώτη φάση γίνεται ανάλυση του ηχητικού σήματος στο πεδίο της συχνότητας από τον κοχλία του αφτιού, ο οποίος διανέμει διακριτά συχνοτικά ερεθίσματα σε διακριτά νευρικά “κανάλια”. Εν συνεχεία, το νευρικό τμήμα του ακουστικού συστήματος αναλύει τα δοσμένα ερεθίσματα στο πεδίο του χρόνου με τη μέθοδο της αυτοσυσχέτισης. Το pitch είναι ένα καθαρά υποκειμενικό χαρακτηριστικό του ήχου το οποίο όμως σχετίζεται άμεσα με τη συχνότητα. Συγκεκριμένα η έννοια της συχνότητας χρησιμοποιείται για να μετρήσει επιστημονικά το pitch. Στην περίπτωση ενός καθαρού ημιτονοειδούς ηχητικού σήματος η συχνότητά του ταυτίζεται με το pitch. Το καθαρό ημίτονο όμως δεν συναντάται στη φύση. Η μελέτη πιο σύνθετων περιοδικών ήχων αποκαλύπτει στο συχνοτικό τους φάσμα πολλαπλές κορυφές. Ένας περιοδικός ήχος μπορεί να αναλυθεί ως άθροισμα ημιτονοειδών σύμφωνα με την εξίσωση

$$x(t) = \sum_{i=1}^{I(t)} A_i(t) \exp(j\phi_i(t)) \quad \text{with} \quad \phi_i(t) = \int_{-\infty}^t \omega_i(\tau) d\tau$$

όπου,  $\omega$ , είναι η συχνότητα του ημιτονοειδούς. Η μικρότερη συχνότητα που συναντάται στην ανάλυση είναι η θεμελιώδης συχνότητα του ήχου (F0). Οι υπόλοιπες συχνότητες ονομάζονται αρμονικές και αντιστοιχούν σε μοτίβα ταλάντωσης της ηχητικής πηγής. Οι αρμονικές ενός περιοδικού ήχου σχετίζονται με τη θεμελιώδη συχνότητα με ακέραιους λόγους. Στην εικόνα 7 παρουσιάζεται η σχέση μεταξύ της θεμελιώδους συχνότητας και των αρμονικών μιας χορδής σε ταλάντωση. Στο πρώτο σχήμα απεικονίζεται η βασική ταλάντωση της χορδής, η οποία αντιστοιχεί στην θεμελιώδη συχνότητα. Το μεγαλύτερο μέρος της ενέργειας του παραγόμενου ήχου οφείλεται στην ταλάντωση αυτήν. Ταυτόχρονα όμως η χορδή ταλαντώνεται και με τους τρόπους που παρουσιάζονται στα σχήματα 2-4. Οι επιμέρους ταλαντώσεις παράγουν τις αρμονικές 2<sup>ης</sup>, 3<sup>ης</sup> και 4<sup>ης</sup> τάξης, με συχνότητες ακέραια πολλαπλάσια της αρχικής.

Fig.1

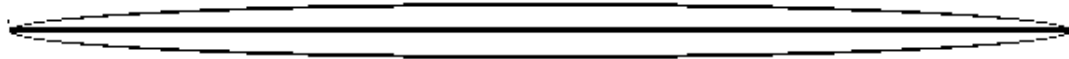


Fig.2



Fig.3

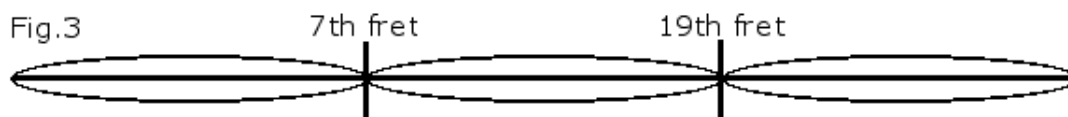


Fig.4



### Εικόνα 7: Θεμελιώδης συχνότητα και αρμονικές σε παλλόμενη χορδή

Το ανθρώπινο αυτί αναγνωρίζει αύξηση του pitch κατά μια οκτάβα όταν διπλασιάζεται η θεμελιώδης συχνότητα ενός ήχου. Η αντίληψη αυτή αλλάζει όμως ανάλογα με την συχνοτική περιοχή στην οποία βρίσκεται η θεμελιώδης. Έτσι για συχνότητες κάτω από 1000Hz διπλασιασμός της θεμελιώδους γίνεται αντιληπτός ως μεταβολή του pitch ελάχιστα μικρότερη από μια οκτάβα, ενώ για συχνότητες πάνω από 5000Hz γίνεται αντιληπτός ως μεταβολή του pitch ελάχιστα μεγαλύτερη από μια οκτάβα. Διαφορές στο pitch γίνονται αντιληπτές και όταν μεταβάλλεται η ένταση ενός ήχου δεδομένου ότι δεν μεταβάλλεται η θεμελιώδης συχνότητα.

Από τα παραπάνω διαπιστώνουμε ότι όσο αυξάνεται η πολυπλοκότητα και το αρμονικό περιεχόμενο ενός ήχου τόσο πιο δύσκολο είναι να ορίσουμε ένα μετρήσιμο μέγεθος που να σχετίζεται με την αντίληψη του pitch. Το πρόβλημα γίνεται ακόμα πιο σύνθετο όταν ένας ήχος είναι ψευδο-περιοδικός ή όταν η έντασή του μεταβάλλεται σε σχέση με το χρόνο ή όταν περιέχει δύο και παραπάνω θεμελιώδεις συχνότητες (πολυφωνία). Επιπλέον αξίζει να αναφερθεί ότι υπάρχουν περιοδικοί ήχοι οι οποίοι όμως δεν διαθέτουν αντιληπτό pitch ενώ συναντάμε και μη περιοδικούς ήχους οι οποίοι προκαλούν στην ανθρώπινη ακοή την ψευδαίσθηση μιας θεμελιώδους συχνότητας. Παρόλα αυτά όταν μελετάμε μονοφωνικούς ψευδο-περιοδικούς ήχους

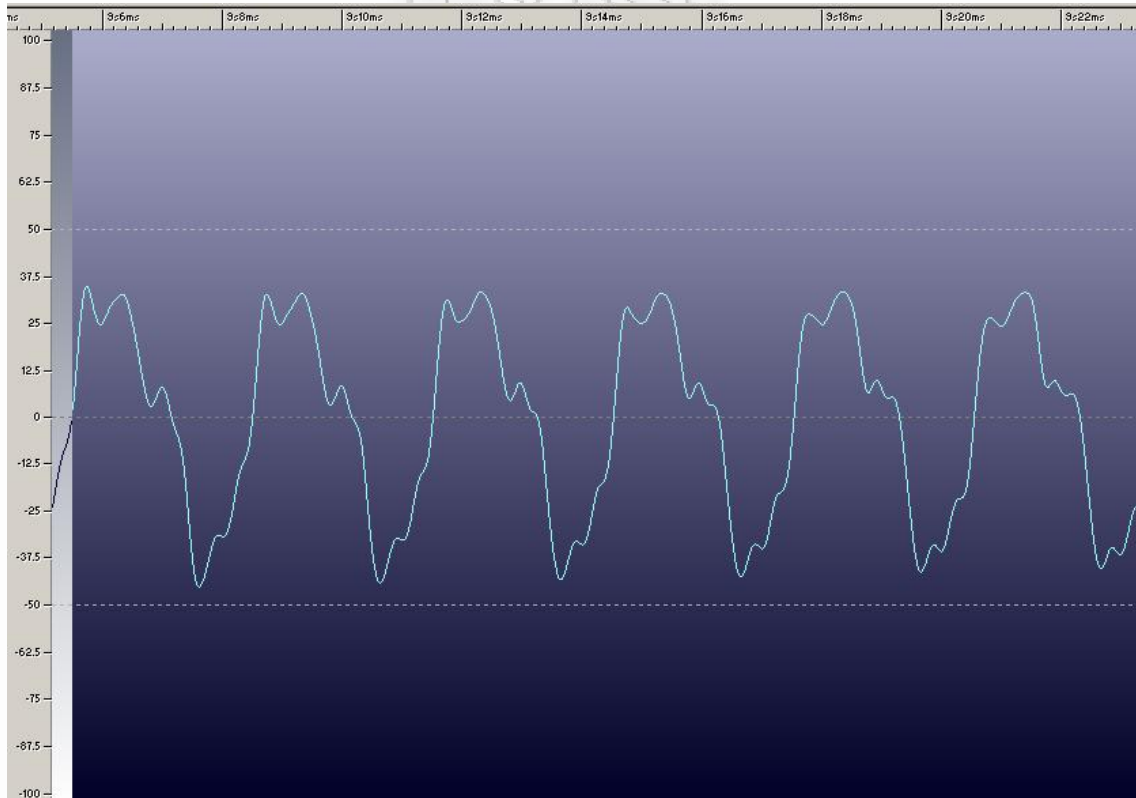
με λογικές μεταβολές στην ένταση και τη συχνότητά τους είναι αρκετά ασφαλές να ταυτίσουμε το pitch με την θεμελιώδη συχνότητα του ήχου.

### 1.2.2 Αλγόριθμοι ανίχνευσης του pitch

Από τα παραπάνω διαπιστώνουμε ότι η ανίχνευση του pitch ακόμη και σε απλά μονοφωνικά ψευδο-περιοδικά σήματα (και κατά συνέπεια η ανίχνευση της θεμελιώδους συχνότητας τους) δεν είναι απλό πρόβλημα. Με την πάροδο χρόνων έρευνας και ανάπτυξης στον τομέα έχουν προταθεί αρκετοί αλγόριθμοι για την επίλυσή του, κάθε ένας από τους οποίους διαθέτει πλεονεκτήματα και μειονεκτήματα ανάλογα με την φύση της εφαρμογής που μας ενδιαφέρει. Οι αλγόριθμοι ανίχνευσης του pitch διακρίνονται αρχικά με βάση το αν εξετάζουν το σήμα στο πεδίο του χρόνου ή της συχνότητας.

- Zero Crossing Rate

Ο zero crossing rate είναι ένας πολύ απλός αλγόριθμος για τον υπολογισμό του pitch δοθέντος σήματος. Λειτουργεί στο πεδίο του χρόνου και βασίζει τη λειτουργία του στην μέτρηση του αριθμού των μεταβάσεων του σήματος από την αρνητική στη θετική περιοχή τιμών. Ο αριθμός των διελεύσεων από το μηδέν σχετίζεται άμεσα με τον αριθμό των επαναλήψεων της κυματομορφής στο χρόνο. Έτσι η μελέτη του μοτίβου των διελεύσεων από το μηδέν μπορεί τελικά να δώσει πληροφορίες για την περίοδο του σήματος και κατά συνέπεια να ανιχνεύσει την θεμελιώδη συχνότητά του. Η απόδοση του αλγορίθμου είναι καλή για πολύ απλά σήματα που δεν περιέχουν σημαντικό ποσοστό ενέργειας στις αρμονικές τους. Όπως φαίνεται στην εικόνα 8, το μοτίβο των διελεύσεων από το μηδέν ανιχνεύεται απλά μεταξύ των επαναλήψεων των περιόδων του σήματος.



Εικόνα 8: Περιοδική Κυματομορφή Χαμηλής Πολυπλοκότητας

Όταν το υπό μελέτη σήμα όμως περιέχει θόρυβο ή ενέργεια σε υψηλές συχνότητες, το μοτίβο των διελεύσεων γίνεται χαοτικό και η εξαγωγή χρήσιμων πληροφοριών από αυτό είναι αρκετά δύσκολη και αναξιόπιστη. Η απόδοσή του μπορεί να βελτιωθεί αρκετά με χρήση προεπεξεργασίας στο σήμα με κατάλληλες μεθόδους (πχ χρήση centre clipping για την αφαίρεση των προβληματικών συχνοτήτων).



**Εικόνα 9: Περιοδική Κυματομορφή Υψηλής Πολυπλοκότητας**

Ο zero crossing rate αλγόριθμος έχει το πλεονέκτημα ότι υλοποιείται πολύ απλά και δεν απαιτεί υψηλή υπολογιστική ισχύ. Συνήθως χρησιμοποιείται σε συνεργασία με άλλους αλγορίθμους για να παρέχει επιπλέον πληροφορία για το σήμα και να βελτιώσει την σταθερή λειτουργία της εφαρμογής. Παραλλαγή του αλγορίθμου αποτελεί και ο Peak Rate αλγόριθμος που μετράει τις θετικές (ή αρνητικές) κορυφές στη μονάδα του χρόνου.

- Autocorrelation Function

Στόχος των μεθόδων που βασίζονται στην αυτοσυσχέτιση (autocorrelation) είναι να εντοπίσουν ομοιότητες μεταξύ ενός τμήματος του σήματος και ενός ολισθημένου στο χρόνο αντιγράφου αυτού. Εφόσον εντοπιστούν ομοιότητες μπορούμε να υποθέσουμε ότι αντιστοιχούν σε διαφορετικές περιόδους του σήματος και να υπολογίσουμε από την χρονική μετατόπιση την περίοδο αυτού.

Η συνάρτηση αυτοσυσχέτισης ενός σήματος ορίζεται από την παρακάτω εξίσωση:



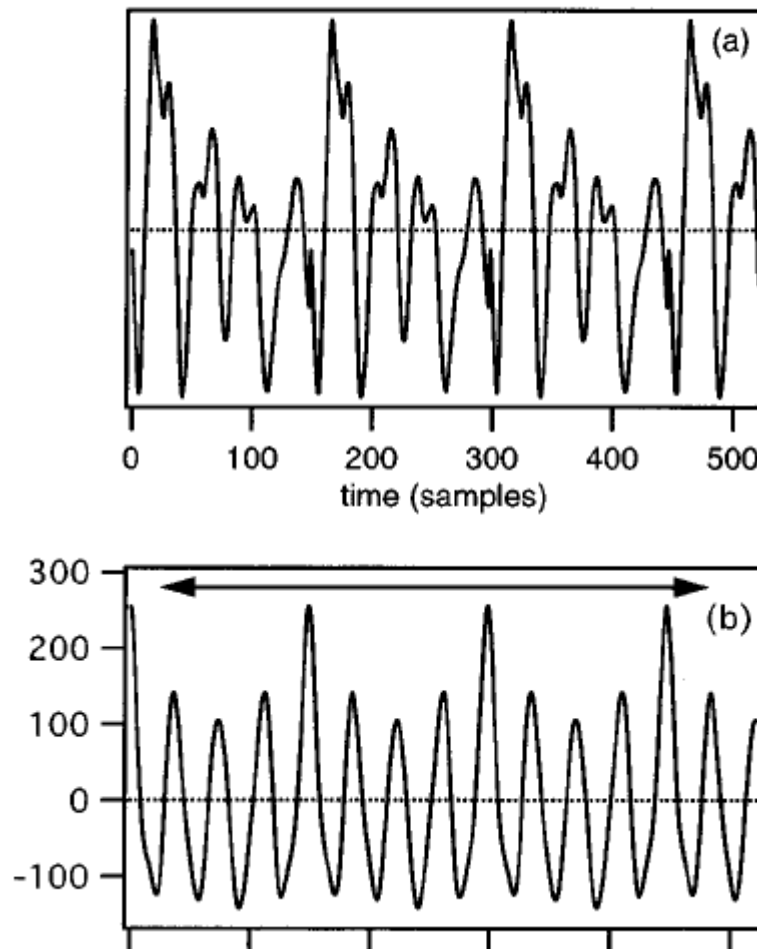
$$R_x(m) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(n)x(n+m)$$

και προκύπτει από το κανονικοποιημένο άθροισμα των γινομένων του σήματος με τον μετατοπισμένο στο χρόνο εαυτό του. Η συνάρτηση στη γενική μορφή παρατηρούμε ότι εκτείνεται σε όλη τη διάρκεια του σήματος και κατά συνέπεια είναι μη εφαρμόσιμη σε ένα πραγματικό σύστημα. Από την γενική εξίσωση αυτή όμως αποδεικνύεται εύκολα ότι εφόσον το σήμα  $X(n)$  είναι περιοδικό ( $X(n) = X(n+T)$ ) τότε και η συνάρτηση αυτοσυσχέτισης θα είναι και αυτή περιοδική και μάλιστα η περίοδος των δύο θα είναι η ίδια ( $R_x(m) = R_x(m+T)$ ).

Η χρήση μεθόδων αυτοσυσχέτισης χρησιμοποιεί τεχνικές ανάλυσης περιορισμένης διάρκειας. Το υπό εξέταση σήμα χωρίζεται σε τμήματα πεπερασμένου μήκους από τα οποία υπολογίζεται η συνάρτηση αυτοσυσχέτισης περιορισμένης διάρκειας σύμφωνα με την εξίσωση:

$$R_x(m) = \frac{1}{N} \sum_{n=0}^{N-1} [x(n+l)w(n)][x(n+l+m)w(n+m)]$$

Από την εξίσωση παρατηρούμε ότι το υπολογιζόμενο άθροισμα είναι πλέον οριοθετημένο από το παράθυρο  $w(n)$  που εφαρμόζεται στο σήμα. Η τιμή της συνάρτησης αυτοσυσχέτισης μεγιστοποιείται όταν η απόσταση της μετατόπισης του σήματος  $m$  πολλαπλάσιο της περιόδου του σήματος. Εντοπίζοντας λοιπόν το πρώτο μέγιστο της συνάρτησης  $R_x(m)$  βρίσκουμε την θεμελιώδη συχνότητα. Στην εικόνα παρουσιάζεται τμήμα ενός σήματος και η συνάρτηση αυτοσυσχέτισής του για περιορισμένο χρόνο



Εικόνα 10: Κυματομορφή και ACF

Ο αλγόριθμος ανίχνευσης του pitch με χρήση της συνάρτησης αυτοσυσχέτισης δίνει ικανοποιητικά αποτελέσματα, ιδίως στις χαμηλές περιοχές συχνότητας. Για τον λόγο αυτόν χρησιμοποιείται ευρέως σε εφαρμογές επεξεργασίας ομιλίας, στις οποίες το συχνотικό περιεχόμενο είναι περιορισμένο. Επιπλέον ο αλγόριθμος λόγω της φύσης του παρουσιάζει ανοχή στον θόρυβο. Βασικό του μειονέκτημα είναι η εμφάνιση λαθών οκτάβας, όπου εντοπίζεται ομοιότητα στο σήμα αλλά με το δεύτερο πολλαπλάσιο της περιόδου και δίνεται η τιμή της επόμενης οκτάβας (διπλάσια θεμελιώδης συχνότητα). Ο αλγόριθμος απαιτεί αρκετή υπολογιστική ισχύ, καθώς σε κάθε παράθυρο προς επεξεργασία εκτελούνται πολλές προσθέσεις και πολλαπλασιασμοί. Η υπολογιστική του πολυπλοκότητα μπορεί να ελατωθεί με χρήση χαμηλότερων ρυθμών δειγματοληψίας, εφόσον οι απαιτήσεις μας σε ανάλυση pitch δεν είναι μεγάλες. Έχει παρατηρηθεί ότι ο αλγόριθμος λειτουργεί εξίσου καλά με ρυθμούς δειγματοληψίας ίσους με το  $\frac{1}{4}$  της αρχικής. Τέλος υπάρχουν τεχνικές περεταιίρω επιτάχυνσης των υπολογισμών, με χρήση πληροφορίας από προηγούμενες εκτελέσεις του αλγορίθμου και τοπικό μόνο υπολογισμό της συνάρτησης αυτοσυσχέτισης.

Και αυτός ο αλγόριθμος ωφελείται από χρήση κατάλληλης προ-επεξεργασίας στο σήμα. Χρήσιμες τεχνικές είναι το harmonic damping στο οποίο χρησιμοποιείται ένα low-pass φίλτρο για να αφαιρέσει το συχνотικό περιεχόμενο των αρμονικών υψηλής τάξης που βρίσκονται εκτός των ορίων της εφαρμογής, η χρήση centre clipping και infinite peak clipping που περιγράφονται από της παρακάτω εξισώσεις αντίστοιχα:

$$y(n) = clc[x(n)] = \begin{cases} (x(n) - C_L), & x(n) \geq C_L \\ 0, & |x(n)| < C_L \\ (x(n) + C_L), & x(n) \leq -C_L \end{cases}$$

$$y(n) = \text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq C_L \\ 0, & |x(n)| < C_L \\ -1, & x(n) \leq -C_L \end{cases}$$

Μετατρέποντας το σήμα με χρήση των παραπάνω εξισώσεων επιτυγχάνεται ο περιορισμός της πληροφορίας που περιέχει και είναι μη σχετική με τον στόχο του αλγορίθμου (ανίχνευση ομοιότητας μεταξύ περιόδων). Ο centre clipping αλγόριθμος διατηρεί μόνο τις κορυφές του προς επεξεργασία σήματος, ενώ ο infinite peak clipping το μετατρέπει σε τετραγωνικό παλμό, διατηρώντας όμως τη συχνότητα του αρχικού σήματος.

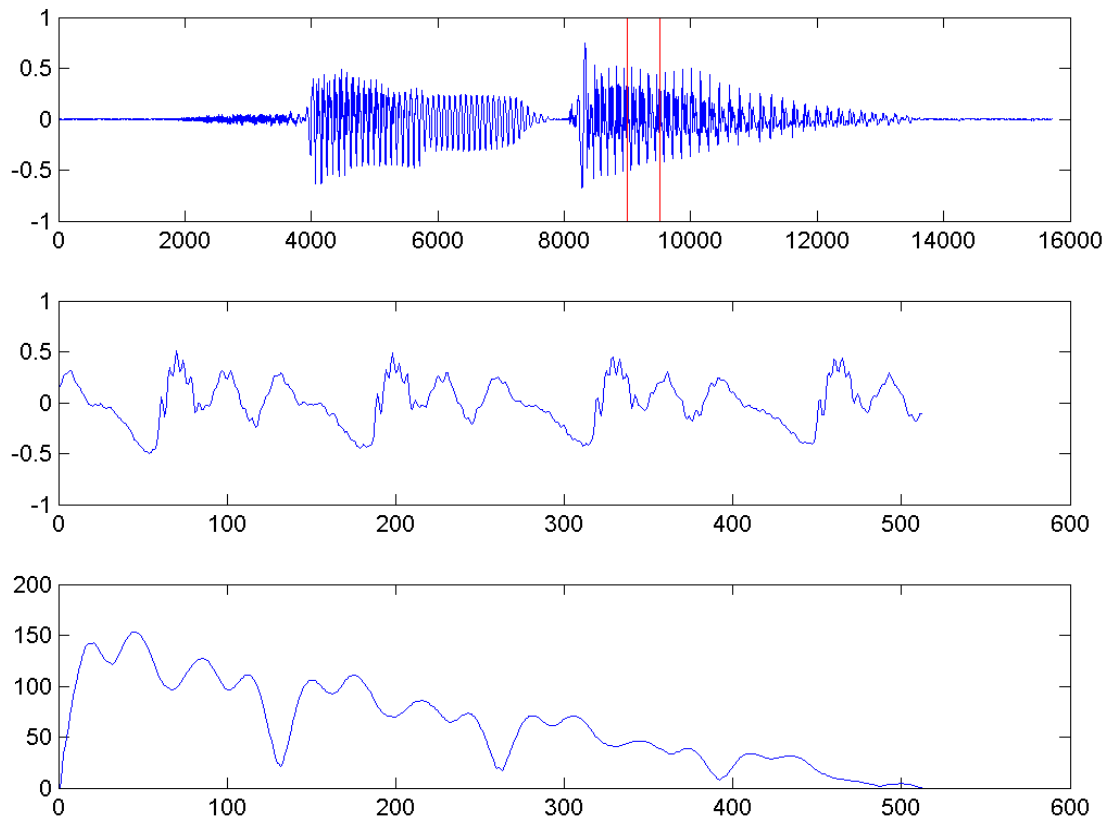
- Average Magnitude Difference Function (AMDF)

Ο AMDF αλγόριθμος αποτελεί υποκατηγορία αλγορίθμου αυτοσυσχέτισης. Δουλεύει και αυτός στο πεδίο του χρόνου και προσπαθεί να εντοπίσει ομοιότητα μεταξύ δύο περιόδων του σήματος. Η ανάγκη για την ανάπτυξη του AMDF προκύπτει από το γεγονός ότι ο πολλαπλασιασμός, που χρησιμοποιείται στην εξίσωση αυτοσυσχέτισης είναι ακριβή υπολογιστικά πράξη. Στόχος λοιπόν είναι να βρεθεί μια εξίσωση με παρόμοια χαρακτηριστικά ώστε να μειωθεί ο υπολογιστικός φόρτος του συστήματος. Η εξίσωση που χρησιμοποιεί ο AMDF είναι η παρακάτω:

$$D_m = \frac{1}{L} \sum_{n=1}^L |x(n) - x(n-m)|, \quad m = 0, 1, \dots, m_{\max}$$

Το σήμα μετατοπίζεται στον χρόνο και αφαιρείται ανά δείγμα κατά απόλυτη τιμή από τον εαυτό του. Σε περίπτωση που παρουσιάζεται μεγάλη ομοιότητα η τιμή της συνάρτησης ελαττώνεται. Ο αλγόριθμος λοιπόν, αφού δημιουργήσει την  $D_m$  για ένα παράθυρο δειγμάτων

αναζητά τοπικά ελάχιστα. Το πρώτο ελάχιστο που θα εντοπιστεί αντιστοιχεί στην θεμελιώδη συχνότητα. Η λειτουργία του αλγορίθμου παρουσιάζεται στην εικόνα 12



**Εικόνα 11: Κυματομορφή και AMDF**

Στο πρώτο τμήμα απεικονίζεται το αρχικό το συνολικό σήμα με το προς επεξεργασία παράθυρο σημειωμένο. Στο δεύτερο τμήμα το παράθυρο στο οποίο θα εφαρμοστεί ο αλγόριθμος. Στο τρίτο απεικονίζεται η AMDF συνάρτηση για το συγκεκριμένο παράθυρο, όπου φαίνεται καθαρά η ύπαρξη τοπικού ελαχίστου μετά από μια περίοδο του σήματος.

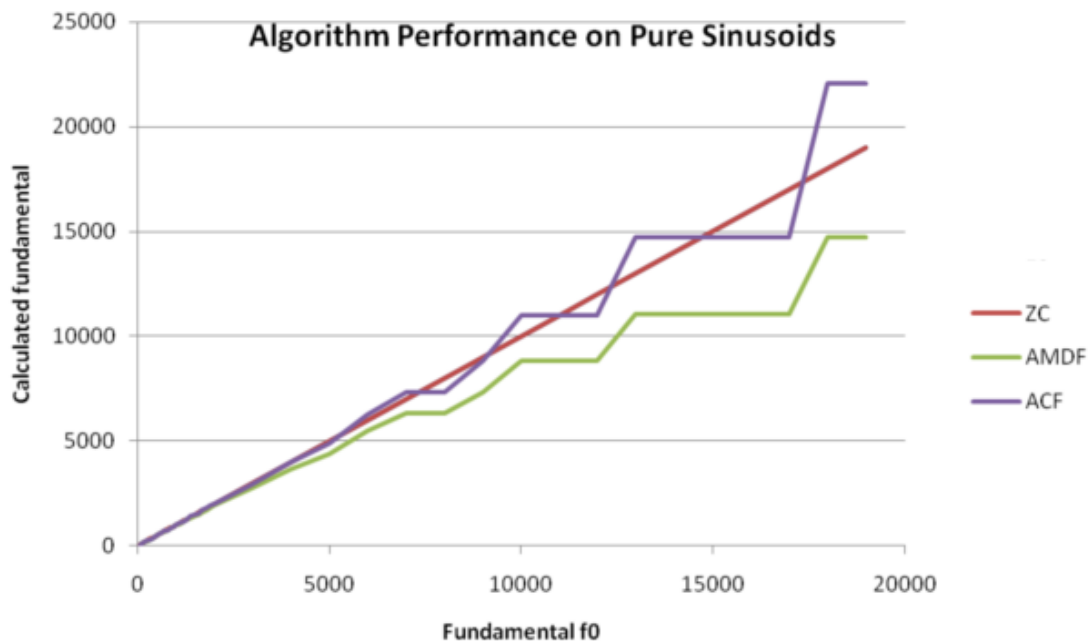
Αν και δεν χρησιμοποιεί χρονοβόρους πολλαπλασιασμούς, ο AMDF αλγόριθμος παραμένει αρκετά απαιτητικός υπολογιστικά, εξαιτίας του όγκου των πράξεων που εκτελούνται ανά παράθυρο. Οι ίδιες τεχνικές επιτάχυνσης που αναφέρθηκαν για τον autocorrelation αλγόριθμο μπορούν να εφαρμοστούν και εδώ. Η απόδοσή του κυμαίνεται στα ίδια επίπεδα με τον αλγόριθμο αυτοσυσχέτισης.

Μελέτη των Kobayashi και Shimamura το 1955 έδειξε ότι η συνάρτηση autocorrelation και η συνάρτηση AMDF είναι στατιστικά ανεξάρτητες και μπορούν να χρησιμοποιηθούν ταυτόχρονα για να προσφέρουν ακόμα μεγαλύτερη ανοχή στο θόρυβο. Η συνάρτηση που πρότειναν ονομάζεται Weighted Autocorrelation Function (WACF) και είναι η

$$f(T) = R_x(T) / D(T) + k$$

όπου  $R_x(T)$  η συνάρτηση autocorrelation και  $D(T)$  η AMDF συνάρτηση. Τα βέλτιστα αποτελέσματα μετρήθηκαν με τιμή της παραμέτρου  $k$  ίσης με 1.

Η εικόνα 12 παρουσιάζει συγκριτικά τα αποτελέσματα των αλγορίθμων zero crossing rate, autocorrelation και AMDF για ημιτονοειδή σήματα.



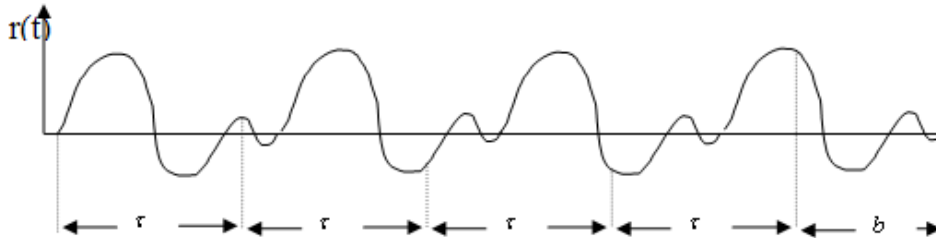
**Εικόνα 12: Σύγκριση Time – Domain Pitch Detection Αλγορίθμων ως προς την απόδοση**

Παρατηρούμε ότι για συχνότητες μέχρι 6 kHz περίπου και οι τρεις αλγόριθμοι ανιχνεύουν σχετικά σωστά την θεμελιώδη συχνότητα. Από εκεί και στην ανώτερη συχνοτική περιοχή η απόδοση πέφτει σε βαθμό που οι αλγόριθμοι γίνονται δύσχρηστοι και αναξιόπιστοι.

- Time Domain Maximum Likelihood

Ο αλγόριθμος καλύτερης πιθανότητας του πεδίου του χρόνου προτάθηκε από τον Noll το 1969. Ο αλγόριθμος διαιρεί ένα παράθυρο ενός περιοδικού σήματος  $r(t)$  μήκους  $T$  σε  $N$  τμήματα μήκους  $\tau$  (με  $1 < \tau < T$ ) όπως στην εικόνα. Στο τέλος του παραθύρου απομένει υπόλοιπο μήκους  $b$ .

$$T = N \cdot \tau + b, \quad 0 < b < \tau$$



Εικόνα 13: Αλγόριθμος Time – Domain Maximum Likelihood

Στη συνέχεια προσθέτουμε τα επιμέρους τμήματα για διάφορες τιμές του  $\tau$  σύμφωνα με την παρακάτω σχέση:

$$r(t, \tau) = \begin{cases} \frac{1}{N+1} \sum_{n=0}^N r(t+n\tau) & 0 \leq t \leq b \\ \frac{1}{N} \sum_{n=0}^N r(t+n\tau) & b \leq t \leq \tau \end{cases}$$

Όταν το  $\tau$  ισούται με την θεμελιώδη περίοδο του σήματος η συνάρτηση  $r(t, \tau)$  θα γίνει μέγιστη. Δημιουργούμε λοιπόν την συνάρτηση  $J(\tau)$  για την γενική περίπτωση

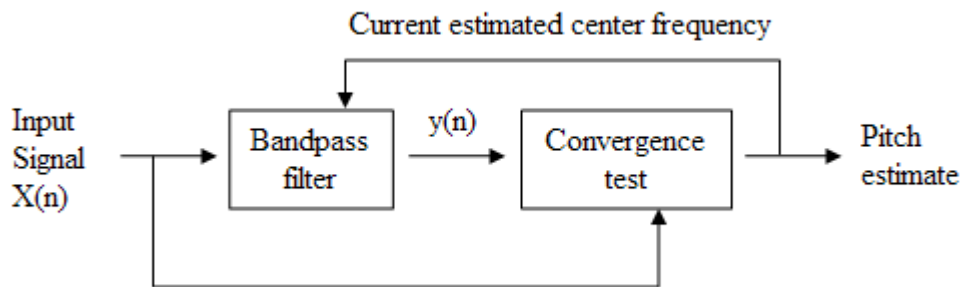
$$J(\tau) = (N+1) \sum_{t=0}^b r^2(t, \tau) + N \sum_{t=b+1}^{\tau} r^2(t, \tau)$$

και στη συνέχεια βρίσκουμε το μέγιστό της για να πάρουμε την θεμελιώδη συχνότητα. Η συνάρτηση  $J(\tau)$  δίνει τοπικά μέγιστα σε όλα τα ακέραια πολλαπλάσια του  $\tau$ . Εντοπίζοντας το μικρότερο λαμβάνουμε την θεμελιώδη συχνότητα.

- Adaptive Filtering Method

Προσαρμοζόμενο ονομάζεται ένα φίλτρο όταν μεταβάλλει την απόκριση συχνότητάς του με βάση μια παράμετρο, η οποία του δίνεται ως είσοδος, συνήθως προερχόμενη από κάποιον αλγόριθμο. Η ανάγκη χρήσης ενός τέτοιου φίλτρου γίνεται προφανής όταν ορισμένες παράμετροι του προς επίλυση προβλήματος δεν είναι εξ' αρχής γνωστές. Τα προσαρμοζόμενα φίλτρα είναι συνήθως ψηφιακά φίλτρα, λόγω της μεγάλης πολυπλοκότητας υλοποίησής τους στον αναλογικό κόσμο.

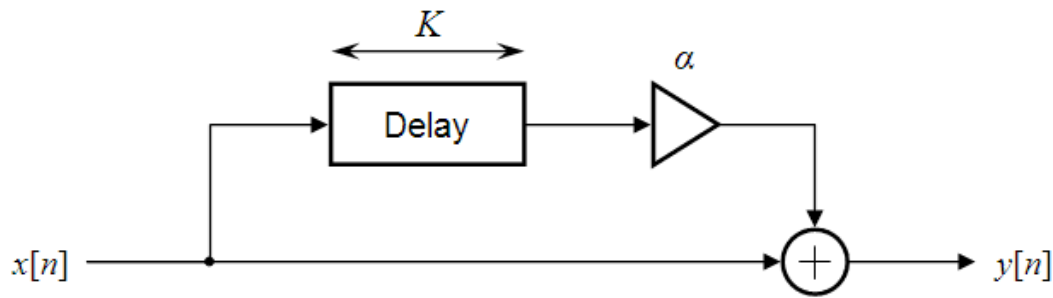
Στην περίπτωση ανίχνευσης του pitch μπορούν να χρησιμοποιηθούν προσαρμοζόμενα φίλτρα με σκοπό, μέσω της μεταβολής μιας παραμέτρου, να ανιχνευθεί ομοιότητα του σήματος με σήμα γνωστής συχνότητας. Η μέθοδος λειτουργεί στο πεδίο του χρόνου. Μια πρώτη περίπτωση χρήσης προσαρμοζόμενου φίλτρου χρησιμοποιεί ένα στενό ζωνοπερατό φίλτρο (narrow bandpass filter). Ένα ζωνοπερατό φίλτρο συνδυάζει ένα ανωπερατό και ένα κατωπερατό φίλτρο με σκοπό την διέλευση μιας περιορισμένης συχνοτικά περιοχής του σήματος εισόδου. Η μεταβαλλόμενη παράμετρος στο συγκεκριμένο φίλτρο είναι η κεντρική συχνότητα διέλευσης του σήματος. Η έξοδος του φίλτρου οδηγείται σε σύγκριση με το αρχικό, αφιλτράριστο σήμα, σε μια διάταξη που φαίνεται στην εικόνα 14



**Εικόνα 14: Αλγόριθμος Adaptive Filtering Με Χρήση Bandpass φίλτρου**

Στη συνέχεια το αποτέλεσμα της σύγκρισης δίνεται ως παράμετρος στο φίλτρο για να μεταβάλλει την συχνότητα διέλευσής του. Όταν η είσοδος είναι ένα περιοδικό σήμα και η συχνότητα διέλευσης του φίλτρου διαφέρει από αυτήν του εισερχόμενου σήματος, το φίλτρο θα αποκόψει το μεγαλύτερο μέρος του αρχικού σήματος και η ομοιότητα στην διάταξη σύγκρισης θα είναι περιορισμένη. Η διαδικασία επαναλαμβάνεται μέχρι τα δύο σήματα να συγκλίνουν, οπότε και η συχνότητα διέλευσης του φίλτρου θα ταυτίζεται με τη θεμελιώδη συχνότητα του σήματος.

Εναλλακτικά, μπορεί να χρησιμοποιηθεί ένα φίλτρο χτένας (comb filter). Ένα comb filter προσθέτει στην έξοδό του το αρχικό σήμα με μία καθυστερημένη στο χρόνο και ανεστραμμένη εκδοχή του.



Εικόνα 15: Αλγόριθμος Adaptive Filtering Με Χρήση Comb φίλτρου

Με τον τρόπο αυτόν, και με μεταβαλλόμενη παράμετρο την εισαγόμενη καθυστέρηση, προσπαθούμε να εντοπίσουμε το σημείο που ελαχιστοποιεί το σήμα εξόδου. Η καθυστέρηση στην συγκεκριμένη περίπτωση δίνει την θεμελιώδη συχνότητα του σήματος.

Η χρήση προσαρμοζόμενων φίλτρων παράγει αποτελέσματα συγκρίσιμα με αυτά των άλλων μεθόδων του πεδίου του χρόνου.

- Harmonic Product Spectrum (HPS)

Ο αλγόριθμος HPS εξετάζει το σήμα στο πεδίο της συχνότητας. Όλοι οι αλγόριθμοι που λειτουργούν στο πεδίο αυτό χρησιμοποιούν μετατροπή του σήματος με χρήση μετασχηματισμού Fourier διακριτού χρόνου, εφαρμοζόμενου σε περιορισμένα τμήματα του σήματος που λαμβάνονται με χρήση μιας συνάρτησης παραθύρου. Αν το σήμα είναι περιοδικό, θα εμφανίσει κορυφές που θα αντιστοιχούν στην θεμελιώδη συχνότητα και στις αρμονικές του (πολλαπλάσια αυτής). Στόχος των αλγορίθμων του πεδίου της συχνότητας είναι να διακρίνουν ποια κορυφή του μετασχηματισμού του σήματος αντιστοιχεί στη θεμελιώδη συχνότητα.

Ο HPS αλγόριθμος εντοπίζει την μέγιστη σύμπτωση αρμονικών σύμφωνα με τις παρακάτω εξισώσεις:

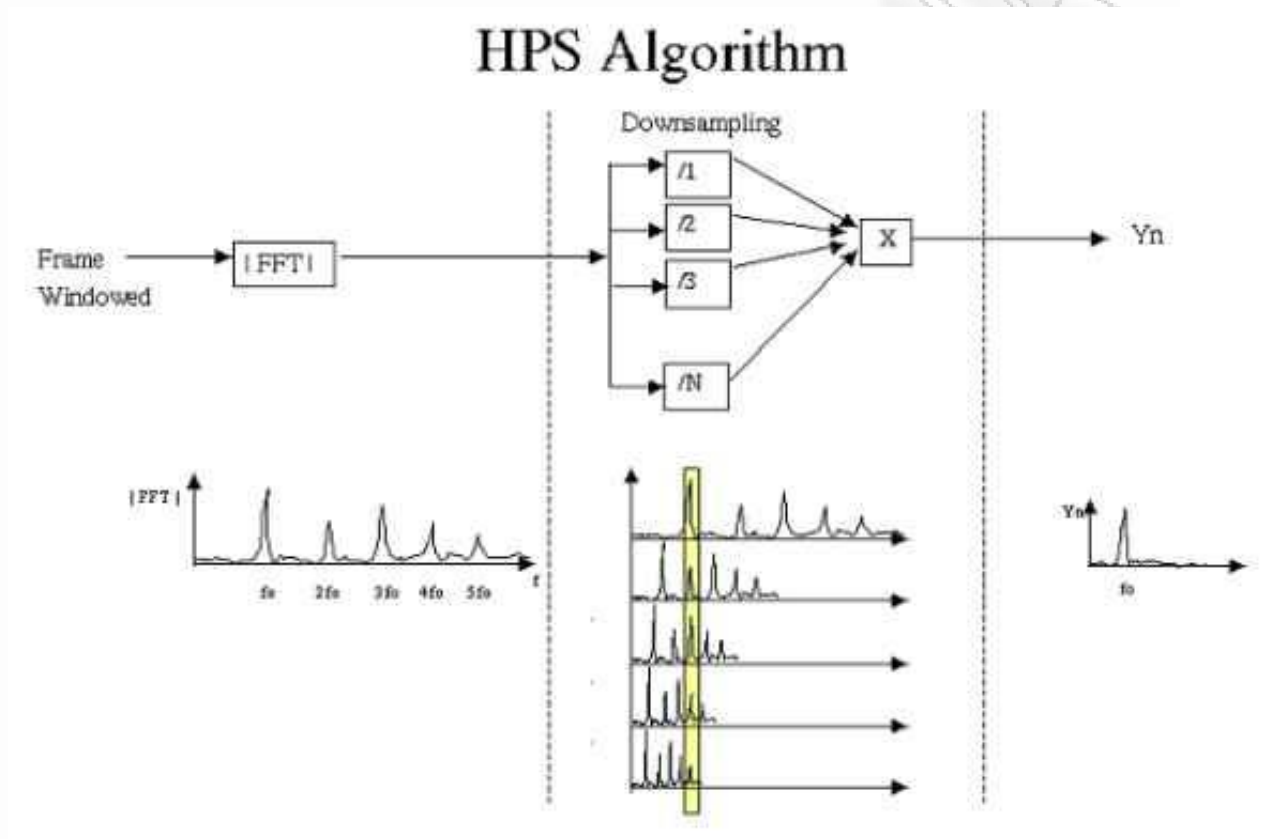
$$Y(\omega) = \prod_{r=1}^R |X(\omega r)|$$

$$\hat{Y} = \max_{\omega_i} \{Y(\omega_i)\}$$

Έτσι, για κάθε ακολουθία συχνοτήτων που αντιστοιχεί σε ένα παράθυρο του σήματος, παίρνουμε το γινόμενο R πολλαπλασίων της κορυφής που αντιστοιχεί στη συχνότητα  $\omega$ . Αυτό επιτυγχάνεται συνήθως με downsampling της ακολουθίας το οποίο προκαλεί συμπίεση αυτής και κατά συνέπεια ευθυγραμμίζει τις κορυφές που αντιστοιχούν στις αρμονικές. Με



πολλαπλασιασμό των επιμέρους συμπιεσμένων ακολουθιών, προκύπτει, ως ξεκάθαρο μέγιστο, η κορυφή που αντιστοιχεί στη θεμελιώδη συχνότητα.



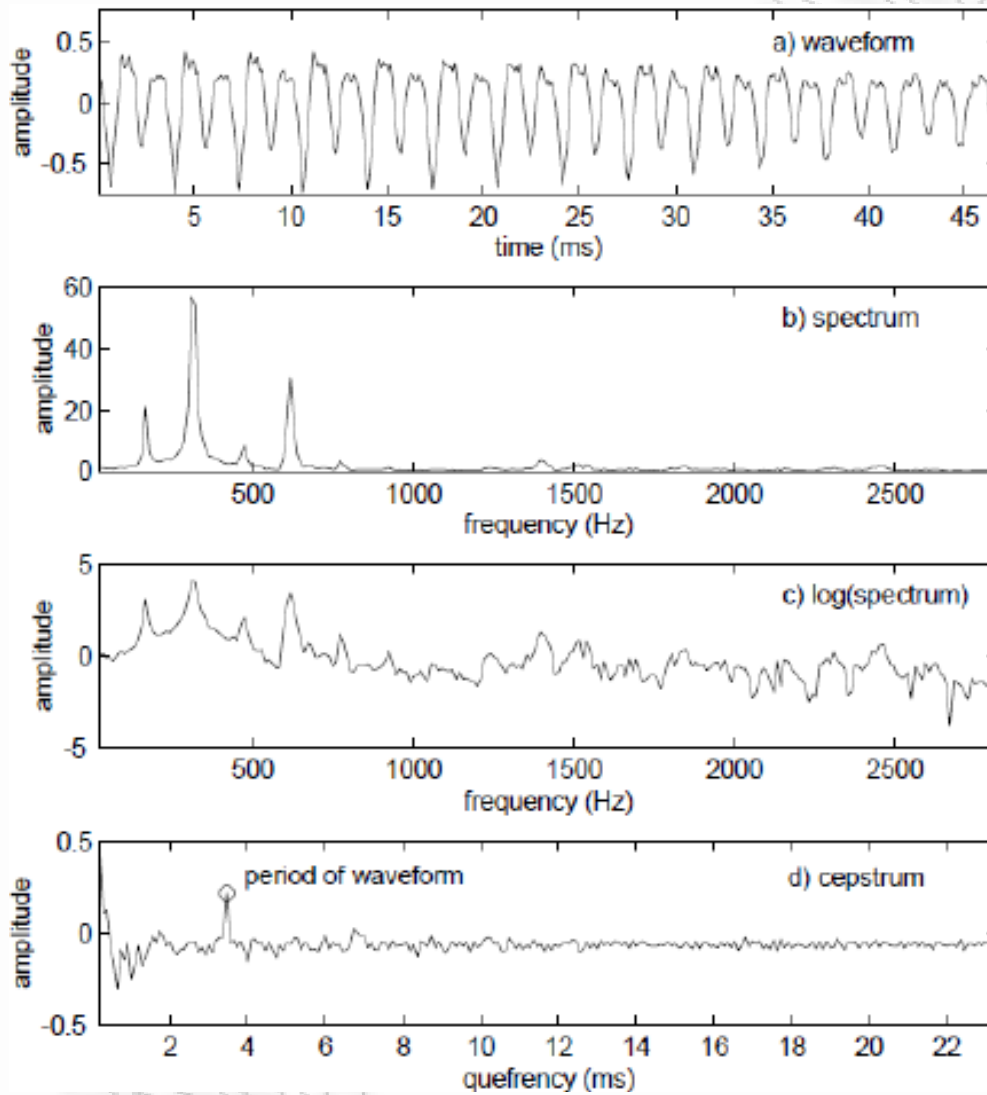
**Εικόνα 16: Αλγόριθμος Harmonic Product Spectrum**

Ο HPS αλγόριθμος είναι υπολογιστικά ελαφρύς και απλός στην υλοποίησή του. Επιπλέον εμφανίζει ανοχή στο θόρυβο και δουλεύει καλά για μεγάλο εύρος σημάτων. Τα αρνητικά του περιλαμβάνουν την συνήθη αναγνώριση της διπλάσιας συχνότητας της θεμελιώδους και την μειωμένη ανάλυση στην χαμηλή συχνότητα περιοχή. Η ανάλυση μπορεί να βελτιωθεί με εφαρμογή μετασχηματισμού Fourier περισσότερων σημείων, κάτι που όμως αυξάνει τον υπολογιστικό φόρτο.

- Cepstrum analysis

Η βασική ιδέα της ανάλυσης cepstrum είναι η εφαρμογή ενός δεύτερου μετασχηματισμού Fourier στο δεκαδικό λογάριθμο της αρχικής συχνοτικής ακολουθίας. Αν το αρχικό σήμα είναι περιοδικό, η ανάλυση Fourier αυτού θα είναι επίσης περιοδική και η εφαρμογή του δεύτερου

μετασχηματισμού θα εμφανίσει μια κορυφή στην περίοδο της συχνότητας. Με τον τρόπο αυτόν απομονώνεται η θεμελιώδης συχνότητα του αρχικού σήματος.

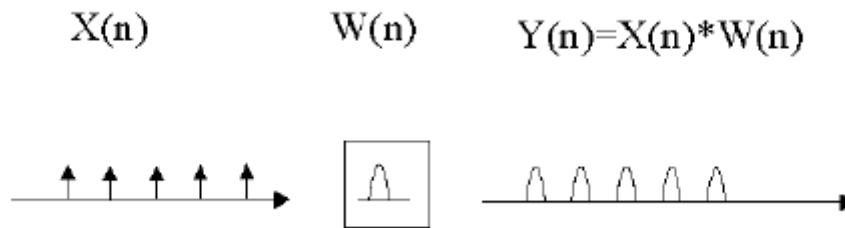


Εικόνα 17: Ανάλυση Cepstrum

Η ανάλυση cepstrum αποτελεί τον πιο διαδεδομένο αλγόριθμο ανίχνευσης pitch στην ανάλυση ομιλίας, καθώς παρέχει αξιοπιστία και ανοχή στον θόρυβο.

- Frequency Domain Maximum Likelihood

Ο αλγόριθμος καλύτερης πιθανότητας του πεδίου της συχνότητας προτάθηκε από τον Noll το 1969 και αναζητά, μέσα από ένα σύνολο ιδεατών συχνοτικών ακολουθιών, αυτήν που ταιριάζει καλύτερα με την συχνοτική ακολουθία που προκύπτει από τον μετασχηματισμό του σήματος εισόδου. Η κάθε ιδεατή συχνοτική ακολουθία προκύπτει από τη συνέλιξη μιας ακολουθίας κρουστικών αποκρίσεων που ξεκινά από την συχνότητα  $\omega$ , με το φάσμα του παραθύρου του εξεταζόμενου σήματος εισόδου.



Εικόνα 18: Frequency Domain Maximum Likelihood

Ο αλγόριθμος έχει σαν στόχο να ελαχιστοποιήσει το σφάλμα μεταξύ του φάσματος εισόδου ( $Y$ ) και των ιδεατών φασματικών ακολουθιών με θεμελιώδη συχνότητα  $\omega$  ( $Y_\omega$ ), όπως αυτό ορίζεται από την παρακάτω εξίσωση,

$$\begin{aligned} E(\omega) &= \|Y - \tilde{Y}_\omega\|^2 \\ &= \|Y\|^2 + \|\tilde{Y}_\omega\|^2 - 2Y\tilde{Y}_\omega^T \end{aligned}$$

$$\hat{Y} = \min_{\omega} \{E(\omega)\} = \max_{\omega} \{Y\tilde{Y}_\omega^T\}$$

Η μέθοδος της καλύτερης πιθανότητας έχει σχετικά μικρή υπολογιστική πολυπλοκότητα καθώς επιτρέπει μικρότερα μεγέθη μετασχηματισμών, όφελος όμως που αντισταθμίζεται από μειωμένη ανάλυση στην ανίχνευση. Επιπρόσθετα, είναι περισσότερο επιρρεπής στο θόρυβο από τις περισσότερες μεθόδους που παρουσιάστηκαν. Τέλος επειδή βασίζεται σε ένα σύνολο δεδομένων σταθερών συχνοτικών ακολουθιών, δεν λειτουργεί καλά όταν η συχνότητα του εισερχόμενου σήματος βρίσκεται στο ενδιάμεσο αυτών. Για τον λόγο αυτόν ο αλγόριθμος προτιμάται σε πηγές που διαθέτουν κβαντισμένες στη συχνότητα εξόδους (πληκτροφόρα όργανα, ξύλινα πνευστά).

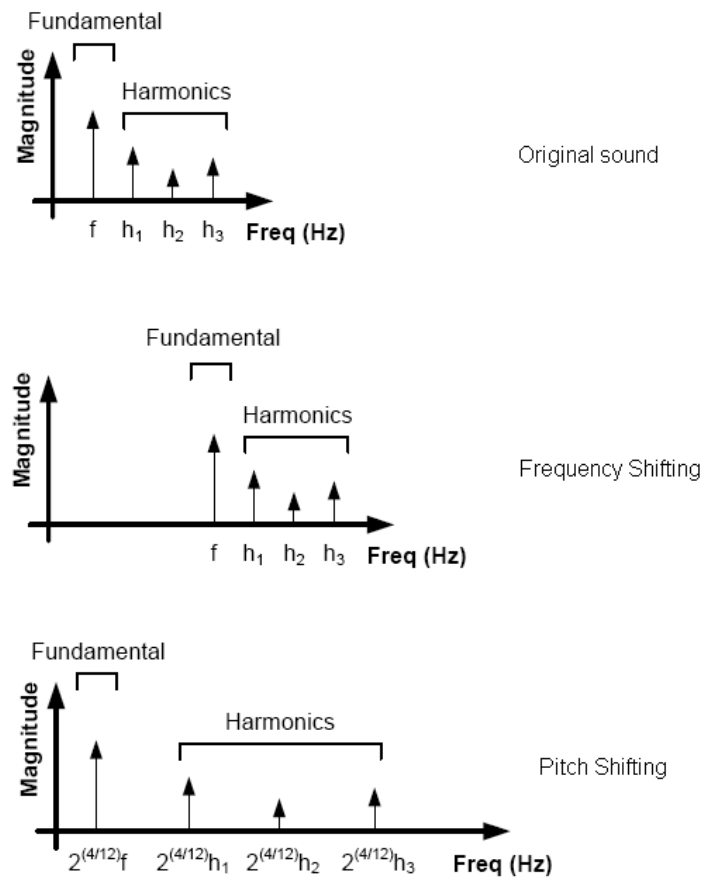
- Neural Network Based Detection

Η ανίχνευση του pitch μπορεί να πραγματοποιηθεί και με χρήση στατιστικών μεθόδων (χρήση και εκπαίδευση ενός κατάλληλου νευρωνικού δικτύου). Τα νευρωνικά δίκτυα προσπαθούν να μοντελοποιήσουν και να μιμηθούν την δομή και τη λειτουργία του ανθρώπινου εγκεφάλου. Ένα νευρωνικό δίκτυο είναι μια διάταξη που αποτελείται από κόμβους και συνδέσεις μεταξύ αυτών στις οποίες αντιστοιχούν μεταβαλλόμενα βάρη. Σε κάθε κόμβο τα εισερχόμενα σήματα πολλαπλασιασμένα με τα αντίστοιχα βάρη προστίθενται και, αν ικανοποιούν μια συγκεκριμένη συνάρτηση μεταφοράς μεταφέρονται στο επόμενο επίπεδο κόμβων.

Πριν η διάταξη του δικτύου τεθεί σε λειτουργία, εφαρμόζεται μια διαδικασία εκπαίδευσης του δικτύου. Αυτό δέχεται εισόδους ενώ ταυτόχρονα του προτείνεται και η ορθή έξοδος. Μέσω της διαδικασίας αυτής, τα βάρη στις συνδέσεις μεταξύ των κόμβων αναπροσαρμόζονται έτσι ώστε η έξοδος του δικτύου να ταυτίζεται με την δεδομένη ορθή έξοδο. Με την διαδικασία αυτήν στοχεύουμε στο να προγραμματίσουμε τη λειτουργικότητα που επιθυμούμε στο δίκτυο, με σκοπό να την διατηρήσει και εκτός του συνόλου εισόδων εκπαίδευσης. Αν και συχνά τέτοιες διατάξεις λειτουργούν, δεν συντελούν στην κατανόηση του προβλήματος και δεν προσφέρουν την βεβαιότητα ότι το δίκτυο θα λειτουργήσει σε κάθε περίπτωση, ειδικά όταν το στάδιο εκπαίδευσης του δικτύου είναι περιορισμένο.

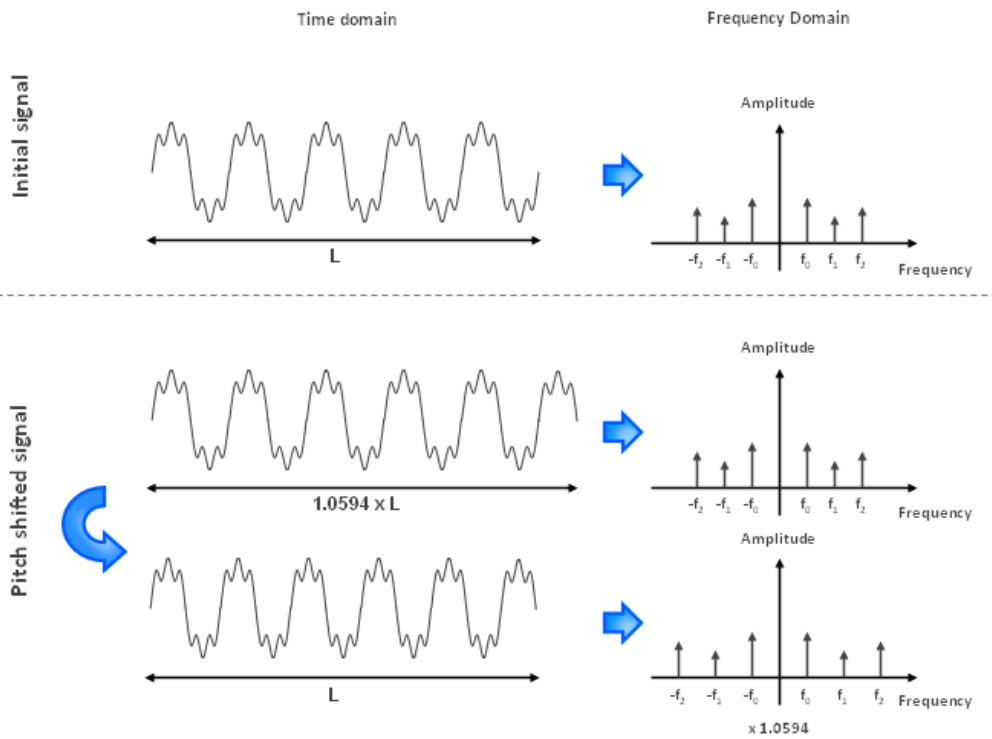
### 1.2.3 Αλγόριθμοι διόρθωσης του pitch

Η διόρθωση του pitch γίνεται με χρήση ενός pitch shift αλγορίθμου. Οι αλγόριθμοι αυτοί έχουν σαν στόχο την μετατόπιση της θεμελιώδους συχνότητας του επεξεργαζόμενου ήχου. Είναι σημαντικό να γίνει διάκριση μεταξύ pitch shifting και frequency shifting. Κατά την διαδικασία του frequency shifting μετατοπίζουμε το συχνотικό περιεχόμενο ενός ήχου, ενώ κατά την διαδικασία του pitch shifting το διαστέλλουμε. Η σημαντική διαφορά έγκειται στο γεγονός ότι το Pitch shifting διατηρεί τη σχέση μεταξύ των αρμονικών (οι οποίες πρέπει πάντα να είναι ακέραια πολλαπλάσια της θεμελιώδους συχνότητας). Η διάκριση γίνεται αντιληπτή από την εικόνα 19.

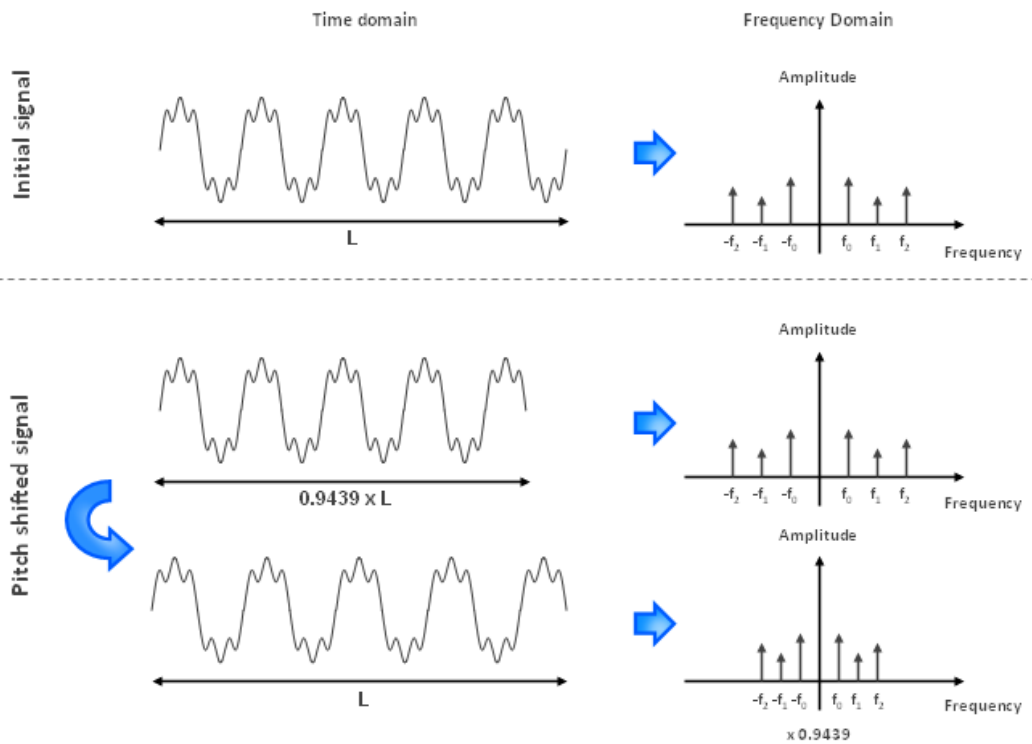


**Εικόνα 19: Διάκριση Pitch / Frequency Shifting**

Η πλειονότητα των pitch shifting αλγορίθμων βασίζονται στην τεχνική της αλλαγής ρυθμού δειγματοληψίας (resampling). Σύμφωνα με την τεχνική αυτήν, αν ένας ήχος έχει ηχογραφηθεί με ρυθμό δειγματοληψίας  $\alpha$ , και αναπαραχθεί με ρυθμό δειγματοληψίας  $\beta$ , θα μετατοπισθεί η θεμελιώδης συχνότητά του (ανάλογα με το λόγο  $\alpha/\beta$ ) αλλά ταυτόχρονα θα μεταβληθεί και η διάρκειά του. Το φαινόμενο αυτό ονομάζεται time / pitch scaling. Από τα παραπάνω διαπιστώνουμε ότι η αλλαγή ρυθμού δειγματοληψίας από μόνη της αποτελεί μη αποδεκτή λύση. Για να διατηρηθεί η διάρκεια του σήματος και ταυτόχρονα να μετατοπιστεί η θεμελιώδης συχνότητά του απαιτείται να γίνει χρονική συμπίεση ή επέκταση (time compression / expansion). Με τον τρόπο αυτόν μεταβάλλουμε την διάρκεια του σήματος χωρίς να επηρεάσουμε το pitch του, και στη συνέχεια εφαρμόζουμε resampling, με σκοπό να το μετατοπίσουμε στην επιθυμητή τιμή. Η διαδικασία του time expansion και time compression απεικονίζεται στις εικόνες 20 και 21.



Εικόνα 20: Time Expansion



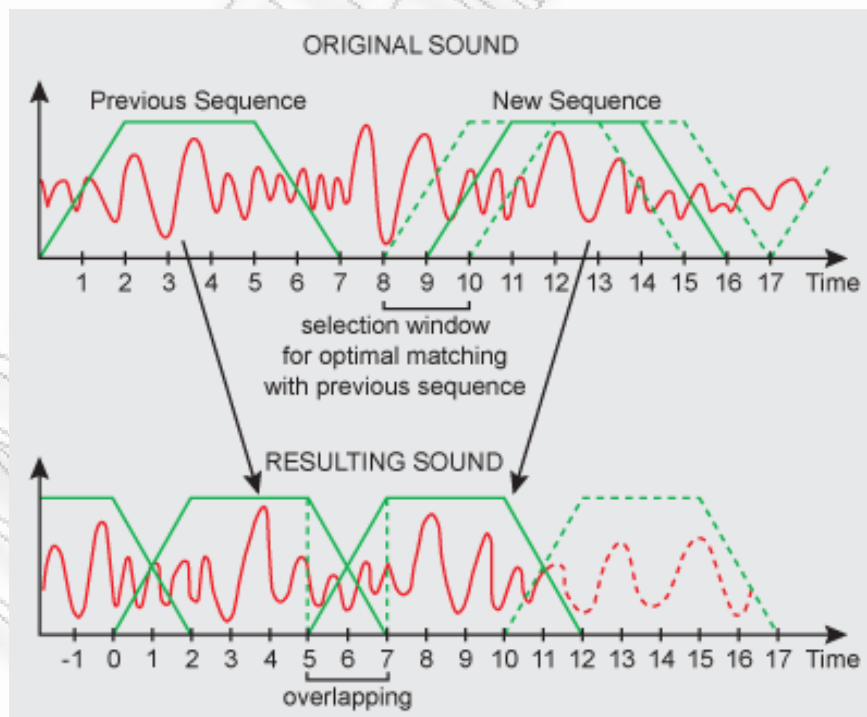
Εικόνα 21: Time Compression

- Synchronous OverLap Add (SOLA)

Ο αλγόριθμος SOLA λειτουργεί διαιρώντας τα ηχητικά δεδομένα σε μικρές ακολουθίες και στη συνέχεια ενώνοντας τα ξανά αγνοώντας ή επαναλαμβάνοντας δεδομένα μεταξύ των ακολουθιών, ανάλογα με το αν στοχεύουμε στο να συμπιέσουμε ή να διαστέλλουμε το χρόνο αναπαραγωγής. Το μεγαλύτερο πρόβλημα που δημιουργείται από την προαναφερθείσα διαδικασία είναι η εμφάνιση ασυνεχειών στα σημεία ένωσης των διαδοχικών ακολουθιών, οι οποίες γίνονται αντιληπτές από τον ακροατή ως παραμόρφωση και υποβαθμίζουν την ποιότητα του σήματος. Για την εξάλειψη των ασυνεχειών η ένωση των ακολουθιών γίνεται με σταδιακή επικάλυψη και πρόσθεσή τους. Με τον τρόπο αυτόν εξασφαλίζεται ομαλή μετάβαση μεταξύ των επιμέρους ακολουθιών και η τελική τους σύνθεση σε ένα ομαλό σήμα εξόδου.

Η απλούστερη υλοποίηση του SOLA αλγορίθμου επιλέγει τυχαίο μέγεθος ακολουθίας και επιλογή της σε ίσα διαστήματα από το σήμα εισόδου. Η υλοποίηση αυτή όμως δεν παράγει καλά αποτελέσματα καθώς τα τμήματα των ακολουθιών που επικαλύπτονται, παρά την όποια εξομάλυνση γίνεται, δημιουργούν ασυνέχειες και υποβαθμίζουν την ποιότητα του ήχου. Το πρόβλημα γίνεται πιο έντονο όσο μεγαλώνει ο λόγος συμπίεσης/ διαστολής του χρόνου.

Για ικανοποιητικά αποτελέσματα, πρέπει η επιλογή των ακολουθιών να γίνει με τέτοιο τρόπο ώστε οι γειτονικές ακολουθίες να μοιάζουν όσο το δυνατόν περισσότερο στα τμήματα που θα επικαλυφθούν. Για να γίνει αυτό πρέπει η επιλογή της κάθε ακολουθίας να γίνει από ένα σύνολο υποψήφιων ακολουθιών εντός ενός παραθύρου επιλογής, έτσι ώστε να επιλεγεί ακολουθία που θα πληρεί τις προδιαγραφές ομοιότητας που έχουν τεθεί. Ο πιο συνηθισμένος τρόπος επιλογής είναι ο υπολογισμός της συσχέτισης (cross-correlation) μεταξύ της προηγούμενης και των πιθανών επόμενων ακολουθιών που βρίσκονται στο παράθυρο επιλογής. Εφόσον λοιπόν υπάρχει σχετική ομοιότητα, η διαδικασία επικάλυψης και πρόσθεσης δεν παράγει σημαντικές ασυνέχειες.



**Εικόνα 22: Αλγόριθμος Synchronous OverLap Add**

Ο αλγόριθμος SOLA δίνει ικανοποιητικά αποτελέσματα για μικρούς λόγους μεταβολής του χρόνου. Σε πολύπλοκες και πολυφωνικές κυματομορφές δίνει καλό αποτέλεσμα για αύξηση

/ μείωση μέχρι 15%. Χαρακτηριστικές παρενέργειές του είναι η παρουσία της αίσθησης του βάθους στο παραγόμενο αποτέλεσμα (reverberating effect) και η ελαφριά διαταραχή της αίσθησης του ρυθμού. Συγχρονισμός των ακολουθιών στο ρυθμικό μοτίβο του ήχου και στην θεμελιώδη συχνότητά του (εφόσον υπάρχει) ελαττώνουν τα παραπάνω προβλήματα. Για απλούστερους μονοφωνικούς και περιοδικούς ήχους το αποδεκτό time compression / expansion μπορεί να φτάσει μέχρι και το 100% του χρόνου.

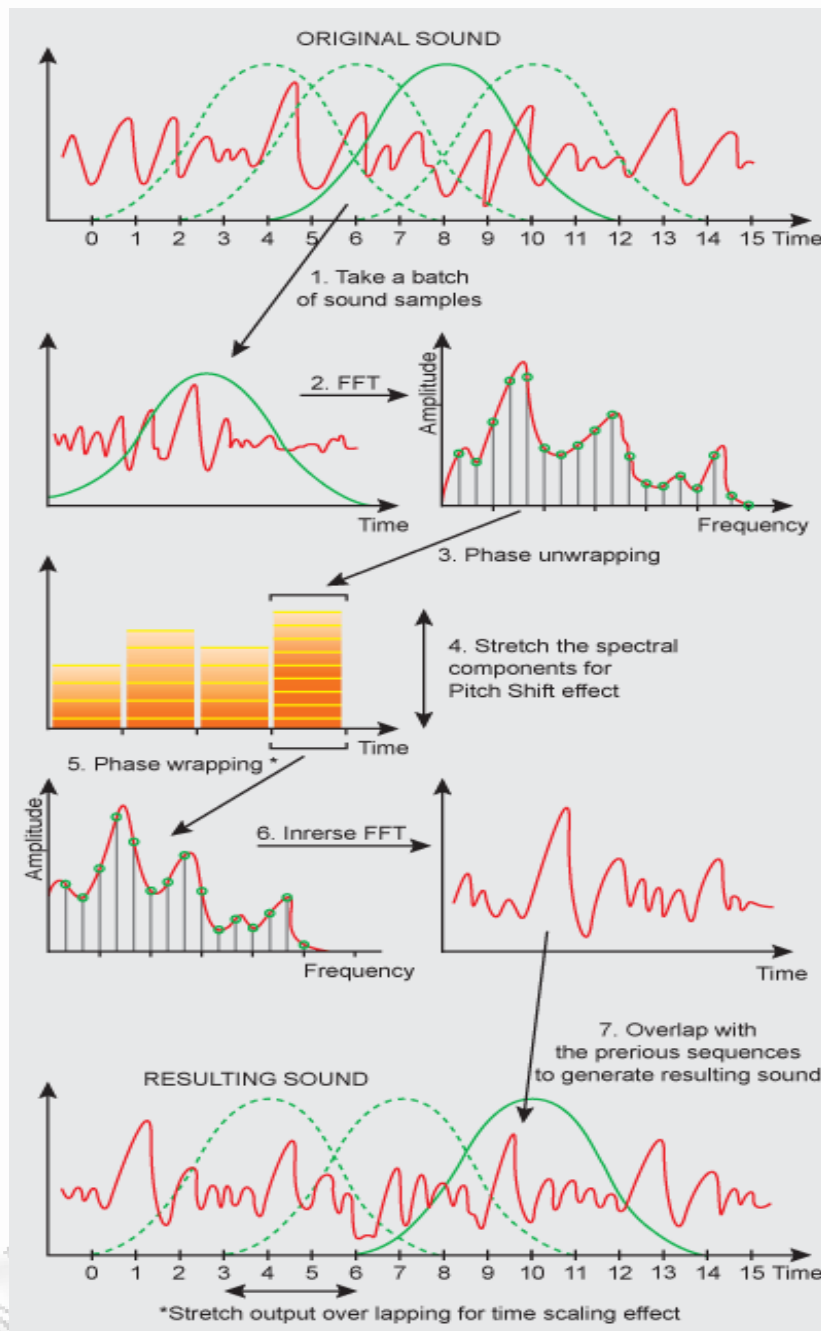
- Time Domain Harmonic Scaling (TDHS)

Ο TDHS αλγόριθμος αποτελεί παραλλαγή του SOLA αλγορίθμου και προτάθηκε από τους Rabiner και Schafer το 1978. Η βασική του ιδέα έγκειται στο γεγονός ότι εφόσον γνωρίζουμε ή μπορούμε να υπολογίσουμε την θεμελιώδη συχνότητα ενός σήματος (μέσω εφαρμογής ενός pitch detection αλγορίθμου), μπορούμε να επιλέξουμε κατάλληλα τις SOLA ακολουθίες ώστε τα επικαλυπτόμενα τμήματά τους να αποτελούν ακέραια πολλαπλάσια της συχνότητας αυτής. Με τον τρόπο αυτόν εξασφαλίζουμε πολύ μεγάλη ομοιότητα μεταξύ τους και εξαλείφουμε την πιθανότητα εμφάνισης ασυνεχειών.

- Phase Vocoder

Ο αλγόριθμος Phase Vocoder προτάθηκε από τους Flanagan και Golden το 1966 και υλοποιήθηκε ψηφιακά από τον Portnoff το 1976. Ο αλγόριθμος λειτουργεί στο πεδίο της συχνότητας και χρησιμοποιεί τον μετασχηματισμό Fourier (FFT) για την μετατροπή. Όπως παρουσιάζεται και στην εικόνα 23 ο αλγόριθμος αποτελείται από τρία στάδια επεξεργασίας: το στάδιο της ανάλυσης κατά Fourier σε ακολουθία πλάτους / συχνότητας, το στάδιο της μετατροπής της ακολουθίας και το στάδιο σύνθεσης μέσω του αντίστροφου μετασχηματισμού σε δείγματα στο πεδίο του χρόνου.





**Εικόνα 23: Αλγόριθμος Phase Vocoder**

Στο στάδιο της ανάλυσης το σήμα εισόδου διαιρείται σε επικαλυπτόμενα τμήματα, σε κάθε ένα από τα οποία εφαρμόζεται μια συνάρτηση παραθύρου (συνήθως παράθυρο Hann). Στη συνέχεια σε κάθε τμήμα εφαρμόζεται ταχύς μετασχηματισμός Fourier. Ο μετασχηματισμός Fourier μετατρέπει τα δεδομένα σε μία ακολουθία τιμών φάσης και πλάτους, κατανεμημένων σε συχνοτικές περιοχές.

Στο επόμενο στάδιο εφαρμόζεται η διαδικασία του ξεδιπλώματος της φάσης (phase unwrapping) κατά την οποία οι πληροφορίες φάσης και πλάτους ανά συγκεκριμένες συχνοτικές περιοχές μετατρέπονται σε ακριβή δεδομένα πλάτους συχνότητας, χρησιμοποιώντας δεδομένα της φάσης από την επεξεργασία των προηγούμενων παραθύρων. Τα δεδομένα αυτά μπορούμε πλέον να τροποποιήσουμε για να μεταβάλλουμε την χρονική διάρκεια ή κατεύθυνση το pitch.

Στο στάδιο της σύνθεσης εφαρμόζουμε τους αντίστροφους μετασχηματισμούς phase unwrapping και Fourier. Αν χρησιμοποιήσουμε διαφορετική αναλογία επικάλυψης των παραθύρων από την αρχική, επιτυγχάνουμε time compression/expansion. Αν όχι ανάλογα με τις μετατροπές του δεύτερου σταδίου μπορούμε να μεταβάλλουμε το pitch.

Για να επιτύχουμε βέλτιστα αποτελέσματα πρέπει να πάρουμε μετασχηματισμό Fourier μεγάλου μεγέθους (4096 ή 8192 samples) και τα τμήματα στα οποία θα διαιρέσουμε την ακολουθία θα πρέπει να παρουσιάζουν μεγάλο ποσοστό επικάλυψης. Αν στα παραπάνω συνυπολογίσουμε και την πολυπλοκότητα του phase unwrapping σταδίου, στο οποίο απαιτούνται μαθηματικού υπολογισμοί κινητής υποδιαστολής, συμπεραίνουμε ότι ο αλγόριθμος phase vocoder είναι αρκετά απαιτητικός υπολογιστικά.

Το σημαντικότερο όφελος του phase vocoder αλγορίθμου είναι η έλλειψη του reverberating φαινομένου που εμφανίζεται στις time domain προσεγγίσεις. Παρόλα αυτά ο αλγόριθμος προκαλεί παραμόρφωση φάσης η οποία στερεί την οξύτητα και την καθαρότητα του ήχου, ιδίως σε σήματα με έντονο attack (ξεκίνημα).

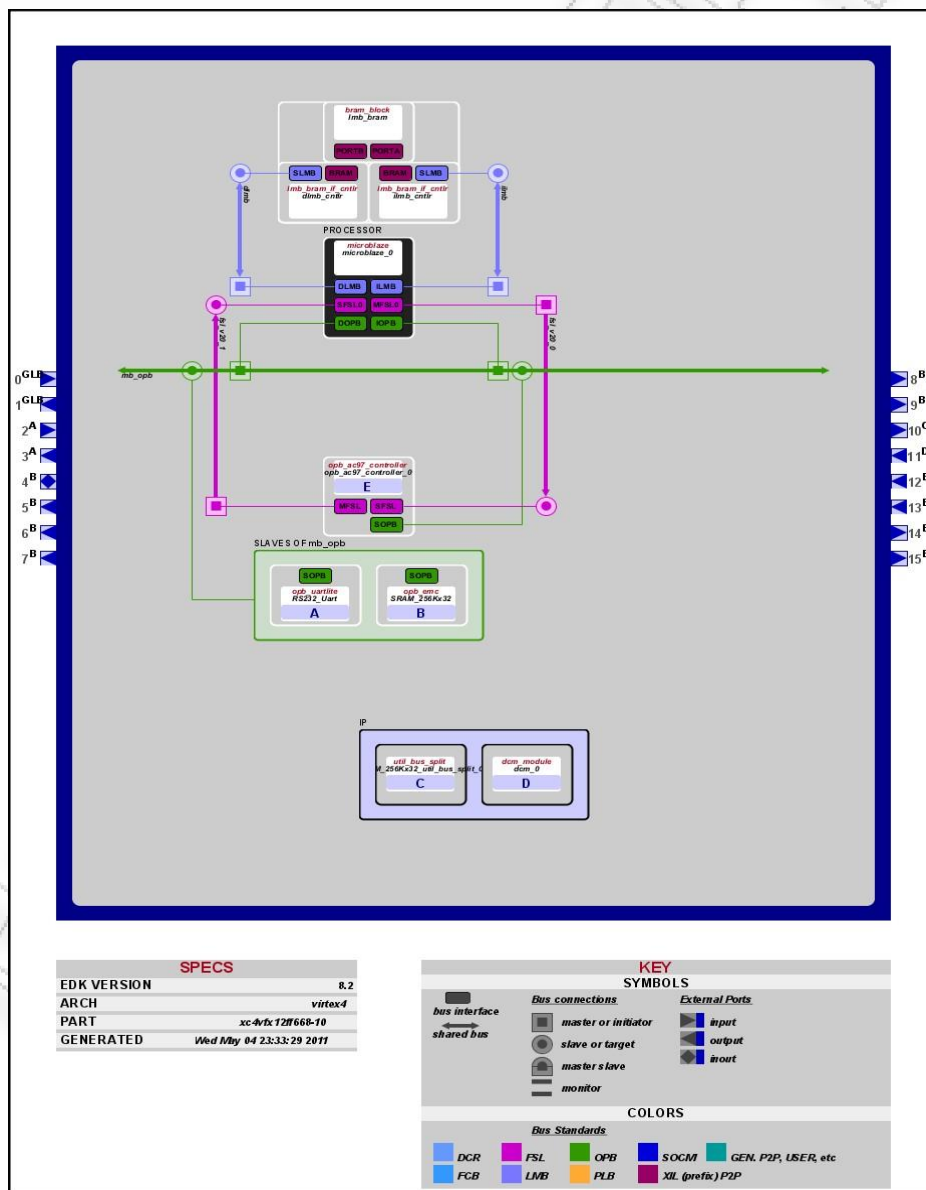
#### 1.2.4 Εφαρμογές της ανίχνευσης και διόρθωσης του pitch

Η ανάπτυξη αξιόπιστων αλγορίθμων ανίχνευσης και μεταβολής του pitch είναι χρήσιμη σε μεγάλο εύρος εφαρμογών. Η εκτίμηση του pitch χρησιμοποιείται στην επεξεργασία ομιλίας και βοηθάει στην επίλυση προβλημάτων όπως η αναγνώριση ομιλίας μέσω ταυτοποίησης φθόγγων και αναγνώριση της προσωδίας (μετρικά και ακουστικά χαρακτηριστικά της γλώσσας). Μπορούν επίσης μέσω αυτής να εκτιμηθούν και άλλα χαρακτηριστικά του προφορικού λόγου όπως το ύφος και η διαφορά ερώτησης/κατάφασης. Επιπρόσθετα δίνει τη δυνατότητα αναγνώρισης λεξικών κατηγοριών σε τονικές γλώσσες. Η διόρθωση / μεταβολή του pitch βρίσκει εφαρμογές στη σύνθεση τεχνητής ομιλίας, με σκοπό την παραγωγή όσο το δυνατόν φυσικότερου λόγου από τη μηχανή. Με τον τρόπο αυτόν διευκολύνεται η αλληλεπίδραση ανθρώπου / μηχανής με χρήση της γλώσσας, ενός μέσου απόλυτα φυσικού προς τον χρήστη.

Ένα άλλο ευρύτατο πεδίο χρήσης των προαναφερθέντων μεθόδων είναι οι μουσικές / καλλιτεχνικές εφαρμογές. Η ανίχνευση του pitch έχει καταστήσει εφικτές εφαρμογές όπως η μεταγραφή της μουσικής από ένα όργανο (ή τη φωνή) σε παρτιτούρα σε πραγματικό χρόνο. Μέσω τέτοιων μεθόδων μάλιστα είναι δυνατή η αποτύπωση μελωδιών με μικροτονικά χαρακτηριστικά (ιαπωνική / μεσανατολική μουσική), κάτι που δεν είναι δυνατόν να γίνει με χρήση της καθιερωμένης μουσικής σήμανσης. Άλλο ένα πεδίο που ευνοείται από την ανάπτυξη αξιόπιστων και αποδοτικών αλγορίθμων ανίχνευσης της θεμελιώδους συχνότητας είναι η μετατροπή ήχου σε σήματα MIDI. Το MIDI (Musical Instrument Digital Interface) είναι ένα ευρύτατα διαδεδομένο πρωτόκολλο διασύνδεσης ψηφιακών μουσικών οργάνων και συσκευών ελέγχου. Η μετατροπή οποιουδήποτε μουσικού ηχητικού σήματος σε δεδομένα MIDI σε πραγματικό χρόνο προσφέρει πρωτόγνωρες καλλιτεχνικές δυνατότητες έκφρασης στον χρήστη / μουσικό ενώ ταυτόχρονα διευκολύνει την αλληλεπίδρασή του με τα σύγχρονα ψηφιακά συστήματα παραγωγής μουσικής. Η χρήση του πρωτοκόλλου MIDI σε συνδυασμό με τις δυνατότητες μεταβολής/ διόρθωσης του pitch του οργάνου δίνουν στον μουσικό απόλυτο έλεγχο πάνω στην καλλιτεχνική δημιουργία. Τέλος η μηχανική αναγνώριση του pitch, πληροφορίας που μέχρι τώρα ήταν προνόμιο της ανθρώπινης αντίληψης, οδηγεί σε αυτοματοποιημένη επεξεργασία ηχητικών δεδομένων μέσω δημιουργίας και αποθήκευσης μεταδεδομένων για χρήση σε υπηρεσίες αυτόματης αναζήτησης περιεχομένου.

## Κεφάλαιο 2: Περιγραφή Υλικού

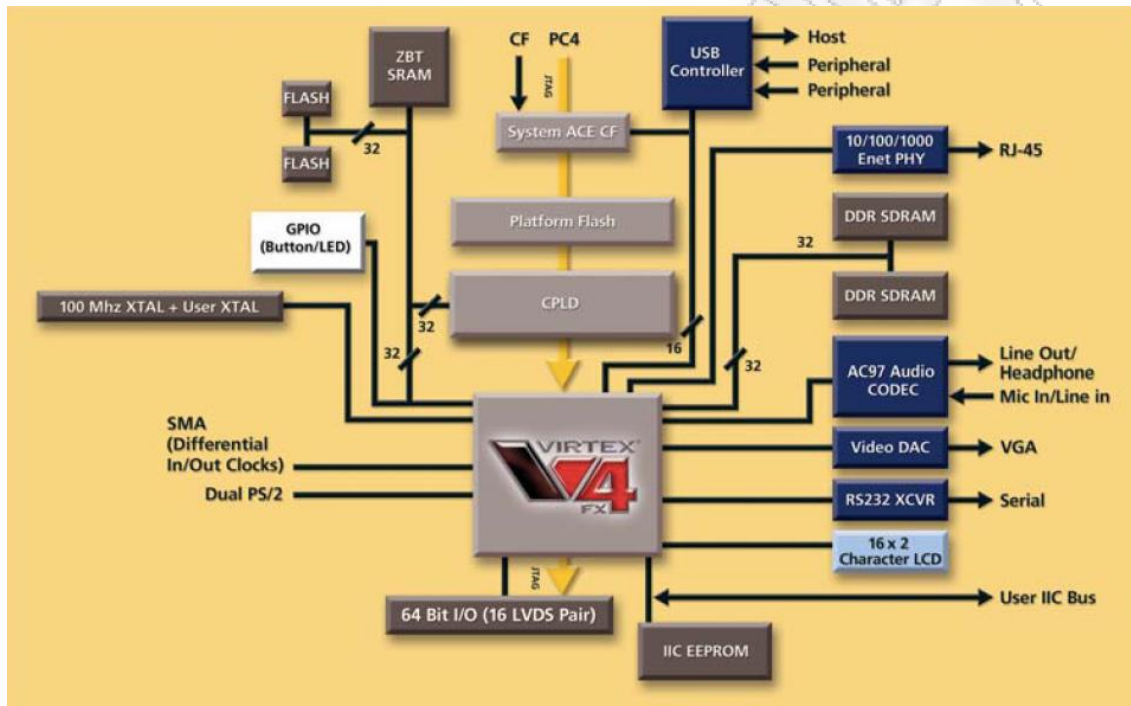
Στο παρόν κεφάλαιο θα ασχοληθούμε με την ανάπτυξη και την περιγραφή του υλικού, πάνω στο οποίο δομήθηκε η εφαρμογή ανίχνευσης και διόρθωσης θεμελιωδών συχνοτήτων μουσικών σημάτων. Η ανάπτυξη ενός σταθερού και ευέλικτου υλικού αποτελεί σημείο κλειδί για την ανάπτυξη αποδοτικών ενσωματωμένων συστημάτων. Ο σωστός και έγκαιρος καθορισμός προδιαγραφών για την εφαρμογή θα οδηγήσει στην ανάπτυξη του σκελετού του ενσωματωμένου συστήματος, ο οποίος θα αποτελέσει πλατφόρμα εκτέλεσης, παροχής πόρων απλά και εύρεσης διεξόδων από το λογισμικό. Το υλικό του ενσωματωμένου συστήματος βρίσκεται στην πλειοψηφία του μέσα στην συσκευή FPGA. Το FPGA περιλαμβάνει τον επεξεργαστή, την μνήμη εντολών και δεδομένων για την εκτέλεση του κώδικα, τον ελεγκτή του AC'97 Audio Codec, τον ελεγκτή της εξωτερικής μνήμης RAM, καθώς και τους διαύλους διασύνδεσης αυτών OPB, FSL και LMB. Εκτός του FPGA βρίσκεται το ολοκληρωμένο κύκλωμα LM4550 (AC '97 Codec) και η μνήμη SRAM 512k x 16 bit. Η δομή του υλικού παρουσιάζεται στην εικόνα 24



Εικόνα 24: Δομή του υλικού του ενσωματωμένου συστήματος

## 2.1 Αναπτυξιακή Πλακέτα Xilinx ML403

Η ανάπτυξη του συστήματος έγινε εξολοκλήρου στην αναπτυξιακή πλακέτα ML403 της εταιρίας Xilinx. Η πλακέτα αποτελεί ιδανική πλατφόρμα για σχεδίαση και γρήγορη πρωτοτυποποίηση ενσωματωμένων συστημάτων, καθώς παρέχει πλήθος επιλογών περιφερειακών συσκευών και τρόπων διασύνδεσης με άλλα ψηφιακά συστήματα. Η δομή της πλακέτας παρουσιάζεται στην εικόνα 25



Εικόνα 25: Δομή της πλακέτας ML403

Παρατηρούμε ότι στην καρδιά του συστήματος βρίσκεται ένα FPGA της οικογένειας VIRTEX 4 (CX4VFX12). Το FPGA παρέχει δυνατότητες σχεδιασμού ενός ολοκληρωμένου system-on-a-chip ενσωματωμένου συστήματος καθώς περιλαμβάνει έναν ενσωματωμένο hard-wired επεξεργαστή αρχιτεκτονικής Power PC (IBM PPC 405 RISK), έχει την δυνατότητα να φιλοξενήσει σχεδιασμούς βασισμένους στον soft-core εμσωματωμένο επεξεργαστή MicroBlaze, παρέχει αρκετά block μνήμης, κυκλώματα οδήγησης ρολογιού, και πλήθος πυλών και κυκλωμάτων για υλοποίηση custom λογικής. Ταυτόχρονα η πλακέτα προσφέρει πλήθος περιφερειακών συσκευών (μνήμες, δυνατότητες διασύνδεσης) προσφέροντας πραγματική ευελιξία στην ανάπτυξη ενσωματωμένων εφαρμογών.

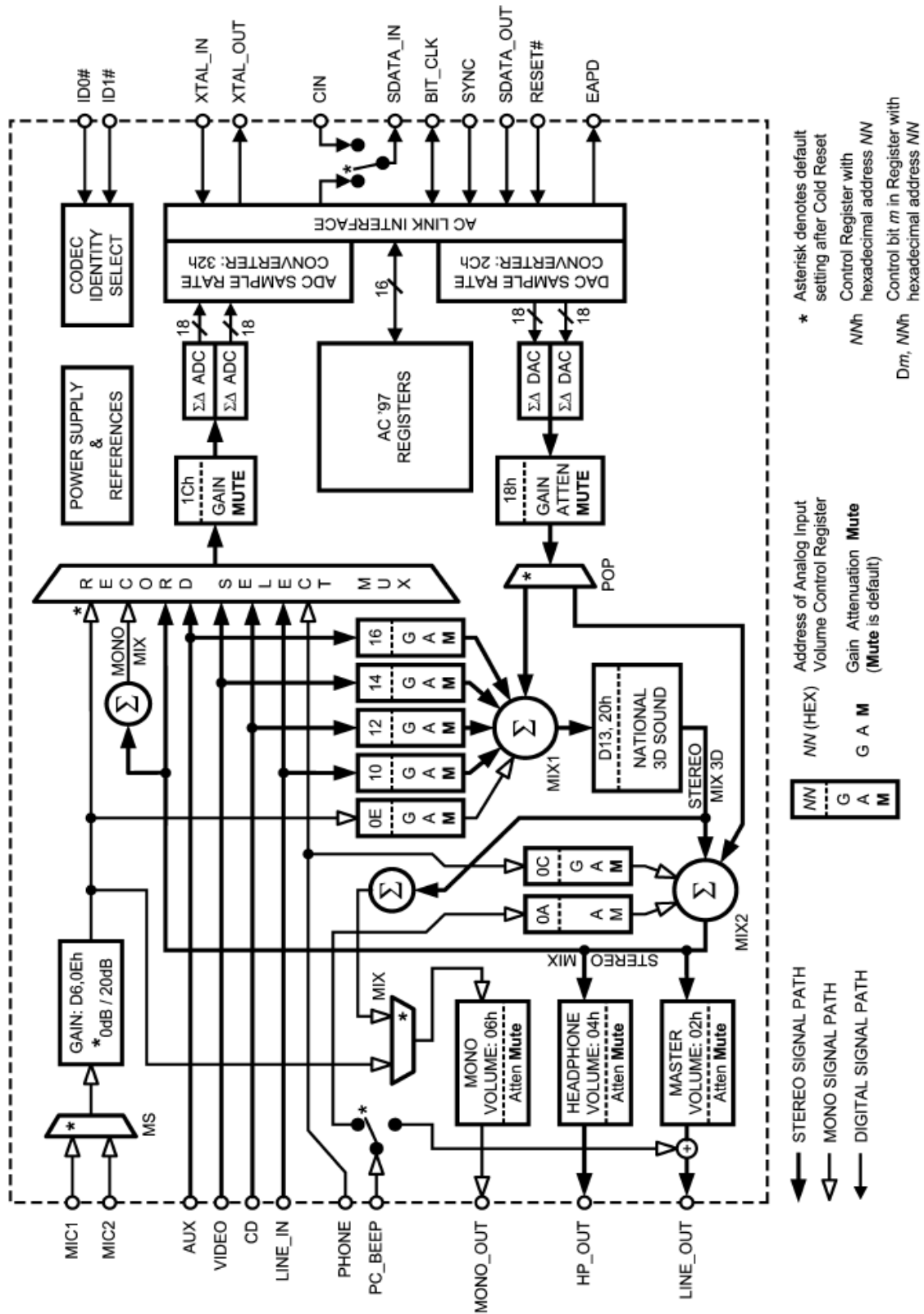
Τέλος η πλακέτα συνοδεύεται και συμπληρώνεται από τα περιβάλλοντα ανάπτυξης της Xilinx, Integrated Software Environment και Embedded Development Kit που ολοκληρώνουν την σχεδιαστική εμπειρία.

## 2.2 AC '97 Audio Codec

Το AC '97 (Audio Codec '97) αποτελεί πρότυπο που αναπτύχθηκε από την Intel το 1997, για την σχεδίαση ολοκληρωμένων κυκλωμάτων χειρισμού ηχητικών σημάτων σε υπολογιστικά συστήματα. Η έννοια Codec προέρχεται από τον συγκερασμό των λέξεων encode / decode και στην συγκεκριμένη περίπτωση χρησιμοποιήθηκε για την περιγραφή της διαδικασίας μετατροπής του ήχου από τον αναλογικό στον ψηφιακό κόσμο. Το AC '97 περιγράφει ένα πλήρως λειτουργικό σύστημα χειρισμού αναλογικών ηχητικών σημάτων, μετατροπής τους σε υψηλής ποιότητας ψηφιακά σήματα και διασύνδεσης με ένα υπολογιστικό περιβάλλον. Το πρότυπο υποστηρίζει λειτουργία σε σταθερό ρυθμό δειγματοληψίας 48kHz για δύο μέχρι έξι κανάλια ήχου (ενώ προβλέπει και καταστάσεις λειτουργίας με ρυθμούς δειγματοληψίας μέχρι 96kHz, με περιορισμό στα δύο κανάλια) και ανάλυση μέχρι 20 bit. Ολοκληρωμένα κυκλώματα σχεδιασμένα σύμφωνα με το πρότυπο AC '97 χρησιμοποιήθηκαν ευρέως στην κατασκευή καρτών ήχου, modems και μητρικών καρτών για προσωπικούς υπολογιστές μέχρι και το 2004, οπότε και το πρότυπο αντικαταστάθηκε από το High Definition Audio (HD Audio) της ίδιας εταιρίας. Το πρότυπο HD Audio βασίζεται στην αρχιτεκτονική του AC '97 αλλά επεκτείνει τις δυνατότητές του προσφέροντας δυνατότητες για υψηλότερο ρυθμό δειγματοληψίας (έως 192kHz), υψηλότερη ανάλυση (32-bit) και αναπαραγωγή μέχρι 8 καναλιών ήχου ταυτόχρονα. Η αναπτυξιακή πλακέτα Xilinx ML403 χρησιμοποιεί το ολοκληρωμένο κύκλωμα LM4550 της εταιρίας National Semiconductors το οποίο είναι σχεδιασμένο σύμφωνα με το πρότυπο AC '97 Revision 2.1.

### 2.2.1 Χαρακτηριστικά και Δομή του LM4550 Audio Codec

Το LM4550 ολοκληρωμένο κύκλωμα, σύμφωνα με την αρχιτεκτονική που ορίζεται στο AC '97 πρότυπο, παρέχει πλήρως διαχωρισμένες αναλογικές και ψηφιακές λειτουργίες στο σύστημα που το χρησιμοποιεί, προσφέροντας με τον τρόπο αυτόν ευελιξία στην σχεδίαση του συστήματος και αύξηση της απόδοσής του. Το κύκλωμα διαθέτει αναλογικό μονοπάτι πλήρως αμφίδρομης λειτουργίας που αποτελείται από στερεοφωνικούς μετατροπείς αναλογικού-σε-ψηφιακό (ADC) και ψηφιακού-σε-αναλογικό (DAC) των 18 bit, καθώς και δύο αναλογικούς μείκτες. Οι μετατροπείς έχουν την δυνατότητα να λειτουργούν με διαφορετικούς ρυθμούς δειγματοληψίας εισόδου / εξόδου, που κυμαίνονται από 4kHz έως 48kHz με ανάλυση 1Hz. Το κύκλωμα υποστηρίζει 4 στερεοφωνικές και 4 μονοφωνικές εισόδους, κάθε μία εκ των οποίων διαθέτει ξεχωριστές λειτουργίες ελέγχου κέρδους ενίσχυσης (gain), εξασθένησης (attenuation) και αποσιώπησης (mute). Οι αναλογικοί μείκτες του κυκλώματος οδηγούν δύο στερεοφωνικές και μία μονοφωνική έξοδο, με ξεχωριστές λειτουργίες ελέγχου εξασθένησης και αποσιώπησης. Τέλος στο αναλογικό τμήμα του κυκλώματος περιλαμβάνεται στερεοφωνικός ενισχυτής ακουστικών ισχύος 50mW στην μία έξοδό του, καθώς και λειτουργία National 3D Sound για εμπλουτισμό του ήχου. Η δομή του κυκλώματος παρουσιάζεται στην εικόνα 26



Εικόνα 26: Δομή του ολοκληρωμένου LM4550

Το LM4550 κύκλωμα υποστηρίζει ταυτόχρονη ροή ηχητικών δεδομένων προς και από το ψηφιακό σύστημα, μέσω του ψηφιακού πρωτοκόλλου AC-Link, για δύο κανάλια εισόδου και δύο κανάλια εξόδου. Ο έλεγχος της ροής των σημάτων στο αναλογικό και το ψηφιακό τμήμα του κυκλώματος γίνεται μέσω αρχείου 27 καταχωρητών των 16-bit οι οποίοι παρουσιάζονται στον πίνακα:

**LM4550 Register Map**

REG	Name	D15	D14	D13	D12	D11	D10	D9	D8	D7	D6	D5	D4	D3	D2	D1	D0	Default
00h	Reset	X	0	0	0	1	1	0	1	0	1	0	1	0	0	0	0	0D50h
02h	Master Volume	Mute	X	X	ML4	ML3	ML2	MIL1	ML0	X	X	X	MR4	MR3	MR2	MR1	MR0	8000h
04h	Headphone Volume	Mute	X	X	ML4	ML3	ML2	MIL1	ML0	X	X	X	MR4	MR3	MR2	MR1	MR0	8000h
06h	Mono Volume	Mute	X	X	X	X	X	X	X	X	X	X	MM4	MM3	MM2	MM1	MM0	8000h
0Ah	PC_Beep Volume	Mute	X	X	X	X	X	X	X	X	X	X	PV3	PV2	PV1	PV0	X	0000h
0Ch	Phone Volume	Mute	X	X	X	X	X	X	X	X	X	X	GN4	GN3	GN2	GN1	GN0	8008h
0Eh	Mic Volume	Mute	X	X	X	X	X	X	X	X	20dB	X	GN4	GN3	GN2	GN1	GN0	8008h
10h	Line In Volume	Mute	X	X	GL4	GL3	GL2	GL1	GL0	X	X	X	GR4	GR3	GR2	GR1	GR0	8808h
12h	CD Volume	Mute	X	X	GL4	GL3	GL2	GL1	GL0	X	X	X	GR4	GR3	GR2	GR1	GR0	8808h
14h	Video Volume	Mute	X	X	GL4	GL3	GL2	GL1	GL0	X	X	X	GR4	GR3	GR2	GR1	GR0	8808h
16h	Aux Volume	Mute	X	X	GL4	GL3	GL2	GL1	GL0	X	X	X	GR4	GR3	GR2	GR1	GR0	8808h
18h	PCM Out Volume	Mute	X	X	GL4	GL3	GL2	GL1	GL0	X	X	X	GR4	GR3	GR2	GR1	GR0	8808h
1Ah	Record Select	X	X	X	X	X	SL2	SL1	SL0	X	X	X	X	X	SR2	SR1	SR0	0000h
1Ch	Record Gain	Mute	X	X	X	GL3	GL2	GL1	GL0	X	X	X	X	GR3	GR2	GR1	GR0	8000h
20h	General Purpose	POP	X	X	X	X	X	MIX	MS	LPBK	X	X	X	X	X	X	X	0000h
22h	3D Control (Read Only)	X	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0101h
24h	Reserved	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	0000h
26h	Powerdown Ctrl/Stat	EAPD	PR6	PR5	PR4	PR3	PR2	PR1	PR0	X	X	X	X	REF	ANL	DAC	ADC	000Xh
28h	Extended Audio ID	ID1	ID0	X	X	X	X	AMAP	0	0	0	X	X	0	X	0	VRA	X201h
2Ah	Extended Audio Control/Status	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	VRA	0000h
2Ch	PCM DAC Rate	SR15	SR14	SR13	SR12	SR11	SR10	SR9	SR8	SR7	SR6	SR5	SR4	SR3	SR2	SR1	SR0	BB80h
32h	PCM ADC Rate	SR15	SR14	SR13	SR12	SR11	SR10	SR9	SR8	SR7	SR6	SR5	SR4	SR3	SR2	SR1	SR0	BB80h
5Ah	Vendor Reserved 1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	0000h
74h	Chain-In Control	X	X	X	X	X	X	X	X	X	X	X	X	X	X	ID1	ID0	000Xh
7Ah	Vendor Reserved 2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	0000h
7Ch	Vendor ID1	0	1	0	0	1	1	1	0	0	1	0	1	0	0	1	1	4E53h
7Eh	Vendor ID2	0	1	0	0	0	0	1	1	0	1	0	1	0	0	0	0	4350h

Όλες οι στερεοφωνικές εισόδους του κυκλώματος (AUX, VIDEO, CD και LINE\_IN) καθώς και δύο μονοφωνικές εισόδους (MIC και PHONE) μπορούν να οδηγηθούν στον 18-bit στερεοφωνικό ADC για να οδηγηθούν στο ψηφιακό τμήμα του κυκλώματος. Η επιλογή της εισόδου γίνεται μέσω του αναλογικού πολυπλέκτη RECORD SELECT MUX που ελέγχεται από τον καταχωρητή Record Select (1Ah). Στην είσοδο MIC του πολυπλέκτη οδηγείται, μετά από επιλογή μέσω του D8 bit του καταχωρητή General Purpose (20h), μία εκ των δύο μονοφωνικών εισόδων MIC1, MIC2. Οι εισόδους αυτές είναι σχεδιασμένες για σύνδεση με δυναμικό μικρόφωνο το οποίο δίνει από τη φύση του ασθενές σήμα και έτσι οδηγούνται μετά την επιλογή τους σε κύκλωμα προενίσχυσης που δίνει αύξηση κατά 20db, ελεγχόμενο από το D6 bit του Mic Volume καταχωρητή (0Eh). Ως επιπλέον εισόδους στον πολυπλέκτη δίνονται το στερεοφωνικό σήμα STEREO MIX (που αποτελεί έξοδο του αναλογικού μείκτη MIX2) και μια μονοφωνική μετατροπή αυτού (MONO MIX). Αξίζει να αναφερθεί ότι η επιλογή πηγής για κάθε κανάλι γίνεται ανεξάρτητα σύμφωνα με τον πίνακα

#### RECORD SELECT REGISTER (1Ah)

SL2:SL0	Source for Left Channel ADC
0	*Mic input
1	CD input (L)
2	VIDEO input (L)
3	AUX input (L)
4	LINE_IN input (L)
5	Stereo Mix (L)
6	Mono Mix
7	PHONE input

SR2:SR0	Source for Right Channel ADC
0	*Mic input
1	CD input (R)
2	VIDEO input (R)
3	AUX input (R)
4	LINE_IN input (R)
5	Stereo Mix (R)
6	Mono Mix
7	PHONE input

Default: 0000h
----------------

Ο καταχωρητής Record Gain (1Ch) ελέγχει τις λειτουργίες κέρδους και αποσιώπησης για την επιλεγμένη από τον πολυπλέκτη είσοδο στον ADC. Το κέρδος ενίσχυσης του σήματος έχει εύρος από 0 db έως 22,5db και ορίζεται σε βήματα του 1,5 db σύμφωνα με τον πίνακα



**Record Gain Register (1Ch)**

Mute	Gx3:Gx0	Function
0	1111	22.5dB gain
0	0000	0dB gain
1	XXXX	*mute
Default: 8000h		

Η ροή του σήματος από το ψηφιακό τμήμα του κυκλώματος προς το αναλογικό ξεκινά με την είσοδο του σήματος εξόδου από το AC-link στον στερεοφωνικό 18-bit DAC. Η έξοδος του οδηγείται σε κύκλωμα ελέγχου έντασης GAM (gain, attenuation, mute) μέσω του καταχωρητή PCM Out Volume (18h). Το σήμα εν συνεχεία οδηγείται ως είσοδος σε έναν από τους αναλογικούς μείκτες MIX1, MIX2, ανάλογα με την τιμή του D15 bit του General Purpose καταχωρητή (20h).

Ο αναλογικός μείκτης MIX1 δέχεται ως εισόδους τις τέσσερις στερεοφωνικές εισόδους του κυκλώματος (AUX, VIDEO, CD, LINE\_IN), την στερεοφωνική έξοδο του DAC και την μονοφωνική είσοδο MIC μετά την επιλογή και την προενίσχυσή του. Όλες οι εισοδοί στον μείκτη διέρχονται μέσω κυκλωμάτων ελέγχου έντασης GAM, ελεγχόμενων από τους καταχωρητές Mixer Input Volume Registers (0Ah – 18h). Τα κυκλώματα GAM είναι υπεύθυνα για την υλοποίηση των λειτουργιών κέρδους ενίσχυσης (έως 12db), εξασθένησης (έως 34.5 db) και αποσιώπησης του σήματος. Ο έλεγχος της έντασης γίνεται σε βήματα του 1,5 db σύμφωνα με τον πίνακα

**MIXER INPUT VOLUME REGISTERS (Index 0Ch – 18h)**

Mute	Gx4:Gx0	Function
0	0 0000	+12dB gain
0	0 1000	0dB gain
0	1 1111	34.5dB attenuation
1	X XXXX	*mute
Default: 8008h (mono registers) 8808h (stereo registers)		

Ο αναλογικός μείκτης MIX2 δέχεται ως εισόδους τη στερεοφωνική έξοδο του DAC και την έξοδο του MIX1 και τις μονοφωνικές εισόδους PHONE (μέσω κυκλώματος GAM ελεγχόμενου από τον καταχωρητή Phone Volume – 0Ch) και PC\_BEEP μέσω κυκλώματος εξασθένησης ως και 45 db σε βήματα των 3 db. Υπεύθυνος για τον έλεγχό του είναι ο καταχωρητής PC\_Beep Volume (0Ah) όπως παρουσιάζεται από τον πίνακα.

**PC BEEP VOLUME REGISTER (0Ah)**

Mute	PV3:PV0	Function
0	0000	*0dB attenuation
0	1111	45dB attenuation
1	XXXX	mute
Default: 0000h		

Η έξοδος του μείκτη MIX2 δίνεται ως είσοδος, όπως αναφέρθηκε, στον πολυπλέκτη RECORD SELECT MUX και ταυτόχρονα οδηγεί και τις δύο στερεοφωνικές εξόδους του κυκλώματος, τις LINE\_OUT και HP\_OUT. Η μονοφωνική έξοδος MONO\_OUT οδηγείται (με επιλογή από το D9 bit του General Purpose καταχωρητή) είτε από την έξοδο του MIX1, είτε από την είσοδο MIC. Ο τελικός έλεγχος έντασης των εξόδων γίνεται μέσω κυκλωμάτων εξασθένησης και αποσιώπησης που παρέχουν έλεγχο έντασης από 0db έως -46.5 db σε βήματα του 1,5 db. Οι καταχωρητές που ελέγχουν τα κυκλώματα είναι οι Master Volume (02h), Headphone Volume (04h) και Mono Volume (06h) και ο έλεγχος γίνεται σύμφωνα με τον πίνακα

**VOLUME REGISTERS (02h – 06h)**

Mute	Mx4:Mx0	Function
0	0 0000	0dB attenuation
0	1 1111	46.5dB attenuation
1	X XXXX	*mute
Default: 8000h		

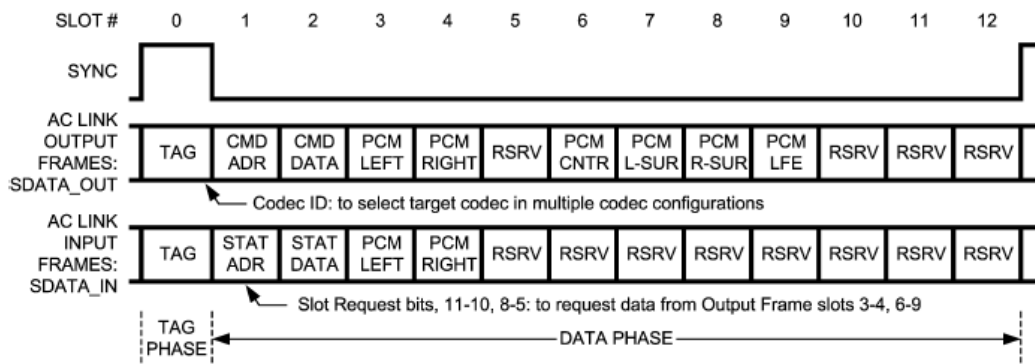
Η επιλογή του ρυθμού δειγματοληψίας για την λειτουργία των μετατροπέων DAC και ADC γίνεται από τους καταχωρητές PCM DAC Rate (2Ch) και PCM ADC Rate (32h), στους οποίους γράφεται η τιμή που επιθυμούμε σε Hz, μέσα στο εύρος 4kHz – 48kHz. Για την επιλογή οποιουδήποτε ρυθμού δειγματοληψίας (ανεξάρτητου μεταξύ DAC και ADC) πρέπει να ενεργοποιηθεί η λειτουργία Variable Rate Audio (VRA) μέσω του D0 bit του Extended Audio Control \ Status καταχωρητή (2Ah). Σε περίπτωση που το VRA bit έχει την τιμή 0 η τιμή των καταχωρητών PCM DAC Rate και PCM ADC Rate αγνοείται και ο ρυθμός δειγματοληψίας καθορίζεται από τον αριθμό εισερχομένων και εξερχομένων πλαισίων του AC-link που περιέχουν έγκυρα δεδομένα, σε σχέση με τον συνολικό αριθμό πλαισίων που θα μεταφερθούν (και που αντιστοιχεί στον μέγιστο ρυθμό δειγματοληψίας – 48kHz).

Από την περιγραφή της δομής του κυκλώματος LM4550 διαπιστώνουμε εύκολα ότι πρόκειται για ένα αρκετά ευέλικτο σύστημα χειρισμού εισερχόμενων και εξερχόμενων ηχητικών ροών. Η ποικιλία επιλογών δρομολόγησης των σημάτων και οι λεπτομερείς δυνατότητες ελέγχου που δίνονται για το κάθε ένα από αυτά οδηγούν εύκολα στην ενσωμάτωσή του σε πλήθος εφαρμογών με διαφορετικές προδιαγραφές.

**2.2.2 Σειριακό πρωτόκολλο διασύνδεσης AC-link**

Το AC-link είναι ένα σειριακό πρωτόκολλο διασύνδεσης του AC '97 Audio Codec ολοκληρωμένου με έναν ψηφιακό ελεγκτή. Αποτελείται από 5 σήματα, τα SDATA\_IN (output), SDATA\_OUT (input), SYNC (input), BIT\_CLK (output / input) και RESET# (input). Το πρωτόκολλο περιγράφει μια σειριακή, αμφίδρομη και σύγχρονη ροή δεδομένων. Χρησιμοποιεί

χρονική πολυπλεξία για να μεταφέρει μέσω αυτής της ροής πολλαπλά κανάλια ηχητικών δεδομένων, καθώς και δεδομένα ελέγχου προς τους καταχωρητές του codec. Τα δεδομένα προς μεταφορά διαιρούνται σε πλαίσια των 256 bits, αποτελούμενα από 13 τμήματα. Το πρώτο τμήμα (TAG slot) αποτελείται από 16 bits και τα επόμενα 12 (DATA slots) από 20 bits. Τα πλαίσια εξόδου δίνονται από τον ελεγκτή ως είσοδος στο codec μέσω του σήματος SDATA\_OUT. Τα πλαίσια εισόδου περνούν από το codec στον ελεγκτή μέσω του σήματος SDATA\_IN. Ο συγχρονισμός των πλαισίων γίνεται από το σήμα SYNC που οδηγεί ο ελεγκτής με βάση το σήμα BIT\_CLK, το οποίο λαμβάνει από το codec. Το σήμα BIT\_CLK έχει σταθερή συχνότητα 12,288 MHz και παράγεται εσωτερικά από το codec (με διαίρεση του clock συχνότητας 24,576 MHz που απαιτεί για τη λειτουργία του), όταν αυτό λειτουργεί ως πρωτεύον. Σε περίπτωση διασύνδεσης πολλαπλών codecs, τα δευτερεύοντα λαμβάνουν το σήμα BIT\_CLK ως είσοδο από το πρωτεύον. Όταν το σήμα SYNC έχει την λογική τιμή 1 και το BIT\_CLK εμφανίσει ανοδική ακμή, σηματοδοτείται έναρξη μεταφοράς των πλαισίων εισόδου και εξόδου. Ο ελεγκτής μηδενίζει το σήμα SYNC στο τέλος της μεταφοράς του TAG slot.



Εικόνα 27: Δομή των πλαισίων εξόδου / εισόδου

Στην εικόνα 27 παρουσιάζεται η δομή των πλαισίων εισόδου και εξόδου σύμφωνα με το πρωτόκολλο AC-link, όπως αυτό υλοποιείται στο LM4550 ολοκληρωμένο κύκλωμα. Τα πλαίσια εξόδου αποτελούνται από τα παρακάτω τμήματα:

- Τμήμα 0 (TAG): μεταφέρει πληροφορία για το αν το πλαίσιο και τα επιμέρους τμήματά του περιέχουν έγκυρα δεδομένα, καθώς και τον κωδικό αναγνώρισης του ολοκληρωμένου. Αποτελείται από 16 bit, η λειτουργία των οποίων παρουσιάζεται στον πίνακα

Bit	Description	Comment	Bit	Description	Comment
15	Valid Frame	1 = Valid data in at least one slot.	9	Left DAC data in Slot 6	1 = Valid PCM Data in Slot 6 (Secondary 3 mode; Center Channel audio)
14	Control register address	1 = Valid Control Address in Slot 1 (Primary codec only)	8	Left DAC data in Slot 7	1 = Valid PCM Data in Slot 7 (Secondary 2 mode; Left Surround Channel audio)
13	Control register data	1 = Valid Control Data in Slot 2 (Primary codec only)	7	Right DAC data in Slot 8	1 = Valid PCM Data in Slot 8 (Secondary 2 mode; Right Surround Channel audio)
12	Left DAC data in Slot 3	1 = Valid PCM Data in Slot 3 (Primary & Secondary 1 modes; Left Channel audio)	6	Right DAC data in Slot 9	1 = Valid PCM Data in Slot 9 (Secondary 3 mode; LFE Channel audio)
11	Right DAC data in Slot 4	1 = Valid PCM Data in Slot 4 (Primary & Secondary 1 modes; Right Channel audio)	5:2	Not Used	Controller should stuff these slots with "0"s
10	Not Used	Controller should stuff this slot with "0"s	1,0	Codec ID (ID1, ID0)	The Codec ID (Table 1) selects the target codec in a multi-codec system to receive the control address and data carried in the Output Frame

- Τμήματα 1,2 (Control Address, Control Data) : περιέχουν την διεύθυνση του καταχωρητή ελέγχου προς εγγραφή / ανάγνωση και τα δεδομένα της ενέργειας αντίστοιχα

SLOT 1, OUTPUT FRAME

Bits	Description	Comment
19	Read/Write	1 = Read 0 = Write
18:12	Register Address	Identifies the Status/Command register for read/write
11:0	Reserved	Controller should set to "0"

SLOT 2, OUTPUT FRAME

Bits	Description	Comment
19:4	Control Register Write Data	Controller should stuff with zeros if operation is "read"
3:0	Reserved	Set to "0"

- Τμήματα 3,4,6-9 (PCM Left, PCM Right, PCM Cntr, PCM L-Sur, PCM R-Sur, PCM LFE) : περιέχουν PCM ηχητικά δεδομένα των 6 καναλιών ενός τυπικού surround συστήματος προς αναπαραγωγή. Για την πλήρη αξιοποίηση όλων σε μια μεταφορά απαιτείται διάταξη με τρία ολοκληρωμένα LM4550, καθώς το κάθε ένα δίνει στην έξοδο του δύο μόνο κανάλια. Αξίζει επιπλέον να σημειωθεί ότι, αν και το πρότυπο του AC '97 ορίζει ως μέγιστη ανάλυση τα 20 bit (κάτι το οποίο γίνεται προφανές και από το μέγεθος των τμημάτων του κάθε πλαισίου), το LM4550 διαθέτει 18 bit μετατροπείς ADC και DAC. Τα LSB των τμημάτων συνεπώς μηδενίζονται.

SLOTS 3 &amp; 4, OUTPUT FRAME

Bits	Description	Comment
19:0	PCM Audio Data (Left /Right Channels)	Slots used to stream data to DAC when codec is in Primary or Secondary 1 modes. Set unused bits to "0"

SLOTS 6 &amp; 9, OUTPUT FRAME

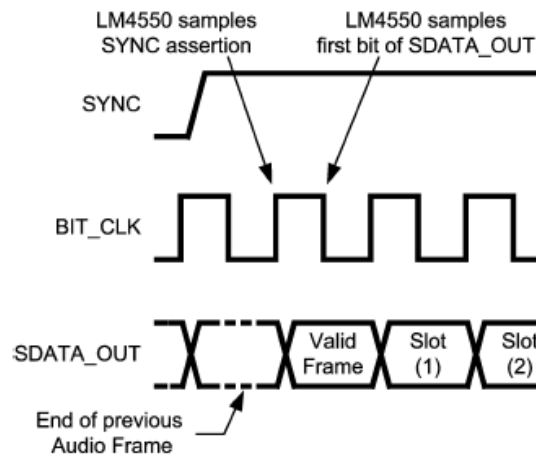
Bits	Description	Comment
19:0	PCM Audio Data (Center/ LFE Surround)	Slots used to stream data to DAC when codec is in Secondary 3 mode. Set unused bits to "0"

SLOTS 7 &amp; 8, OUTPUT FRAME

Bits	Description	Comment
19:0	PCM Audio Data (Left/Right Surround)	Slots used to stream data to DAC when codec is in Secondary 2 mode. Set unused bits to "0"

- Τμήματα 5,10-12: δεν χρησιμοποιούνται από το LM4550 καθώς αντιστοιχούν σε μη υλοποιημένα χαρακτηριστικά του προτύπου AC '97 (modem δεδομένα, ρυθμοί δειγματοληψίας μεγαλύτεροι από 48kHz – το πρότυπο υποστηρίζει ρυθμό δειγματοληψίας έως 96 kHz).

Η έναρξη της λήψης από το codec του TAG τμήματος ξεκινά στην πρώτη ανοδική ακμή του BIT\_CLK σήματος εφόσον το SYNC βρίσκεται στο λογικό 1. Η ανίχνευση του επιθυμητού SYNC σηματοδοτεί λήψη των έγκυρων δεδομένων από την γραμμή SDATA\_OUT σε κάθε καθοδική ακμή του BIT\_CLK και για την προκαθορισμένη διάρκεια του πλαισίου.



**Εικόνα 28: Χρονισμός του πλαισίου εξόδου**

Τα πλαίσια εισόδου αντίστοιχα αποτελούνται από τα παρακάτω τμήματα:

- Τμήμα 0 (TAG): περιέχει πληροφορία για το αν το codec είναι έτοιμο να στείλει δεδομένα στον ελεγκτή, καθώς και για το αν τα επιμέρους τμήματα περιέχουν έγκυρα δεδομένα

**SLOT 0, INPUT FRAME**

Bit	Description	Comment
15	Codec Ready Bit	1 = AC Link Interface Ready
14	Slot 1 data valid	1 = Valid Status Address or Slot Request
13	Slot 2 data valid	1 = Valid Status Data
12	Slot 3 data valid	1 = Valid PCM Data (Left ADC)
11	Slot 4 data valid	1 = Valid PCM Data (Right ADC)

- Τμήματα 1,2 (Status Address, Status Data) : περιέχουν την διεύθυνση του καταχωρητή ελέγχου προς ανάγνωση και τα δεδομένα της ενέργειας αντίστοιχα

**SLOT 1, INPUT FRAME**

Bits	Description	Comment
19	Reserved	Stuffed with "0" by LM4550
18:12	Status Register Index	Echo of the requested Status Register address.

**SLOT 2, INPUT FRAME**

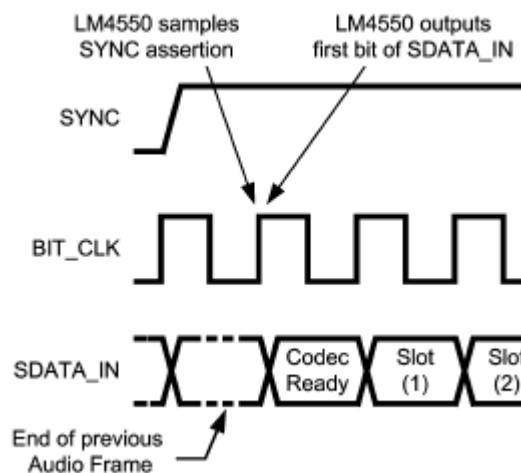
Bits	Description	Comment
19:4	Status Data	Data read from a codec control/status register. Stuffed with "0"s if no read-request in previous frame.
3:0	Reserved	Stuffed with "0"s by LM4550

- Τμήματα 3,4 (PCM Left, PCM Right) : περιέχουν PCM ηχητικά δεδομένα των 2 καναλιών που εισέρχονται από τον ADC μετατροπέα.

SLOT 3, INPUT FRAME			SLOT 4, INPUT FRAME		
Bits	Description	Comment	Bits	Description	Comment
19:2	PCM Record Left Channel data	18-bit PCM audio sample from left ADC	19:2	PCM Record Right Channel data	18-bit PCM audio sample from right ADC
1:0	Reserved	Stuffed with "0"s by LM4550	1:0	Reserved	Stuffed with "0"s by LM4550

- Τμήματα 5-12: δεν χρησιμοποιούνται από το LM4550 καθώς αντιστοιχούν σε μη υλοποιημένα χαρακτηριστικά του προτύπου AC '97

Ο χρονισμός αποστολής των πλαισίων εισόδου από το codec χρησιμοποιεί τις ανοδικές μόνο ακμές του BIT\_CLK σήματος και για την ανίχνευση της στάθμης του SYNC και για την αποστολή των έγκυρων δεδομένων όπως φαίνεται στην εικόνα 29.



Εικόνα 29: Χρονισμός του πλαισίου εισόδου

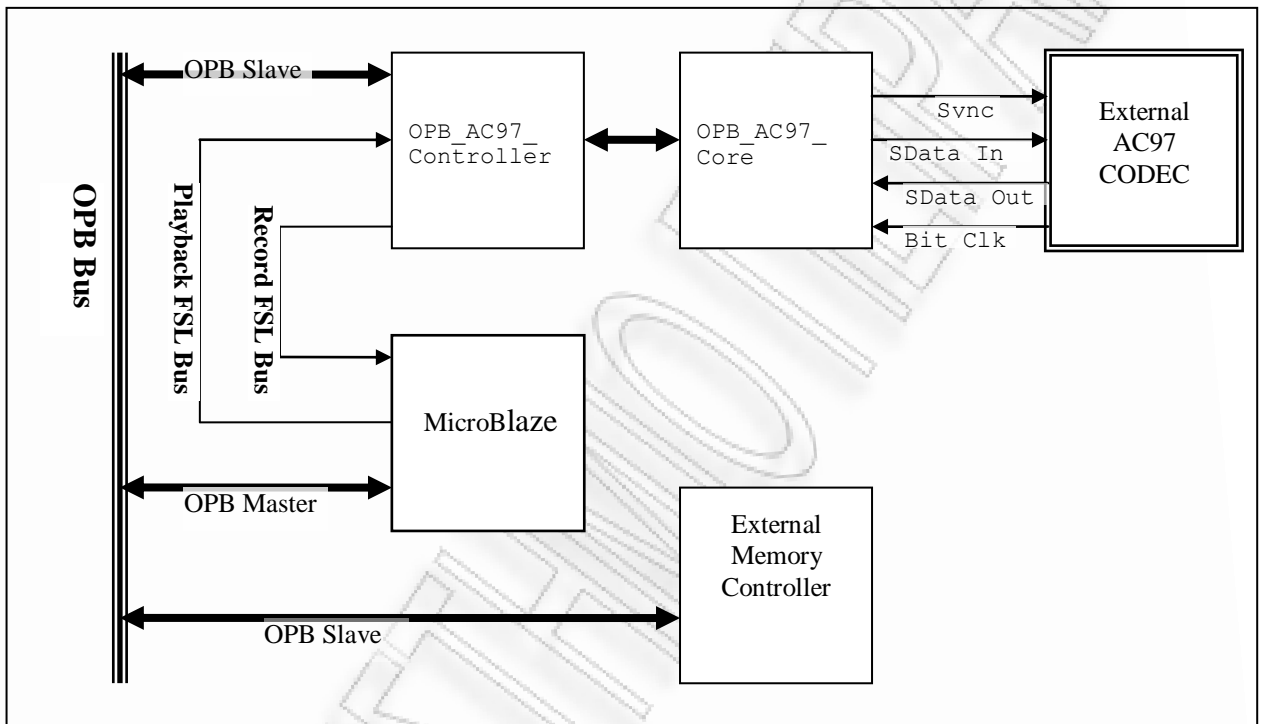
### 2.2.3 Ελεγκτής OPB\_AC97\_Controller

Για τη διασύνδεση του AC '97 Audio Codec σε ένα υπολογιστικό σύστημα απαιτείται η χρήση ενός ελεγκτή, ο οποίος αναλαμβάνει την μετατροπή του AC-link πρωτοκόλλου στο πρωτόκολλο του διαύλου σύνδεσης με τον επεξεργαστή. Η σωστή ανάπτυξη και χρήση ενός τέτοιου περιφερειακού, απαλλάσσει τον σχεδιαστή από τις λεπτομέρειες και τις ιδιαιτερότητες της υλοποίησης ενός μη διαδεδομένου πρωτοκόλλου όπως το AC-link, διευκολύνοντας σε μεγάλο βαθμό το έργο του. Στην συγκεκριμένη περίπτωση η ύπαρξη δύο σημάτων χρονισμού στο σύστημα (το ρολόι του επεξεργαστή και το παραγόμενο BIT\_CLK από το codec) θα έκανε την διασύνδεσή του ακόμα πιο απαιτητική.

Κατά την ανάπτυξη του ενσωματωμένου συστήματος ανίχνευσης και διόρθωσης των θεμελιωδών συχνοτήτων μουσικών σημάτων χρησιμοποιήθηκε για την διασύνδεση του LM4550 ολοκληρωμένου κυκλώματος με τον επεξεργαστή το περιφερειακό OPB\_AC97\_Controller version 3.10a. Ο ελεγκτής βασίστηκε στην έκδοση του περιφερειακού 1.00a που παρέχεται από την Xilinx για ανάπτυξη ενσωματωμένων συστημάτων, και αναπτύχθηκε περαιτέρω από το Electrical And Computer Engineering Department του πανεπιστημίου του Τορόντο. Στην νέα έκδοση του ελεγκτή υλοποιήθηκε πλήρως η μετατροπή και η χρήση του πρωτοκόλλου Fast Simplex Link Bus (FSL) ταυτόχρονα με την διασύνδεση στον δίαυλο OPB. Με τον τρόπο αυτόν ο έλεγχος του ολοκληρωμένου γίνεται μέσω του OPB διαύλου, ενώ τα ηχητικά δεδομένα

μεταφέρονται είτε μέσω του OPB, είτε μέσω του FSL. Επιπλέον στην νέα έκδοση διορθώθηκαν προβλήματα που παρουσιάζονταν στην ορθή χρήση διακοπών για την λήψη των ηχητικών δεδομένων μέσω του OPB διαύλου.

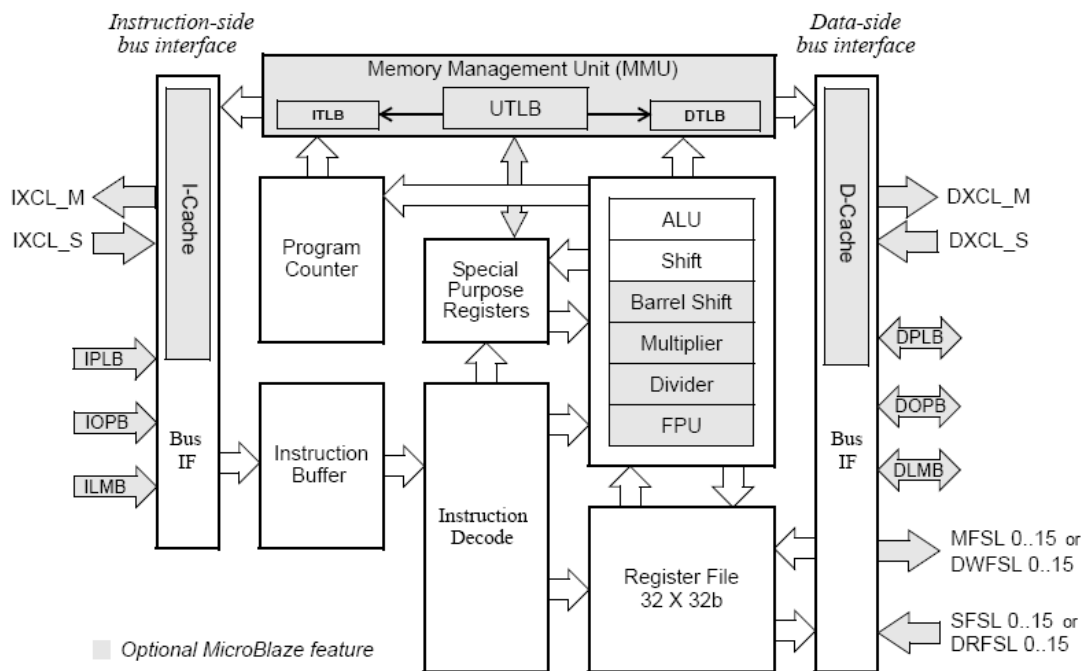
Ο ελεγκτής αποτελείται από δύο διακριτά τμήματα, τον OPB\_AC97\_Core που είναι υπεύθυνος για τον έλεγχο της σειριακής ροής δεδομένων - μέσω του AC-link – μεταξύ του FPGA και του LM4550 ολοκληρωμένου, και τον OPB\_AC97\_Controller που υλοποιεί την μεταφορά των δεδομένων στον επεξεργαστή μέσω των OPB και FSL διαύλων. Μια τυπική διασύνδεση του ελεγκτή σε ένα ενσωματωμένο σύστημα παρουσιάζεται στην εικόνα 30



Εικόνα 30: Τυπική διασύνδεση του OPB\_AC97\_controller σε ενσωματωμένο σύστημα

### 2.3 Ενσωματωμένος Επεξεργαστής MicroBlaze

Ο μικροεπεξεργαστής MicroBlaze είναι ένας επεξεργαστής αρχιτεκτονικής περιορισμένου σετ εντολών(RISC) που αναπτύχθηκε για χρήση ως ενσωματωμένος επεξεργαστής σε FPGA's της εταιρίας Xilinx. Η βασική του δομή παρουσιάζεται στην εικόνα 31



Εικόνα 31: Δομή του MicroBlaze

Βασικό χαρακτηριστικό του σχεδιασμού του είναι η παραμετροποίησή του που δίνει στον σχεδιαστή του ενσωματωμένου συστήματος την δυνατότητα να επιλέξει τα προς υλοποίηση χαρακτηριστικά και τις υπομονάδες του επεξεργαστή που τον ενδιαφέρουν, ενώ ταυτόχρονα δεν σπαταλά τους πόρους της πλατφόρμας ανάπτυξης.

Ο επεξεργαστής διαθέτει 32 32-bit καταχωρητές, 32bit εύρος διαύλων εντολών και δεδομένων, δύο τρόπους διευθυνσιοδότησης καθώς και αρχιτεκτονική διοχέτευσης τριών ή πέντε σταδίων. Παρέχει την δυνατότητα σύνδεσης με τους διαύλους LMB, OPB, PLB και FSL. Η στιβαρή αρχιτεκτονική του, οι ποικίλες δυνατότητες διασύνδεσης που παρέχει η υψηλή του απόδοση σε συνδυασμό με το μικρό του μέγεθος τον καθιστούν ιδανικό για χρήση σε ένα περιβάλλον ανάπτυξης ενσωματωμένων συστημάτων.



### **Κεφάλαιο 3: Ανάπτυξη Εφαρμογής Ανίχνευσης και Διόρθωσης Θεμελιώδους Συχνότητας Μουσικών Σημάτων**

Στο κεφάλαιο αυτό θα ασχοληθούμε με την ανάπτυξη της εφαρμογής της ανίχνευσης και διόρθωσης της θεμελιώδους συχνότητας μουσικών σημάτων η οποία θα εκτελεστεί στο ενσωματωμένο σύστημα που περιγράφηκε και υλοποιήθηκε στο προηγούμενο κεφάλαιο. Ο κύκλος ανάπτυξης μιας εφαρμογής στο πλαίσιο σχεδίασης ενός ενσωματωμένου συστήματος δεν είναι απλή υπόθεση, καθώς περιλαμβάνει πλήθος διαδικασιών και αποφάσεων που πρέπει να ληφθούν. Ξεκινά από τη μελέτη του προβλήματος, και των προτεινόμενων στη βιβλιογραφία αλγορίθμων για την επίλυσή του, τον ορισμό των προδιαγραφών και των ορίων της εφαρμογής, μελετά εναλλακτικές υλοποιήσεις υλικού και λογισμικού και καταλήγει στην συγγραφή του κώδικα, στον έλεγχο και τη βελτιστοποίηση του συστήματος. Το σημαντικότερο πρόβλημα που αντιμετωπίζει ο σχεδιαστής, ειδικά στα πρώτα στάδια ανάπτυξης ενός τέτοιου συστήματος, είναι το γεγονός ότι το υλικό δεν είναι ακόμα πλήρως καθορισμένο, με αποτέλεσμα ο σχεδιασμός της εφαρμογής να γίνεται απαιτητικός. Επιπλέον λήψη λανθασμένων αποφάσεων στα πρώτα στάδια της ανάπτυξης μπορούν εύκολα να οδηγήσουν σε μεγάλες δυσκολίες κατά την ολοκλήρωσή της. Παρά, όμως, τα όποια προβλήματα παρουσιάζει η ανάπτυξη ενός ενσωματωμένου συστήματος, η απλότητα της δομής του, και η διαφανής συνεργασία υλικού και λογισμικού οδηγούν σε ευέλικτες υλοποιήσεις.

Στόχος της προς υλοποίηση εφαρμογής, όπως γίνεται άμεσα κατανοητό από την περιγραφή της, είναι σε πρώτη φάση η ανίχνευσή των θεμελιωδών συχνοτήτων που περιέχονται σε μια μουσική φράση, και τέλος η διόρθωση όσων από αυτές αποκλίνουν από μια αρχικά καθορισμένη δομή. Το σύστημα πρέπει να δίνει συνεπή αποτελέσματα και να λειτουργεί σταθερά σε πραγματικές συνθήκες, καθώς στόχος του είναι η χρήση σε καλλιτεχνικές εφαρμογές. Πολύ σημαντικό χαρακτηριστικό του πρέπει να είναι η ανοχή στον θόρυβο και η παραμετροποίησή του για την καλύτερη προσαρμογή του σε δεδομένη είσοδο. Τέλος επειδή η εφαρμογή αφορά μουσικά δεδομένα, η ποιότητα του σήματος εισόδου και εξόδου πρέπει να είναι υψηλή, ακόμα και αν αυτό αποβαίνει σε βάρος της απόδοσής του.

Η μελέτη των αλγορίθμων ανίχνευσης και μεταβολής του pitch που πραγματοποιήθηκε στα πρώτα στάδια της ανάπτυξης έδειξε ότι η υλοποίηση του συστήματος, σύμφωνα με τις προδιαγραφές που τέθηκαν, σε πραγματικό χρόνο δεν είναι εφικτή. Η προϋπόθεση της διατήρησης της υψηλής ποιότητας του ήχου, καθώς και οι απαιτήσεις ακρίβειας των αλγορίθμων για χρήση σε μουσικές εφαρμογές, οδηγούν στην επιλογή υψηλού ρυθμού δειγματοληψίας κατά την μετατροπή του ηχητικού σήματος στον ψηφιακό κόσμο, κάτι που αυξάνει κατά πολύ τον υπολογιστικό φόρτο του συστήματος. Επιπλέον περιορισμοί του υλικού στην χρήση πράξεων κινητής υποδιαστολής επέβαλαν επιλογή αλγορίθμων που απέτρεπαν την ανάπτυξη σε πραγματικό χρόνο. Δεδομένων των περιορισμών στο υλικό, η μη υλοποίηση σε πραγματικό χρόνο δεν αποτελεί έκπληξη.

Άλλος ένας περιορισμός που τέθηκε κατά την σχεδίαση της εφαρμογής ήταν ο περιορισμός του εύρους των σημάτων που θα λαμβάνει το σύστημα ως είσοδο. Όπως αναφέρθηκε στο κεφάλαιο 1, το πρόβλημα αναγνώρισης της θεμελιώδους συχνότητας ενός σήματος, ακόμα και αν πρόκειται για απλό, μονοφωνικό σήμα, δεν είναι τετριμμένο. Το αποτέλεσμα ποικίλει ακόμα περισσότερο αν συνυπολογίσουμε προβλήματα αλλοίωσης του σήματος από θόρυβο. Η παραμετροποίηση ενός αλγορίθμου δίνει λύση σε ένα βαθμό για δεδομένο υποσύνολο εισόδων, αλλά η λύση του προβλήματος στην γενική περίπτωση είναι μάλλον αδύνατη. Φυσική συνέπεια των παραπάνω είναι ο περιορισμός της εφαρμογής σε ένα αυστηρά καθορισμένο σύνολο εισόδων, που όμως δεν θα διαταράσσει την προδιαγραφή για χρήση σε πραγματικές συνθήκες. Η επιλογή που έγινε ήταν να περιοριστεί η εφαρμογή σε ανάλυση και επεξεργασία σημάτων που προέρχονται από ένα συγκεκριμένο μονοφωνικό όργανο (ηλεκτρική κιθάρα) με έμφαση σε μια συγκεκριμένη συχνοτικά περιοχή του οργάνου.

### 3.1 Χαρακτηριστικά του Ήχου της Ηλεκτρικής Κιθάρας

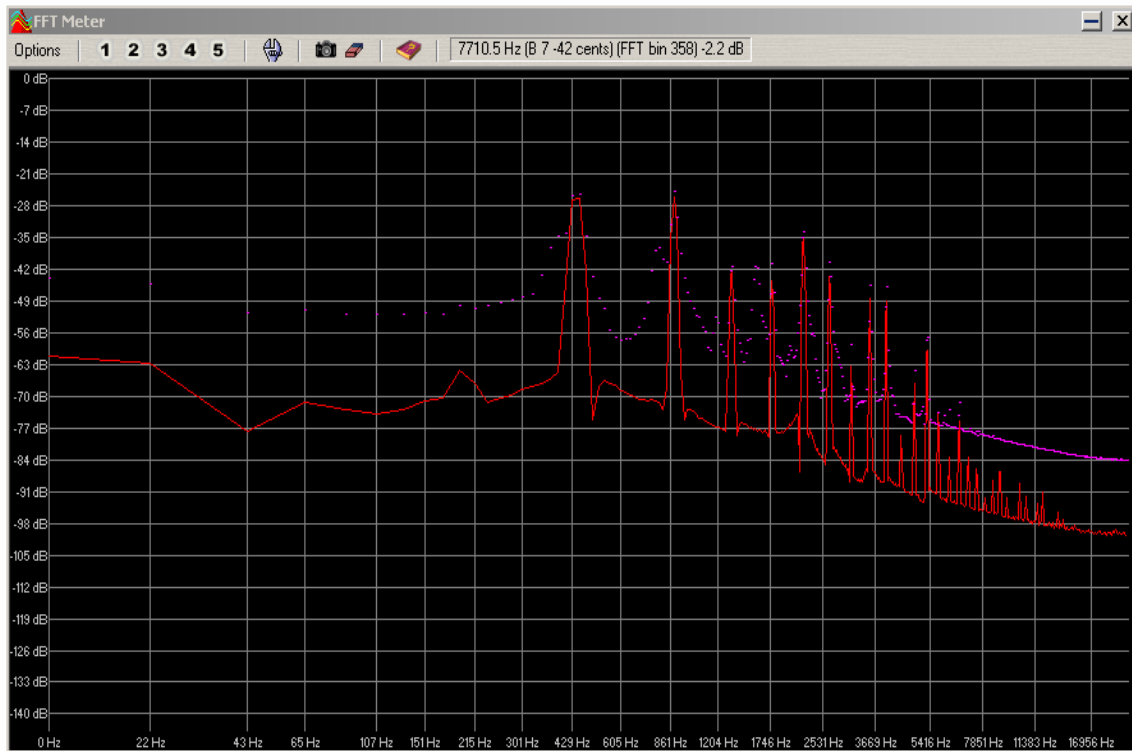
Η ηλεκτρική κιθάρα αποτελεί ίσως το πιο διαδεδομένο ηλεκτρικό όργανο στις μέρες μας. Οι πρώτες ηλεκτρικές κιθάρες με την σημερινή τους μορφή κατασκευάστηκαν τη δεκαετία του '30 (αν και μελέτες με ηλεκτρικά όργανα είχαν εμφανιστεί από το 1910 περίπου), και η χρήση τους διευρύνθηκε τη δεκαετία του '50. Η ηλεκτρική κιθάρα αποτελεί έγχορδο όργανο με ξύλινο συμπαγές συνήθως σώμα και λαιμό. Η έκτασή της είναι 4 περίπου οκτάβες και στο καθιερωμένο κούρδισμα αποδίδει νότες με εύρος θεμελιωδών συχνοτήτων από 82,4 Hz (E2) μέχρι 1174,7 Hz (D6). Η έκταση της κιθάρας και οι συχνότητες στις οποίες αυτή αντιστοιχεί απεικονίζονται στην εικόνα 32

Note	Hz	Note	Hz	Note	Hz	Note	Hz	Note	Hz	Note	Hz	Note	Hz
C1	32.7	C2	65.4	C3	130.8	C4	261.6	C5	523.3	C6	1046.5	C7	2093.0
C#1	34.6	C#2	69.3	C#3	138.6	C#4	277.2	C#5	554.4	C#6	1108.7	C#7	2217.5
D1	36.7	D2	73.4	D3	146.8	D4	293.7	D5	587.3	D6	1174.7	D7	2349.3
D#1	38.9	D#2	77.8	D#3	155.6	D#4	311.1	D#5	622.3	D#6	1244.5	D#7	2489.0
E1	41.2	E2	82.4	E3	164.8	E4	329.6	E5	659.3	E6	1318.5	E7	2637.0
F1	43.7	F2	87.3	F3	174.6	F4	349.2	F5	698.5	F6	1396.9	F7	2793.8
F#1	46.2	F#2	92.5	F#3	185.0	F#4	370.0	F#5	740.0	F#6	1480.0	F#7	2960.0
G1	49.0	G2	98.0	G3	196.0	G4	392.0	G5	784.0	G6	1568.0	G7	3136.0
G#1	51.9	G#2	103.8	G#3	207.7	G#4	415.3	G#5	830.6	G#6	1661.2	G#7	3322.4
A1	55.0	A2	110.0	A3	220.0	A4	440.0	A5	880.0	A6	1760.0	A7	3520.0
A#1	58.3	A#2	116.5	A#3	233.1	A#4	466.2	A#5	932.3	A#6	1864.7	A#7	3729.3
B1	61.7	B2	123.5	B3	246.9	B4	493.9	B5	987.8	B6	1975.5	B7	3951.1

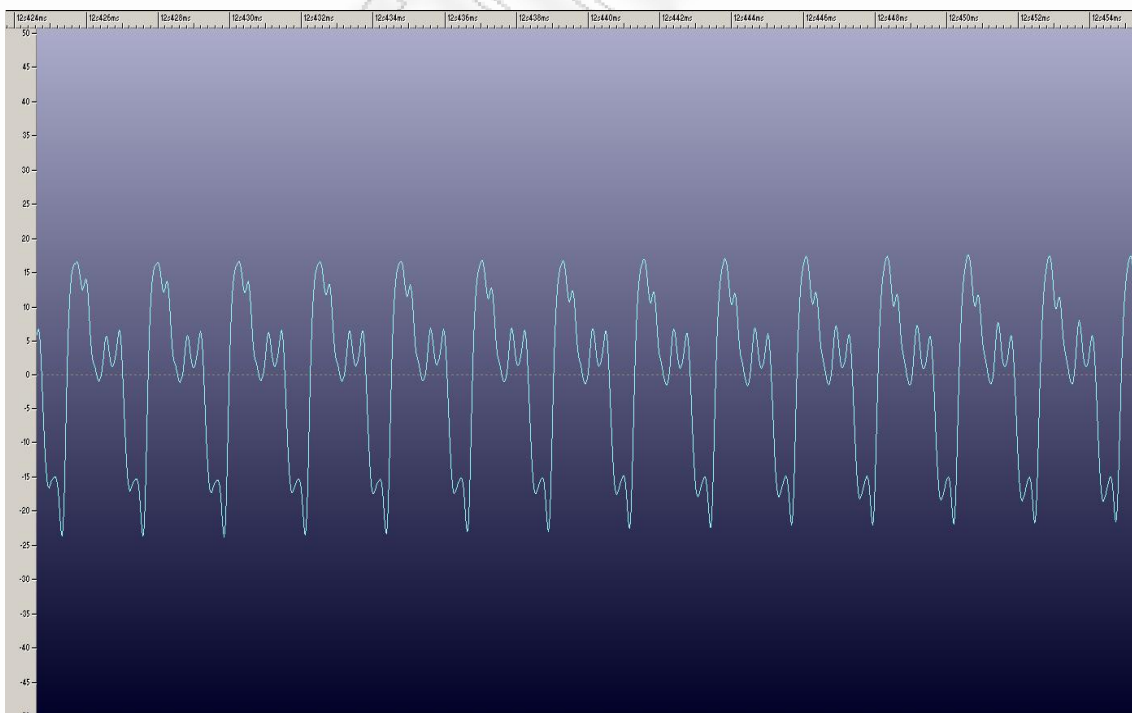
Εικόνα 32: Συχνотική έκταση της ηλεκτρικής κιθάρας

Η λειτουργία της ηλεκτρικής κιθάρας βασίζεται στο φαινόμενο της ηλεκτρικής επαγωγής. Η κιθάρα διαθέτει έξι μεταλλικές χορδές οι οποίες όταν διεγείρονται από τον μουσικό, πάλλονται μέσα στο πεδίο μιας μαγνητικής διάταξης. Η διάταξη αποτελείται από κάποιο μαγνητικό υλικό τυλιγμένο με ένα πηνίο. Η επαναλαμβανόμενη ταλάντωση των χορδών μέσα στο μαγνητικό πεδίο επάγει στο πηνίο (ασθενές) ηλεκτρικό ρεύμα με περιοδική τάση ανάλογη με τα χαρακτηριστικά ταλάντωσης της χορδής. Με τον τρόπο αυτόν μετατρέπεται το ηχητικό σήμα της χορδής σε ηλεκτρικό. Εξαιτίας της ασθενούς φύσης του ηλεκτρικού σήματος απαιτείται συνήθως κάποια μορφή προενίσχυσής του για την περετέρω επεξεργασία του.

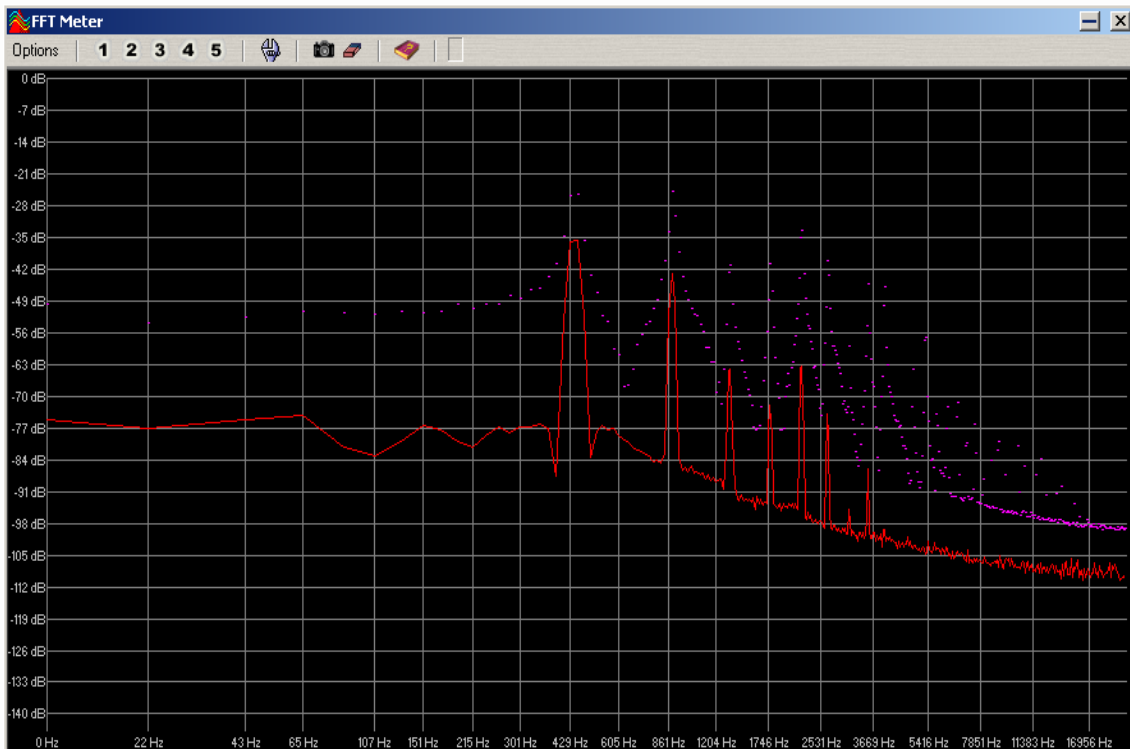
Η ανάλυση μιας απλής νότας της ηλεκτρικής κιθάρας στο πεδίο του χρόνου και στο πεδίο της συχνότητας δίνει σημαντικές πληροφορίες για τη φύση του ήχου της. Στις εικόνες παρατηρούμε το φάσμα και την κυματομορφή της ίδιας νότας σε δύο διαφορετικές χρονικές στιγμές: στο ξεκίνημά της, λίγο χρόνο μετά τη διέγερση και στο σβήσιμό της. Παρατηρούμε ότι κατά το ξεκίνημα η νότα παρουσιάζει σημαντικό ποσοστό της ενέργειάς της στην θεμελιώδη συχνότητα, και συγκρίσιμο ποσοστό στην πρώτη αρμονική της. Ταυτόχρονα το αρμονικό της περιεχόμενο είναι αρκετά πλούσιο και εύκολα διακρίνονται οι αρμονικές μέχρι και 8<sup>ης</sup> τάξης. Κατά το σβήσιμο της νότας οι περισσότερες αρμονικές έχουν εξασθενήσει, η θεμελιώδης έχει σημαντικά μεγαλύτερο ποσοστό ενέργειας από τις αρμονικές ενώ έντονη είναι η παρουσία αρμονικών μέχρι και 5<sup>ης</sup> τάξης. Στο πεδίο του χρόνου τα πράγματα είναι πιο απλά. Η κυματομορφή της πρόσφατα διεγερμένης νότας είναι πολύπλοκη, παρουσιάζει αρκετές επιμέρους διακυμάνσεις αλλά η ομοιότητα μεταξύ των διαδοχικών περιόδων του σήματος είναι εμφανής. Η μελέτη της εξασθενημένης νότας παρουσιάζει ακόμα μεγαλύτερη ομοιότητα μεταξύ των περιόδων σε βαθμό που πλησιάζει πλέον ένα ημιτονοειδές σήμα.



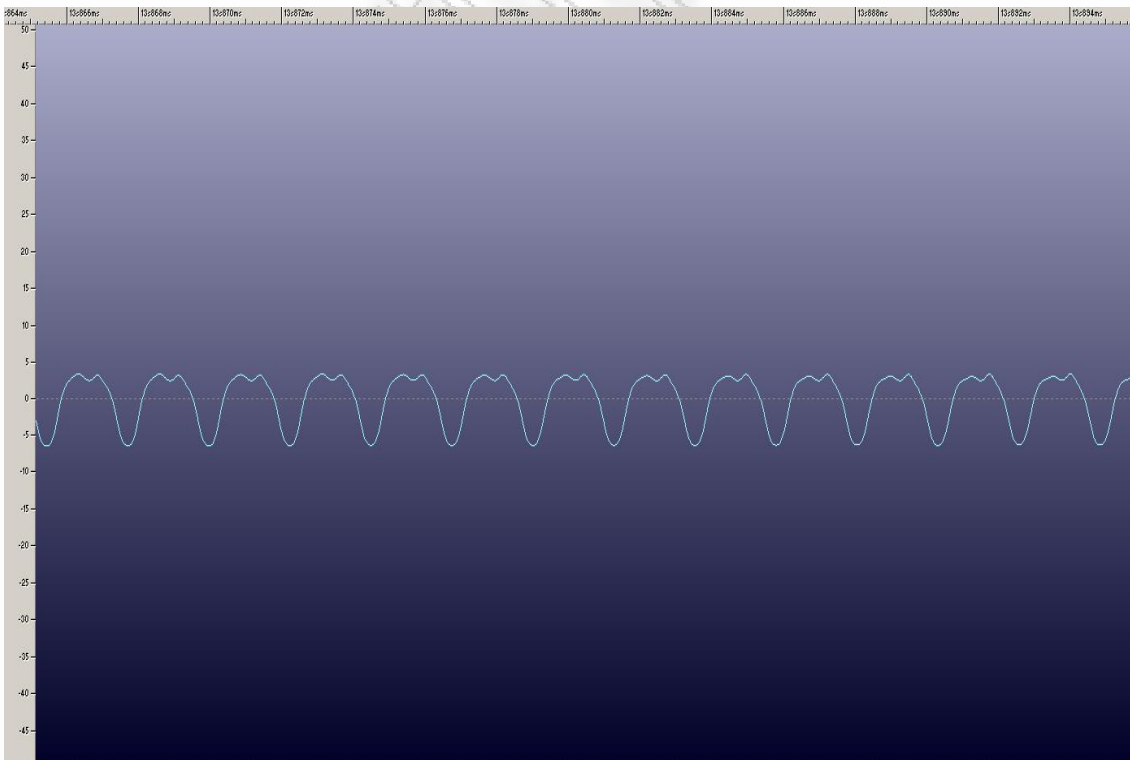
**Εικόνα 33: Συχνοτικό φάσμα πρόσφατα διεγερμένης χορδής ηλεκτρικής κιθάρας**



**Εικόνα 34: Κυματομορφή πρόσφατα διεγερμένης χορδής ηλεκτρικής κιθάρας**

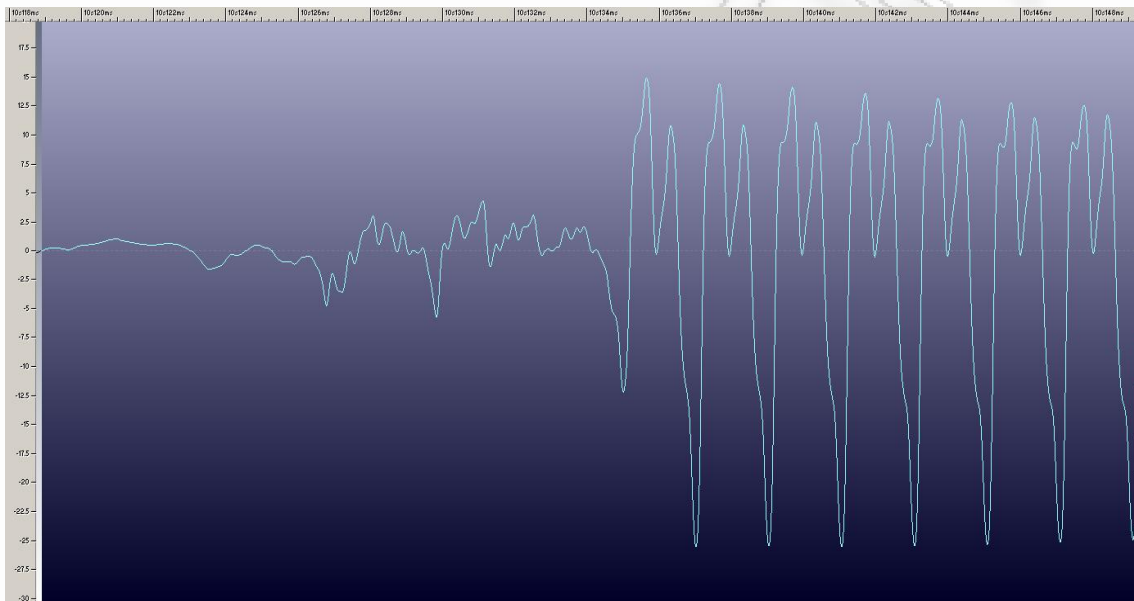


**Εικόνα 35: Συχνотικό φάσμα εξασθενημένης χορδής ηλεκτρικής κιθάρας**



**Εικόνα 36: Κυματομορφή εξασθενημένης χορδής ηλεκτρικής κιθάρας**

Σε ότι αφορά την ένταση ο ήχος της ηλεκτρικής κιθάρας χαρακτηρίζεται από έντονη ατάκα (το διάστημα από την αρχή της ταλάντωσης μέχρι την σταθεροποίηση της κυματομορφής σε περιοδική) και σχετικά μεγάλη απόσβεση. Η νότα δεν έχει διάρκεια καθώς η διέγερσή της γίνεται με κρούση (σε αντίθεση για παράδειγμα με την νότα που παράγει ένα βιολί, όπου η διέγερση της χορδής με το δοξάρι είναι συνεχής). Ένα στιγμιότυπο της κυματομορφής της αρχής της διέγερσης της χορδής φαίνεται στην εικόνα 37:



**Εικόνα 37: Κυματομορφή ατάκας χορδής ηλεκτρικής κιθάρας**

Τέλος η ταλάντωση της χορδής μιας ηλεκτρικής κιθάρας σε πραγματικές συνθήκες διαφέρει κατά πολύ από την απλή περιγραφή της. Ο θόρυβος που κάνει η πένα στη χορδή (και οι αντανακλάσεις του στα όριά της), ο κρουστός ήχος που κάνει το χέρι όταν πατά το τάστο και όταν το αφήνει και άλλα παρόμοια φαινόμενα αποτελούν παράγοντες που δυσκολεύουν το έργο της αναγνώρισης της θεμελιώδους συχνότητας του οργάνου.

Τα προαναφερθέντα χαρακτηριστικά του ήχου της ηλεκτρικής κιθάρας έπαιξαν σημαντικότατο ρόλο κατά την ανάπτυξη της εφαρμογής σε ότι αφορά την επιλογή των αλγορίθμων και την απαιτούμενη επεξεργασία.

### **3.2 Δομή Μουσικών Φράσεων και Διάταξη Νοτών σε Κλίμακες**

Τα βασικά χαρακτηριστικό μιας μουσικής φράσης είναι ο ρυθμός και η μελωδία. Ως μελωδία ορίζεται η διάταξη ενός αριθμού από νότες με βάση την θεμελιώδη συχνότητά τους. Όμως, αν και η συχνότητα μιας νότας αποτελεί συνεχές φυσικό μέγεθος, σε ένα μουσικό πλαίσιο δεν χρησιμοποιούνται νότες με αυθαίρετες συχνότητες, αλλά προσεκτικά διαλεγμένες από ένα σύνολο, ώστε το αποτέλεσμα να είναι αισθητικά όμορφο. Η διάταξη των νοτών του συνόλου με βάση τη θεμελιώδη συχνότητά τους, συνήθως σε διάστημα μιας οκτάβας (που αντιστοιχεί σε διπλασιασμό της θεμελιώδους συχνότητας της πρώτης νότας) αποτελεί μια μουσική κλίμακα. Ο παραπάνω ορισμός είναι αρκετά ασαφής, καθώς στην τέχνη οι ορισμοί χάνουν εν μέρει το νόημα τους, αλλά δίνει μια αρκετά καλή βάση για την ανάπτυξη μιας μουσικής εφαρμογής. Η

συνέχεια που χαρακτηρίζει ένα μέγεθος σαν τη θεμελιώδη συχνότητα, κβαντίζεται σε διακριτά μεγέθη τα οποία μπορούν να οριστούν και να μετρηθούν.

Στην θεωρία της δυτικής μουσικής, που αποτελεί την πιο μελετημένη, τεκμηριωμένη και διαδεδομένη μουσική σημειογραφία, έχουν επικρατήσει οι οκτατονικές κλίμακες, βασισμένες σε ένα δωδεκατονικό πλαίσιο. Αυτό σημαίνει ότι κάθε οκτάβα διαιρείται σε δώδεκα ίσα συχνοτικά διαστήματα (κάθε ένα από τα οποία ονομάζεται ημιτόνιο) και με βάση αυτά σχηματίζουμε κλίμακες αποτελούμενες από οκτώ νότες (συμπεριλαμβανομένης και της νότας της οκτάβας). Στις πιο συνηθισμένες από τις ευρωπαϊκές κλίμακες δεν χρησιμοποιείται μεταξύ δύο διαδοχικών νοτών διάστημα μεγαλύτερο από έναν τόνο. Ο ορισμός της μουσικής κλίμακας παρέχει μια καθορισμένη δομή πάνω στην οποία βασίζεται η διόρθωση της θεμελιώδους συχνότητας, όπως αυτή ανιχνεύεται από το ενσωματωμένο σύστημα. Η επιλογή των οκτατονικών κλιμάκων έγινε για λόγους απλότητας και εξοικείωσης με το ηχητικό τους περιεχόμενο. Με παρόμοια λογική θα μπορούσαμε να χρησιμοποιήσουμε οποιαδήποτε κλίμακα, αρκεί να ορίσουμε συχνοτικά τις σχέσεις μεταξύ των νοτών που την αποτελούν.

### 3.3 Ηχογράφηση Σήματος Εισόδου

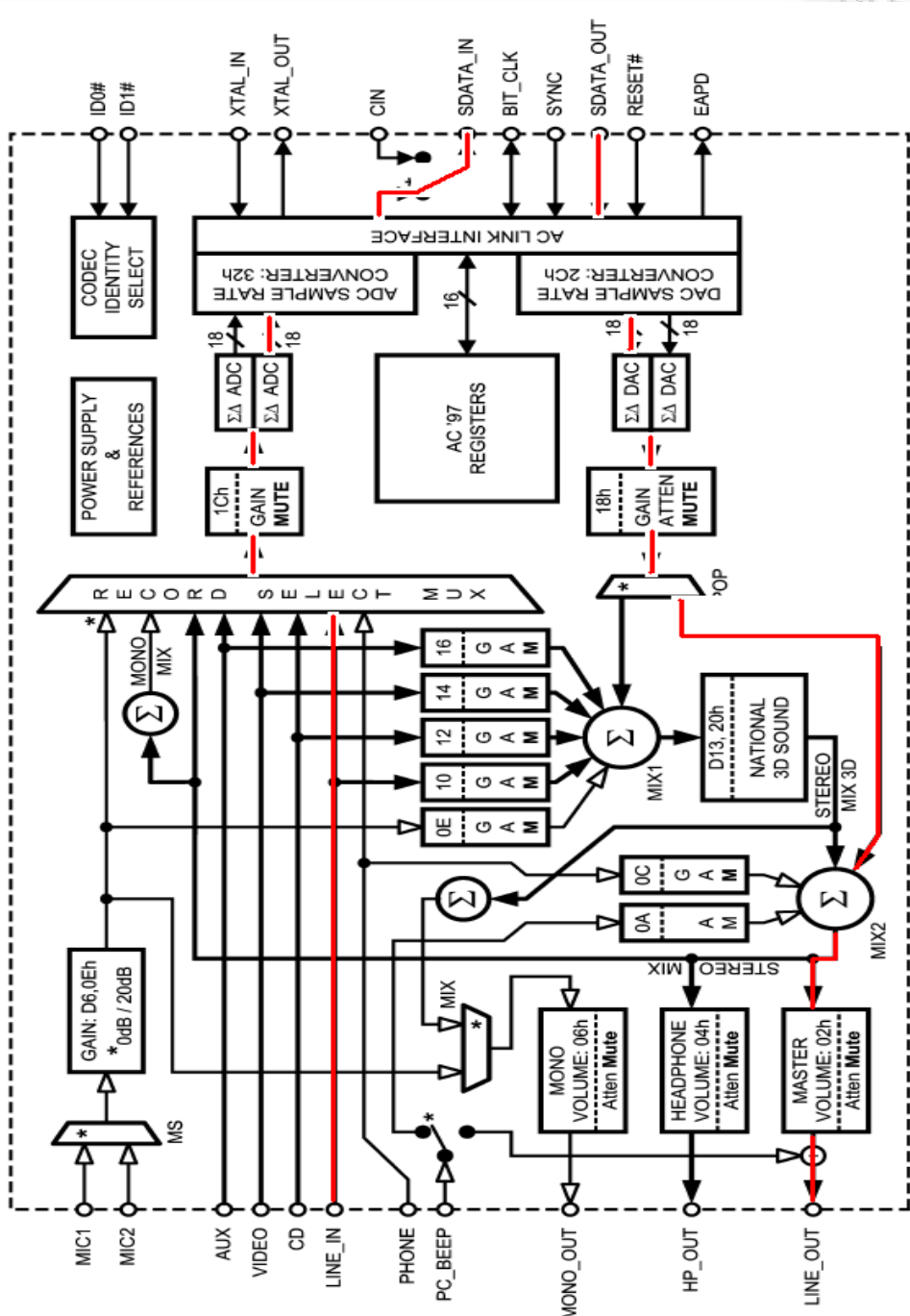
Ο κύκλος ανάπτυξης ενός ενσωματωμένου συστήματος ξεκινά με τον καθορισμό των προδιαγραφών των συστατικών του. Για τον λόγο αυτόν ξεκινάμε την περιγραφή της εφαρμογής με τα χαρακτηριστικά του ηχητικού σήματος εισόδου. Το σύστημα λαμβάνει το σήμα της ηλεκτρικής κιθάρας σε αναλογική μορφή από την LINE\_IN είσοδο του AC '97 Audio Codec. Ειδική μέριμνα λαμβάνεται πριν την ηχογράφηση, ώστε το τροφοδοτούμενο σήμα να έχει περάσει από στάδιο προενίσχυσης, και να έχει αρκετή ένταση ώστε να μην αλλοιώνεται από τον θόρυβο κατά την μεταφορά. Ταυτόχρονα η έντασή του ελέγχεται ώστε να μην εμφανιστεί clipping distortion. Το sound-to-noise ratio που προσφέρουν οι 18 – bit μετατροπείς του κυκλώματος δεν είναι αρκετά μικρό για να εξασφαλίσει την απόλυτα πιστή μετατροπή του σε ψηφιακό. Το στάδιο της ηχογράφησης αποτελεί το πλέον ευάλωτο στο θόρυβο σημείο της εφαρμογής.

Πρωταρχικό βήμα της ηχογράφησης του σήματος εισόδου αποτελεί η αρχικοποίηση του AC '97 Audio Codec με βάση τα επιθυμητά για την εφαρμογή χαρακτηριστικά. Ο ρυθμός δειγματοληψίας εισόδου και εξόδου ορίζεται στα 44,1kHz. Η τιμή αυτή επιλέγεται καθώς παρέχει αρκετά καλή ποιότητα ήχου για μουσικές εφαρμογές, παρέχει ικανοποιητική ανάλυση στην ανίχνευση της θεμελιώδους συχνότητας με βάση τον αλγόριθμο που θα υλοποιηθεί, στην περιοχή του οργάνου που εξετάζεται, και έχει ικανοποιητικά περιθώρια αύξησης και μείωσης του, κατά την φάση της διόρθωσης της συχνότητας, μέχρι να φτάσουμε στον μέγιστο ρυθμό δειγματοληψίας των 48kHz που επιτρέπει το ολοκληρωμένο. Επιλογή του ρυθμού των 48kHz δεν θα επέτρεπε διόρθωση του σήματος προς υψηλότερες τιμές της θεμελιώδους συχνότητας, ενώ επιλογή του αμέσως προηγούμενου καθιερωμένου ρυθμού των 22.05 kHz δεν δίνει ικανοποιητική ανάλυση κατά την ανίχνευση του pitch για μουσική (θα αποτελούσε ίσως καλή επιλογή αν η επεξεργασία αφορούσε φωνή). Η ανάλυση του ψηφιακού σήματος που παράγεται κατά την μετατροπή είναι 16 – bit. Δεν προτιμάται η παρερχόμενη από τον μετατροπέα ADC ανάλυση των 18 – bit για λόγους συμμετρίας, καθώς θα οδηγούσε σε σπατάλη πόρων του 32 – bit συστήματος, ενώ το ποιοτικό κέρδος δεν είναι αρκετά μεγάλο. Τέλος τα PCM δεδομένα που παρέχονται από το codec βρίσκονται σε μορφή προσημασμένων ακεραίων.

Το AC '97 Audio Codec διαθέτει στερεοφωνικούς μετατροπείς, και το σήμα εισόδου, αν και μονοφωνικό πολυπλέκεται κατά την μεταφορά του χρονικά με ένα ακόμα κανάλι ηχητικών δεδομένων (το οποίο δεν μεταφέρει χρήσιμη ηχητική πληροφορία). Η ύπαρξη του δεύτερου καναλιού δεν μπορεί να αποφευχθεί, αλλά δεν επηρεάζει την εφαρμογή, καθώς αυτό αγνοείται από το λογισμικό κατά την λήψη του.

Με βάση τα παραπάνω χαρακτηριστικά του σήματος εισόδου αρχικοποιούμε το audio codec μέσω της ρουτίνας init\_sound (44100) που ορίζεται στους οδηγούς λογισμικού του ελεγκτή OPB\_AC97\_Controller. Η ρουτίνα θέτει στους μετατροπείς τον επιθυμητό ρυθμό δειγματοληψίας, αφού ενεργοποιήσει πρώτα την λειτουργία VRA. Στην συνέχεια ορίζει το αναλογικό μονοπάτι του σήματος όπως αυτό φαίνεται στην εικόνα 38, φροντίζοντας να

αποσιωπήσει όλα τα μονοπάτια που δεν χρησιμοποιούνται. Επειδή έχει ληφθεί μέριμνα για τις όσο γίνεται σωστές στάθμες του σήματος, αυτό δεν ενισχύεται, ούτε εξασθενεί σε κανένα κύκλωμα. Το επεξεργασμένο από το σύστημα σήμα όταν μεταφερθεί κατά την αναπαραγωγή στο codec, θα ακολουθήσει την ενδεδειγμένη πορεία προς την έξοδο LINE\_OUT



**Εικόνα 38: Αναλογικό μονοπάτι ηχογράφησης αναπαραγωγής στο LM4550**

Μόλις ολοκληρωθεί η αρχικοποίηση του codec, οι οδηγοί μέσω της ρουτίνας `record_enable()` ενεργοποιούν την διαδικασία εγγραφής. Μόλις αυτό γίνει, το λογισμικό λαμβάνει τα ηχητικά δεδομένα από το codec (αγνοώντας το ένα κανάλι όπως προαναφέρθηκε) και αποθηκεύει το σήμα στην μνήμη RAM. Η μνήμη έχει χωρητικότητα 512k x 16 bit που αντιστοιχεί σε εγγραφή 11,8 δευτερολέπτων ήχου με τα προαναφερθέντα χαρακτηριστικά.

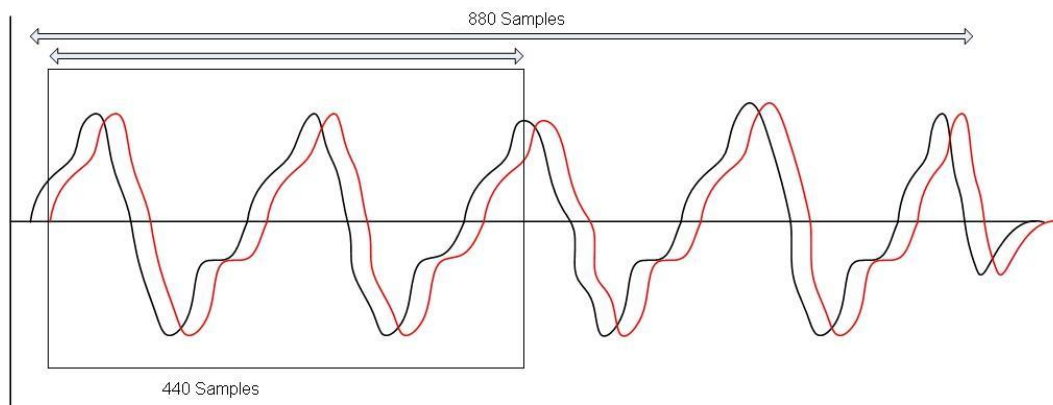
### 3.4: Αλγόριθμος Ανίχνευσης Θεμελιώδους Συχνότητας

Η επιλογή του κατάλληλου αλγορίθμου ανίχνευσης της θεμελιώδους συχνότητας, αποτελεί το πλέον κομβικό σημείο της εφαρμογής, καθώς η πιθανή διόρθωση της που θα γίνει στη συνέχεια βασίζεται σε δεδομένα που εξάγονται από αυτόν. Πιθανό λάθος στην εκτίμηση της θεμελιώδους συχνότητας οδηγεί σε αστοχία του συστήματος. Από την περιγραφή των χαρακτηριστικών του ήχου της ηλεκτρικής κιθάρας αλλά και από τα μειονεκτήματα του AC' 97 γίνεται αντιληπτό ότι απαιτείται η επιλογή ενός αλγορίθμου που προσφέρει ανοχή στο θόρυβο και αγνοεί όσο είναι δυνατόν το περίπλοκο συχνοτικό περιεχόμενο και την ισχυρή ατάκα της παλλόμενης χορδής της κιθάρας. Επιπλέον, η έλλειψη δυνατοτήτων περίπλοκων πράξεων από το υλικό ευνοεί έναν αλγόριθμο με όσο το δυνατόν απλούστερη υλοποίηση.

Με βάση τα προαναφερθέντα επελέγη μια παραλλαγή του αλγορίθμου Average Magnitude Difference Function (AMDF). Κατά την υλοποίηση του αλγορίθμου η επεξεργασία των ηχητικών δεδομένων και η ανάλυση που οδηγεί στην αναγνώριση της θεμελιώδους συχνότητας γίνεται σε πλαίσια σταθερού μήκους  $L$ . Σε κάθε πλαίσιο υπολογίζεται η συνάρτηση:

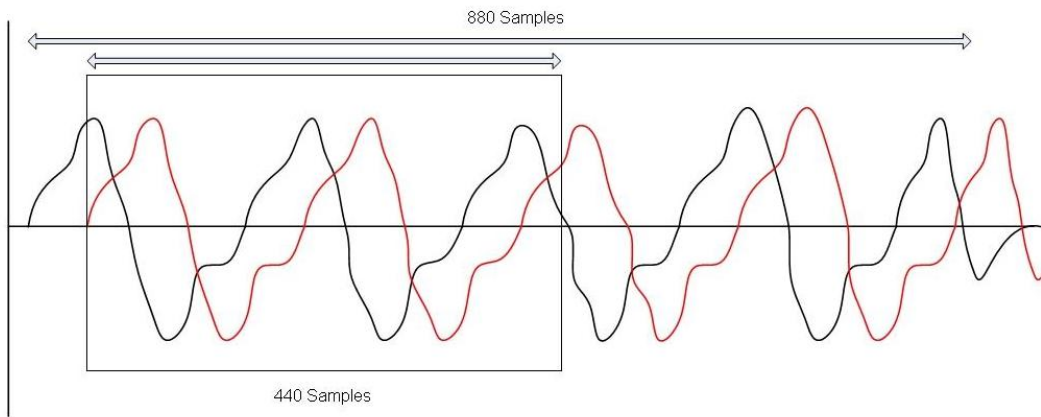
$$D_m = \sum_{n=1}^{\frac{L}{2}} |x(n) - x(n+m)|, \quad m_1 \geq m \geq \frac{L}{2}$$

για τιμές του  $m$  μεταξύ της παραμέτρου  $m_1$  και του μισού μήκους του πλαισίου. Με τον τρόπο αυτόν πετυχαίνουμε μια κανονικοποιημένη συνάρτηση  $D_m$  χωρίς την ανάγκη για διαίρεση με το μήκος του επικαλυπτόμενου παραθύρου όπως στον κλασικό AMDF. Η συνάρτηση  $D_m$  εκφράζει την αυτοσυσχέτιση του σήματος  $x(n)$  με μια μετατοπισμένη κατά  $m$  εκδοχή του εαυτού του. Η συνάρτηση διαθέτει όλα τα χαρακτηριστικά μιας τυπικής συνάρτησης αυτοσυσχέτισης. Είναι περιοδική με περίοδο ίδια με αυτή του σήματος  $x(n)$  και εμφανίζει τοπικά ελάχιστα σε όλα τα ακέραια πολλαπλάσια της περιόδου αυτής. Εντοπίζοντας λοιπόν το πρώτο τοπικό ελάχιστο που παρουσιάζει η συνάρτηση  $D_m$ , υπολογίζουμε την θεμελιώδη συχνότητα. Η εκτέλεση του αλγορίθμου παρουσιάζεται γραφικά στις εικόνες 39 - 47:

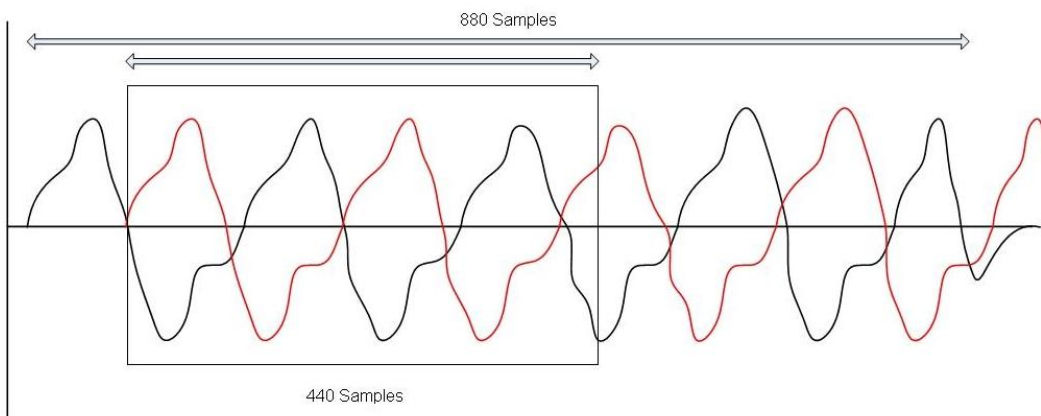


Εικόνα 39: Πρώτο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο

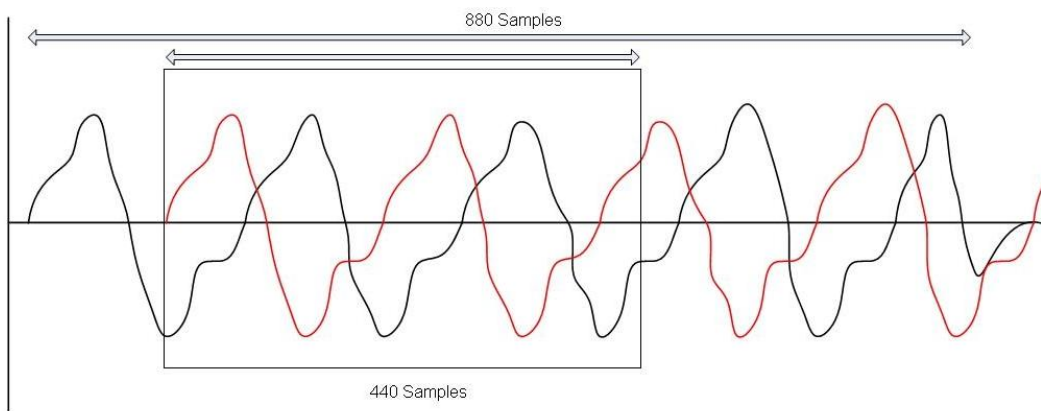




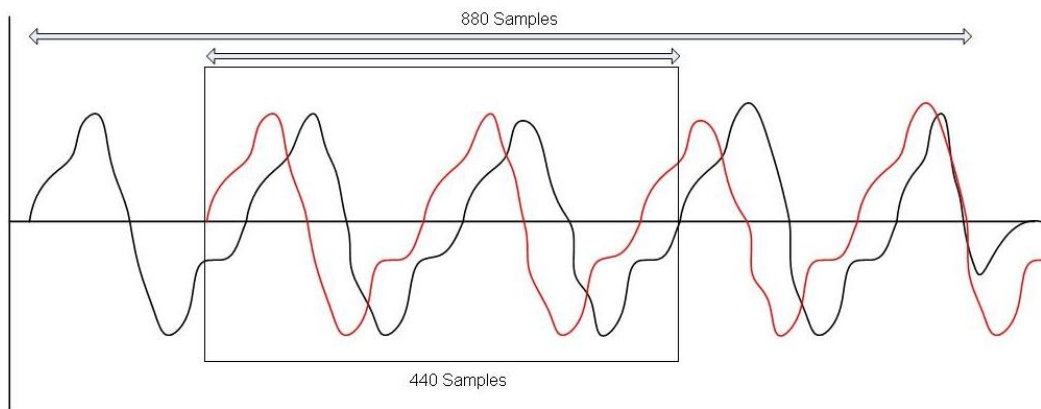
**Εικόνα 40: Δεύτερο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο**



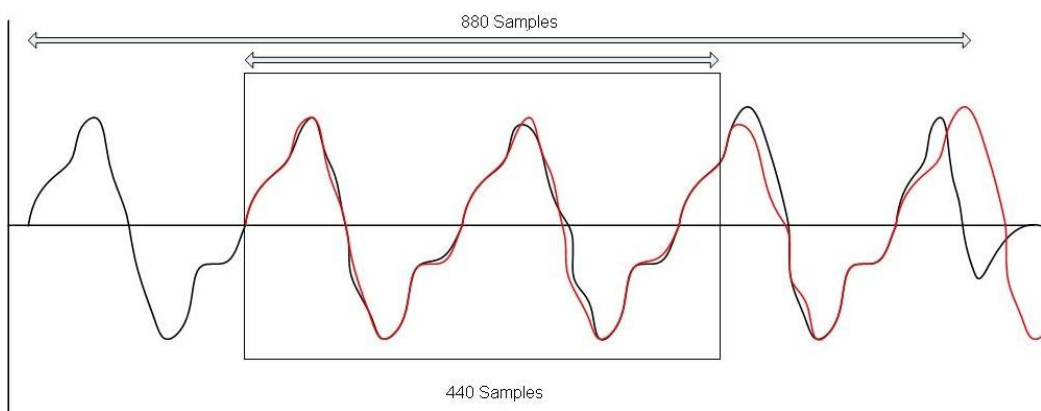
**Εικόνα 41: Τρίτο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο**



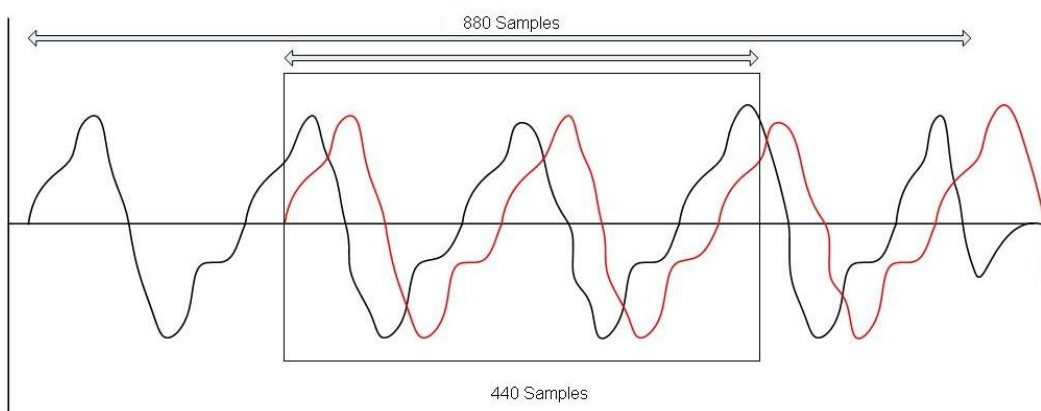
**Εικόνα 42: Τέταρτο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο**



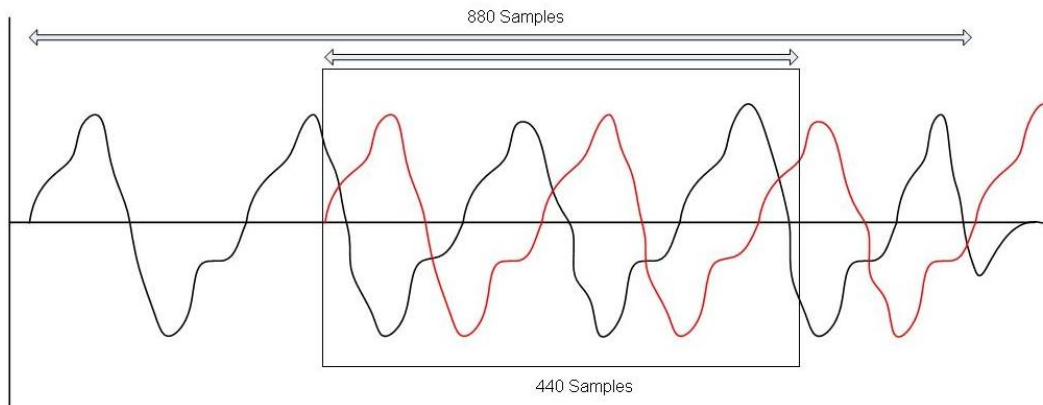
**Εικόνα 43: Πέμπτο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο**



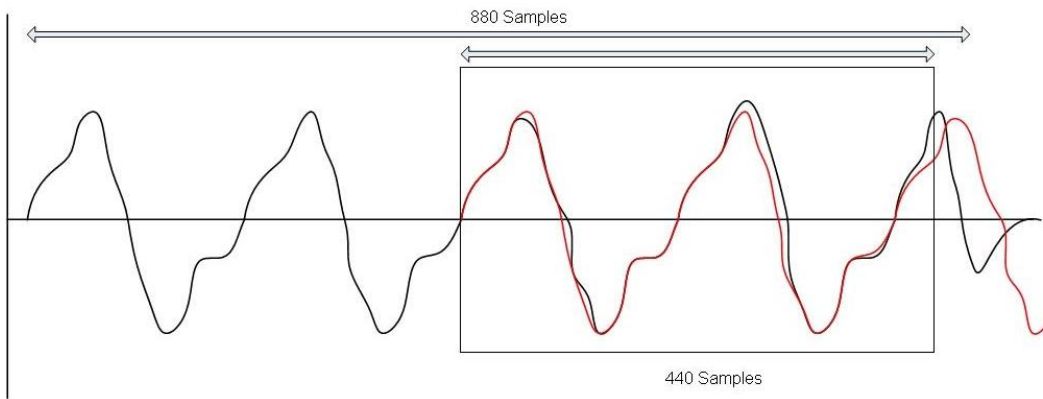
**Εικόνα 44: Έκτο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο**



**Εικόνα 45: Έβδομο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο**

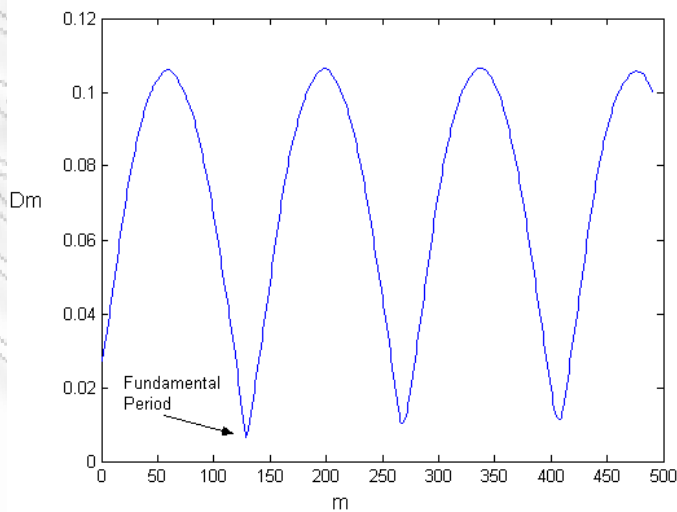


**Εικόνα 46: Όγδοο στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο**



**Εικόνα 47: Τελικό στάδιο υπολογισμού της AMDF συνάρτησης σε ένα πλαίσιο**

Η γραφική αναπαράσταση της συνάρτησης  $D_m$  εικονίζεται παρακάτω. Είναι εύκολο να παρατηρηθούν τα τοπικά ελάχιστα που αντιστοιχούν σε οκτάβες της θεμελιώδους συχνότητας.



**Εικόνα 48: Τυπική μορφή AMDF συνάρτησης σε ένα πλαίσιο**

Πολύ σημαντικό ρόλο στα αποτελέσματα της ανίχνευσης παίζει η ρύθμιση των παραμέτρων του αλγορίθμου και πιο συγκεκριμένα, ο καθορισμός του μήκους του πλαισίου L και η παράμετρος  $m_1$ . Ο αλγόριθμος υπολογίζει την θεμελιώδη συχνότητα ως αποτέλεσμα υπολογισμού της περιόδου στην οποία το σήμα εμφανίζει την μεγαλύτερη ομοιότητα με τον εαυτό του. Ο υπολογισμός της περιόδου μετρίεται σε απόσταση μετατόπισης σε δείγματα. Εφόσον ο ρυθμός δειγματοληψίας είναι σταθερός (και ίσος με 44,1kHz) είναι εύκολα αντιληπτό ότι η συχνότητα f υπολογίζεται σύμφωνα με τον τύπο:

$$f = \text{sample\_rate} / \#\text{samples}$$

όπου #samples είναι ο αριθμός των δειγμάτων που αντιστοιχεί στην ιδεατή μετατόπιση. Από τον παραπάνω τύπο και για τις ακραίες τιμές της παραμέτρου #samples (από 1 έως L/2) υπολογίζεται ο πίνακας που περιέχει τις τιμές συχνοτήτων που ανιχνεύει ο αλγόριθμος στον δεδομένο ρυθμό δειγματοληψίας. Από αυτόν τον πίνακα ορίζουμε και τις παραμέτρους L και  $m_1$  του αλγορίθμου, οι οποίες αν δοθούν στην θέση της μεταβλητής #samples δίνουν την ελάχιστη και την μέγιστη συχνότητα αντίστοιχα που αναγνωρίζει ο αλγόριθμος. Ο πίνακας απεικονίζει τις ανιχνευόμενες τιμές συχνοτήτων για ρυθμό δειγματοληψίας 44,1kHz στο εύρος  $40 \leq m \leq 200$ .

200	220,5	A3	150	294	D4	100	441	A4	50	882	A5
199	221,608		149	295,9732		99	445,4545		49	900	
198	222,7273		148	297,973		98	450		48	918,75	
197	223,8579		147	300		97	454,6392		47	938,2979	A#5
196	225		146	302,0548		96	459,375		46	958,6957	
195	226,1538		145	304,1379		95	464,2105	A#4	45	980	B5
194	227,3196		144	306,25		94	469,1489		44	1002,273	
193	228,4974		143	308,3916		93	474,1935		43	1025,581	
192	229,6875		142	310,5634	D#4	92	479,3478		42	1050	C6
191	230,8901		141	312,766		91	484,6154		41	1075,61	
190	232,1053		140	315		90	490		40	1102,5	C#6
189	233,3333	A#3	139	317,2662		89	495,5056	B4			
188	234,5745		138	319,5652		88	501,1364				
187	235,8289		137	321,8978		87	506,8966				
186	237,0968		136	324,2647		86	512,7907				
185	238,3784		135	326,6667		85	518,8235				
184	239,6739		134	329,1045	E4	84	525	C5			
183	240,9836		133	331,5789		83	531,3253				
182	242,3077		132	334,0909		82	537,8049				
181	243,6464		131	336,6412		81	544,4444				
180	245		130	339,2308		80	551,25	C#5			
179	246,3687	B3	129	341,8605		79	558,2278				
178	247,7528		128	344,5313		78	565,3846				
177	249,1525		127	347,2441		77	572,7273				
176	250,5682		126	350	F4	76	580,2632				
175	252		125	352,8		75	588	D5			
174	253,4483		124	355,6452		74	595,9459				
173	254,9133		123	358,5366		73	604,1096				
172	256,3953		122	361,4754		72	612,5				
171	257,8947		121	364,4628		71	621,1268	D#5			
170	259,4118		120	367,5		70	630				
169	260,9467	C4	119	370,5882	F#4	69	639,1304				
168	262,5		118	373,7288		68	648,5294				
167	264,0719		117	376,9231		67	658,209	E5			
166	265,6627		116	380,1724		66	668,1818				
165	267,2727		115	383,4783		65	678,4615				
164	268,9024		114	386,8421		64	689,0625				
163	270,5521		113	390,2655		63	700	F5			
162	272,2222		112	393,75	G4	62	711,2903				
161	273,913		111	397,2973		61	722,9508				
160	275,625		110	400,9091		60	735	F#5			
159	277,3585	C#4	109	404,5872		59	747,4576				
158	279,1139		108	408,3333		58	760,3448				
157	280,8917		107	412,1495		57	773,6842				
156	282,6923		106	416,0377	G#4	56	787,5	G5			
155	284,5161		105	420		55	801,8182				
154	286,3636		104	424,0385		54	816,6667				
153	288,2353		103	428,1553		53	832,0755	G#5			
152	290,1316		102	432,3529		52	848,0769				
151	292,053		101	436,6337		51	864,7059				

Στην υλοποίηση της εφαρμογής ανίχνευσης της θεμελιώδους συχνότητας για το σχεδιαζόμενο ενσωματωμένο σύστημα, το εύρος του πλαισίου ορίστηκε στα 880 δείγματα (αντιστοιχεί σε 20 ms ηχητικών δεδομένων) και η παράμετρος  $m_1$  ορίστηκε στα 40 δείγματα. Έτσι το θεωρητικό εύρος ανίχνευσης συχνοτήτων είναι από 100 Hz έως 1102,5 Hz

Ο αλγόριθμος αν και δεν χρησιμοποιεί βαριές υπολογιστικά πράξεις, είναι αρκετά απαιτητικός, καθώς σε κάθε πλαίσιο, για κάθε μετατόπιση  $m$  υπολογίζεται το άθροισμα  $L / 2$  απόλυτων διαφορών. Επιπλέον, αν και δεν χρησιμοποιούνται ολισθήσεις για την χρονική μετατόπιση του σήματος, αλλά μεταφορές από διαφορετικά τμήματα της μνήμης, το υπολογιστικό κόστος των μεταφορών δεν είναι αμελητέο. Αξίζει να σημειωθεί ότι η μνήμη RAM είναι το μόνο τμήμα του συστήματος στην παρούσα φάση επεξεργασίας το οποίο βρίσκεται εκτός του FPGA. Παρόλα αυτά χρησιμοποιούνται ορισμένες τεχνικές επιτάχυνσής του οι οποίες βελτιώνουν κατά πολύ την απόδοσή του.

Η πρώτη βασίζεται στην παρατήρηση ότι εφόσον αναζητούμε το πρώτο τοπικό ελάχιστο, δεν είναι απαραίτητο να υπολογίσουμε την συνάρτηση  $D_m$  για όλες τις τιμές του  $m$  μέχρι το  $L / 2$ . Αρκεί να ορίσουμε ένα κριτήριο για το τι συνιστά τοπικό ελάχιστο, το οποίο μόλις επαληθευτεί, τερματίζει η εκτέλεση του αλγορίθμου για το δεδομένο πλαίσιο. Η υλοποίηση αυτής της ιδέας γίνεται με την εισαγωγή της παραμέτρου  $threshold$ . Η παράμετρος αυτή, όπως ίσως γίνεται αντιληπτό από το όνομά της, συνιστά ένα κατώφλι, κάτω από το οποίο όταν βρεθεί τοπικό ελάχιστο, αυτό σηματοδοτεί την ολοκλήρωση της εκτέλεσης του αλγορίθμου για το δεδομένο πλαίσιο. Πειραματικές εκτελέσεις του αλγορίθμου με μεταβαλλόμενη παράμετρο  $threshold$  έδειξαν ότι η ιδέα δουλεύει αρκετά καλά και μπορεί να χρησιμοποιηθεί για να προσαρμόσει τον αλγόριθμο σε ποικιλία σημάτων εισόδου.

Η δεύτερη τεχνική επιτάχυνσης του αλγορίθμου βασίζεται στην παραδοχή ότι, μεταξύ διαδοχικών πλαισίων, για ένα όργανο όπως η ηλεκτρική κιθάρα δεν είναι πιθανό να παρατηρείται μεγάλη αλλαγή μεταξύ των θεμελιωδών συχνοτήτων που αντιστοιχούν σε αυτά. Έτσι, εφόσον σε ένα πλαίσιο ανιχνεύεται μια θεμελιώδης συχνότητα, η εκτέλεση του αλγορίθμου του επόμενου πλαισίου θα ξεκινήσει όχι από την αρχή, αλλά από τιμή του  $m$  κοντινή στην ανιχνευθείσα συχνότητα. Η χρήση των δύο τεχνικών επιτάχυνσης δίνει καλά αποτελέσματα αναφορικά με την απόδοση του αλγορίθμου, αλλά δεν εγγυάται τα αποτελέσματα αυτά καθώς η πιθανή επιτάχυνση εξαρτάται άμεσα από τα δεδομένα εισόδου.

### 3.5 Ανίχνευση Νοτών Μουσικής Φράσης

Η εκτέλεση του αλγορίθμου που περιγράφηκε στην προηγούμενη ενότητα δίνει ως αποτέλεσμα την ανιχνευθείσα θεμελιώδη συχνότητα για κάθε πλαίσιο ηχητικών δεδομένων, διάρκειας 20 ms. Λόγω όμως της φύσης του αλγορίθμου και του προβλήματος γενικότερα, είναι γεγονός ότι δεν έχουμε τρόπο να εξακριβώσουμε αν η ανιχνευθείσα τιμή για κάθε πλαίσιο αντιστοιχεί πράγματι στην κανονική του θεμελιώδη συχνότητα, ή είναι απλά μια λανθασμένη εκτίμηση. Αν σε αυτό συνυπολογίσουμε την χασομική φύση του οργάνου ή την αλλοίωση από τον θόρυβο κατά τη διαδρομή του ηχητικού σήματος στα αναλογικά και ψηφιακά κυκλώματα, συμπεραίνουμε ότι απαιτείται μια μέθοδος απομόνωσης τιμών που δεν συμβαδίζουν με τις γειτονικές τους.

Επιπλέον η διόρθωση της θεμελιώδους συχνότητας που ανιχνεύθηκε σε τμήμα του ηχητικού σήματος δεν έχει νόημα να γίνει για μικρές διάρκειες ηχητικών δεδομένων καθώς η μεταβολή δεν θα γίνει εύκολα αντιληπτή. Για την προετοιμασία της διόρθωσης της συχνότητας και της αναπαραγωγής των ηχητικών δεδομένων, τα αποτελέσματα του αλγορίθμου ανίχνευσης της συχνότητας εξομαλύνονται με χρήση ενός φίλτρου στατιστικού μέσου ( $median\ filter$ ) πλάτους 5 και διάστασης 1. Το φίλτρο στατιστικού μέσου πλάτους 5, σαρώνει σειριακά τα αποτελέσματα της ανίχνευσης θεμελιώδους συχνότητας των ηχητικών δειγμάτων σε ομάδες των 5. Οι τιμές κάθε ομάδας διατάσσονται από την μικρότερη στη μεγαλύτερη και επιλέγεται η μεσαία. Στη συνέχεια το φίλτρο προχωράει στην επόμενη ομάδα τιμών μέχρι την πλήρη σάρωση όλων των αποτελεσμάτων. Με τον τρόπο αυτόν παράγεται ένα νέο σύνολο τιμών το οποίο παρουσιάζει εξομάλυνση ακραίων τιμών. Πειραματικά δεδομένα από συνεχείς εκτελέσεις του αλγορίθμου αναγνώρισης συχνότητας έδειξαν ότι παρατηρούνται ακραίες τιμές στο τέλος

κάθε νότας της κιθάρας. Η πιθανότερη αιτία του φαινομένου είναι η αλλοίωση του ήχου από τα χέρια του μουσικού που σηκώνονται από τις χορδές. Το φίλτρο στατιστικού μέσου απομονώνει τέτοιες ακραίες και μικρής διάρκειας τιμές. Η υλοποίηση του φίλτρου στατιστικού μέσου σε λογισμικό έγινε με χρήση του αλγορίθμου quicksort της standard βιβλιοθήκης της C για την διάταξη των τιμών.

Μετά την εφαρμογή του φίλτρου στατιστικού μέσου, τα όμοια συνεχόμενα αποτελέσματα ομαδοποιούνται. Θεωρούμε ότι όταν η ίδια τιμή θεμελιώδους συχνότητας έχει ανιχνευθεί για πάνω από τρία συνεχόμενα πλαίσια ηχητικών δεδομένων (διάρκεια ήχου 60ms), έχουμε μια νότα της μουσικής φράσης, η οποία στο επόμενο στάδιο της εφαρμογής θα εξεταστεί αν ανήκει στην δεδομένη μουσική κλίμακα και θα αναπαραχθεί ως έχει, ή αν θα διορθωθεί.

### 3.6: Αλγόριθμος Διόρθωσης της Θεμελιώδους Συχνότητας και Αναπαραγωγή του Ηχητικού Σήματος

Ως διόρθωση της θεμελιώδους συχνότητας ορίζουμε τον μετασχηματισμό του ηχητικού σήματος έτσι ώστε η νέα του θεμελιώδης συχνότητα να αντιστοιχεί σε μια καθορισμένη δομή. Όπως αναφέρθηκε προηγουμένως, χρησιμοποιούμε μια δυτική οκτατονική κλίμακα και συγκεκριμένα την Ντο Μείζονα. Ο πίνακας παρουσιάζει τις συχνότητες που αντιστοιχούν στις νότες που περιλαμβάνει η Ντο Μείζονα κλίμακα σε εύρος μιας οκτάβας με πορτοκαλί χρώμα. Από τις υπόλοιπες πιθανές ανιχνεύσιμες συχνότητες, όσες σημειώνονται με πράσινο χρώμα διορθώνονται στην χαμηλότερη πιο κοντινή νότα της κλίμακας, ενώ όσες σημειώνονται με μπλέ χρώμα, διορθώνονται στην υψηλότερη πιο κοντινή νότα της κλίμακας.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	200		220.5 A3			150		294 D4			100	44100	441 A4			50		882 A5	
2	199		221,608			149		295,9732			99	43606	445,4545			49		900	
3	198		222,7273			148		297,973			98	43111	450			48		918,75	
4	197		223,8579			147		300			97	42616	454,6392			47		938,2979 A#5	
5	196		225			146		302,0548			96	42121	459,375			46		958,6957	
6	195		226,1538			145		304,1379			95	41626	464,2105 A#4			45		980 B5	
7	194		227,3196			144		306,25			94	46724	469,1489			44		1002,273	
8	193		228,4974			143		308,3916			93	46200	474,1935			43		1025,581	
9	192		229,6875			142		310,5634 D#4			92	45675	479,3478			42		1050 C6	
10	191		230,8901			141		312,766			91	45150	484,6154			41		1075,61	
11	190		232,1053			140		315			90	44625	490			40		1102,5 C#6	
12	189		233,3333 A#5			139		317,2662			89	44100	495,5056 B4						
13	188		234,5745			138		319,5652			88	43482	501,1364						
14	187		235,8289			137		321,8978			87	42865	506,8966						
15	186		237,0968			136		324,2647			86	45378	512,7907						
16	185		238,3784			135		326,6667			85	44739	518,8236						
17	184		239,6739			134	44100	529,1045 E4			84	44100	525 C5						
18	183		240,9836			133	43792	531,5789			83	43480	531,3253						
19	182		242,3077			132	43483	534,0909			82	42862	537,9049						
20	181		243,6464			131	43174	536,6412			81	42244	544,4444						
21	180		245			130	42865	539,2306			80	41626	551,25 C#5						
22	179		246,3687 B3			129	45378	541,8605			79	46724	558,2270						
23	178		247,7528			128	44952	544,5313			78	46068	565,3846						
24	177		249,1525			127	44526	547,2441			77	45412	572,7273						
25	176		250,5682			126	44100	550 F4			76	44756	580,2632						
26	175		252			125	43744	552,8			75	44100	588 D5						
27	174		253,4483			124	43391	555,6452			74	43480	595,9459						
28	173		254,9133			123	43038	558,5396			73	42862	604,1096						
29	172		256,3953			122	42685	561,4754			72	42244	612,5						
30	171		257,8947			121	42332	564,4628			71	41626	621,1268 D#5						
31	170		259,4118			120	41979	567,5			70	46724	630						
32	169		260,9467 C4			119	41626	570,5882 F#4			69	45850	639,1304						
33	168		262,5			118	46724	573,7288			68	44975	648,5294						
34	167		264,0719			117	46265	576,9231			67	44100	658,209 E5						
35	166		265,6627			116	45848	580,1724			66		668,1818						
36	165		267,2727			115	45411	583,4783			65		678,4615						
37	164		268,9024			114	44974	586,8421			64		689,0625						
38	163		270,5521			113	44537	590,2655			63		700 F5						
39	162		272,2222			112	44100	593,75 G4			62		711,2903						
40	161		273,913			111	43686	597,2973			61		722,9508						
41	160		275,625			110	43274	600,9091			60		735 F#5						
42	159		277,3585 C#4			109	42862	604,5872			59		747,4576						
43	158		279,1139			108	42450	608,3333			58		760,3448						
44	157		280,8917			107	42038	612,1495			57		773,6842						
45	156		282,6923			106	41626	616,0377 G#4			56		787,5 G5						
46	155		284,5161			105	46724	620			55		801,8182						
47	154		286,3636			104	46200	624,0395			54		816,6667						
48	153		288,2353			103	45675	628,1553			53		832,0755 G#5						
49	152		290,1316			102	45150	632,3529			52		848,0769						
50	151		292,053			101	44625	636,6337			51		864,7059						
51																			

Η διόρθωση της θεμελιώδους συχνότητας γίνεται σε ευρύτερα χρονικά πλαίσια από ό τι η ανίχνευσή της, η οποία εκτελούνταν αυστηρά για κάθε 20 ms ηχητικών δεδομένων. Ακολουθώντας την ομαδοποίηση των πλαισίων σε νότες, ο αλγόριθμος επεξεργάζεται κάθε νότα ξεχωριστά, ανεξάρτητα από την διάρκειά της. Ο αλγόριθμος που χρησιμοποιείται ανήκει στην οικογένεια αλγορίθμων Time Domain Harmonic Scaling, και χρησιμοποιεί την πληροφορία της ανίχνευσης της θεμελιώδους συχνότητας από το προηγούμενο στάδιο για να παράγει ένα ομαλό ηχητικά αποτέλεσμα.

Ο αλγόριθμος βασίζεται στην τεχνική του time – pitch scaling που περιγράφηκε λεπτομερώς στο κεφάλαιο 1. Για την παραγωγή ομαλού ηχητικού αποτελέσματος χωρίς ασυνέχειες πρέπει κατά την εφαρμογή της συμπίεσης ή εξάπλωσης της χρονικής διάρκειας των δεδομένων, τα επικαλυπτόμενα τμήματά τους να ταιριάζουν όσο το δυνατόν περισσότερο. Ο αλγόριθμος χρησιμοποιεί την ανιχνευμένη θεμελιώδη συχνότητα για να εξασφαλίσει ότι, αν η επικαλυπτόμενη περιοχή αντιστοιχεί ακριβώς σε μία περίοδο του σήματος, η πιθανότητα ομοιότητας είναι αρκετά μεγάλη.

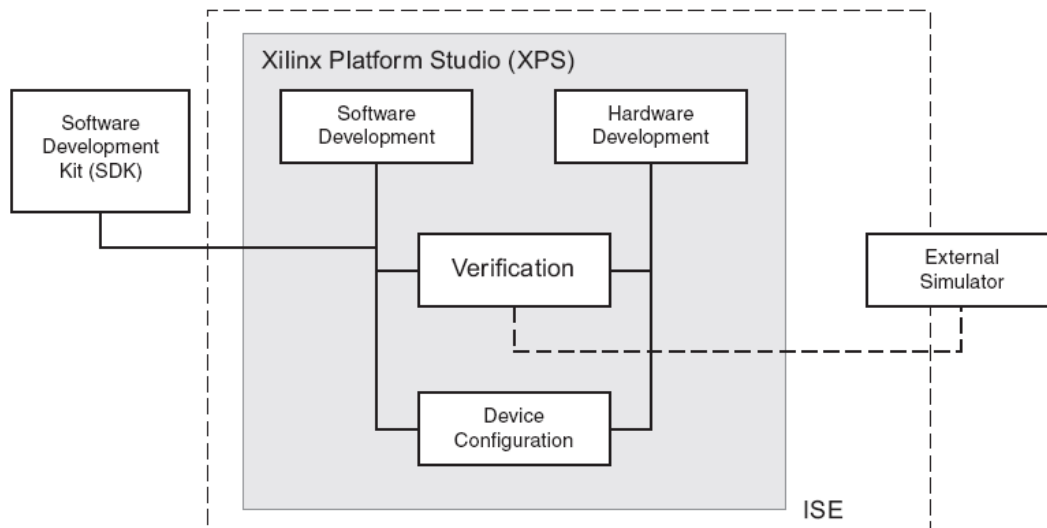
Το ενδιαφέρον τμήμα της υλοποίησης του αλγορίθμου βρίσκεται στο γεγονός ότι το τελικό βήμα της αλλαγής του ρυθμού δειγματοληψίας για την μετατόπιση της θεμελιώδους συχνότητας του σήματος, υλοποιείται στο υλικό. Χρησιμοποιώντας την VRA λειτουργία του AC '97 codec και ενώ η ηχογράφηση των δεδομένων έγινε με σταθερό ρυθμό δειγματοληψίας στα 44,1 kHz, υπολογίζουμε το άλμα της μετατόπισης της θεμελιώδους συχνότητας και στέλνουμε εντολή στο codec να εκτελέσει αναπαραγωγή των δεδομένων με διαφορετικό ρυθμό δειγματοληψίας. Η ενέργεια αυτή σε συνδυασμό με την μεταβολή του χρόνου διάρκειας του ηχητικού σήματος αναπαράγει στην έξοδο του codec το διορθωμένο ηχητικό σήμα εξόδου.

## Κεφάλαιο 4: Σχεδιαστική Ροή, Εργαλεία υλοποίησης, Συμπεράσματα

Το ενσωματωμένο σύστημα ανίχνευσης και διόρθωσης θεμελιωδών συχνοτήτων μουσικών σημάτων σχεδιάστηκε και υλοποιήθηκε στο εργαλείο Embedded Development Kit (EDK) Version 8.1i της εταιρίας Xilinx. Το EDK αποτελεί ένα ολοκληρωμένο περιβάλλον ανάπτυξης ενσωματωμένων συστημάτων καθώς περιλαμβάνει μια πλήρη συλλογή εργαλείων σχεδίασης, υλοποίησης, προγραμματισμού, αποσφαλμάτωσης και πρωτότυποποίησης ενός λειτουργικού ενσωματωμένου συστήματος σε ένα FPGA. Ταυτόχρονα προσφέρει και μια βάση δεδομένων πυρήνων IP για εύκολη και γρήγορη ανάπτυξη πλήθους συστημάτων, διαφορετικών προδιαγραφών. Το περιβάλλον EDK προσφέρει στον σχεδιαστή ευκολίες καθώς ευνοεί την διαδικασία ανάπτυξης ενός συστήματος, δομώντας τη πάνω σε συγκεκριμένες σχεδιαστικές ροές, ανάλογα με τις προδιαγραφές του.

### 4.1: Ολοκληρωμένο Περιβάλλον Σχεδίασης EDK

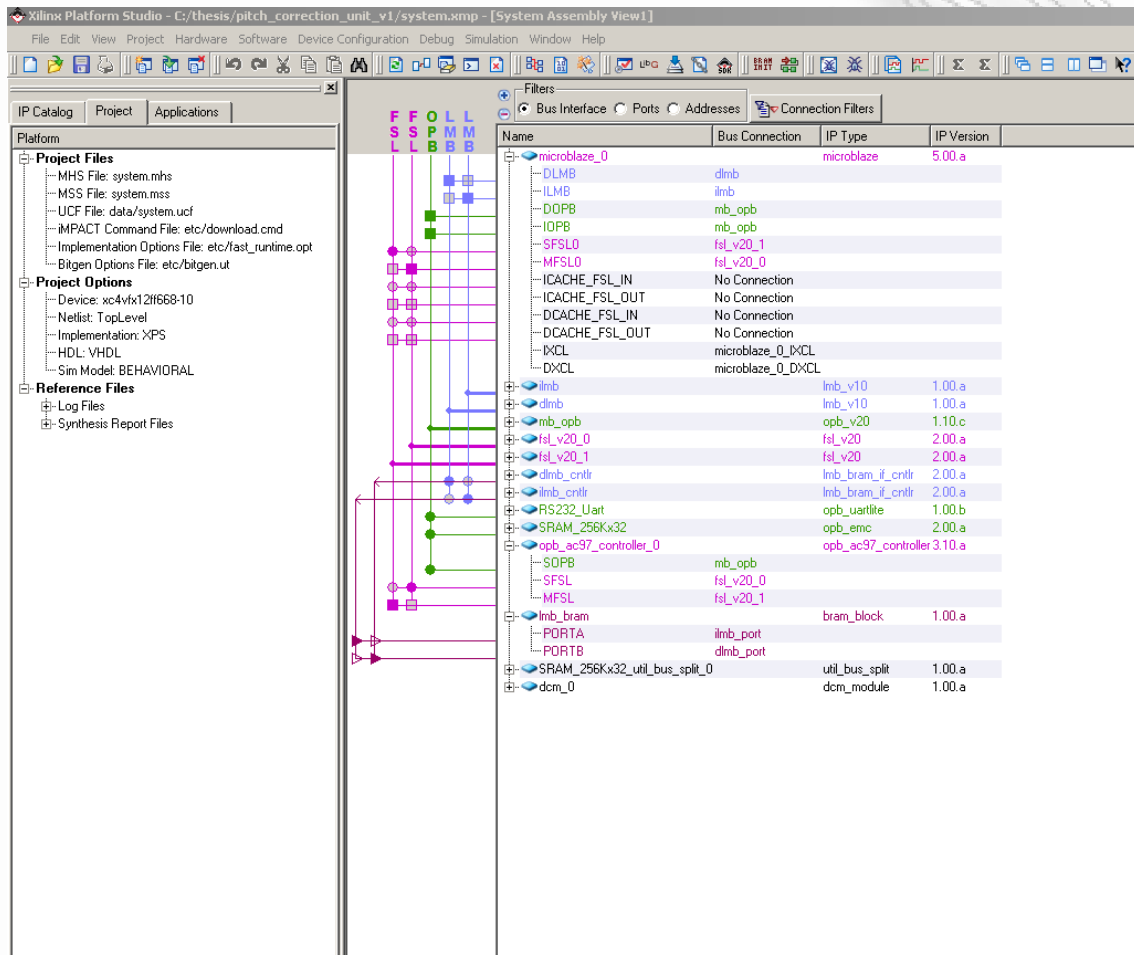
Στην εικόνα παρουσιάζεται η βασική δομή του περιβάλλοντος του EDK. Η σχεδίαση ξεκινά από το Xilinx Platform Studio, το οποίο είναι υπεύθυνο για την ανάπτυξη της βασικής πλατφόρμας του υλικού, πάνω στο οποίο θα χτιστεί το σύστημα. Το XPS περιβάλλον παρέχει στον σχεδιαστή εποπτεία του συστήματος, κάποια περιορισμένα εργαλεία ανάπτυξης εφαρμογών, σε περίπτωση που αυτό απαιτεί hardware / software co-development, καθώς και επιλογές για επιβεβαίωση της ορθής λειτουργίας του συστήματος, σε συνεργασία με κάποιο third-party πρόγραμμα εξομοίωσης. Δίνεται η δυνατότητα χρήσης έτοιμων, επαληθευμένων και αρκετά παραμετροποιήσιμων IP πυρήνων, ενώ παρέχονται και τα κατάλληλα εργαλεία σχεδίασης και χρήσης custom IP cores. Το XPS χρησιμοποιεί τα εργαλεία του ISE, παρέχοντας ένα γνώριμο περιβάλλον σε σχεδιαστές με εμπειρία στη χρήση του.



Εικόνα 49: Δομή του Embedded Development Kit



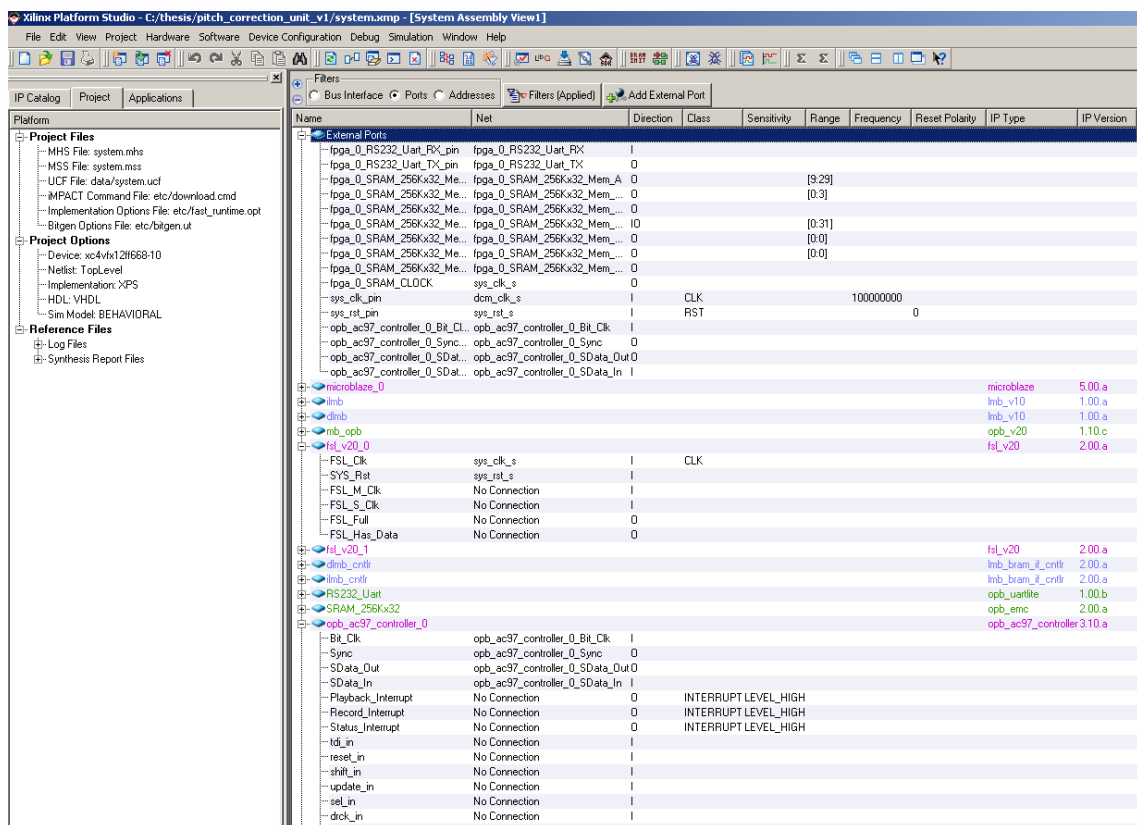
Όταν ολοκληρωθεί η σχεδίαση του υλικού, δίνεται η δυνατότητα χρήσης του Software Development Kit (SDK) που αποτελεί ένα πιο πλήρες περιβάλλον ανάπτυξης και αποσφαλμάτωσης λογισμικού. Το SDK παρέχει δυνατότητες αποσφαλμάτωσης σε πραγματικό χρόνο ή offline με χρήση ιδεατών συστημάτων εξομοίωσης. Παρέχονται δυνατότητες βελτιστοποίησης του κώδικα και ανάλυσης του με εξαγωγή μετρικών σχετικά με επιμέρους χρόνους εκτέλεσης των τμημάτων του.



Εικόνα 50: Στιγμιότυπο Χρήσης του XPS

Στην εικόνα 50 παρουσιάζεται στιγμιότυπο από την χρήση του περιβάλλοντος XPS κατά την ανάπτυξη του συστήματος. Απεικονίζεται η δομή του ενσωματωμένου συστήματος και η διασύνδεση των επιμέρους περιφερειακών στον επεξεργαστή μέσω των OPB, FSI και LMB διαύλων.

Στην εικόνα 51 εμφανίζεται η όψη διαχείρισης των ports των επιμέρους υπομονάδων του συστήματος. Η πραγματοποίηση αλλαγών η εύκολη αναδιάρθρωση του συστήματος αποτελεί χρήσιμο εργαλείο για τον σχεδιαστή, ιδιαίτερα στα πρώτα στάδια ανάπτυξης, όταν και το υλικό δεν έχει ολοκληρωθεί.



Εικόνα 51: Στιγμιότυπο Χρήσης του XPS

Η διαδικασία της σύνθεσης και η χρονική ανάλυση των εργαλείων έδειξαν μέγιστη συχνότητα λειτουργίας 103,749 MHz, αποτέλεσμα θετικό, καθώς η συχνότητα λειτουργίας του επηρεργαστή και του OPB διαύλου είχαν οριστεί στα 100 MHz.

**Timing Summary:**

-----

Speed Grade: -10

Minimum period: 9.639ns (Maximum Frequency: 103.749MHz)

Minimum input arrival time before clock: 3.748ns

Maximum output required time after clock: 4.867ns

Maximum combinational path delay: No path found

**Timing Detail:**

-----

All values displayed in nanoseconds (ns)

Το utilization της συσκευής μετά την διαδικασία υλοποίησης και την αντιστοίχιση του σχεδιασμού στην συσκευή είναι σχετικά χαμηλό, αποτέλεσμα αναμενόμενο καθώς δεν σχεδιάστηκε κάποια υπομονάδα σε υλικό.

## Device Utilization Summary:

Number of BUFGs	2 out of 32	6%
Number of DCM_ADVs	1 out of 4	25%
Number of DSP48s	3 out of 32	9%
Number of ILOGICs	34 out of 320	10%
Number of External IOBs	70 out of 320	21%
Number of LOCed IOBs	70 out of 70	100%
Number of OLOGICs	62 out of 320	19%
Number of RAMB16s	10 out of 36	27%
Number of Slices	1827 out of 5472	33%
Number of SLICEMs	156 out of 2736	5%

Overall effort level (-ol): High  
 Placer effort level (-pl): High  
 Placer cost table entry (-t): 1  
 Router effort level (-rl): High

## 4.2: Συμπεράσματα - Βελτιώσεις

Η διαδικασία ανάπτυξης ενός ενσωματωμένου συστήματος από την αρχή είναι μια εμπειρία μοναδική για κάποιον μηχανικό, καθώς η επιτυχία ή η αποτυχία του συστήματος, η φιλικότητα προς τον χρήστη, η καλή απόδοση και η σταθερή λειτουργία εξαρτώνται από πλήθος μικρών και μεγάλων σχεδιαστικών αποφάσεων σε κάθε στάδιο του κύκλου κατασκευής. Παρά τη όποια πολυπλοκότητα όμως, παρά την έλλειψη εμπειρίας σε παρόμοιες σχεδιαστικές ροές και συστήματα, ο μηχανικός αντιλαμβάνεται την δύναμη και την ευελιξία που του παρέχεται από ένα περιβάλλον σχεδιασμού ενσωματωμένων συστημάτων, στο οποίο δεν υπάρχουν περιορισμοί σταθερού υλικού ή επιλογών αρχιτεκτονικής, και το μόνο όριο είναι η φαντασία του και η σχεδιαστική του ικανότητα.

Σε ότι αφορά το ενσωματωμένο σύστημα ανίχνευσης και διόρθωσης της θεμελιώδους συχνότητας μουσικών σημάτων, θα παρουσίαζε μεγάλο ενδιαφέρον η υλοποίηση του αλγορίθμου ανίχνευσης AMDF σε υλικό, καθώς δίνεται η ευκαιρία εκμετάλλευσης της μεγάλης παραλληλίας των πράξεων του αλγορίθμου. Επιπρόσθετα, θα άξιζε να ερευνηθούν εναλλακτικοί τρόποι προστασίας του συστήματος από τον θόρυβο (χρήση ενός δεύτερου αλγορίθμου ανίχνευσης θεμελιώδους συχνότητας για περαιτέρω πληροφορία, εξάλειψη του AC '97 για recording και ψηφιακή είσοδος δεδομένων). Τέλος θα ήταν ενδιαφέρον να μελετηθεί η εφαρμογή στην πλήρη έκταση της ηλεκτρικής κιθάρας.

## **Βιβλιογραφία**

- [1] FPGA – based Implementation of Signal Processing Systems, Roger Woods, John McAllister, Gaye Lightbody, Ying Yi, 2008
- [2] Pitch Neural Coding And Perception, Christofer J. Plack, Andrew J. Oxenham, Richard R. Fay, Arthur N Popper, 2005
- [3] Pitch Perception Models, Alain de Cheveigné
- [4] A Comparative Performance Study Of Several Pitch Detection Algorithms, Lawrence R. Rabiner
- [5] Pitch Detection Algorithm: Autocorrelation Method And AMDF, Li Tan, Montri Karnjanadecha
- [6] Efficient Pitch Detection Techniques For Interactive Music, Patricio De La Cuadra, Aaron Master, Craig Saap
- [7] Audio Time Scale Modification Using A Hybrid Time-Frequency Domain Approach, David Dorran, Eugene Coyle
- [8] Audio Codec '97 Revision 2.3 , Intel 2002
- [9] LM 4550 Datasheet, National Semiconductors, 2004
- [10] EDK: Concepts Tools And Techniques, Xilinx 2007
- [11] Microblaze Processor Reference Guide, Xilinx, 2008

## Παράρτημα Α: Κώδικας Εφαρμογής

```

#include "xparameters.h"
#include "xbasic_types.h" /* Xilinx basic types for device drivers
*/
#include "xac97.h"
#include "xac97_l.h"
#include "xac97_i.h"
#include "xuartlite_1.h"
#include "fsl.h"
#include "xstatus.h"
#include "stdlib.h"

#define TRUE 1
#define FALSE 0
#define printf xil_printf

/* qsort int comparison function */
int int_cmp(const void *a, const void *b)
{
    const int *ia = (const int *)a; // casting pointer types
    const int *ib = (const int *)b;
    return *ia - *ib;
    /* integer comparison: returns negative if b > a
    and positive if a > b */
}

Xuint16* sound = (Xuint16*)XPAR_SRAM_256KX32_MEM0_BASEADDR ;

XAC97 Status97;

int main (void)
{
    XStatus status1;
    XStatus status2;
    printf("%c[2J",27);

    //-----AC97 Initialization-----
    status1 = XAC97_Initialize (&Status97,
XPAR_OPB_AC97_CONTROLLER_0_DEVICE_ID );

```

```

    if (status1 != XST_SUCCESS)
    {printf("fail\n\r");
return XST_FAILURE;
}
else printf("AC97 Initialization...Success\n\r");

status2 = XAC97_SelfTest (&Status97 );
if (status2 != XST_SUCCESS)
{printf("fail\n\r");
return XST_FAILURE;
}
else printf("AC97 Self Test...Success\n\r");

//-----Recording-----
-----

Xuint16 soundbyte;
Xuint16 soundbyte2;

printf("Starting Application\n\r");

init_sound (44100);
printf("AC97 Configured At 44100 Hz\n\r");

record_enable();
printf("Recording Enabled\n\r");

printf("Recording\n\r" );

while( sound < (Xuint16*)XPAR_SRAM_256KX32_MEM0_HIGHADDR )
{
    getfsl (soundbyte, 0);
    getfsl (soundbyte2, 0); //to soundbyte 2 periexei
to left channel pou agnoeitai
    *sound = soundbyte;
    sound++;
}

printf("recorded\n\r" );
record_disable();

```

```

//-----Picth Detection Software-----
-----
unsigned int sample_dist=13;          // deixnei thn periodo
unsigned int T_prev=0;

unsigned int bitwise_sum=0;
unsigned int sum=0;
unsigned int previous_sum=0;
unsigned int count=0;                //counter mesa sto frame
unsigned int count2=0;              //counter gia to frame_T array
unsigned int window_offset=0;      //offset between windows
unsigned int threshold=0x00300000; //threshold
stable 0x00150000
unsigned int freq;

unsigned int frame_T[200];          //array pou apothikeuei thn
periodo kathe frame
unsigned int frame_T_filtered[200]; //array pou apothikeuei thn
periodo kathe frame meta to median filtering, to sample rate gia to
correction, kai thn target freq

for (count2=0; count2<200; count2++) //arxikopoihsh tou array
{
    frame_T[count2]= 0;
    frame_T_filtered[count2]= 0;
}

unsigned int scale_array[68][2] = { {134, 44100},          //C
major E4 - E5

    {133, 43792},
    {132, 43483},
    {131, 43174},
    {130, 42865},
    {129, 45378},
    {128, 44952},
    {127, 44526},
    {126, 44100},
    {125, 43744},
    {124, 43391},
    {123, 43038},
    {122, 42685},
    {121, 42332},

```

{120, 41979},  
{119, 41626},  
{118, 46724},  
{117, 46285},  
{116, 45848},  
{115, 45411},  
{114, 44974},  
{113, 44537},  
{112, 44100},  
{111, 43686},  
{110, 43274},  
{109, 42862},  
{108, 42450},  
{107, 42038},  
{106, 41626},  
{105, 46724},  
{104, 46200},  
{103, 45675},  
{102, 45150},  
{101, 44625},  
{100, 44100},  
{99, 43606},  
{98, 43111},  
{97, 42616},  
{96, 42121},  
{95, 41626},  
{94, 46724},  
{93, 46200},



```
{ 92, 45675},  
{ 91, 45150},  
{ 90, 44625},  
{ 89, 44100},  
{ 88, 43482},  
{ 87, 42865},  
{ 86, 45378},  
{ 85, 44739},  
{ 84, 44100},  
{ 83, 43480},  
{ 82, 42862},  
{ 81, 42244},  
{ 80, 41626},  
{ 79, 46724},  
{ 78, 46068},  
{ 77, 45412},  
{ 76, 44756},  
{ 75, 44100},  
{ 74, 43480},  
{ 73, 42862},  
{ 72, 42244},  
{ 71, 41626},  
{ 70, 46724},  
{ 69, 45850},  
{ 68, 44975},  
{ 67, 44100} };
```

```
count2=0;

printf("Detecting Pitch...\n\r" );

while (window_offset <= 176050) //200 windows
{
    sample_dist=40;
    previous_sum = threshold;

    //if (T_prev > 56)
    //    sample_dist = T_prev - 30; //we assume that the
frequency

    Xuint16* set1_ptr = (Xuint16*)XPAR_SRAM_256KX32_MEM0_BASEADDR +
window_offset;
    Xuint16* set2_ptr = (Xuint16*)XPAR_SRAM_256KX32_MEM0_BASEADDR +
window_offset + sample_dist;

    while (sample_dist < 440)
    {

        sample_dist++;
        set1_ptr = (Xuint16*)XPAR_SRAM_256KX32_MEM0_BASEADDR +
window_offset; //o set1_ptr sthn arxh tou window
        set2_ptr = (Xuint16*)XPAR_SRAM_256KX32_MEM0_BASEADDR +
window_offset + sample_dist;

        sum = 0;
        for(count=1; count<=440; count++) //
        {
            if (*set1_ptr >= *set2_ptr)
                bitwise_sum = *set1_ptr - *set2_ptr;
            else
                bitwise_sum = *set2_ptr - *set1_ptr;

            sum = sum + bitwise_sum;

            set1_ptr++;
            set2_ptr++;
        }

        if ((sum < threshold) && (sum < previous_sum))
            previous_sum = sum;

        if ((sum < threshold) && (sum > previous_sum))
            break;
    }

    if ((sample_dist == 440) || (sample_dist<=48))

        {
```

```

        T_prev = 0;
        freq=0;
    }
    else
    {
        T_prev = sample_dist - 1;
        frame_T[count2] = T_prev;
        //freq = 44100/T_prev;
    }

    //printf("%d / ", freq);

window_offset = window_offset + 880;
count2++;

}
/*
//-----print pitch array-----
for (count2=0; count2<200; count2++)
    {
        printf("%d / ", frame_T[count2]);
    }

//-----median filtering-----
*/

int filter_window[5] = {0,0,0,0,0};
int filter_count = 0;

filter_count=0;
{
    filter_window[0] = frame_T[filter_count];
    filter_window[1] = frame_T[filter_count];
    filter_window[2] = frame_T[filter_count];
    filter_window[3] = frame_T[filter_count+1];
    filter_window[4] = frame_T[filter_count+2];

    qsort(filter_window, 5, sizeof(int), int_cmp);

    frame_T_filtered[filter_count] = filter_window[2];

}

filter_count=1;
{

```

```
filter_window[0] = frame_T[filter_count-1];
filter_window[1] = frame_T[filter_count-1];
filter_window[2] = frame_T[filter_count];
filter_window[3] = frame_T[filter_count+1];
filter_window[4] = frame_T[filter_count+2];

qsort(filter_window, 5, sizeof(int), int_cmp);

frame_T_filtered[filter_count] = filter_window[2];

}

for (filter_count=2; filter_count < 198; filter_count++)
{
filter_window[0] = frame_T[filter_count-2];
filter_window[1] = frame_T[filter_count-1];
filter_window[2] = frame_T[filter_count];
filter_window[3] = frame_T[filter_count+1];
filter_window[4] = frame_T[filter_count+2];

qsort(filter_window, 5, sizeof(int), int_cmp);

frame_T_filtered[filter_count] = filter_window[2];

}

filter_count=198;
{
filter_window[0] = frame_T[filter_count-2];
filter_window[1] = frame_T[filter_count-1];
filter_window[2] = frame_T[filter_count];
filter_window[3] = frame_T[filter_count+1];
filter_window[4] = frame_T[filter_count+1];

qsort(filter_window, 5, sizeof(int), int_cmp);

frame_T_filtered[filter_count] = filter_window[2];

}

filter_count=199;
{
filter_window[0] = frame_T[filter_count-2];
filter_window[1] = frame_T[filter_count-1];
filter_window[2] = frame_T[filter_count];
filter_window[3] = frame_T[filter_count];
filter_window[4] = frame_T[filter_count];

qsort(filter_window, 5, sizeof(int), int_cmp);

frame_T_filtered[filter_count] = filter_window[2];
```

```

    }

//-----note grouping & corrected
//-----
playback_enable();
printf ("\n\r");

count = 0;
count2=0;
sound = (Xuint16*)XPAR_SRAM_256KX32_MEM0_BASEADDR;
unsigned int note_on = 0;
unsigned int correction_rate=44100;
unsigned int l=0;

for (count2 = 0; count2<198; count2++)
    {
        if ( (frame_T_filtered[count2]!=0) &&
            (frame_T_filtered[count2]==frame_T_filtered[count2+1]) &&
            (frame_T_filtered[count2]==frame_T_filtered[count2+2]) )
            {
                note_on = 2;
                if
                (frame_T_filtered[count2]!=frame_T_filtered[count2-1])
                {
                    //while (!(XAC97_mGetStatusReg(XPAR_OPB_AC97_CONTROLLER_0_BASEADDR) & XAC97_CODEC_RDY)){
                    for (l=0; l <68; l++)
                        {
                            if
                            (frame_T_filtered[count2]==scale_array[l][0])
                                {
                                    correction_rate
                                }
                                break;
                            }
                            else
                                correction_rate
                                = 44100;
                        }
                }
            }
    }

```

```

WriteAC97Reg(XPAR_OPB_AC97_CONTROLLER_0_BASEADDR, 0x2C,
correction_rate);
        if (correction_rate == 44100)
            {
                printf("\n\rNote Detected,
Frequency %d", (44100/frame_T_filtered[count2]));
            }

        if (correction_rate < 44100)
            {
                printf("\n\rNote Detected,
Frequency %d, Shifted Down", (44100/frame_T_filtered[count2]) );
            }

        if (correction_rate > 44100)
            {
                printf("\n\rNote Detected,
Frequency %d, Shifted Up", (44100/frame_T_filtered[count2]) );
            }
    }

//-----
Playback 1 frame-----

        for (count=0; count < 880; count++)
            {
                soundbyte = *sound;

                putfsl(soundbyte, 0);
                putfsl(0x0000, 0); //to left
channel mhdemizetai

                sound++;
            }
    }

else
    if (note_on == 2)
        {
            note_on = 1;

            //-----
-----Playback 1 frame-----

```

```

        for (count=0; count < 880; count++)
        {
            soundbyte = *sound;

            putfsl(soundbyte, 0);
            putfsl(0x0000, 0); //to left
            sound++;
        }
    }
else
    if (note_on == 1)
    {
        note_on= 0;

        -----//-----
        -----Playback 1 frame-----

        for (count=0; count < 880;
count++)
        {
            soundbyte = *sound;

            putfsl(soundbyte, 0);
            putfsl(0x0000, 0);
            sound++;
        }

        {
            //while (!(XAC97_mGetStatusReg(XPAR_OPB_AC97_CONTROLLER_0_BASEADDR) & XAC97_CODEC_RDY)){}

            WriteAC97Reg(XPAR_OPB_AC97_CONTROLLER_0_BASEADDR, 0x2C, 44100);
            //printf("off");
        }

    }

}

else

```

```

        {
            note_on = 0;

            //-----
-----Playback 1 frame-----

            for (count=0; count < 880; count++)
            {
                soundbyte = *sound;

                putfsl(soundbyte, 0);
                putfsl(0x0000, 0); //to left
channel mhdenizetai
                sound++;
            }
        }

    }

    sound = (Xuint16*)XPAR_SRAM_256KX32_MEM0_BASEADDR;
    playback_disable();

    //-----print filtered
pitch array-----
    printf ("\n\r");
    for (count2=0; count2<200; count2++)
    {
        printf("%d / ",frame_T_filtered[count2]);
    }
    printf ("\n\r");

    while (1) {}
    return 0;
}

```