



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ

ΥΠΟΛΟΓΙΣΤΩΝ

---

ΕΝΣΩΜΑΤΩΜΕΝΟ ΣΥΣΤΗΜΑ  
ΕΠΕΞΕΡΓΑΣΙΑΣ ΨΗΦΙΑΚΟΥ ΗΧΟΥ

---

Τριαντάφυλλος Μουρτζανός

*Επιτροπή*

Επιβλέπων:

Μέλος Επιτροπής:

Μέλος Επιτροπής:

Καθηγητής Απόστολος Δόλλας

Καθηγητής Διονυσιος Πνευματικάτος

Καθηγητής Μιχάλης Ζερβάκης

Χανιά, Οκτώβριος 2016

# Περίληψη

Στον ψηφιακό ήχο η πιο διαδεδομένη μορφή αποθήκευσης της μουσικής χωρίς συμπίεση είναι η cd ποιότητα στα 44.1kHz και 16bit. Για αρκετούς μουσικόφιλους η ποιότητα του cd δε μπορεί να προσφέρει τη μέγιστη ακουστική εμπειρία που μπορεί να λάβει ο ακροατής και υποστηρίζουν την ανωτερότητα των αρχείων ήχου υψηλής ανάλυσης, μεγαλύτερης συχνότητας δειγματοληψίας δηλαδή και ε-ύρους bit, ανοίγοντας έτσι μια δημόσια συζήτηση η οποία ξεκίνησε στα τέλη της δεκαετίας του '90 με την εμφάνιση των πρώτων μουσικών αρχείων σε ποιότητα Dvd και συνεχίζεται μέχρι και σήμερα.

Η παρούσα διπλωματική στοχεύει στην βελτίωση της cd ποιότητας με την εφαρμογή μαθηματικών μεθόδων Interpolation πάνω στο αρχικό δειγματοληπτημένο σήμα χωρίς την αλλαγή του ρυθμού δειγματοληψίας ώστε στη συνέχεια να σχεδιαστεί μια Hardware υλοποίηση η οποία θα το προσφέρει αυτό σε πραγματικό χρόνο. Αφού μελετήθηκαν οι δυνατότητες που έχει το ανθρώπινο αυτί και το ρόλο που παίζει ο τομέας της ψυχοακουστικής κατά την ακουστική εμπειρία, σχεδιάστηκαν και μοντελοποιήθηκαν δυο μέθοδοι Interpolation, η Linear και η Cubic Spline.

Αφού υλοποιήθηκαν και μελετήθηκε πειραματικά η συμπεριφορά των δυο μεθόδων τόσο οπτικά με το φασματογράφημα και τις κυματομορφές τόσο και ακουστικά με την επιλογή μουσικών δειγμάτων με διαφορετικά χαρακτηριστικά, αξιολογήσαμε τις μεθόδους, συγκρίναμε τις διαφορές για διαφορετικούς ρυθμούς δειγματοληψίας και αναφέραμε τα θετικά και αρνητικά που παρουσιάζουν σε κάθε περίπτωση. Η μέθοδος με το Cubic Spline η οποία είναι και η βασική μέθοδος για το Interpolation σε αυτή τη διπλωματική, έδωσε ενδιαφέρον αποτελέσματα σε σύγκριση με το αρχικό σήμα, ενώ αντίθετα η Linear μέθοδος ήταν σχεδόν πάντα κατώτερη του αρχικού.

Στη συνέχεια προσαρμόστηκε η μέθοδος μας για τη πιθανή υλοποίηση σε Hardware η οποία και σχεδιάστηκε και αξιολογήθηκε.

# Abstract

In digital sound, the most common format to store music without compress is the cd quality on 44.1Khz sample rate and 16bit depth. For many audio-philes, the cd quality can't offer the maximum listening experience that the listener can get and they support the superiority of high-resolution audio files. So, a strong debate which started at the end of the 90's when the first Dvd Quality music files were released is still open to discuss where new arguments appear constantly from both sides.

This thesis diploma aims to improve the standard cd-quality with the application of interpolation mathematical methods on the original .wav file, without resampling the rate so that can be applied in a Hardware implementation. Having studied the capabilities of the human ear and the role played by psychoacoustics during the audio experience, two methods of interpolation were designed and modeled, Linear and Cubic Spline.

Once implemented and experimentally studied the behavior of two methods both visually with spectrogram and waveforms so and the audibility with musical samples selection with different characteristics, we evaluated the methods, we compared the differences for different sampling rates and discussed the positives and negatives that have to each case. The Cubic Spline which is the main interpolation method in this work, gave us interesting results in compare to the original file, while on the other hand the Linear method was almost always worse than the original.

Then our method adjusted for possible implementation in Hardware, which was designed and evaluated.

# Ευχαριστίες

Πρωτίστως θα ήθελα να ευχαριστήσω τον επιβλέπων καθηγητή μου κ. Απόστολο Δόλλα για την ευκαιρία που μου έδωσε να ασχοληθώ με κάτι που με ενδιαφέρει ιδιαίτερα,για τη πρακτική καθοδήγηση κατά την εκπόνηση αυτής της διπλωματικής καθώς και για τη προσπάθεια του για δημιουργία της σωστής νοοτροπίας που πρέπει να έχει ένας μηχανικός,κατα το διάστημα που συνεργαστήκαμε.Θα ήθελα να ευχαριστήσω επίσης τους καθηγητές Διονυσιο Πνευματικάτο και Μιχάλη Ζερβάκη που ως μέλη της εξεταστικής μου επιτροπής συμμετέχουν στην επιτυχή ολοκλήρωση των σπουδών μου.

Ευχαριστώ τους φίλους και συμφοιτητές μου για όλες τις όμορφες στιγμές που περάσαμε όλα αυτά τα χρόνια καθώς και για την πνευματική και πρακτική βοήθεια που μου παρήχαν σε όλη τη διαδρομή της φοίτησης μου,χωρίς αυτούς τα πράγματα θα ήταν πολύ διαφορετικά.

Τέλος,θέλω να ευχαριστήσω την οικογένεια μου που ήταν πάντα δίπλα μου να με στηρίζει,στις καλές και στις δύσκολες στιγμές και να τους αφιερώσω με όλη μου τη καρδιά αυτή τη δουλειά γιατί το αξίζουν.



# Περιεχόμενα

Περίληψη	2
Abstract	3
Ευχαριστίες	4
<b>1 Εισαγωγή</b>	<b>10</b>
1.1 Κίνητρο	10
1.2 Σκοπός και αντικείμενο της διπλωματικής	10
1.3 Συνεισφορά της διπλωματικής	11
1.4 Οργάνωση κεφαλαίων	11
<b>2 Σχετική έρευνα</b>	<b>12</b>
2.1 Εισαγωγή στον ψηφιακό ήχο	12
2.1.1 Αναλογικό και ψηφιακό ηχητικό σήμα	12
2.1.2 Sampling Theory and Nyquist	12
2.1.3 PCM and Bit Depth	13
2.1.4 Quantization and Coding	14
2.1.5 Dithering	14
2.1.6 Stereophonic Sound	14
2.1.7 Ultrasonic Sound	14
2.2 Interpolation	15
2.2.1 Linear Interpolation	15
2.2.2 Cubic Spline Interpolation	16
2.3 Ψηφιακή Επεξεργασία Ηχου	20
2.3.1 Using Matlab to process sound	20
2.3.2 Processing sound with fpga	21
<b>3 Ανθρώπινο αυτί και ψυχοακουστική</b>	<b>23</b>
3.1 Ανθρώπινο αυτί	23
3.2 Ψυχοακουστική	25
3.3 Πείραμα <i>Meyer – Moran</i>	26
3.4 Meta-Analysis	28
<b>4 Μοντελοποίηση</b>	<b>30</b>
4.1 Εισαγωγή στο μοντέλο	30
4.2 Μελετώντας το σπεκτρόγραμμα	31
4.3 Μελετώντας τις κυματομορφές	37
4.4 <i>Resample</i> και κυματομορφές	42
<b>5 Πειραματική επαλήθευση μοντέλου</b>	<b>46</b>
5.1 Ακουστικά <i>Tests</i>	46
5.1.1 Interpolation Tests	46
5.1.2 Resample Tests	47
5.2 Πείραμα με θόρυβο	48
5.3 Μελέτη του <i>SNR</i>	50

5.4	Επαλήθευση λειτουργίας μεθόδου . . . . .	51
5.5	Επαλήθευση μοντέλου <i>Matlab</i> για λειτουργία σε <i>Hardware</i> . .	52
<b>6</b>	<b>Σχεδίαση ενσωματωμένου συστήματος</b>	<b>55</b>
6.1	Εισαγωγή στα Εργαλεία . . . . .	55
6.2	Σχεδίαση του συστήματος . . . . .	55
6.2.1	Σχεδιασμός συστήματος <i>LinearInterpolation</i> . . . . .	55
6.2.2	Σχεδιασμός συστήματος <i>CubicSplineInterpolation</i> . .	58
<b>7</b>	<b>Συμπεράσματα και μελλοντική εργασία</b>	<b>65</b>
7.1	Ανασκόπηση . . . . .	65
7.2	Συμπεράσματα . . . . .	66
7.3	Μελλοντική εργασία . . . . .	67
	<b>Βιβλιογραφία</b>	<b>69</b>

# Κατάλογος Σχημάτων

2.1	Aliasing and Nyquist . . . . .	13
2.2	Data points of the line $y$ are the linear interpolation points . . . . .	16
2.3	Floating-Point vs Fixed-Point . . . . .	22
3.1	Perceived Human Hearing . . . . .	23
3.2	Το όριο του ακουστικού πόνου . . . . .	24
3.3	Κοχλία . . . . .	24
3.4	ABX Comparator . . . . .	25
3.5	Mayer-Moran Setup . . . . .	26
3.6	Meta-Analysis Study list . . . . .	28
4.1	Το μοντέλο της Matlab . . . . .	31
4.2	Journey into Sound waveform . . . . .	32
4.3	Spectrograms of 44.1Khz and 48Khz of the Original Audio File . . . . .	33
4.4	Spectrograms of 44.1Khz and 48Khz of the Linear Interpolation . . . . .	33
4.5	Spectrograms of 44.1Khz and 48Khz of the Cubic Spline . . . . .	33
4.6	Spectrograms of 66.15khz and 96Khz of the Original Audio . . . . .	34
4.7	Spectrograms of 66.15khz and 96Khz of the Linear Interpolation . . . . .	34
4.8	Spectrograms of 66.15khz and 96Khz of the Cubic Spline . . . . .	35
4.9	Spectrograms of 192Khz Original and Cubic Spline . . . . .	35
4.10	Metallica-Enter Sandman Dual Channel Waveform . . . . .	36
4.11	Spectrogram of the Original and of Cubic Spline for whole track . . . . .	36
4.12	Africa Mono Drum Loop Waveform . . . . .	37
4.13	Waveform of the Mono Sample,Cubic and Original Comparison . . . . .	38
4.14	Cubic and Linear comparison of the Mono Sample . . . . .	39
4.15	Waveform of the Stereo Sample,Cubic and Original Comparison . . . . .	40
4.16	Whole song 64 samples Waveform . . . . .	41
4.17	Whole song 256 samples Waveform . . . . .	42
4.18	Cubic Spline and Original Comparison 44.1Khz . . . . .	43
4.19	Cubic Spline and Original Comparison 48Khz . . . . .	43
4.20	Cubic Spline and Original Comparison 66.15Khz . . . . .	44
4.21	Cubic Spline and Original Comparison 96Khz . . . . .	44
4.22	Cubic Spline and Original Comparison 192Khz . . . . .	45
5.1	Το σύστημα εφαρμογής θορύβου και στη συνέχεια Interpolation στο αρχικό "καθαρό" σήμα . . . . .	48
5.2	Cubic,Linear,Original Noise Waveform . . . . .	49
5.3	Επόμενο στιγμιότυπο των μεθόδων μας μετά την εφαρμογή θορύβου. . . . .	50
5.4	Spline Methods Comparison . . . . .	52
5.5	4 sample Method Comparison . . . . .	53
5.6	4 sample Method Comparison 2 . . . . .	54
5.7	4 sample Method Spectrogram . . . . .	54
6.1	Linear Interpolation Simulink Diagram . . . . .	57
6.2	Top Level Cubic Spline Interpolation Diagram . . . . .	58

6.3	The Path of the Sound inside the System . . . . .	60
6.4	HDelta Module . . . . .	61
6.5	Sigma Module . . . . .	62
6.6	BCD Module . . . . .	63
6.7	Interpolation Module . . . . .	64

# Κατάλογος Πινάκων

5.1	Πίνακας των SNR μετά την εφαρμογή θορύβου . . . . .	51
-----	---	----

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Κίνητρο

Τα τελευταία χρόνια στο χώρο της μουσικής, η ευρεία διάδοση του .mp3 ως πρότυπο ψηφιακής κωδικοποίησης έχει άμεσες συνέπειες στο τρόπο που αποθηκεύουμε και αναπαράγουμε τον ήχο μας. Η εξέλιξη της τεχνολογίας στις συσκευές αναπαραγωγής καθώς και η ανάπτυξη της ευρυζωνικότητας σε παγκόσμια κλίμακα κατέστησαν το Mp3 ως πρώτη προτίμηση στα ψηφιακά αρχεία ήχου.

Η επιτυχία του .mp3 ως πρότυπο, βασίζεται στο ότι δεσμεύει αρκετά μικρό χώρο σε δεδομένα από τις άμεσες μεθόδους λόγω συμπίεσης, χρησιμοποιώντας ψυχοακουστικά μοντέλα που εξαλείφουν τους ήχους που δεν μπορεί να αντιληφθεί το ανθρώπινο αυτί. Αυτό έχει ως αποτέλεσμα να έχουμε μεν πολλά περισσότερα αρχεία στα μέσα αποθήκευσης αλλά να χάνουμε σε ποιότητα, έστω και αν εκ πρώτης όψης δεν καταλαβαίνουμε τη διαφορά. Είναι δεδομένο πως με το .mp3 αλλά και με συνδρομητικές υπηρεσίες streaming μουσικής τύπου Spotify, πως έχουμε θυσιάσει την ποιότητα στο βωμό της ευκολίας.

Πολλοί απαιτητικοί ακροατές δεν έχουν πειστεί για την αξία της ψηφιοποίησης του ήχου είτε για πρακτικούς λόγους όπως η ποιότητα, είτε ενίοτε για συναισθηματικούς. Η βασική διαμάχη μεταξύ των υπέρμαχων του βινυλίου, του αναλογικού ήχου δηλαδή έναντι αυτών του cd, είναι η ποιότητα. Οι λάτρεις των δίσκων υποστηρίζουν πως η “ζεστασιά” που βγάζει το βινύλιο ακούγεται καλύτερα από κάθε ψηφιακή μορφή, καθώς η ποιότητα δε μετριέται σε συχνότητες δειγματοληψίας και βάθος bits, αλλά κατευθείαν με το ανθρώπινο αυτί. Από την άλλη πλευρά, οι λάτρεις της ψηφιακής μορφής αντιτείνουν την καθαρότητα του ήχου, την συνεχή ροή χωρίς ατέλειες και “σκρατς” ως επιχειρήματα περί ανωτερότητας σε ποιότητα.

Συνεχώς νέες συσκευές και πρότυπα, υπόσχονται καλύτερη ποιότητα και υψηλή ανάλυση στο κοινό, άλλα με χαμηλό και άλλα με πολύ υψηλό κόστος. Κατα κοινή ομολογία η χρήση ενός DAC (digital-to-analog converter) κατά την αναπαραγωγή των ψηφιακών αρχείων εγγυάται καλύτερη ποιότητα από το standard chip μιας ενσωματωμένης κάρτας ήχου σε έναν υπολογιστή ή ενός hifi της μαζικής αγοράς. Κατα πόσο όμως μπορεί κάποιος είναι διατεθειμένος να διαθέσει επιπλέον χρήματα για κάτι που το έχει ήδη;

Απο επιστημονικής πλευράς το θετικό με την ψηφιοποίηση του αναλογικού ήχου, είναι πως με τη μέθοδο της δειγματοληψίας μπορούμε να επεξεργαστούμε κάθε κομμάτι του ήχου όπως επιθυμούμε, άρα αυτό ίσως μας δώσει τη δυνατότητα να αναλύσουμε τη ποιότητα του αρχείου και να πειραματιστούμε πάνω σε αυτή.

### 1.2 Σκοπός και αντικείμενο της διπλωματικής

Η διπλωματική εργασία που παρουσιάζεται αφορά την έρευνα και την εφαρμογή μαθηματικών μοντέλων και συναρτήσεων σε αρχεία ήχου χωρίς συμπίεση

με στόχο την βελτίωση της ποιότητας, τόσο μαθηματικά όσο και ακουστικά. Με τη μελέτη των αποτελεσμάτων του πειραματικού μοντέλου στοχεύουμε να δημιουργήσουμε ένα ενσωματωμένο σύστημα το οποίο επεξεργάζεται τον ψηφιακό ήχο.

Η έρευνα και το πειραματικό κομμάτι έγινε σε περιβάλλον Matlab και η προσομοίωση του στο Stimulink της Matlab μαζί με το εργαλείο Xilinx. Σκοπός μας είναι να μελετήσουμε τη συμπεριφορά του ήχου standard ποιότητας cd μέσω τεχνικών resampling και interpolation και να δούμε κατα πόσο επηρεάζεται η ποιότητα, όπως μπορεί να την αντιληφθεί το γυμνό αυτί. Αφού καταλήξαμε πως το ιδανικό μοντέλο αριθμητικής ανάλυσης πάνω στα samples μας είναι το Cubic Spline Interpolation προσομοιώσαμε τη σχεδίαση με σκοπό την ενσωμάτωση της σε μια fpga.

Αντιλαμβανόμενοι τις τάσεις της σύγχρονης μουσικής τεχνολογίας περι hifi fidelity ήχου, μελετάμε την αποτελεσματικότητα των προτάσεων με υψηλό resolution και samplerate όσον αφορά την ανθρώπινη ακοή και πως συμβάλει στο τελική άποψη περι ποιότητας ο τομέας της ψυχοακουστικής.

### 1.3 Συνεισφορά της διπλωματικής

Μέσω αυτής της διπλωματικής, προτείνουμε μια μαθηματική λύση για την βελτίωση της ποιότητας της μουσικής που θα μπορούσε να χρησιμοποιηθεί είτε στη δημιουργία υλικού που θα επεξεργάζεται τον ήχο (κάρτα ήχου, digital converter) είτε στη δημιουργία λογισμικού, νέου προτύπου ψηφιακού ήχου η studio effect που θα χρησιμοποιείται απο επαγγελματίες ηχολήπτες και παραγωγούς εως απλούς τυπικούς ακροατές.

### 1.4 Οργάνωση κεφαλαίων

Η δομή της παρούσας διπλωματικής είναι η παρακάτω:

Στο κεφάλαιο 2 γίνεται αναφορά στον ψηφιακό ήχο και τα στοιχεία αυτού που μελετήσαμε κατα την εκπόνηση της εργασίας. Στη συνέχεια γίνεται μαθηματική ανάλυση στις μεθόδους Interpolation που χρησιμοποιήθηκαν και στο πως γίνεται η επεξεργασία ήχου στη πλατφόρμα Matlab αλλά και τι συμβαίνει με τον ήχο σε μια FPGA .

Στο κεφάλαιο 3 γίνεται ανάλυση στο πως αντιλαμβάνεται το ανθρώπινο αυτί τον ήχο και στο ρόλο της ψυχοακουστικής σε αυτό. Στη συνέχεια γίνεται μελέτη των πειραμάτων που μελετάνε την ακουστικότητα ατόμων μέσω Blind Tests και στο πως εξαρτάται η προπόνηση του αυτιού με την αναγνώριση της ποιότητας του ήχου.

Στο κεφάλαιο 4 αναφέρεται η μελέτη της μοντελοποίησης του πειράματος, όπως αυτό εκπονήθηκε σε περιβάλλον Matlab μελετώντας και αξιολογώντας κυματομορφές και σπεκτρογράμματα με η χωρίς την εφαρμογή interpolation και σε διαφορετικούς ρυθμούς δειγματοληψίας.

Στο κεφάλαιο 5 γίνεται μια ακουστική αξιολόγηση και πειραματική επαλήθευση του μοντέλου.

Στο κεφάλαιο 6 αναφέρεται στο πως το παρον μοντέλο μπορεί να υλοποιηθεί σε μια FPGA ώστε να δημιουργηθεί ένα πλήρες ενσωματωμένο σύστημα και τους κινδύνους και τα προβλήματα που μπορούν να εμφανισούν σε μια τέτοια μοντελοποίηση.

Στο κεφάλαιο 7 πραγματοποιείται μια σύνοψη της διπλωματικής εργασίας και εξάγονται συμπεράσματα, αναφέρονται τα προβλήματα που παρουσιάστηκαν κατα την εκπόνηση της και προτείνονται μελλοντικές επεκτάσεις γι αυτήν.

## Κεφάλαιο 2

# Σχετική έρευνα

### 2.1 Εισαγωγή στον ψηφιακό ήχο

Ο ήχος είναι ένα συνεχές αναλογικό κύμα που μεταδίδεται στον αέρα και γίνεται αντιληπτό στον άνθρωπο μέσω της αίσθησης της ακοής. Η τεχνολογία έχει καταφέρει να συλλάβει και να συμπίεσει τον ήχο σε ψηφιακά αρχεία. Η μέθοδος με την οποία έχει γίνει η αναπαράσταση του αναλογικού ήχου σε ψηφιακό εγγυάται και την ποιότητα του ήχου.

#### 2.1.1 Αναλογικό και ψηφιακό ηχητικό σήμα

Σαν αναλογικό σήμα αντιλαμβανόμαστε μια ρέουσα πληροφορία που λαμβάνει συνεχές τιμές, συναρτήσει του χρόνου σε κανονικό ή και ακανόνιστο μοτίβο. Ο ήχος που ταξιδεύει στον αέρα και μεταφέρει τη διακύμανση της πίεσης των μορίων του αέρα που προκαλούνται από μία ταλάντωση (μηχανική) σε ένα περιορισμένο εύρος συχνοτήτων που αντιλαμβάνεται το ανθρώπινο αυτί, είναι ο αναλογικός ήχος. Όταν πάλλεται η πηγή του ήχου, στέλνει διαφορετικές συχνότητες στον αέρα, κάθε μια από τις οποίες παράγει ένα κύμα. Η ποιότητα του ήχου εξαρτάται από το συνδυασμό διαφορετικών συχνοτήτων σε πολλαπλά κύματα.

Στην μουσική τεχνολογία ως αναλογική καταγραφή ήχου ορίζουμε την φυσική αντιπροσώπευση του αρχικού σήματος που κωδικοποιείται σε ένα υπόστρωμα όπως το αυλάκι ενός δίσκου βινυλίου ή το μαγνητικό πεδίο μια μαγνητοταινίας.

Το ψηφιακό σήμα είναι η αναπαράσταση του αναλογικού σήματος σαν μια διακριτή ακολουθία από bits σε ένα ψηφιακό σύστημα. Η αναπαράσταση δεν αποθηκεύει όλο το αρχικό κύμα αλλά σημεία αυτού, καθώς “τεμαχίζεται” με τη μέθοδο της δειγματοληψίας (sampling) και ένα μέσο επίπεδο εύρους εφαρμόζεται σε κάθε δείγμα. Αυτό επιτρέπει την εύκολη επεξεργασία κάθε δείγματος ενώ ολόκληρο του κύματος καθώς και την μείωση του χώρου που χρειάζεται για την αποθήκευση στα επελεγμένα μέσα (cd, dvd, usb). Η συσκευή αναπαραγωγής διαβάζει αυτά τα σημεία και τα ερμηνεύει ώστε να δημιουργήσει εκ νέου ένα αναλογικό ηχητικό κύμα.

#### 2.1.2 Sampling Theory and Nyquist

Για να μετατραπεί ένα σήμα από αναλογικό σε ψηφιακό πρέπει να περάσει από τα στάδια της δειγματοληψίας (sampling), της κβάντισης (quantization) και της κωδικοποίησης (coding). Το θεώρημα της δειγματοληψίας συνδέει συνεχές σήματα στο πεδίο του χρόνου (αναλογικά) με διακριτά σήματα στο πεδίο της συχνότητας (ψηφιακά).

Αφού επιλέξουμε έναν ρυθμό για τη δειγματοληψία μας (sampling rate) λαμβάνουμε τιμές από το συνεχές σήμα σε δείγματα σε χρονικές στιγμές που ισχύουν μεταξύ τους. Η συχνότητα δειγματοληψίας μας δείχνει πόσα δείγματα λαμβάνουμε σε 1 δευτερόλεπτο και η απόσταση μεταξύ δύο δειγμάτων μας δείχνει τη περίοδο δειγματοληψίας όπως επιβεβαιώνει και ο τύπος της συχνότητας



$$Fs = \frac{1}{Ts}.$$

Ο πιο συχνός και διαδεδομένος ρυθμός δειγματοληψίας είναι τα 44.100HZ (CD quality), αυτό διότι σύμφωνα με τη Nyquist Sampling Theory του Dr. Nyquist, το sampling rate χρειάζεται να υπερβαίνει το συνολικό εύρος του σήματος κατά διπλάσια τιμή για αποφύγουμε το aliasing.

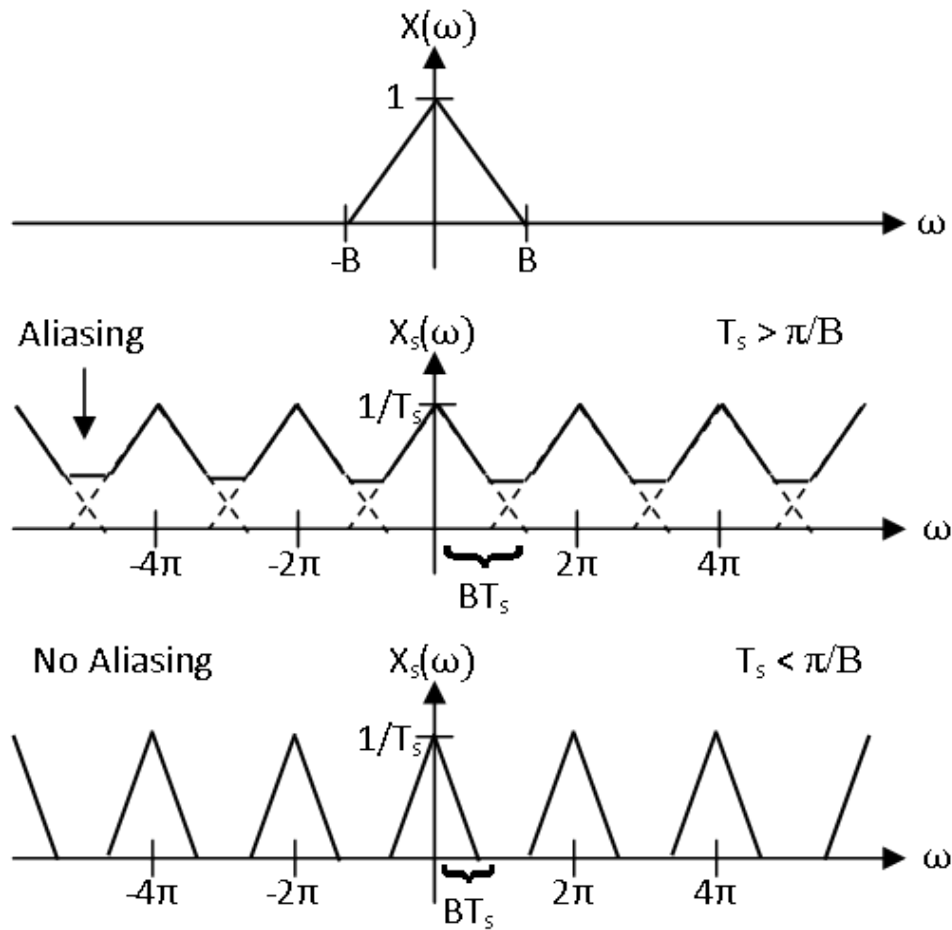


Figure 2.1: Όταν τα δείγματα πέφτουν το ένα πάνω στο άλλο με αποτέλεσμα το τελικό σήμα να είναι δυσανάγνωστο και να διαφέρει από το αρχικό το ονομάζουμε aliasing.

Έτσι λοιπόν, με δεδομένο πως το ανθρώπινο αυτί σπάνια αντιλαμβάνεται ήχους άνω των 20KHZ ήρθαμε στη σύμβαση των 44.1kHz, που θεωρητικά μετά την μετατροπή από έναν DAC και την αναπαραγωγή των ψηφιακών αρχείων μουσικής ο άνθρωπος θα τα ακούσει χωρίς απώλειες.

### 2.1.3 PCM and Bit Depth

Στην μουσική τεχνολογία, ως PCM (Pulse-code modulation) ορίζουμε την πιο διαδεδομένη μέθοδο με την οποία αναπαρίσταται ψηφιακά τα δειγματοληπτημένα αναλογικά σήματα. Η PCM μέθοδος χρησιμοποιείται στο σύνολο σχεδόν της ψηφιακής αποθήκευσης ήχου χωρίς απώλειες καθώς σε αυτήν βασίζονται τα πρότυπα .wav και .aiff.

Τα βασικά στοιχεία που τη χαρακτηρίζουν είναι το sampling rate και το bit depth δηλαδή το πόσα bits πληροφορίας υπάρχουν σε κάθε sample. Κάθε sample αποθηκεύεται σαν ψηφιακή πληροφορία δηλαδή σαν δυαδική πληροφορία.

Το bit depth αναπαριστά το πλάτος (amplitude) του αναλογικού ήχου και μας δείχνει την ανάλυση (resolution) του κάθε δείγματος. Στη μουσική τεχνολογία χρησιμοποιούμε 16, 24 και 32 bit με τα 16bit να είναι και τα πιο συνηθισμένα.

τιμές που μπορεί να δεχτεί το κάθε δείγμα είναι το bit depth σαν δύναμη του 2. Για 16bit λοιπόν έχουμε  $2^{16}=65536$  τιμές.

#### 2.1.4 Quantization and Coding

Με το sampling έχουμε ένα σύνολο διακριτών τιμών μιάς ακολουθίας απο πραγματικούς αριθμούς. Η κβάντιση(quantization) μεταφέρει αυτούς τους αριθμούς στο κοντινότερο σημείο απο ένα προκαθορισμένο πλήθος από διακριτές στάθμες, με τη βοήθεια του bit depth, μετρώντας το πλάτος(amplitude) της κυματορφής και βρίσκοντας την κοντινότερη τιμή του bit στο πλάτος του σημείου απο το sampling.

Με αυτό το τρόπο κάνουμε το εύρος του σήματος διακριτό ώστε το κβαντισμένο σήμα να παίρνει μόνο διακριτές και πεπερασμένες τιμές. Η διαδικασία της κβάντισης σε αντίθεση με το sampling (όπου υπο ορισμένες συνθήκες η πλήρη ανακατασκευή του σήματος είναι πιθανή) είναι μη αναστρέψιμη και έχει σαν αποτέλεσμα το χάσιμο πληροφορίας.

Το κβαντισμένο σήμα κωδικοποιείται σε ψηφιακές λέξεις ανάλογα με το bit depth ώστε στο τέλος να έχουμε μια πλήρη ψηφιακή πληροφορία στην οποία αποθηκεύουμε τον ήχο μας.

#### 2.1.5 Dithering

Η κβάντιση του σήματος μπορεί ορισμένες φορές να προκαλέσει κάποιες απότομες ασυνέχειες σε αυτό. Μια τεχνική αντιμετώπισης αυτού του προβλήματος είναι το dithering ή αλλιώς pseudorandom noise quantization. Με το dithering, προσθέτουμε στο σήμα μια μικρή ποσότητα ασυσχέτιστου λευκού, χαμηλού επιπέδου τυχαίου θορύβου ακριβώς πριν την κβάντιση.

Στόχος κάθε συστήματος ψηφιακού ήχου είναι η εξάλειψη κάθε είδους παραμόρφωσης(disortion) ώστε το τελικό σήμα να είναι γραμμικό. Η χρησιμοποίηση του κατάλληλου dither σε κάθε μη τετριμμένη(non trivial) διαδικασία μας βοηθά να αποφύγουμε κάθε μη-γραμμική παραμόρφωση που δημιουργεί η κβάντιση ως αποτέλεσμα των φαινομένων του truncation και του rounding.

Η τεχνική αυτή ήταν διαδεδομένη στα πρώτα χρόνια των ψηφιακών ηχογραφήσεων αλλά χρησιμοποιείται ακόμα σήμερα για να επανορθώσουμε για έναν κακής ποιότητας DAC(ψηφιακό σε αναλογικό μετατροπέα).

#### 2.1.6 Stereophonic Sound

Το cd χρησιμοποιεί στερεοφωνικό δικάναλο ήχο. Τα κανάλια είναι μεταξύ τους ανεξάρτητα και όταν ο ήχος αναπαράγεται απο δύο ή και περισσότερα ηχεία(ή stereo ακουστικά) δίνεται η ψευδαίσθηση πως ο ήχος έρχεται απο διάφορες κατευθύνσεις.

Για να παραχθεί stereo ήχος, κατά την ηχογράφηση και παραγωγή τοποθετούνται δύο μικρόφωνα για να λάβουν τη πηγή ήχου και στη συνέχεια οι ήχοι που ελήφθησαν θα γίνουν mix σε ένα ηχητικό αρχείο όπου κάποια στοιχεία θα βρίσκονται στο αριστερό κανάλι και κάποια στο δεξί.

#### 2.1.7 Ultrasonic Sound

Οι ήχοι τους οποίους δε μπορεί να αντιληφθεί το ανθρώπινο αυτί(η ακουστικότητα του ανθρώπινου αυτιού καλύπτεται ενδελεχώς στο κεφάλαιο 3) γιατί βρίσκονται σε συχνότητες μεγαλύτερες απο τις ακουστικές δυνατότητες του, ονομάζονται Ultrasonic sounds ή UltraSounds. Όσον αφορά τη μουσική και για τη ποιότητα του cd, δεχόμαστε ως ultrasonic κάθε ήχο που δε μπορεί να καλύψει το cd quality. Οι ήχοι αυτοί φτάνουν στο αυτί αλλά αυτό δεν μπορεί να δεχθεί κάποιο ερέθισμα ώστε να τους αναπαράγει.

Σύμφωνα με την(αμφισβητούμενη ομολογουμένως επιστημονικά) έρευνα του Tsutomu Ōhashi, αναφέρεται το Hypersonic effect ,ένα φαινόμενο όπου ναι μεν το ανθρώπινο αυτί δεν μπορεί να αντιληφθούν ευσεινήδητα τα ultrasounds ,η απουσία τους απο έναν ήχο προκαλεί φυσιολογικές και ψυχολογικές αντιδράσεις στον ακροατή.Αν ισχύει δηλαδή αυτό το φαινόμενο,πρακτικά η επιλογή ενός υψηλού ρυθμού δειγματοληψίας σε μια μουσική ηχογράφηση,σε συνδυασμό με την μη αποκοπή και είσοδο στο mastering των υπέρηχων δειγμάτων,μπορεί να προκαλέσει μια αίσθηση πως το τελικό κομμάτι ήχου είναι ανώτερο ποιοτικά,λόγω του ψυχοακουστικού παράγοντα.

## 2.2 Interpolation

Κάθε κομμάτι ψηφιακού ήχου,χάνει δεδομένα άλλοτε ελάχιστο άλλοτε περισσότερο απο τον αρχικό αναλογικό ήχο.Κατα τη δειγματοληψία(sampling)παίρνουμε διάφορα data points.Με την τεχνικές του interpolation λαμβάνουμε καινούρια data points τα οποία γεμίζουν τα δεδομένα που λείπουν απο το ψηφιακό ηχητικό αρχείο που εξετάζουμε.

Μετά τη διαδικασία του sampling,έχουμε τα δείγματα και θέλουμε να τα μετατρέψουμε πάλι σε έναν ήχο κύματος.Αν απλά ενώσουμε τα σημεία που είναι τα δείγματα μεταξύ τους θα πάρουμε ένα ηχητικό σήμα σαν το αρχικό αλλά με αρκετή παραμόρφωση(disortion).

Για να αποφύγουμε την παραμόρφωση,χρησιμοποιούμε την μέθοδο Interpolation η οποία ενώνει τα σημεία ομαλά μέσα στο διακριτό σύνολο των σημείων που βγάλαμε απο το sampling,μιμούμενη τη συμπεριφορά μιας ηχητικής κυματομορφής.Ετσι ανακτούμε το αρχικό σήμα χωρίς απώλειες και σφάλματα διασφαλίζοντας την αρχική ποιότητα.

Όσον αφορά τον ήχο στόχος του interpolation είναι στις ακμές του ήχου να εφαρμόσουμε καμπύλες(curve fitting) ώστε να έχουμε μια πιο απαλή και ομαλή αίσθηση του ήχου.Για να γίνει αυτό πρέπει να εντοπίσουμε μέσω της τεχνικής μας τα σημεία εκείνα στα οποία θα εφαρμόζουμε την καμπύλη η οποία κατασκευάζεται μέσω μαθηματικών εξισώσεων ανάλογα της μεθόδου που χρησιμοποιούμε.

Συγκεκριμένα,παίρνουμε δύο σημεία απο το αρχείο που έχουμε δειγματοληπτήσει και εφαρμόζουμε μια εξίσωση ανάμεσα σε αυτά τα σημεία για να υπολογίσουμε που βρίσκεται ένα τρίτο σημείο μεταξύ τους. Υπάρχουν πολλές τεχνικές για data interpolation ,ωστόσο σε αυτή τη διπλωματική θα χρησιμοποιήσουμε δύο τεχνικές,την linear interpolation και την cubic spline interpolation[10].

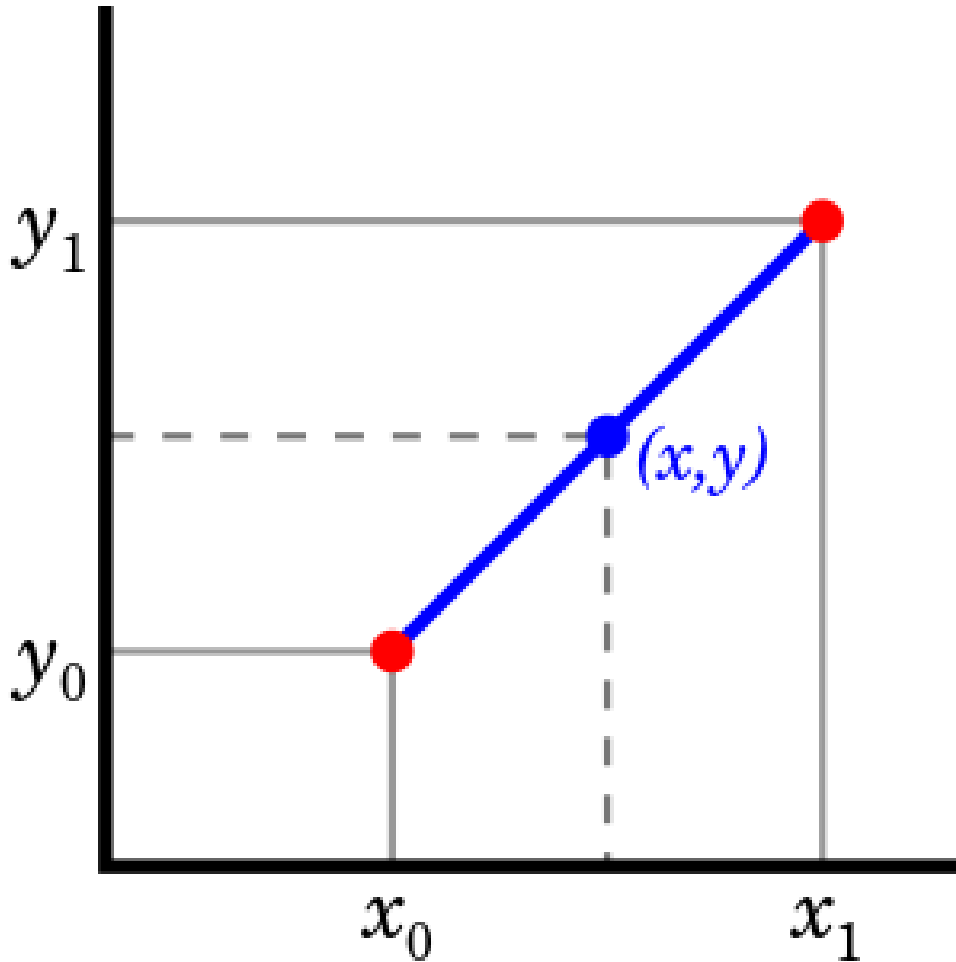
### 2.2.1 Linear Interpolation

Η γραμμική μέθοδος interpolation είναι και η πιο απλή απο τις μεθόδους.Ανάμεσα σε δύο γνωστά σημεία με συντεταγμένες  $A(x_0, y_0)$ ,  $B(x_1, y_1)$  ,η μέθοδος υπολογίζει την κλίση της ευθείας γραμμής που τέμνει τον άξωνα  $yy'$  και περνάει απο αυτά τα δύο σημεία.Για μια τιμή  $x$  στο μεσοδιάστημα  $(x_0, x_1)$  η τιμή του  $y$  υπολογίζεται μέσω της εξίσωσης

$$y = y_0 + (x - x_0) \frac{y_1 - y_0}{x_1 - x_0} \quad (2.1)$$

διότι ισχύει πως

$$\frac{y - y_0}{y_1 - y_0} = \frac{x - x_0}{x_1 - x_0} \quad (2.2)$$



Σχήμα 2.2: Data points of the line  $y$  are the linear interpolation points

Στην υλοποίηση μας αναπαριστούμε στον άξωνα  $xx'$  τα σημεία του χρόνου  $t$  και στον άξωνα  $yy'$  τα δείγματα(samples) μας. Το σημείο  $(x, y)$  είναι τα interpolated δεδομένα για της δοθείσες τιμές. Επαναλαμβάνουμε τη διαδικασία για όλα τα δείγματα σε κάθε χρονική στιγμή για να πάρουμε όλα τα δεδομένα interpolated.

### 2.2.2 Cubic Spline Interpolation

Όπως και στη Linear Interpolation έτσι και στη Cubic Spline, η παρεμβάλλουσα(interpolant) είναι μία τμηματική συνάρτηση(piecewise function). Ο βαθμός του πολυωνύμου περιορίζεται σε τρίτου βαθμού και σε αντίθεση με τη Linear Interpolation οι καμπύλες της είναι πιο ομαλές πράγμα που αποδίδει πιο πιστά τη συμπεριφορά των ηχητικών κυμάτων.

Αν το spline είναι μια συνάρτηση που συμβολίζεται με  $s(x)$  και αν οι κλίσεις της είναι μικρές, η δεύτερη παράγωγος  $s''(x)$  προσεγγίζει την καμπυλότητα και το διαφορικό μήκος τόξου προσεγγίζεται από την  $dx$ . Η ενέργεια του γραμμικού spline είναι αναλογική του ολοκληρώματος  $\int s''(x)^2 dx$ . Όταν μας οι κόμβοι  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  είναι πεπερασμένοι, η interpolating spline είναι μια συνάρτηση τέτοια ώστε  $s(x_i) = y_i (i = 1, 2, \dots, n)$  και η ενέργεια  $\int_{x_1}^{x_n} s''(x)^2 dx$  ελαχιστοποιείται.

Εστω πως οι  $s$  και  $s'$  είναι συνεχείς στο διάστημα  $[x_1, x_n]$ . Για τη κατασκευή του cubic spline για κάθε υποδιάστημα  $[x_i, x_{i+1}]$  για  $(i=1, 2, \dots, n)$  ορίζουμε

$$\begin{aligned}
h_i &= x_{i+1} - x_i \\
w &= \frac{x - x_i}{h_i} \\
\bar{w} &= 1 - w
\end{aligned}$$

και το spline μας δίνεται απο το

$$s(x) = wy_{i+1} + \bar{w}y_i + h_i^2[(w^3 - w)\sigma_{i+1} + (\bar{w}^3 - \bar{w})\sigma_i] \quad (2.3)$$

Ψπολογίζοντας τις τρεις πρώτες παραγώγους για το  $s(x)$  ακολουθώντας τον κανόνα της αλυσίδας και το γεγονός πως  $w' = \frac{1}{h_i}$  και  $\bar{w}' = -\frac{1}{h_i}$  έχουμε:

$$\begin{aligned}
s'(x) &= \frac{y_{i+1} - y_i}{h_i} + h_i[(3w^2 - 1)\sigma_{i+1} - (3\bar{w}^2 - 1)\sigma_i] \\
s''(x) &= 6w\sigma_{i+1} + 6\bar{w}\sigma_i \\
s'''(x) &= \frac{6(\sigma_{i+1} - \sigma_i)}{h_i}
\end{aligned}$$

Αντικαθιστώντας το τύπο του  $h_i$  και για  $x = x_i$  έχουμε:

$$\sigma_i = \frac{s''(x_i)}{6} \quad (2.4)$$

Απο αυτή τη σχέση ορίζουμε την έννοια της  $\sigma_i$  αλλά όχι την τιμή της. Ορίζουμε ως σταθερά Δέλτα όπου οι ποσότητες της είναι προσεγγίσεις ως προς τις πρώτες παραγώγους:

$$\Delta_i = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \quad (2.5)$$

με

$$\begin{aligned}
\Delta_i^2 &= \frac{\Delta_{i+1} - \Delta_i}{x_{i+2} - x_i} \\
\Delta_i^3 &= \frac{\Delta_{i+1}^2 - \Delta_i^2}{x_{i+3} - x_i}
\end{aligned}$$

Εκτιμώντας την  $s'(x)$  στο πάνω άκρο του υποδιαστήματος έχουμε:

$$\begin{aligned}
s'_+(x_i) &= \Delta_i - h_i(\sigma_{i+1} + 2\sigma_i) \\
s'_-(x_i) &= \Delta_i + h_i(2\sigma_{i+1} + \sigma_i)
\end{aligned}$$

όπου για να εξασφαλίσουμε την συνέχεια της  $s'$  επιβάλλουμε το παρακάτω κανόνα:

$$s'_+(x_i) = s'_-(x_i), \quad i = 2, \dots, n-1.$$

Η τιμή της  $s'_-(x_i)$  λαμβάνεται απο το υποδιάστημα  $[x_{i-1}, x_i]$ , ένας τύπος μπορεί

να βρεθεί αν αντικαταστήσουμε όπου  $i$  σε  $i - 1$  στο  $s'_-(x_{i+1})$ . Έτσι έχουμε:

$$\Delta_{i-1} + h_{i-1}(2\sigma_i + \sigma_{i-1}) = \Delta_i - h_i(\sigma_{i+1} + 2\sigma_i)$$

συνεπάγεται

$$h_{i-1}\sigma_{i-1} + 2(h_{i-1} + h_i)\sigma_i + h_i\sigma_{i+1} = \Delta_i - \Delta_{i-1}, \quad i = 2, \dots, n-1. \quad (2.6)$$

Έχουμε λοιπόν ένα σύστημα  $n - 2$  ταυτοχρόνως γραμμικών εξισώσεων με  $n$  αγνώστους. Για να ορίσουμε την interpolating spline, χρειαζόμαστε δύο ακόμα παραμέτρους.

Ορίζουμε τις παραμέτρους  $c_1(x)$  και  $c_n(x)$  ως μοναδικές cubics που περνάνε από τα πρώτα και τα τελευταία τέσσερα data points. Οι συνθήκες που καλύπτουν σε σχέση με το  $s(x)$  και τις τρίτες παραγώγους τους είναι

$$\begin{aligned} s'''(x_1) &= c_1''' \\ s'''(x_n) &= c_n''' \end{aligned}$$

και από το τύπο του Δέλτα έχουμε

$$\begin{aligned} c_1''' &= 6\Delta_1^3 \\ c_n''' &= 6\Delta_{n-3}^3 \end{aligned}$$

Συνεπώς από το τύπο της τρίτης παραγώγου του  $s(x)$  έχουμε

$$\begin{aligned} \frac{\sigma_2 - \sigma_1}{h_1} &= \Delta_1^3 \\ \frac{\sigma_n - \sigma_{n-1}}{h_{n-1}} &= \Delta_{n-3}^3 \end{aligned}$$

Για να έχουμε συμμετρικές εξισώσεις στο τελικό μας σύστημα, πολλαπλασιάζουμε τις δύο τελευταίες με  $h_1^2$  και  $-h_{n-1}^2$  και παίρνουμε

$$\begin{aligned} -h_1\sigma_1 + h_1\sigma_2 &= h_1^2\Delta_1^3 \\ h_{n-1}\sigma_{n-1} - h_{n-1}\sigma_n &= -h_{n-1}^2\Delta_{n-3}^3 \end{aligned}$$

Δημιουργούμε τον παρακάτω πίνακα για το spline με τις παραπάνω συνθήκες με  $n$  γραμμικές εξισώσεις και  $n$  αγνώστους:

$$\begin{pmatrix} -h_1 & h_1 & 0 & 0 & 0 \\ h_1 & 2(h_1 + h_2) & h_2 & 0 & 0 \\ 0 & h_2 & 2(h_2 + h_3) & h_3 & \\ & \ddots & \ddots & \ddots & \\ & & h_{n-2} & 2(h_{n-2} + h_{n-1}) & h_{n-1} \\ & & 0 & h_{n-1} & -h_{n-1} \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \vdots \\ \sigma_{n-1} \\ \sigma_n \end{pmatrix} = \begin{pmatrix} h_1^2\Delta_1^3 \\ \Delta_2 - \Delta_1 \\ \Delta_3 - \Delta_2 \\ \vdots \\ \Delta_{n-1} - \Delta_{n-2} \\ -h_{n-1}^2\Delta_{n-3}^3 \end{pmatrix}$$

Ο πίνακας αυτός είναι τριδιαγώνιος(tridiagonal ) και συμμετρικός.Για κάθε επιλογή απο  $x_1 < x_2 < \dots < x_n$ , ο πίνακας είναι ομαλός(nonsingular ) και κυρίαρχα διαγώνιος(diagonally dominant ).

Ετσι υπάρχει πάντα μια μοναδική λύση  $\sigma_1, \dots, \sigma_n$ .Για κάθε επιλογή του  $x_1, x_2, \dots, x_n$  ο πίνακας των συντελεστών είναι well conditioned .Δεδομένου αυτού και του γεγονότος πως της διαγώνιας κυριαρχίας,μπορούμε να υπολογίσουμε ακριβές λύσεις εφαρμόζοντας Γκαουσιανή απαλοιφή(Gaussian elimination) χωρίς κλιμάκωση(scaling) η περιστροφή(pivoting).

Θα πάρουμε μετά την απαλοιφή τον πίνακα στην εξής μορφή:

$$\begin{pmatrix} \alpha_1 & h_1 & & & 0 \\ & \alpha_2 & h_2 & & \\ & & \alpha_3 & h_3 & \\ & & & \ddots & \\ & & & & \ddots & \\ & & & & & \alpha_n \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \vdots \\ \sigma_n \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_n \end{pmatrix}$$

Οπου τα διαγώνια στοιχεία  $\alpha_i$  υπολογίζονται ως εξής:

$$\begin{aligned} \alpha_1 &= -h_1, \\ \alpha_i &= 2(h_{i-1} + h_i) - \frac{h_{i-1}^2}{a_{i-1}}, \quad i = 2, 3, \dots, n-1, \\ \alpha_n &= -h_{n-1} - \frac{h_{n-1}^2}{a_{n-1}} \end{aligned}$$

Και τα ανομοιογενή στοιχεία  $\beta_i$  υπολογίζονται:

$$\begin{aligned} \beta_1 &= h_1^2 \Delta_1^3, \\ \beta_i &= (\Delta_i - \Delta_{i-1}) - \frac{h_{i-1} \beta_{i-1}}{a_{i-1}}, \quad i = 2, 3, \dots, n-1, \\ \beta_n &= -h_{n-1}^2 \Delta_{n-3}^3 - h_{n-1} \frac{\beta_{n-1}}{a_{n-1}} \end{aligned}$$

Υπολογίζουμε και τα  $\sigma_i$  χρησιμοποιώντας τη μέθοδο Back-Substitution για το γραμμικό μας σύστημα εξισώσεων και έχουμε:

$$\begin{aligned} \sigma_n &= \frac{\beta_n}{\alpha_n} \\ \sigma_i &= \frac{\beta_i - h_i \sigma_{i+1}}{a_i} \quad i = n-1, n-2, \dots, 1 \end{aligned}$$

Επειδή το spline στην εφαρμογή μας θα εκτιμηθεί αρκετές φορές,υπολογίζουμε και αποθηκεύουμε τους συντελεστές του cubic  $b_i, c_i, d_i, i = 1, 2, \dots, n-1$ , σε κάθε διάστημα  $[x_i, x_{i+1}]$  όπου:

$$s(x) = y_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad x_i \leq x \leq x_{i+1} \quad (2.7)$$

Οπου οι συντελεστές  $b, c, d$  δίνονται απο:

$$\begin{aligned} b_i &= \frac{y_{i+1} - y_i}{h_i} - h_i(\sigma_{i+1} + 2\sigma_i) \\ c_i &= 3\sigma_i \\ d_i &= \frac{\sigma_{i+1} - \sigma_i}{h_i} \end{aligned}$$

Για  $i = 1, 2, \dots, n-1$ . Χρησιμοποιούμε αυτή τη μορφή για να αποθηκεύσουμε τα splines ώστε να απλοποιήσουμε διεργασίες όπως οι παράγωγοι και τα ολοκληρώματα.

### Υλοποίηση σε Matlab.

Για να υλοποιήσουμε το παραπάνω σύστημα που θα δημιουργεί cubic splines θα χρειαστούμε δύο συναρτήσεις στο Matlab .Μια που θα δημιουργεί και θα υπολογίζει τους συντελεστές και μια που θα λύνει το τριδιαγώνιο πίνακα.

Στη μεταβλητή  $x_i$  για  $i = 1, 2 \dots n-1$ , αναπαριστούμε το χρόνο  $t$  και στη μεταβλητή  $y_i$  τα δείγματά μας. Το μήκος των  $x, y$  πρέπει να είναι το ίδιο. Δημιουργούμε σύμφωνα με τους παραπάνω τύπους τα  $h_i, \Delta_I$  τα οποία και θα χρειαστούμε να τροποποιήσουμε για να κατασκευάσουμε το τριδιαγώνιο πίνακα του οποίου η λύση μας δίνει τα  $\sigma_i$ .

Αφού καλέσουμε την συνάρτηση που λύνει το τριδιαγώνιο και πάρουμε την λύση του πίνακα, μας μένει να υπολογίσουμε τους συντελεστές  $b_i, c_i, d_i$  και τέλος το  $s(x)$  στο οποίο και αποθηκεύουμε τα δείγματα μας interpolated τα οποία γυρνάει η συνάρτηση μαζί με τον χρόνο.

Η τιμή της μεταβλητής  $x$ , στο σύστημα μας ορίζεται μέσα στο υποδιάστημα όπου  $x_i \leq x \leq x_{i+1}$  με  $i = 1, 2 \dots n-1$ , και θα βρίσκεται πάντα μεταξύ των δύο τελευταίων χρονικών στιγμών. Ενδιαφέρον παρουσιάζει το γεγονός πως αν το  $x$  πάρει κάποια τιμή μεγαλύτερη απο τα όρια στο οποίο βρίσκεται, τότε θα αυξηθούν τα samples. Θα μπορούσε λοιπόν να παίζει το ρόλο του resampler σε κάποια άλλη εφαρμογή η πειραματισμό.

## 2.3 Ψηφιακή Επεξεργασία Ηχου

### 2.3.1 Using Matlab to process sound

Μέσω του Matlab μας δίνεται η δυνατότητα να λάβουμε όλα τα δείγματα απο ένα ψηφιακό αρχείο ήχου για κάθε χρονική στιγμή ανάλογα με το ρυθμό που θα ακολουθήσουμε και να τα τοποθετήσουμε σε ένα πίνακα ώστε να τα επεξεργαστούμε.

Μέσω της συνάρτησης **audioread** διαβάζουμε το αρχείο ήχου και ανακτούμε όλα τα δείγματα και τη συχνότητα δειγματοληψίας. Γνωρίζοντας αυτό και το πλήθος των samples (αφού  $T = \frac{1}{F_s}$ ) γνωρίζουμε και το χρόνο  $t$  τον οποίο και δημιουργούμε χειροκίνητα.

Τα samples που διαβάζουμε απο το αρχείο ήχου, αποθηκεύονται σε ένα πίνακα της μορφής **Number Of Samples x Number Of Channels** .Για παράδειγμα, ένα stereo ηχητικό απόσπασμα τεσσάρων σχεδόν δευτερολέπτων, με συχνότητα δειγματοληψίας 44.100khz θα αποθηκευτεί σε ένα πίνακα double με διαστάσεις 174.102x2.

Η συνάρτηση **resample** μας δίνει τη δυνατότητα να αλλάζουμε το sample rate προσθέτοντας νέα δεδομένα σε μορφή fixed rate .Η συχνότητα θα αλλάξει κατά  $\frac{P}{Q}$  φορές του αυθεντικού. Για παράδειγμα, για το ίδιο απόσπασμα που αναφέραμε στη προηγούμενη παράγραφο, για να κάνουμε upsampling στα 48khz



θα θέσουμε τις μεταβλητές P και Q στις τιμές 160 και 147 αντίστοιχα. Αν θελήσουμε να κάνουμε downsample στο αρχικό rate αρκεί να αντιστρέψουμε τις τιμές των P και Q. Η μέθοδος αυτή ονομάζεται resampling μέσω ενός λογικού κλάσματος (rational fraction).

Η διαδικασία του resampling αποτελεί από μόνη της μια μορφή interpolation στο αρχικό σήμα. Όταν εφαρμόζεται παράγει μια προσέγγιση της αλληλουχίας των samples που διαβάσει από τον πίνακα που τα έχουμε αποθηκεύσει όπως θα ήταν κανονικά αμα στο σήμα είχε εφαρμοστεί sampling σε μεγαλύτερες (ή μικρότερες, στο μοντέλο μας όμως χρησιμοποιούμε μόνο upsampling) τιμές. Το upsampling χρειάζεται ένα χαμηλοπερατό φίλτρο μετά την αύξηση του ρυθμού δεδομένων και το downsampling ακριβώς πριν την αποδεκτίση.

Για να αποθηκεύσουμε τα δεδομένα του ήχου στο οποίο εφαρμόσαμε τη τεχνική του resample θα γράψουμε τον ήχο σε μορφή .wav μέσω της συνάρτησης **audiowrite** και στη συνέχεια θα διαβάσουμε εκ νέου το κομμάτι με την **audioread**. Για το παράδειγμά μας, το resample από 44.1 στα 48 khz θα μας δώσει έναν πίνακα με μέγεθος  $189.499 \times 2$  και αν το πάμε στα 192khz τότε θα γίνει ένας πίνακας μεγέθους  $757.966 \times 2$  με  $\frac{P}{Q} = \frac{640}{147}$ .

Με το Matlab έχουμε πολλές δυνατότητες για παραμετροποιήσεις και πειραματισμό επάνω στον ήχο. Δημιουργώντας συναρτήσεις για τις τεχνικές του interpolation που χρησιμοποιούμε και αναφέραμε παραπάνω, μπορούμε ανα πάσα στιγμή καλώντας τις να παρέμβουμε στα samples και να αξιολογήσουμε το αποτέλεσμα είτε γραφικά εμφανίζοντας τη κυματομορφή μέσω του **plot** ή τη μεταβολή των συχνοτήτων στο χρόνο με το **spectrogram** είτε ηχητικά εσωτερικά του προγράμματος με την εντολή **sound** ή εξωτερικά γράφοντας τον ήχο σε ένα νέο αρχείο μέσω της **audiowrite**.

### 2.3.2 Processing sound with fpga

Το FPGA ή Field Programmable Gate Array είναι ένας τύπος προγραμματιζόμενου ολοκληρωμένου κυκλώματος γενικής χρήσης. Ο προγραμματισμός του γίνεται μέσω μια γλώσσας περιγραφής υλικού (HDL) όπως είναι η VHDL και η Verilog. Διαθέτει έναν τεράστιο αριθμό λογικών πυλών, καταχωρητών μνήμης, απαριθμητών και άλλων διαθέσιμων πόρων διαθέσιμα στο χρήστη.

Ανάμεσα στις εφαρμογές που χρησιμοποιούν FPGA είναι και η Ψηφιακή Επεξεργασία Σήματος. Όταν μια εφαρμογή χρειάζεται αυξημένους πόρους και υπολογιστική δύναμη ώστε να αποδίδει σε υψηλά standards η χρησιμοποίηση μιας fpga έναντι ενός DSP μικροεπεξεργαστή είναι η καλύτερη επιλογή καθώς οι δυνατότητες της πρώτης υπερέχουν. Βέβαια στην επιλογή του σχεδιαστή παίζουν ρόλο και άλλοι παράγοντες όπως η γλώσσα προγραμματισμού, οι περισσότερες υλοποιήσεις είναι γραμμένες σε c και στον DSP μικροεπεξεργαστή μπορεί να γίνει κατευθείαν προγραμματισμός εν αντιθέση με τις FPGA που πρέπει να γίνει μεταφορά σε γλώσσα hdl, η συχνότητα δειγματοληψίας και το data rate όπου όσο μεγαλύτερα είναι τόσο πιο ευκολα θα λειτουργήσει σε μια fpga, οι βιβλιοθήκες που θέλουμε να χρησιμοποιήσουμε, οι conditional operations όπου αν χρειαζόμαστε τέτοιους θα δυσκολευτούμε αρκετά κατά το προγραμματισμό σε vhdl και μια υλοποίηση σε software θα ήταν η ιδανική λύση και τέλος το σε ποια μορφή είναι αποθηκευμένη η πληροφορία μας ώστε να δούμε τι τελεστές θα χρησιμοποιήσουμε, αν θα χρειαστούμε floating point αριθμητική δηλαδή.

Μέχρι πρότινος καμία fpga δεν έδινε τη δυνατότητα στους σχεδιαστές να χρησιμοποιήσουν floating point παρα μόνο fixed-point για συστήματα DSP. Αυτό άλλαξε το 2014 με την εμφάνιση στην αγορά της FPGA της Altera η οποία υποστηρίζει single precision floating point και υπόσχεται ίδιες αποδόσεις με τη χρησιμοποίηση fixed point. Ακολούθησε η Xilinx με τη σειρά 7 των FPGA's το 2015 και πλέον μπορούμε να πούμε πως όλες οι νέες fpga θα υποστηρίζουν το IEEE 754 πρότυπο.

Δεν υπάρχει αμφιβολία πως το floating point είναι αρκετά ανώτερο format απο το fixed point,εδικά όταν θέλουμε να επιλέξουμε την ακρίβεια,το εύρος και τη ταχύτητα που χρειαζόμαστε σε μια σχεδίαση.

Characteristic:	Floating-Point	Fixed-Point
Dynamic Range	Much Larger	Smaller
Resolution	Comparable	
Algorithm Design Time	Much Less	More
Ease of Use	Better	Worse
Compiler Efficiency	Much Better	Much Worse

Figure 2.3: Floating-Point vs Fixed-Point

Μια fpga για να επεξεργαστεί τον ήχο θα πρέπει να είναι εξοπλισμένη με ένα A-D converter στην είσοδο που θα παίρνει τη πληροφορία του ήχου αναλογικά και θα τη μετατρέπει σε ψηφιακή για την επεξεργασία και έναν D-A στην έξοδο όπου θα δίνει το επεξεργασμένο σήμα σε αναλόγική μορφή ώστε να είναι έτοιμο για αναπαραγωγή.Οι καινούριες υλοποιήσεις fpga board συνήθως περιλαμβάνουν ένα FPGA Mezzanine Card (FMC) interface το οποίο δίνει τη δυνατότητα να διευρύνουμε το I/O του συστήματος συνδέοντας(στην περίπτωση μας) μια κάρτα ήχου.Παλαιότερες fpga (και καινούριες βεβαία) ενσωματώνουν τον AC'97 audio codec για τη διαχείριση του ήχου.

Ο AC'97 codec περιλαμβάνει D-A και A-D converters .Είναι ένας αρκετά δημοφιλής codec για την αναπαραγωγή απλού ήχου και χρησιμοποιείται σε αρκετα τσιπ κάρτων ήχου,τα οποία είναι συνήθως ενσωματωμένα σε motherboards.Μπορεί να δώσει σε μια fpga δικάναλο ήχο είσοδο και να πάρει πίσω στην έξοδο επεξεργασμένο ήχο στην ίδια μορφή.Η διασύνδεση του είναι σειριακή,δηλαδή η πληροφορία μεταδίδεται απο και στον codec με 1 bit κάθε φορά,δηλαδή σε κάθε κύκλο του ρολογιού.Υποστηρίζει 18-bit ήχο μέχρι 48khz το οποίο είναι και η βασική του συχνότητα δειγματοληψίας και στις πρώτες εκδόσεις υποστήριζε μόνο αυτο το ρυθμό.Η σύνδεση του είναι η πιο απλή της μορφής Line-In με audio-jacks.Για να συνδέσουμε τον ήχο που δέχεται η είσοδος στο line-in το Codec συνδέεται σε έναν Wrapper ο οποίος συνδέεται στη συνέχεια το σύστημα που επεξεργάζεται τον ήχο ώστε να του δίνει την είσοδο και να παίρνει την έξοδο.Η έξοδος στη συνέχεια δίνεται στη συσκευή αναπαραγωγής και φτάνει στα ηχεία η τα ακουστικά μας.

## Κεφάλαιο 3

# Ανθρώπινο αυτί και ψυχοακουστική

### 3.1 Ανθρώπινο αυτί

Το ανθρώπινο αυτί ακούει μέσω τριχωειδών κυττάρων(hair cells) που είναι τοποθετημένα σε μια μεμβράνη στον κοχλία.Κάθε κύτταρο μπορεί να αντιληφθεί μία στενή ζώνη συχνοτήτων και καθορίζεται ανάλογα με τη θέση της στη μεμβράνη.Οι χαμηλότερες συχνότητες καλύπτονται απο τη κορυφή της μεμβράνης ενώ οι υψηλές απο τη βάση της.Ένας ήχος είναι μη ακουστός όταν δεν υπάρχουν τα τριχωειδή κύτταρα να τον ακούσουν.

Το κάτω και πάνω όριο του ανθρώπινου ακουστικού φάσματος κυμαίνεται απο τα 20hzως τα 20khz.Για να φτάσουμε σε αυτούς τους συγκεκριμενους αριθμούς πειραματικά,μετρούμε αρχικά σε όλο το φάσμα του ήχου το απόλυτο όριο της ακοής των ακροατών που συμμετέχουν στο πείραμα.Ως αποτέλεσμα παίρνουμε μια καμπύλη που μας δείχνει τον πιο σιγανό ήχο που αντιλαμβάνεται το αυτί σε κάθε δωθείσα συχνότητα.Τα πειράματα για να έχουν επιτυχία πρέπει να λαμβάνουν υπόψιν τόσο τη κούραση του αυτιού όσο και την απαιτούμενη συγκέντρωση του ακροατή,γιάυτό χρειάζονται αρκετές παύσεις πράγμα που σημαίνει πως η διαδικασία μπορεί να είναι αρκετά χρονοβόρα.

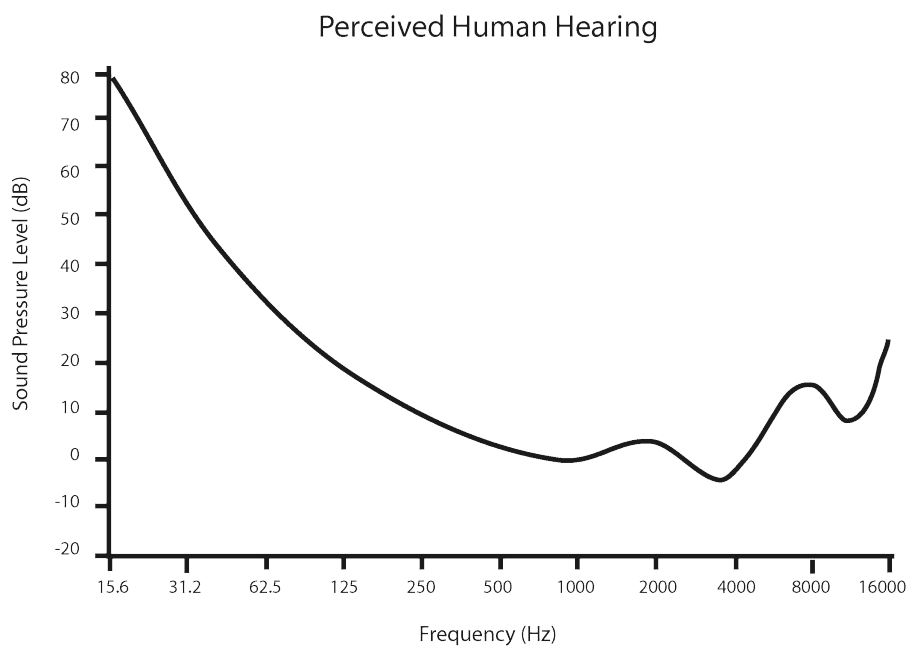
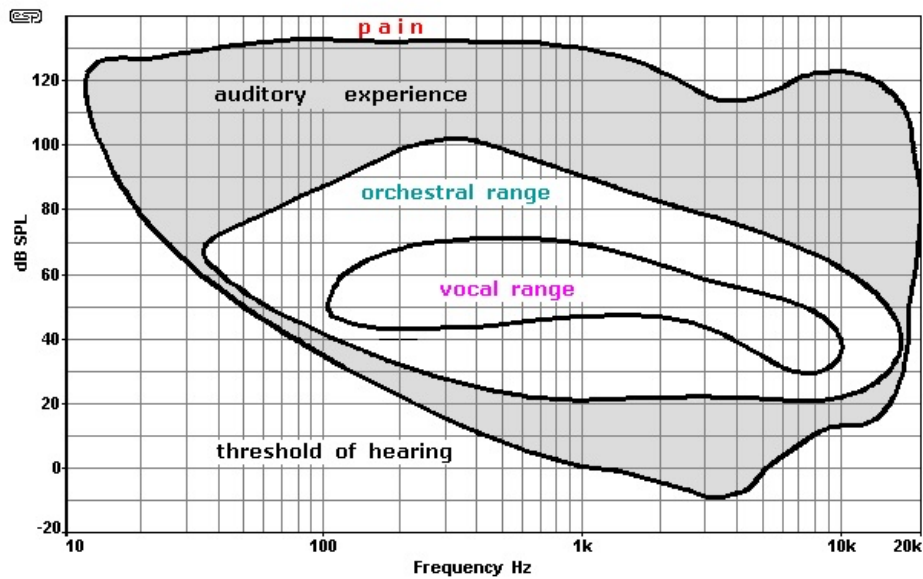


Figure 3.1: Ο βαθμός της ακουστικής πίεσης για κάθε δεδομένη συχνότητα που προκαλεί ένα ηχητικό σήμα

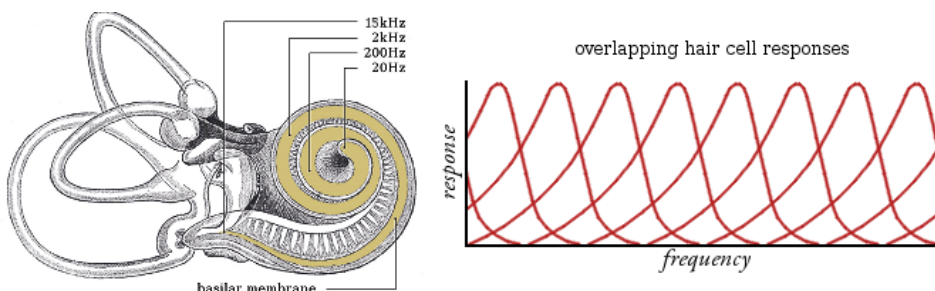
Στη συνέχεια συγκεντρώνουμε στοιχεία για το όριο του ακουστικού πόνου (threshold of pain), το σημείο δηλαδή που το πλάτος του ήχου είναι τόσο υψηλό που το νευρικό σύστημα του αυτιού νιώθει φυσικό πόνο. Η διαδικασία αυτή θέλει αρκετή προσοχή ώστε να μη δημιουργήσει κάποιο μόνιμο πρόβλημα στην ακοή των συμμετεχόντων.

Με αυτόν το τρόπο, το άνω όριο του ακουστικού φάσματος ορίζεται εκεί που η καμπύλη του απόλυτου ορίου της ακοής περνάει από το όριο του ακουστικού πόνου. Για τις χαμηλές συχνότητες ο κοχλίας δουλεύει παρόμοια με μια bass reflex (ανάκλασης χαμηλών συχνοτήτων) καμπίνα που έχουν τα μεγάλα ηχεία. Όπως αυτά έχουν μια οπή στην καμπίνα και οι χαμηλές συχνότητες εκπέμπονται σε φάση με τις χαμηλές που παράγει το ίδιο του woofer, έτσι το helicotrema είναι ένα άνοιγμα στη κορυφή της μεμβράνης το οποίο λειτουργεί σαν μια πόρτα που συντονίζεται κάπου ανάμεσα στα 40 και στα 65Hz, ανάλογα με τον άνθρωπο.



Σχήμα 3.2: Το όριο του ακουστικού πόνου

Σε γενικές γραμμές το όριο 20Hz-20kHz είναι αρκετά μεγάλο και σίγουρα δεν αντιπροσωπεύει τις ακουστικές δυνατότητες των περισσότερων ανθρώπων δε παύει όμως να τις καλύπτει. Το ακουστικό φάσμα μειώνεται συνεχώς με την ηλικία ειδικά στη περιοχή των υψηλών συχνοτήτων. Αυτό συμβαίνει καθώς με τη πάροδο του χρόνου, ο άνθρωπος εκτίθεται όλο και πιο πολύ σε θορύβους, φυσικούς από το περιβάλλον ή τεχνητούς και χάνει όλο και περισσότερα τριχωειδή κύτταρα κοντά στη βάση της μεμβράνης, αυτά δηλαδή που συντονίζονται στις υψηλές συχνότητες. Τα τριχωειδή κύτταρα δεν έχουν την δυνατότητα να αναγεννηθούν.



Σχήμα 3.3: Κοχλίας

Βέβαια, αυτό δεν συμβαίνει σε όλους τους ανθρώπους καθώς αυτή από αυτή διαφέρει κυρίως λόγω γενετικής αλλά και άσκησης. Είναι γεγονός πως ένα νέο, υγιές αυτί ακούει πολύ καλύτερα από ένα μεγαλύτερης ηλικίας ή από κάποιο με κάποια ζημιά. Κάποιοι άνθρωποι έχουν τη δυνατότητα να ακούν χροίες στη μουσική που δεν είναι εύκολα αντιληπτές από τη πλειονότητα των ακροατών. Με ειδική εκπαίδευση λοιπόν ένα υγιές αυτί μπορεί να αποκτήσει ακόμα περισσότερες ακουστικές δυνατότητες από το μέσο όρο.

### 3.2 Ψυχοακουστική

Η ψυχοακουστική είναι κλάδος της ακουστικής επιστήμης και ασχολείται με το τρόπο με τον οποίο αντιλαμβανόμαστε τον ήχο. Με την ψυχοακουστική εμπλέκονται και πολλές άλλες επιστήμες όπως η ψυχολογία, οι νευροεπιστήμες και η γλωσσολογία ως μέσο έρευνας της ανθρώπινης συμπεριφοράς. Η κύρια ιδέα της ψυχοακουστικής επιστήμης και η βασική συνεισφορά της στην μελέτη μας είναι πως ο τρόπος που κάθε άτομο αντιλαμβάνεται τους διάφορους ήχους είναι καθαρά υποκειμενικός.

Στην μουσική τεχνολογία αλλά και στην επιστήμη των υπολογιστών, συναντάμε πολλές εφαρμογές της ψυχοακουστικής. Βασικές είναι στα lossy πρότυπα ήχου mp3 και όπου τα ψυχοακουστικά μοντέλα προσφέρουν τη δυνατότητα για μεγαλύτερη συμπίεση στον ήχο.

Άλλες εφαρμογές βρίσκουμε σε μηχανικούς ήχου και παραγωγούς που καλύπτουν συγκεκριμένες συχνότητες ώστε να δώσουν έμφαση και να ενισχύσουν κάποιες άλλες, σε φθηνά ηχητικά συστήματα που καλύπτουν την έλειψη ποιότητας στη κατασκευή και στον ήχο, στην ηχητική κάλυψη σε θέατρα και σινεμά, ακόμα και σε ηχητικά όπλα και συστήματα ασφαλείας.

Μια προσέγγιση της ψυχοακουστικής επίδρασης στην αντίληψη του ήχου μπορεί να γίνει μέσω ενός ABX test, μια μέθοδος η οποία συγκρίνει δύο επιλογές ηχητικών κυμάτων για να εντοπίσει ανιχνεύσιμες διαφορές μεταξύ τους. Δύο γνωστά δείγματα A και B δίνονται σαν εισοδοί στο σύστημα και ένα άγνωστο δείγμα X το οποίο επιλέγεται τυχαία μεταξύ A και B βγαίνει σαν έξοδος. Ο ακροατής που συμμετέχει στο πείραμα καλείται να αναγνωρίσει αν η έξοδος είναι το A ή το B. Σε ένα προκαθορισμένο βάθος δοκιμών, αν ο ακροατής δε μπορεί να αξιόπιστα το X κατά ένα ποσοστό που να καταρρίπτει την null hypothesis (η οποία θεωρείται true μέχρι αποδείξεως του εναντίου), τότε αυτή δε μπορεί να διαψευστεί και έτσι δεν μπορούμε να αποδείξουμε πως τα δύο δείγματα A και B έχουν κάποια αισθητή διαφορά μεταξύ τους.

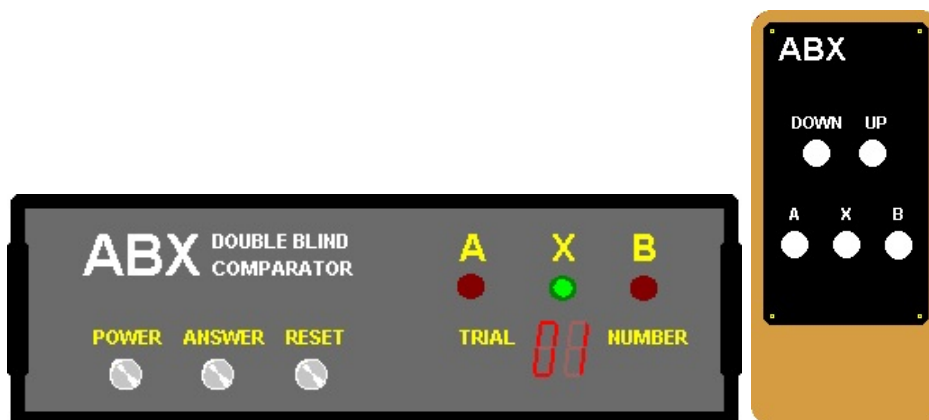


Figure 3.4: ABX Comparator  
To Logic and Display Module αριστερά και το Hand Control Module δεξιά

Ενας ABX comparator είναι ο ABX CS-5 Double Blind Comparator που κατασκευάστηκε το 1980. Τα κουμπιά up και down στο hand module επιλέγουν τον αριθμό των δοκιμών, τα A και B κάνουν τα αντίστοιχα δείγματα να παίξουν και το X κάνει την επιλογή του δείγματος ενεργή ώστε να εμφανιστεί ως αριθμός του trial στο Logic Display Module.

### 3.3 Πείραμα Meyer – Moran

Αναφέραμε πως κάθε άτομο έχει υποκειμενική άποψη επάνω στο τι ακούει, άρα έχει και υποκειμενική άποψη όσον αφορά την ποιότητα του ήχου. Αυτήν την υπόθεση, στηρίζει η πειραματική μελέτη των E. Brad Meyer και David R. Moran μελών του Boston Audio Society [2].

Για πάνω από ένα χρόνο, διεξήγαγαν blind tests σε δεκάδες ακροατές, ανέμεσα τους μηχανικοί ήχου και σπουδαστές μουσικής τεχνολογίας, με σκοπό να μελετήσουν την ικανότητα τους να αναγνωρίσουν την ανώτερη ποιότητα των high-resolution players που αναπαράγουν high-resolution αρχεία ήχου έναντι του cd standard.

Συγκεκριμένα, δημιούργησαν ένα σύστημα αναπαραγωγής βασισμένο σε έναν SACD/DVD-A player\*. Η αναλογική έξοδος δίνει ένα σήμα το οποίο περνάει μέσα από ένα 44kHz/16bit “bottleneck” που παίζεται μέσω ενός γνωστού cd recorder με monitoring πραγματικού χρόνου. Η εναλλαγή μεταξύ των δύο συστημάτων για την πραγματοποίηση των blind tests γίνεται μεταξύ του ABX CS-5 double blind comparator. Στη συνέχεια, πριν ακόμα σταλθεί το σήμα στα ηχεία περνάει μέσα από ένα ρυθμιζόμενο προενισχυτή. Όλα τα υποσυστήματα είναι κορυφαίας ποιότητας. Για την αναπαραγωγή του τελικού σήματος χρησιμοποιούν αρκετά setups σε κάθε πείραμα ανάμεσα τους επαγγελματικά studio monitors και ακριβά και μεγάλα ηλεκτροστατικά ηχεία. Χρησιμοποιήθηκαν αρκετά είδη μουσικής ανάμεσα τους κλασική, jazz, pop, rock και ηλεκτρονική.

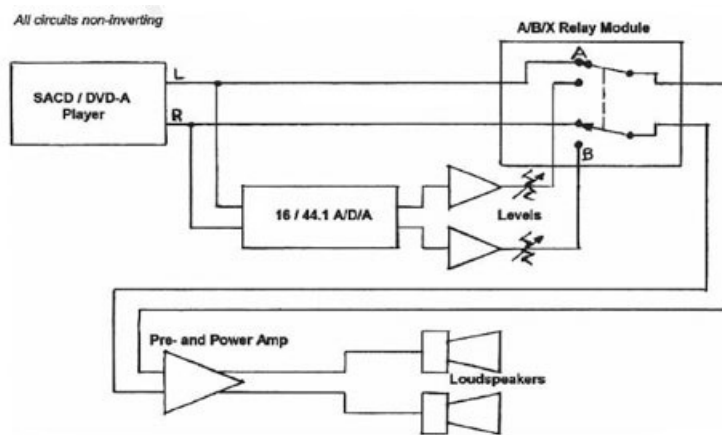


Figure 3.5: The setup of the experiment

Το κίνητρο για αυτό το πείραμα ήταν η γενικότερη κουβέντα τόσο στην επιστημονική κοινότητα όσο και μεταξύ επαγγελματιών του ήχου, περί της επάρκειας και της ποιότητας του cd standard έναντι των high-resolution συστημάτων (όπως του sacd). Γι αυτό επέλεξαν να εξετάσουν την ακουστικότητα του cd standard όταν εισαχθεί σε ένα τέτοιο σύστημα και να αποφανθούν αν πραγματικά πρόκειται για ένα “bottleneck”.

Θεωρητικά, η διαφορά στην ακουστική μεταξύ τους θα έπρεπε να είναι εμφανής, ειδικά αφού το πείραμα διεξήχθη σε σχετικούς με το είδος, σε εκπαιδευμένα δηλαδή αυτιά. Τα ηχητικά πλεονεκτήματα του hi-res συστήματος λόγω του αυξημένου εύρους ζώνης (bandwidth), δε θα έπρεπε να διατηρηθούν με το cd

bottleneck και έτσι οι διαφορές θα έπρεπε να είναι ακουστές ειδικά μετά το re-sampling. Τα αποτελέσματα όμως για την ανιχνευσιμότητα του hi-res ήταν στα επίπεδα της τυχαίας μαντεψιάς. Απο 554 πειράματα σε normal επίπεδα ήχου, τα 276 βρήκαν τη σωστή απάντηση, το 49.82%. Το σημαντικό είναι πως κανένας από τους ακροατές δεν είχε 100% επιτυχία στις απαντήσεις του, με το καλύτερο score να φτάνει το 8/10. Κανένας συσχετισμός που αφορά το φύλο την ηλικία ακόμα και την επαγγελματική ενασχόληση με τη μουσική δεν βρέθηκε στα πειράματα.

Στη συνέχεια οι ερευνητές ως μέρος ενός υποπειράματος, άυξησαν την ένταση του ήχου κατά 14db και έπαιξαν χωρίς κάποιο ηχητικό σήμα τον ήχο από τα ηχεία σε 2 ακροατές. Οι ακροατές ήταν σε θέση να αναγνωρίσουν 100% το cd λόγω του υψηλότερου noise floor που δημιούργησαν αυτή η ενίσχυση στην ένταση του ήχου. Όταν έπαιξαν την μουσική σε αυτά τα επίπεδα και τα 2 συστήματα ακούγονταν αρκετά δυνατά και άσχημα.

Στη πλειοψηφία τους σε όλες τις ηχογραφήσεις που έπαιξαν ο 16/44.1 floor θόρυβος υπερκαλύπτονταν από τη τονικότητα του δωματίου καθώς και τον θόρυβο από την προενίσχυση (preamplifier noise) εκτός από 2 περιπτώσεις. Αυτό σημαίνει πως το cd-standard διαθέτει το bandwidth και το δυναμικό εύρος (dynamic range) για να καλύψει σχεδόν όλες τις ηχογραφήσεις που υπάρχουν. Αρα ακουστικά δεν υπάρχει καμία υπεροχή του High-resolution audio έναντι του cd-standard, ακόμα και οι ηχογραφήσεις που μπορούν να γίνουν ώστε να υποστηρίξουν πλήρως high-resolution ήχο (π.χ. 192kHz 24bit mastering), μπορούν να αναπαραχθούν με την ίδια ακουστικότητα από το cd standard.

Ο πιο σημαντικός παράγοντας σύμφωνα με τους ερευνητές, δεδομένου του πειράματος, για την βελτίωση της ποιότητας του ήχου και της ακουστικότητας της μουσικής βρίσκεται στη ποιότητα του συστήματος αναπαραγωγής και όχι στην ανάλυση των αρχείων ήχου. Η επένδυση σε ποιοτικά ηχεία, ακουστικά και ενισχυτές ενδείκνυται για την βελτίωση της ακουστικής εμπειρίας των απανταχού μουσικόφιλων.

\*SACD (Super Audio CD)/DVD-Α είναι το πρότυπο δίσκου για αποθήκευση αρχείων δίσκου που δημιουργήθηκε το 1999 σε συνεργασία της Sony με την Philips, με σκοπό την αντικατάσταση του CD format αλλά απέτυχε σε επίπεδα αγοράς. Πρόσφερε surround 96kHz ήχο και 256 λεπτά χωρητικότητα αρχείων ήχου.

### Απήχηση και αποδοχή του πειράματος

Η μελέτη των Meyer-Moran δημοσιεύθηκε το 2007[2] και από τότε έχει προκαλέσει ποικίλες αντιδράσεις στην επιστημονική και μουσική κοινότητα. Με τους high-definition players να υπάρχουν στην αγορά από τις αρχές του 2000 και να υπόσχονται μια ανώτερη ακουστική εμπειρία, η μη αποδοχή της υπεροχής τους από τη πλευρά των ερευνητών προκάλεσε τις αντιδράσεις των εταιριών και του αγοραστικού κοινού. Τα αποτελέσματα έχουν αμφισβητηθεί τόσο σε online forums σχετικά με τον ήχο όσο σε ερευνητικές δημοσιεύσεις και αν και έχει περάσει σχεδόν μια δεκαετία, το πείραμα και η μελέτη συνεχίζουν να διχάζουν καθώς αποτελούν ακόμα σημείο αναφοράς.

Το βασικό επιχείρημα όσων διατηρούν αρνητική στάση προς τη μελέτη, είναι πως οι ηχογραφήσεις που ακούστηκαν κατά τη διάρκεια των πειραμάτων, δεν ήταν πραγματικά high-resolution, να έχουν ηχογραφηθεί δηλαδή χρησιμοποιώντας υψηλής ανάλυσης εξοπλισμό, αλλά παλιότερες αναλογικές tape masters όπου είχαν μεταφερθεί σε DSD 64 format. Στο mastering των SACD και των DVD-A ηχογραφήσεων μπορεί να έχουν εφαρμοστεί επιπλέον χαμηλοπερατά



φίλτρα.Επίσης το setup στο οποίο έγιναν τα πειράματα δεν εξηγείται αναλυτικά απο τους συγγραφείς,πράγμα που μπορεί να σημαίνει πως ακόμα και να ήταν high-resolution το ακουστικό υλικό,στα αυτιά των ακροατών δεν έφθανε ως τέτοιο.Παρόλα αυτά μπορούμε να δεχθούμε πως αυτές οι συνθήκες έρχονται πιο κοντά στην τυπική οικιακή ακουστική εμπειρία που μπορεί να ζητά ο μέσος ακροατής και μπορεί να έχει ακριβώς τα ίδια προβλήματα.

### 3.4 Meta-Analysis

Τον Ιούνιο του 2016,δημοσιεύθηκε στο ημερολόγιο που εκδίδει το Audio Engineering Society μια Meta-Analysis του Dr Joshua Reiss [6] όλων των ερευνών που μελετάνε την αντιληπτική αξιολόγηση του High Resolution ήχου απο τον ακροατή.Σκοπός της ανάλυσης είναι να εξάγει ένα συμπέρασμα μέσα απο τη μελέτη 18 δημοσιευμένων πειραμάτων,για τα οποία είχαν επαρκή δεδομένα να επεξεργαστούν,με τη συμμετοχή περισσότερων απο 400 ακροατών σε πάνω απο 12.500 δοκιμές,για το κατα πόσο αντιλαμβάνεται ο ακροατής τη διαφορά μετα-ξύ high resolution και standard ,16 bit,44.1 ή 48khz ήχου.Τα αποτελέσματα έδειξαν μια μικρή αλλά σημαντική στατιστικά ικανότητα των συμμετεχόντων ακροατών στα πειράματα να ξεχωρίζουν το High Resolution περιεχόμενο και αυτό το αποτέλεσμα αυξάνεται ραγδαία όταν εφαρμόστηκε το πείραμα σε εκπαιδευμένα αυτιά.

Για την μελέτη αυτή εφαρμόστηκε για πρώτη φορά η τεχνική Meta-Analysis στον τομέα της μηχανικής του ήχου.Απο τις 80 περίπου δημοσιεύσεις που αφορούν το θέμα απο το 1980 και έπειτα,μετά απο ενδελεχή μελέτη και σαφή επιλογή κριτηρίων,έμειναν τελικά 18 πειράματα προς ανάλυση τα οποία ικανοποιούσαν όλες τις προϋποθέσεις.Στη συνέχεια εφαρμόστηκε μια secondary analysis πάνω στις μελέτες ώστε να ανακτηθούν όλα τα δεδομένα που χρειάζονται για το επόμενο στάδιο(meta-analysis),τα προβλήματα που μπορεί να δημιουργηθούν ως αποτέλεσμα των μεθόδων που χρησιμοποιήθηκαν για την εφαρμογή του πειράματος καθώς και η συλλογή του ποσοστού των σωστών απαντήσεων καθώς και της δυωνυμικής πιθανότητας που σχετίζεται με το πείραμα.Αναλυτικά οι μελέτες που χρησιμοποιήθηκαν μαζί με τα στοιχεία που αναφέραμε βρίσκονται στον παρακάτω πίνακα:

A. List of studies included in meta-analysis. B. Risks of potential biases and other issues in the included studies  
Low risk “-”; unclear risk “?”; high risk “⊠”. The last column identifies if these risks tend strongly towards false positives (Type I errors), false negatives (Type II errors) or neither (Neutral). C. Total number of trials and correct answers for each study, with the associated binomial probability

A. Study			B. Risk of Bias							C. Binomial test			
Study	Year	Main references	Allocation bias	Methodology	Experimental design	Stimuli bias	Attrition bias	Reporting bias	Leading to	# correct	total	percent correct	probability
Plenge	1980	[59]	-	?	⊠	⊠	-	-	Type II errors	1367	2580	52.98%	1.294E-03
Muraoka	1981	[35]	-	?	⊠	?	-	-	Neutral	542	1060	51.13%	0.2400
Oohashi	1991	[43]	-	-	-	-	-	?	Neutral	392	800	49.00%	0.7261
Yoshikawa	1995	[67]	-	-	?	?	?	-	Neutral	85	132	64.39%	5.976E-04
Theiss	1997	[23]	-	?	?	?	⊠	-	Neutral	38	51	74.51%	3.105E-04
Nishiguchi	2003	[64, 68]	-	-	-	⊠	-	-	Type II errors	489	920	53.15%	0.0301
Hamasaki	2004	[62, 64]	-	-	-	?	-	?	Neutral	944	1848	51.08%	0.1821
Nishiguchi	2005	[58]	-	-	-	⊠	-	-	Type II errors	418	864	48.38%	0.8381
Repp	2006	[71]	-	⊠	⊠	⊠	⊠	-	Type II errors	42	86	48.84%	0.6267
Meyer	2007	[63]	⊠	-	?	⊠	⊠	⊠	Type II errors	276	554	49.82%	0.5507
Woszyck	2007	[69]	-	?	-	?	-	?	Type II errors	54	114	47.37%	0.7439
Pras	2010	[66]	?	-	?	?	-	?	Neutral	368	707	52.05%	0.1462
King	2012	[72]	-	⊠	-	?	⊠	?	Type II errors	34	61	55.74%	0.2213
KanetadaA	2013	[24]	?	-	?	-	-	-	Type I errors	62	108	57.41%	0.0743
KanetadaB	2013	[24]	?	-	?	-	-	-	Type I errors	135	224	60.27%	1.281E-03
Jackson	2014	[11, 65]	-	?	-	-	-	-	Neutral	585	960	60.94%	6.352E-12
Mizumachi	2015	[70]	?	-	?	-	-	-	Type I errors	86	136	63.24%	1.279E-03
Jackson	2016	[65]	-	?	-	-	-	-	Neutral	819	1440	56.88%	1.000E-07
Total										6736	12645	53.27%	1.006E-13

Figure 3.6: Meta-Analysis Study list



Τα σημαντικά στοιχεία που προκύπτουν απο το παρών paper ,είναι βέβαια η απόδειξη έστω και με ένα μικρό στατιστικό ποσό,πως είναι δυνατή η αντίληψη μεταξύ standard(cd) και ποιότητας πέρα απο αυτήν και τα συμβατικά επίπεδα(high resolution) αλλά και η σημαντική διαφορά που παρατηρείται στην ακουστική αντίληψη μεταξύ του εκπαιδευμένου και του απαίδευτου αυτιού με τα εκπαιδευμένα αυτιά να μπορούν να ανιληφθούν εως και 60% τις διαφορές μεταξύ formats .Επισημάνθηκαν πως οι μεθόδοι για τις δοκιμές(trials) μπορούν να οδηγήσουν σε Type-2 errors,υπονοώντας πως τα αποτελέσματα θα μπορούσαν να είναι κ υψηλότερα απο αυτά που η στατιστική ανάλυση μας δείχνει.Όσον αφορά την διάρκεια των ηχητικών δειγμάτων που παρουσιάζονται στα tests,ο ιδανικός αριθμός είναι να μην ξεπερνάει το δείγμα τα 15-20 δευτερόλεπτα και οι παύσεις μεταξύ των επαναλήψεων να είναι 1.5 δευτερόλεπτο,έτσι η προσεκτική επιλογή των ηχητικών ερεθισμάτων πρέπει να λαμβάνεται υπόψιν.Οι περισσότερες μελέτες επικεντρώνονται στο sample rate για την ακουστική αντίληψη και όσον αφορά το bit depth,η ερώτηση 24-bit vs 16-bit παραμένει ένα ανοιχτό ερωτηματικό.

## Κεφάλαιο 4

# Μοντελοποίηση

Σε αυτό το κεφάλαιο θα δείξουμε την μοντελοποίηση που κάναμε σε περιβάλλον Matlab. Λαμβάνοντας θεωρητικά υπόψιν όσα αναφέρθηκαν στα προηγούμενα τρία κεφάλαια, θέλουμε να εξετάσουμε και να επαληθεύσουμε το πως συμπεριφέρεται ο ήχος πρακτικά.

Μέσω του Matlab λοιπόν, φτιάξαμε ένα μοντέλο που διαβάζει ένα αρχείο ψηφιακού ήχου .wav ποιότητας cd και εφαρμόσαμε πάνω του τις τεχνικές του resampling και του interpolation, μελετώντας τη συμπεριφορά του γραφικά και ακουστικά.

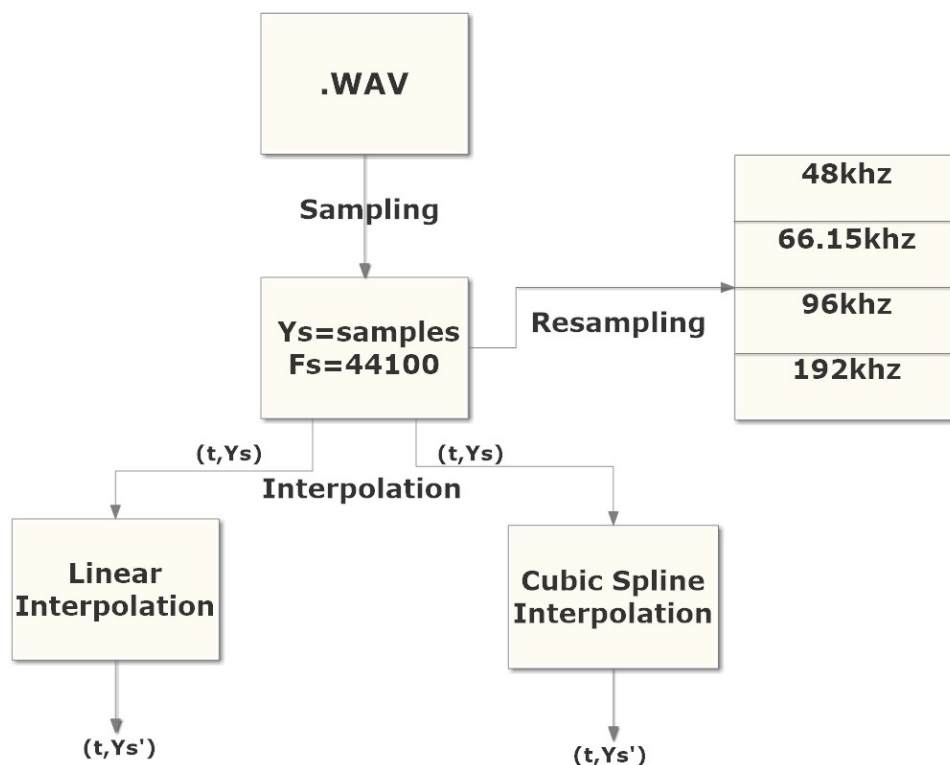
### 4.1 Εισαγωγή στο μοντέλο

Αρχικά διαβάζουμε όλη την ψηφιακή πληροφορία του ήχου (δείγματα και συχνότητα δειγματοληψίας). Αυτή τη δυνατότητα μας τη δίνει η συνάρτηση `audioread` του Matlab. Στη συνέχεια ανάλογα το πλήθος των δειγμάτων και τον ρυθμό δειγματοληψίας, φτιάχνουμε τον χρόνο  $t$  στον οποίο κινείται το κομμάτι. Αφού έχουμε αποκτήσει όλη τη πληροφορία μπορούμε να εφαρμόσουμε την τεχνική του resampling για να μελετήσουμε τη συμπεριφορά του ήχου σε κάθε sample rate.

Να αναφέρουμε, πως τα samples τα οποία διαβάζουμε, κανονικοποιούνται (normalized) και αποθηκεύονται σε ένα πίνακα τύπου double. Αυτό το κάναμε για να μας διευκολύνει στην επεξεργασία και κυρίως στο interpolation μετέπειτα καθώς τα native δεδομένα θα ήταν της μορφής int16. Η διαδικασία του resampling γίνεται βρίσκοντας έναν λογικό αριθμό (ratio of integers)  $P/Q$ , τέτοιο ώστε  $P/Q * 44.100$  (original sample rate) να μας κάνει το επιθυμητό rate που θέλουμε να φτάσουμε και τοποθετώντας αριθμητή και παρονομαστή μαζί με το δειγματοληπτημένο σήμα στη συνάρτηση `resample`.

Για τα stereo αρχεία ήχου, η διαδικασία του interpolation επαναλαμβάνεται καθώς κατά το sampling δεχόμαστε σαν είσοδο έναν πίνακα double της μορφής `Number of Samples * 2`, όπου έχουμε δύο στήλες, μια για κάθε κανάλι. Έτσι καλούμε τη συνάρτηση για τη μέθοδο interpolation που θέλουμε δύο φορές ξεχωριστά δημιουργώντας δύο πίνακες double με ίδιο μέγεθος και τέλος συνθέτουμε τα samples από τους δύο πίνακες για να πάρουμε το τελικό stereo interpolated αρχείο.

Μελετούμε τη κυματομορφή (waveform) και το σπεκτρογράμμο (spectrogram) του σήματος για κάθε sample rate. Το interpolation γίνεται επάνω στα δείγματα και τον χρόνο  $t$  και στις δύο μεθόδους. Στο παρακάτω διάγραμμα μπορούμε να δούμε το πως επεξεργάζεται το αρχείο .wav με τις διαδικασίες του resampling και του interpolation.



Σχήμα 4.1: Το μοντέλο της Matlab

## 4.2 Μελετώντας το σπεκτρόγραμμα

Το σπεκτόγραμμα(spectrogram) μας δίνει τη δυνατότητα να μελετήσουμε το φάσμα των συχνοτήτων για το ηχητικό σήμα μας αναλογικά με το χρόνο. Είναι μια ακριβής εικόνα του ήχου μας και απεικονίζει το πως συμπεριφέρεται το πλάτος για κάθε συνιστώσα συχνοτήτων στο σήμα.

Θα δημιουργήσουμε το σπεκτόγραμμα μας ψηφιακά, χρησιμοποιώντας μετασχηματισμό **Fourier** και ένα παράθυρο **Hamming** για να εξασφαλίσουμε τη καθαρότητα του σήματος με τα εξής χαρακτηριστικά:

Μεγεθος Παραθύρου  $L=1024$

Noverlap limit=512

Number of DFT points=2048

Το κλάσμα  $\frac{Power(dB)}{Frequency(Hz)}$  μας δείχνει την Φασματική Πυκνότητα Ισχύος(Power Spectral Density ) του σήματος κατά το μετασχηματισμό Fourier και απεικονίζει την δύναμη της μεταβολής της ενεργειας ως συνάρτηση της συχνότητας. Οσον αφορά τον ήχο, όσο πιο ομαλές είναι οι μεταβολές της PSD τόσο πιο ομαλός και καθαρός θα ακούγεται ο ήχος μας θεωρητικά. Δε θέλουμε σε καμία περίπτωση να υπερβούμε το όριο του ακουστικού πόνου για το αυτί μας, έτσι όσο πιο χαμηλά κυμαινόμαστε σε κάθε συχνότητα τόσο πιο εύκολα και ξεκούραστα θα μπορεί ο ακροατής να ακούσει το σήμα μας.

Το ηχητικό αρχείο του οποίου το σπεκτρόγραμμα μελετάμε, είναι το **Journey Into Sound**, ένα απόσπασμα 24 δευτερολέπτων που φτιάχθηκε ειδικά στα πλαίσια αυτής της διπλωματικής με συχνότητα δειγματοληψίας τα 44.100Hz στο οποίο έχει εφαρμοστεί 16bit dither στην έξοδο του.

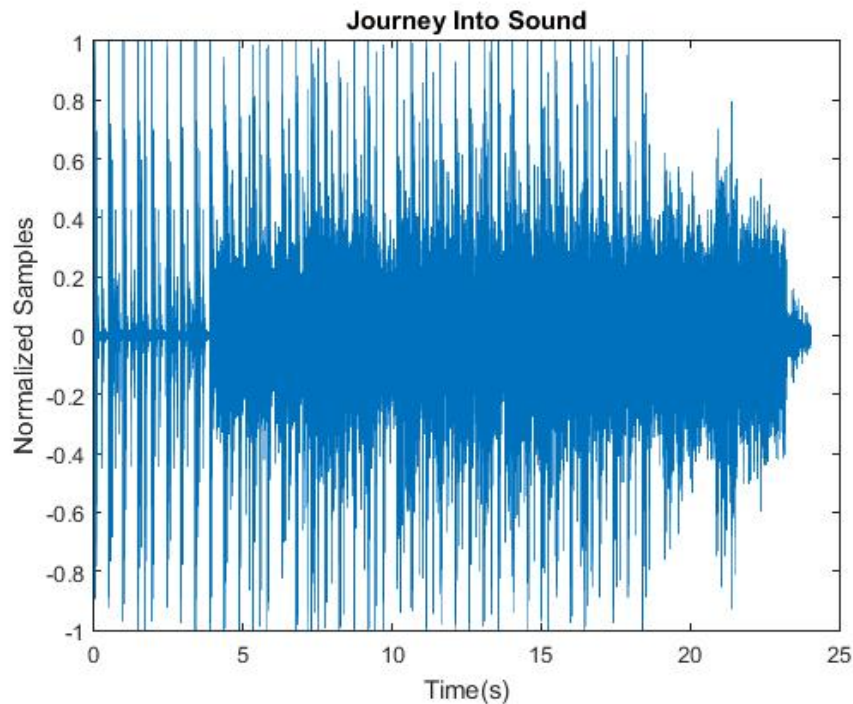


Figure 4.2: Η κυματομορφή του Journey Into Sound

Όπως φαίνεται και στη κυματομορφή, στα πρώτα τέσσερα δευτερόλεπτα ακούμε μόνο τα drums (kick, clap, hats, percussions) και στη συνέχεια αρχίζει το bass που καλύπτει τις χαμηλές συχνότητες, τα φωνητικά που καλύπτουν μεσαίες και υψηλές συχνότητες κατά κύριο λόγο και κάποια keys που καλύπτουν μεσαίες και χαμηλές. Λόγω του ότι είναι 'γεμάτο' σε όλες τις συχνότητες θα μας είναι εύκολο να εντοπίσουμε τις διακυμάνσεις του πλάτους λόγω resampling ή interpolation στο αρχικό σήμα.

### Resampling and Spectrogram

Θα εξετάσουμε τη συμπεριφορά του σπεκτρογράμματος σε διαφορετικές συχνότητες δειγματοληψίας του **Journey Into Sound** εφαρμόζοντας resampling χωρίς να προσθέσουμε κάποια πληροφορία. Τα sample rates που επιλέξαμε είναι τα:

- Sample rate 1 : 44.1kHz Cd quality
- Sample rate 2 : 48kHz Dvd quality
- Sample rate 3 : 66.15kHz Theoretical Best Quality
- Sample rate 4 : 96kHz Super Audio Quality
- Sample rate 5 : 192kHz High Definition Quality

Μελετώντας το σπεκτρόγραμμα του αρχικού μας σήματος στα 44.1kHz, παρατηρούμε το πλάτος να γίνεται εντονότερο στις μεσαίες και υψηλές συχνότητες. Ειδικά στα πρώτα και στα τελευταία δευτερόλεπτα, παρατηρούμε κάποιες αισθητές αιχμές (spikes) στη διακύμανση, σε απαγορευτικά για το αυτί μας επίπεδα (120+ dB/Hz).



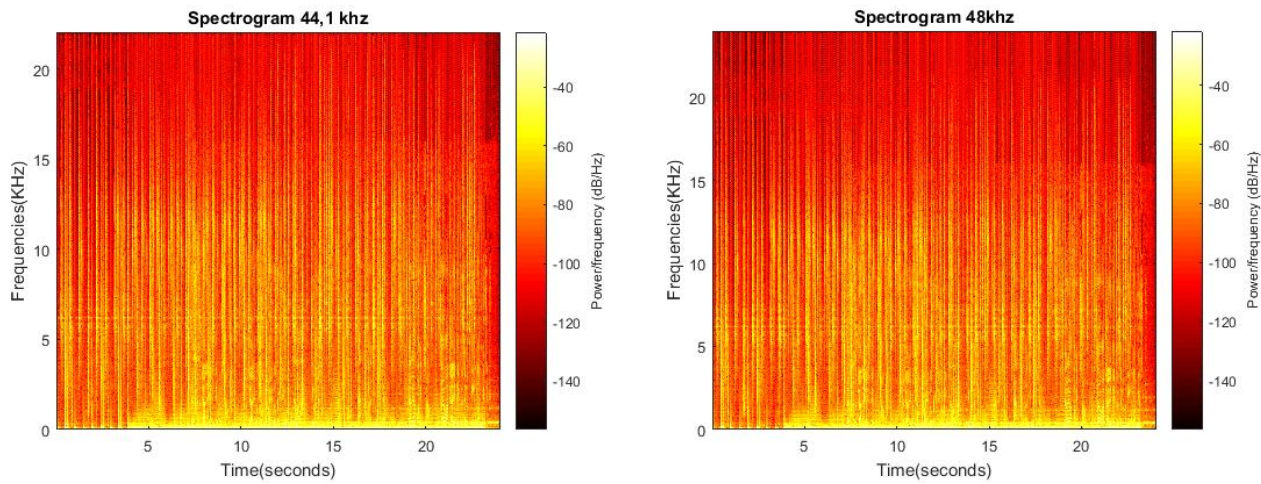


Figure 4.3: Spectrograms of 44.1Khz and 48Khz of the Original Audio File

Με resampling στα 48khz ,παρτηρούμε μια μικρή άυξηση του πλάτους στις υψηλές συχνότητες,ειδικά στα πρώτα τέσσερα δευτερόλεπτα.Στις υπόλοιπες συχνότητες οι διαφορές είναι ελάχιστες και ανευ σημασίας.Οι αιχμές παραμένουν και είναι και πιο έντονες.Παμε να δούμε πως αλλάζει το spectrum με εφαρμογή Linear και Cubic Spline Interpolation:

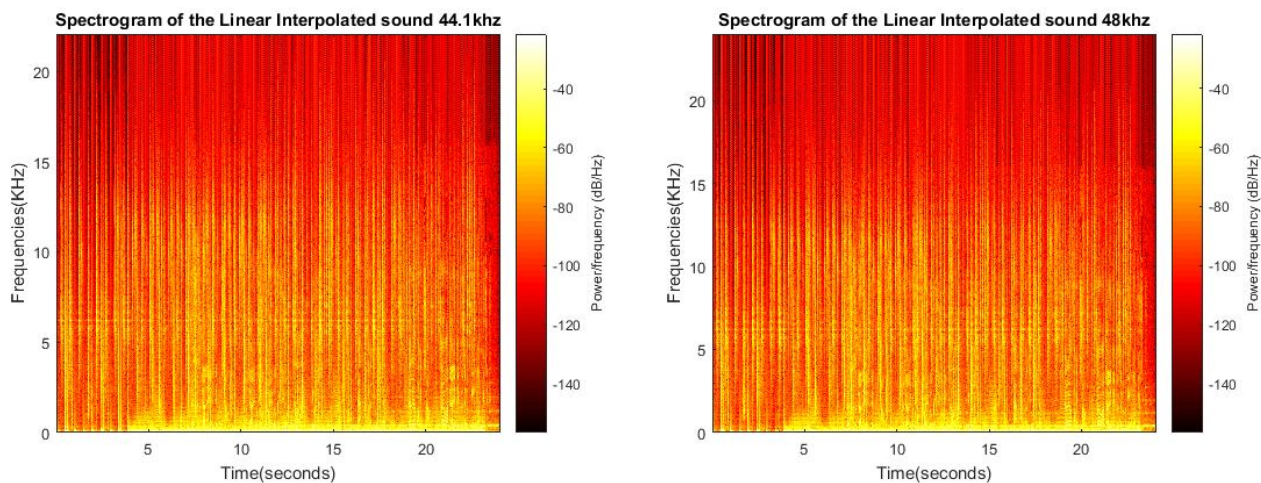


Figure 4.4: Spectrograms of 44.1Khz and 48Khz of the Linear Interpolation

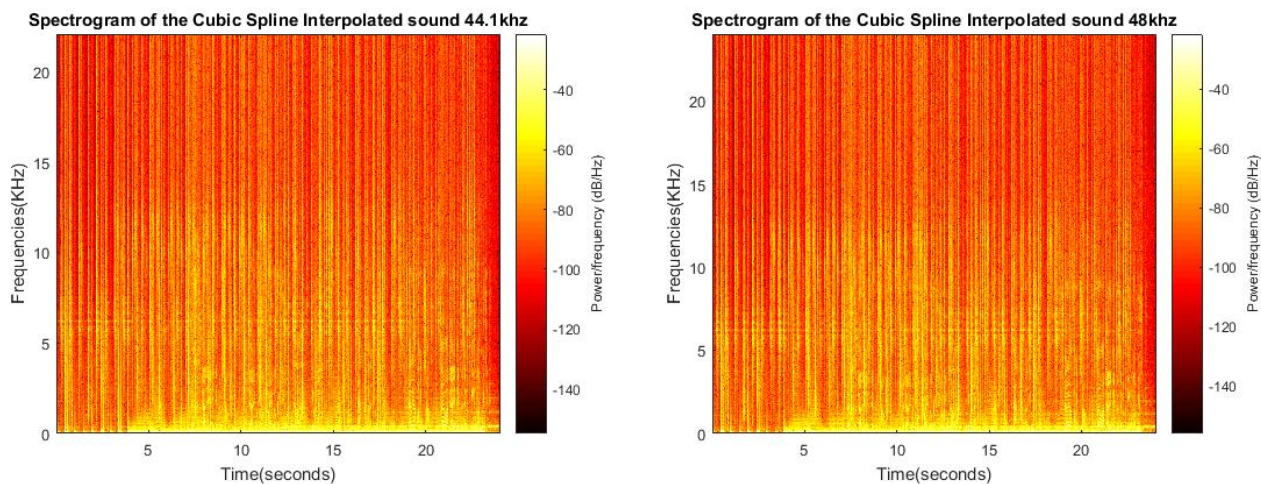


Figure 4.5: Spectrograms of 44.1Khz and 48Khz of the Cubic Spline



Τα σπεκτρογράμματα των δυο μεθόδων έχουν τρομερές διαφορές. Το linear σήμα είναι ελάχιστα χειρότερο του αρχικού όπως φαίνεται ειδικά στις υψηλές συχνότητες. Αντίθετα το σήμα που έχει υποστεί Cubic Spline Interpolation είναι αρκετά καθαρό. Παρατηρούμε πως οι αιχμές που υπήρχαν στο αρχικό και στο linear σήμα έχουν εξαλειφθεί και στις υψηλές συχνότητες το πλάτος έχει πλέον μετατοπιστεί ομαλά. Ελάχιστες είναι και οι διαφορές μεταξύ του interpolated σήματος στα διαφορετικά sample rate. Βλέπουμε πως η εφαρμογή του cubic interpolation κρατάει χαμηλά το μέγεθος της PSD και εξαλείφει τις διακυμάνσεις που παρουσιάστηκαν κατά την εφαρμογή resampling στο σήμα. Αυξάνουμε την συχνότητα δειγματοληψίας και παρατηρούμε εκ νέου τις διακυμάνσεις. Να σημειώσουμε πως το όριο των συχνοτήτων αλλάζει λόγω του αυξημένου εύρους αλλά εμάς μας ενδιαφέρουν κυρίως έως τα 20kHz που ακούει ο άνθρωπος.

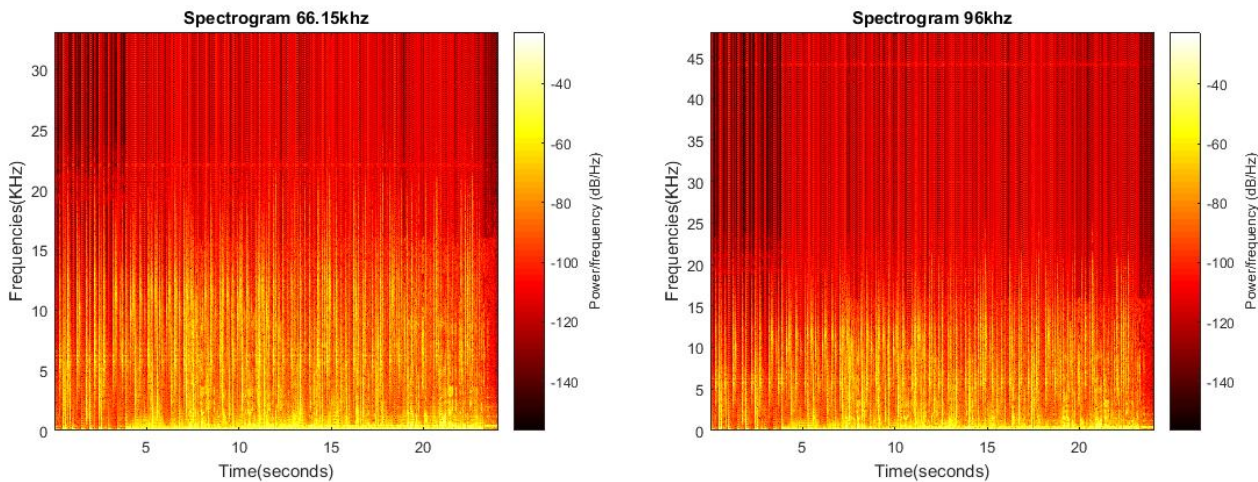


Figure 4.6: Spectrograms of 66.15khz and 96Khz of the Original Audio

Εντοπίζουμε εμφανείς διαφορές στο πλάτος με την αλλαγή στο ρυθμό δειγματοληψίας σε 66.15 και 96 khz. Μέχρι τα 20 khz δε παρατηρούμε κάποια σημαντική διαφορά σε σχέση με τα 44,1 και 48 khz αλλά πάνω από αυτά οι διαφορές είναι εμφανέστερες, όπου όσο αυξάνονται οι συχνότητες τόσο βλέπουμε να μεγαλώνει το πλάτος. Εφαρμόζουμε interpolation για τα δύο αυτά sample rates:

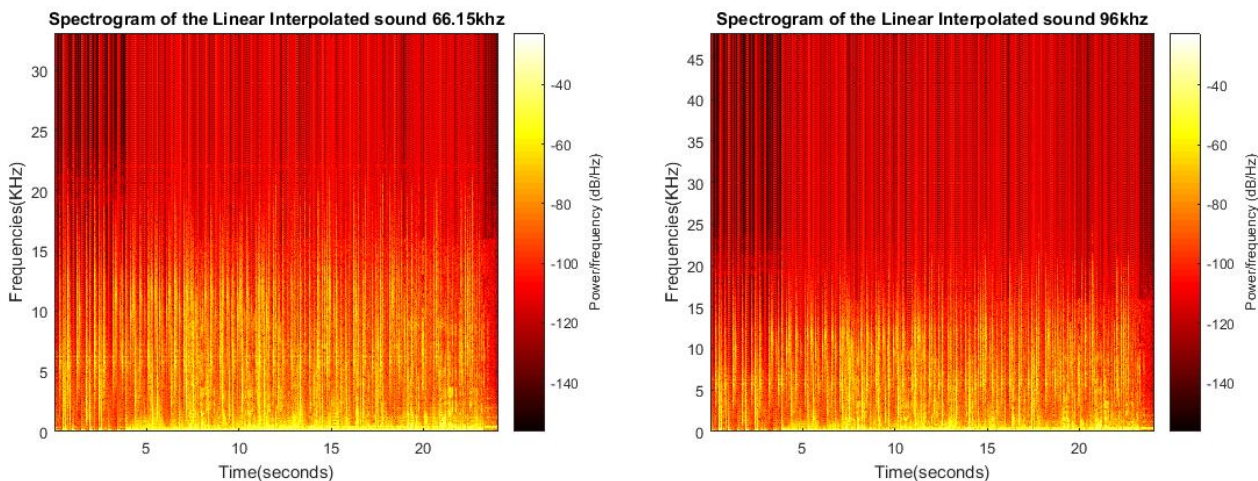


Figure 4.7: Spectrograms of 66.15khz and 96Khz of the Linear Interpolation

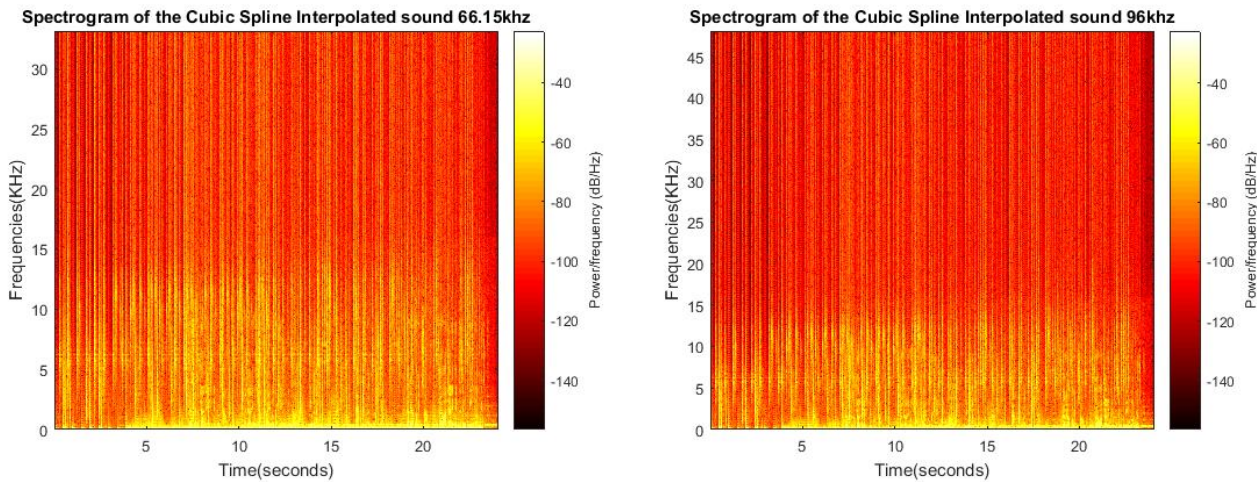


Figure 4.8: Spectrograms of 66.15khz and 96Khz of the Cubic Spline

Η Cubic μέθοδος μας ομαλοποίησε σε τεράστιο βαθμό τα spikes της ενέργειας του σήματος όπως γίνεται εύκολα αντιληπτό με γυμνό μάτι. Αντίθετα η linear μέθοδος συνεχίζει να μη τα πηγαίνει καλά και εμφανίζεται ίδια και χειρότερη από άποψη ενέργειας του αρχικού. Πειραματιστήκαμε με διαφορετικές τιμές στα χαρακτηριστικά του παραθύρου Hamming και του μετασχηματισμού Fourier ώστε να εξετάσουμε αν τυχόν μπορούμε να εντοπίσουμε κάποια διαφορά αλλά οπτικά είχαμε τα ίδια αποτελέσματα για όλες τις τιμές που δώσαμε. Ανεβάζουμε το ρυθμό δειγματοληψίας στα 192kHz και παρατηρούμε ξανά πως το σήμα μας 'καθαρίζει' όταν του εφαρμόσουμε Cubic Spline interpolation σε όλα τα πλάτη συχνότητων.

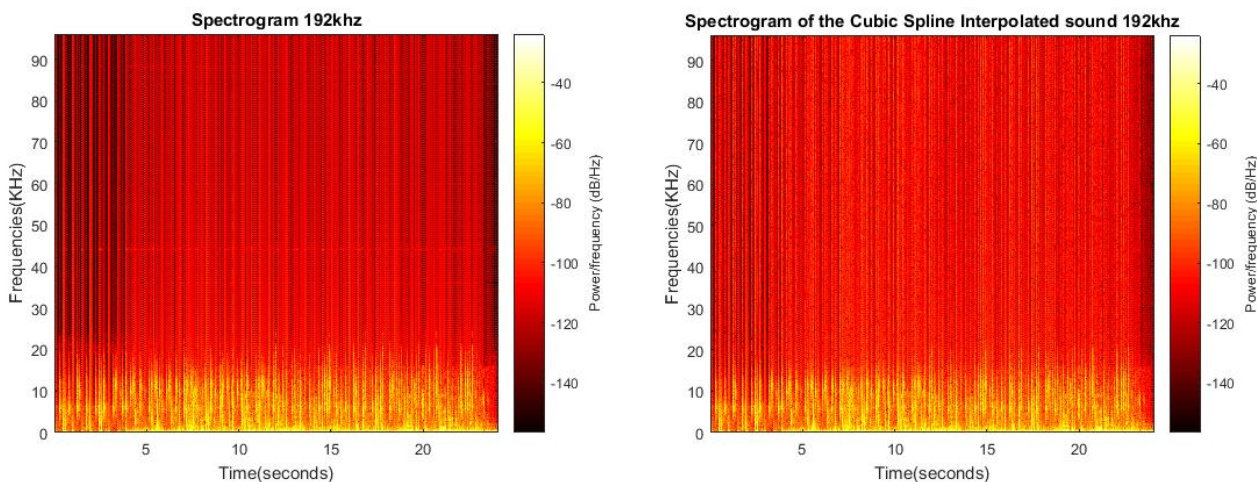
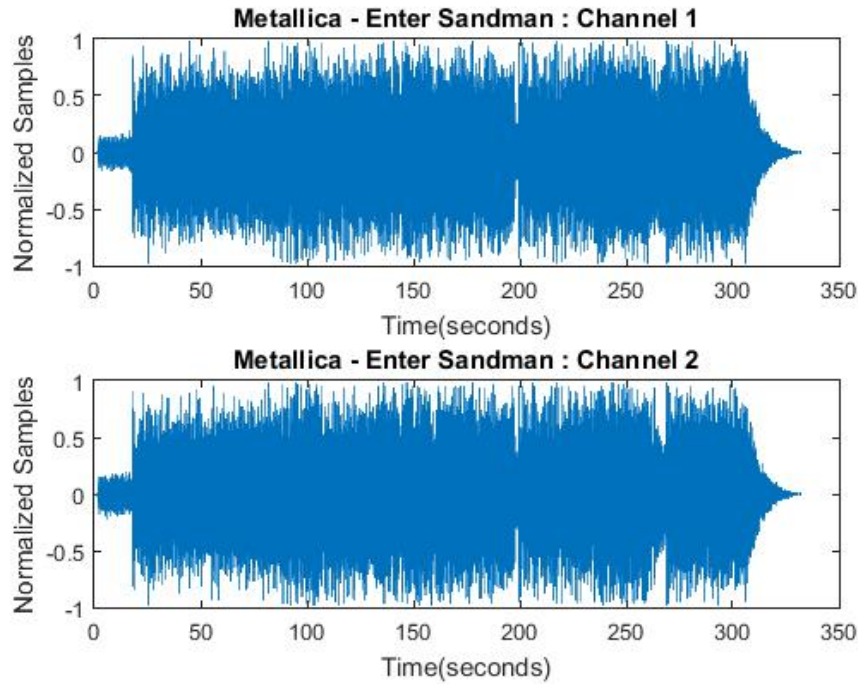


Figure 4.9: Spectrograms of 192Khz Original and Cubic Spline

Για τα μη ακουστικά κομμάτια του ήχου, στα οποία παρατηρούμε και το μεγαλύτερο πρόβλημα από άποψης amplitude, η ομαλοποίηση τους μέσω της μεθόδου μας μπορεί να σημαίνει αν δεχτούμε τον ισχυρισμό περί Hypersonic effect, βελτίωση της ποιότητας και μεγιστοποίηση της ψυχοακουστικής εμπειρίας. Αξίζει να σημειωθεί, πως στο απόσπασμα που μελετήσαμε, το μεγαλύτερο πρόβλημα οπτικά εντοπίζεται στην αρχή και στο τέλος του όπου είτε παίζουν μόνο τα drums, είτε μόνο το μπάσο και φωνητικά και βρίσκεται κατά κύριο λόγο στις υψηλές συχνότητες.



Αυτή τη φορά θα μελετήσουμε ένα ολόκληρο κομμάτι διάρκειας 5.30 λεπτών το οποίο έχει γίνει κατευθειαν rip απο το cd. Εμφανίζουμε μόνο τη μέθοδο με το Cubic και το αρχικό σήμα λόγω του ότι η linear μέθοδος είχε απογοητευτικά αποτελέσματα απο άποψη καθαρότητας σήματος σπεκτρογράμματος.



Σχήμα 4.10: Metallica-Enter Sandman Dual Channel Waveform

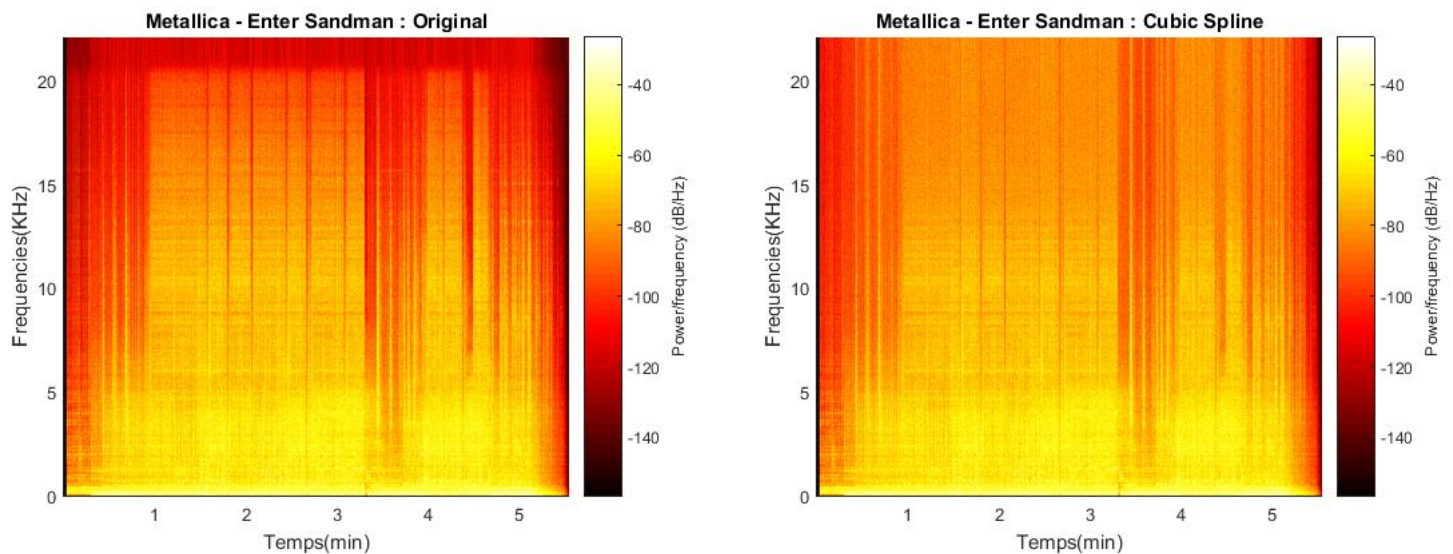


Figure 4.11: Spectrogram of the Original and of Cubic Spline for whole track

Παρατηρούμε εμφανή βελτίωση του σπεκτρογράμματος όπως και στα προηγούμενα παραδείγματα. Εξετάζοντας προσεκτικά σπεκτρογράμματα και κομμάτι, παρατηρούμε πως τα μεγαλύτερα προβλήματα απο άποψης amplitude εμφανίζονται όταν δεν έχουμε πλήρης συμμετοχή όλων των οργάνων (μπάσο, ντραμς, κιθάρες, φωνητικά), πράγμα που φαίνεται απ την αρχή μέχρι και σχεδόν το πρώτο λεπτό και γύρω στα τρία λεπτά και το τέλος. Πρακτικά δεν μπορεί να ακουστεί ο ήχος που δημιουργεί υψηλό PSD όμως η βελτίωση του μπορεί να δημιουργήσει την αίσθηση καλύτερης ποιότητας απο ψυχοακουστικής πλευράς.



### 4.3 Μελετώντας τις κυματομορφές

Μελετήσαμε τη συμπεριφορά του ήχου σε διαφορετικές συχνότητες ανάλογα με το χρόνο. Στη παράγραφο αυτή θα εξετάσουμε το πως επιδρά το interpolation στην κυματομορφή και στις ακμές της. Θεωρητικά, περιμένουμε με την εφαρμογή των Cubic Spline στο σήμα μας, η απεικόνιση των δειγμάτων στην κυματομορφή θα έρθει πιο κοντά σε μορφή καμπύλης και οι ακμές θα γίνουν πιο ομαλές σε σχέση με πριν. Δεδομένου πως το σημείο που γίνεται το interpolation βρίσκεται ανάμεσα σε δύο διαδοχικά samples και λόγω του ρυθμού δειγματοληψίας, η απόσταση μεταξύ τους είναι αρκετά μικρή, είναι πιθανόν η καμπύλη που δημιουργείτε να συνεχίζει να δημιουργεί ακμές όταν εξετάσουμε τα δύο επόμενα σημεία. Βασιζόμαστε στους τρεις συντελεστές που έχουμε θέσει για την καμπυλότητα.

Χρησιμοποιούμε αρχικά ένα mono drum loop διάρκειας 6 δευτερολέπτων:

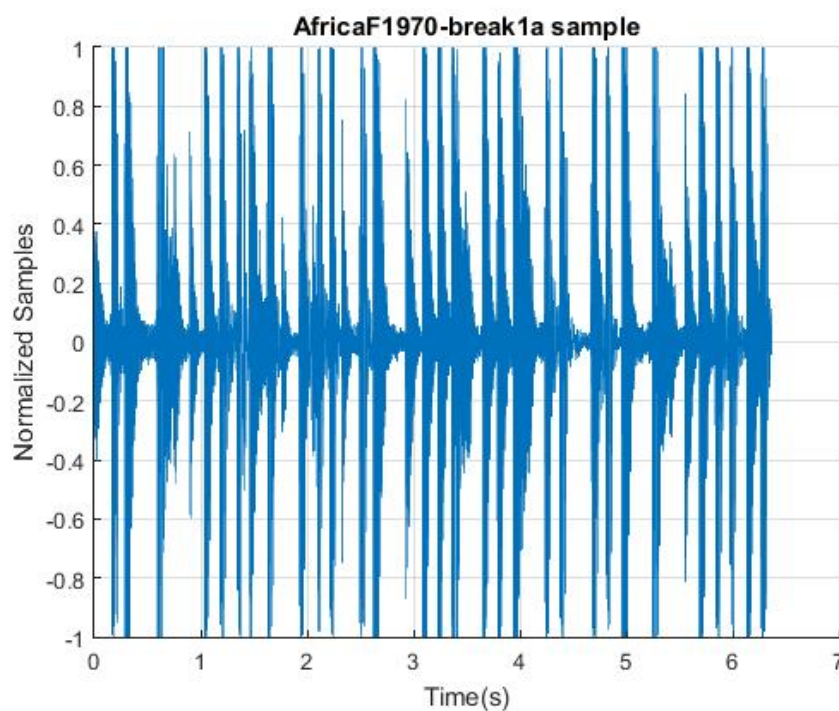


Figure 4.12: Africa Mono Drum Loop Waveform

Απο τα 280.840 δείγματα της λούπας, θα εμφανίσουμε τυχαία ένα απόσπασμα 64 δειγμάτων και θα συγκρίνουμε στην ίδια κυματομορφή τις διαφορές του αρχικού σήματος και του σήματος με interpolation. Με κόκκινο χρώμα απεικονίζουμε τα δείγματα που έχουν υποστεί interpolation και με μπλέ τα αρχικά.

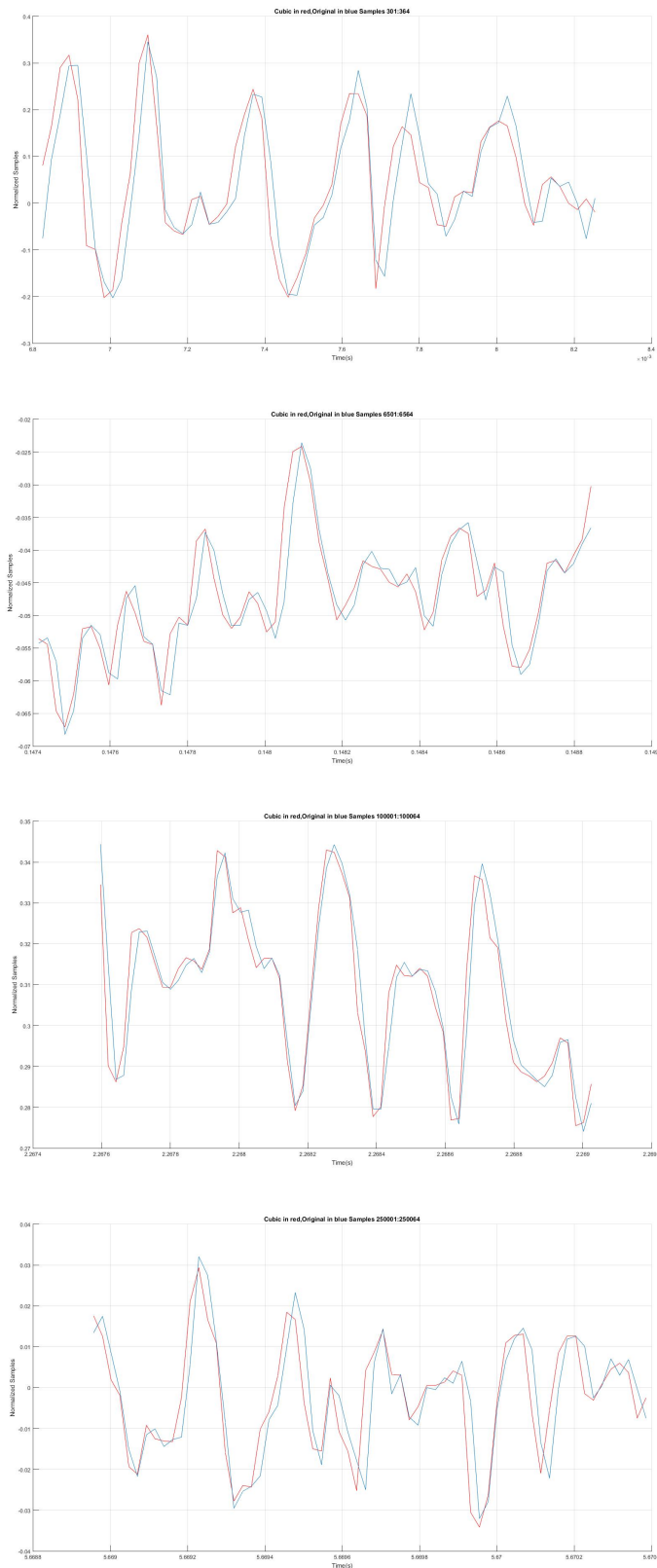


Figure 4.13: Waveform of the Mono Sample,Cubic and Original Comparison

Παρατηρούμε πως πρακτικά δεν καταφέρνουμε σε καμία περίπτωση να σχηματίσουμε τέλειες καμπύλες στις κυματομορφές μας όπως περιμέναμε. Παρόλα αυτά ερχόμαστε πιο κοντά στο σχήμα της καμπυλης απο το αρχικό σήμα. Όσον αφορά τις ακμές, στο αρχικό σήμα παρατηρούμε μια τριγωνική απότομη μορφή που η μέθοδος μας καταφέρνει να εξαλείψει σε αρκετές των περιπτώσεων. Έτσι γίνεται μια πιο ομαλή μετάβαση μεταξύ των δειγμάτων πράγμα που μας αφήνει ικανοποιημένους όσον αφορά το interpolation .

### Linear Interpolation vs Cubic Spline Interpolation

Συγκρίνουμε την μέθοδο cubic spline και αυτή του linear interpolation πάλι για 64 δείγματα:

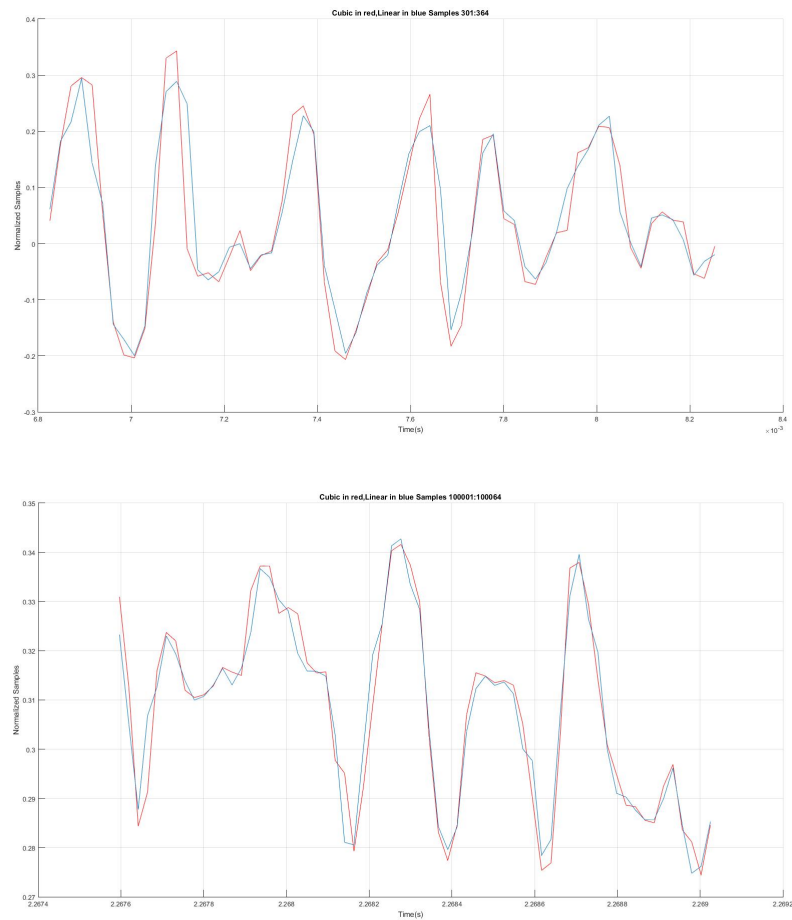


Figure 4.14: Cubic and Linear comparison of the Mono Sample

Η Linear μέθοδος έρχεται πιο κοντά στις κυματομορφές πριν την εφαρμογή interpolation, πράγμα λογικό αν σκεφθεί κανείς πως η cubic spline χρησιμοποιεί τρεις επιπλέον συντελεστές και η linear παίρνει πάντα τη μέση απόσταση μεταξύ δύο samples. Έτσι οι αχμές με την linear είναι πιο "άγριες" στις περισσότερες των περιπτώσεων.

### Stereo Interpolation

Χρησιμοποιούμε μια stereo λούπα διάρκειας τεσσάρων δευτερολέπτων και παρουσιάζουμε για 64 δείγματα τις αλλαγές στις κυματομορφές και για τα δύο κανάλια για κάθε μια από τις μεθόδους μας. Η προσθήκη ενός ακόμη καναλιού θέλουμε να κάνει το Interpolation πιο έντονο όσον αφορά την ακουστική εμπειρία δεδομένου πως ο ήχος θα φθάνει στα αυτιά μας από δύο σημεία.

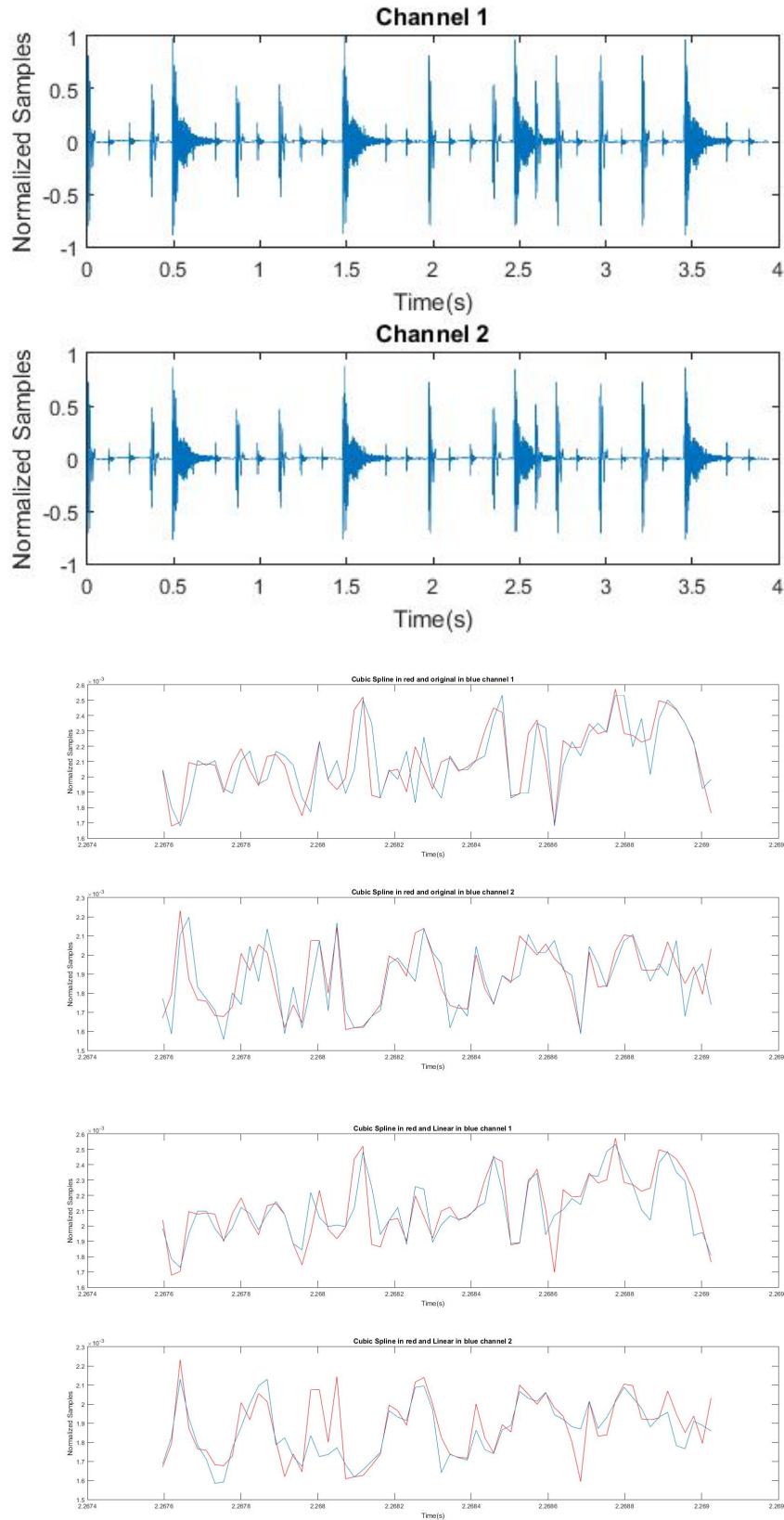


Figure 4.15: Waveform of the Stereo Sample,Cubic and Original Comparison

Σε γενικές γραμμές, παρατηρούμε πως οι κυματομορφές με τη Cubic μέθοδο, έρχονται πιο κοντά για συνεχόμενα δείγματα στην αρχική κυματομορφή πριν το interpolation σε σχέση με τη linear μέθοδο. Αυτό συμβαίνει καθώς η μέθοδος linear θα παίρνει πάντα το μεσαίο σημείο ανάμεσα σε δυο κυματομορφές ενώ η Cubic έχει περισσότερους συντελεστές και άρα μπορεί να εφαρμόζει το

interpolation με καλύτερη προσέγγιση όσον αφορά το σημείο ανάμεσα στα δύο δείγματα που μελετά κάθε φορά.

Χρησιμοποιώντας δύο κανάλια για την αποτύπωση του stereo ήχου, βλέπουμε μέσω των κυματομορφών πόσο διαφορετική είναι η έξοδος του κάθε καναλιού σε σχέση με το άλλο. Το γεγονός πως αν ακούσουμε το κάθε κανάλι ξεχωριστά (σαν mono) δεν καταλαβαίνουμε διαφορές σε σχέση με τη ποιότητα του ήχου παρα μόνο από το που ακούγεται η πηγή (δεξιά ή αριστερά ανάλογα το monitoring) μας δείχνει πως και με το interpolation, η ακουστική διαφορά που μπορεί να διακρίνει ένα απάιδευτο αυτί είναι πολύ δύσκολο να επιτευχθεί ειδικά σε ένα απλό setup ηχείων. Παρόλα αυτά, ακριβώς για το λόγο πως οπτικά διακρίνουμε πιο εύκολα τις διαφορές απότι με το αυτί και για ψυχοακουστικούς λόγους όπως αναφέρθηκαν στο τρίτο κεφάλαιο, το interpolation έχει τη δυνατότητα να βελτιώσει την ακουστική εμπειρία του ακροατή.

### Δοκιμή σε ολόκληρο κομμάτι

Μελετήσαμε τις κυματομορφές του κομματιού που χρησιμοποιήσαμε στο κεφάλαιο 4.2, πρόκειται για το Enter Sandman των Metallica (η επιλογή έγινε εντελώς τυχαία λόγω του ότι βρέθηκε σε αυθεντική ποιότητα cd).

Προβάλλουμε τόσο την αυθεντική κυματομορφή, όσο και τις interpolated, με πράσινο και κόκκινο η Linear και η Cubic αντίστοιχα.

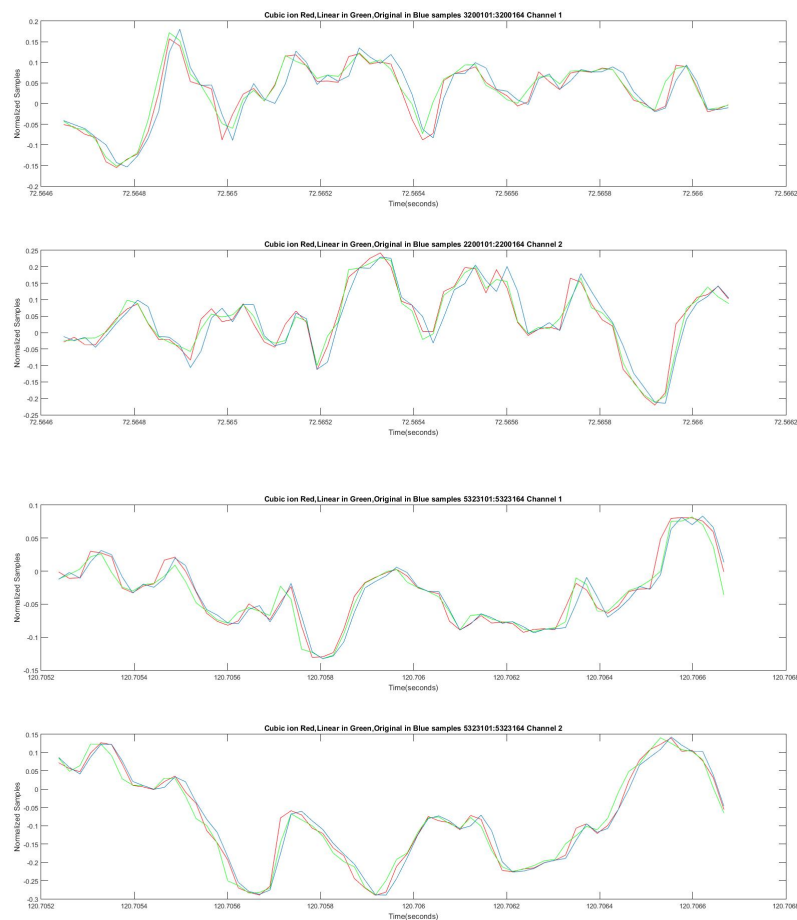


Figure 4.16: 64 δείγματα σε κάθε κανάλι απο το ολόκληρο κομμάτι

Παρατηρούμε ακριβώς τις ίδιες συμπεριφορές από τις μεθόδους μας, όπως και στα προηγούμενα samples που εξετασαμε. Και οι δύο μέθοδοι ομαλοποιούν τις αιχμές στις απότομες κορυφές στις περισσότερες των περιπτώσεων, με την Cubic Spline να προσπαθεί ειδικά όπως φαίνεται στο δεύτερο set καναλιών να δημιουργήσει χαμπύλες. Στο τρίτο set εμφανίζουμε 256 δείγματα αντι για 64.

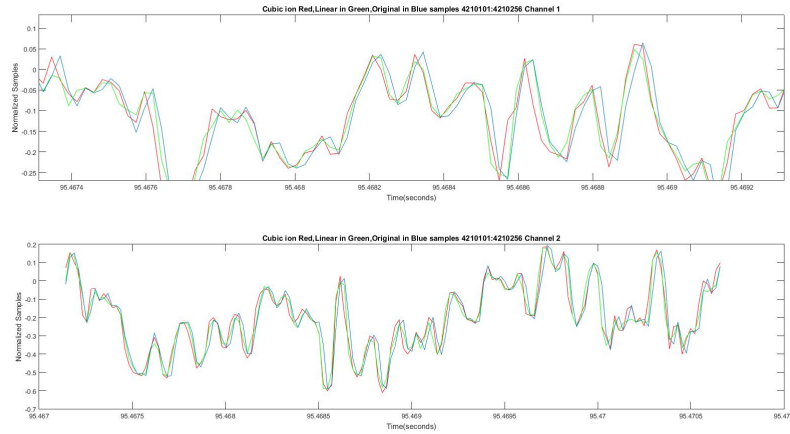


Figure 4.17: 256 δείγματα σε κάθε κανάλι απο το ολόκληρο κομμάτι

#### 4.4 *Resample* και κυματομορφές

Εφαρμόζοντας την τεχνική του resample στο sample μας και στη συνέχεια Cubic Spline-Linear Interpolation προσπαθούμε να ερευνήσουμε τη συμπεριφορά των μεθόδων μας οπτικά με τις κυματομορφές, πέρα από τα καθιερωμένα ακουστικά όρια, εκεί που δεν μπορούμε να ακούσουμε κάτι στη πραγματικότητα με τις ακουστικές μας δυνατότητες.

Συγκεκριμένα χρησιμοποιούμε την τεχνική του Upsampling η οποία στην πραγματικότητα είναι μια μορφή Interpolation στο σύστημα μας όπου παράγουμε μια προσέγγιση της ακολουθίας των samples όπως θα μπορούσαν να δειγματοληφθούν αν το κομμάτι μας βρισκόταν σε υψηλότερη συχνότητα. Δημιουργείται έτσι μια νέα ακολουθία με δείγματα που περιλαμβάνει τα παλιά όπου ανάμεσα τους τοποθετούνται μηδενικά. Στη συνέχεια εφαρμόζεται ένα χαμηλοπερατό φίλτρο το οποίο αντικαθιστά τα μηδενικά ώστε να εξομαλυνθούν τυχόν ασυνέχειες.

Για την συχνότητα 44.100Hz εμφανίζαμε 64 δείγματα στις κυματομορφές μας η οποία κάλυπτε περίπου 15 χιλιοστά του δευτερολέπτου. Για να καλύψουμε την ίδια χρονική περίοδο ενώ έχουμε εφαρμόσει υποσAMPLING κάνοντας τις πράξεις θα χρειαστούμε 70,96,140,279 δείγματα για 48,66.15,96,192 khz αντίστοιχα. Εμφανίζουμε σε κάθε κανάλι ξεχωριστά το σήμα με και χωρίς το Cubic Spline:



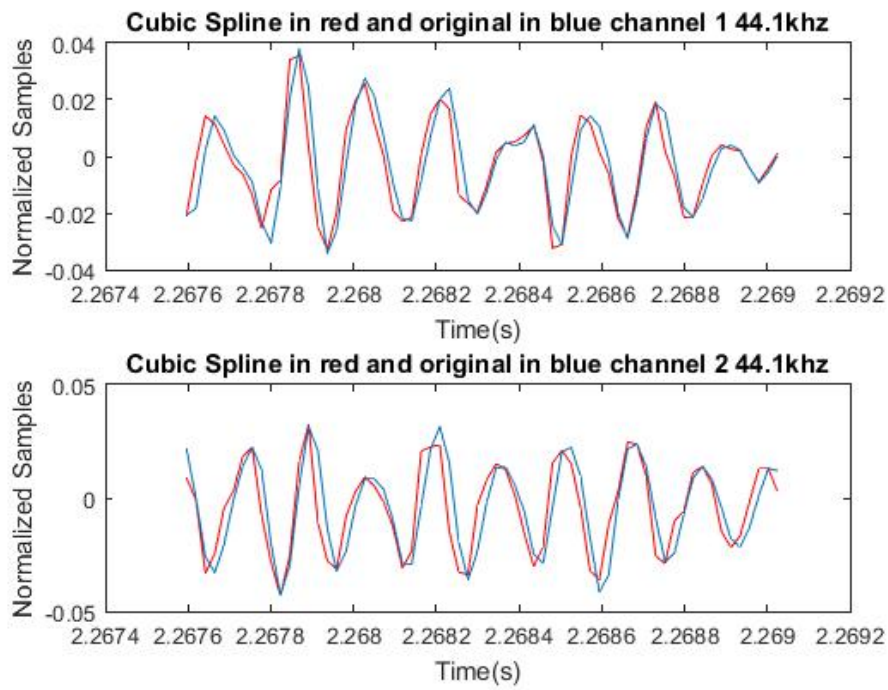


Figure 4.18: Cubic Spline and Original Comparison 44.1Khz

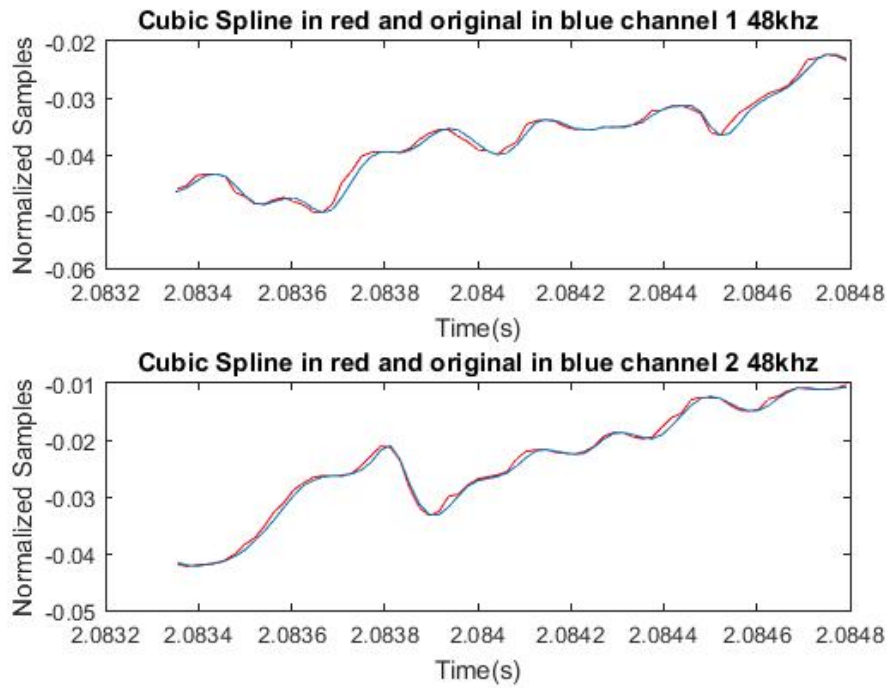


Figure 4.19: Cubic Spline and Original Comparison 48Khz

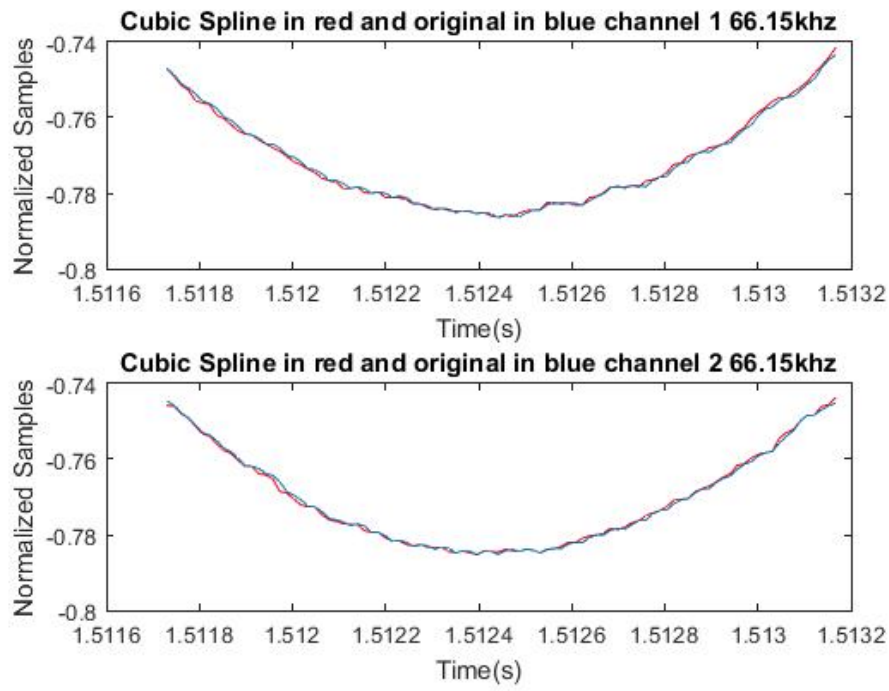


Figure 4.20: Cubic Spline and Original Comparison 66.15Khz

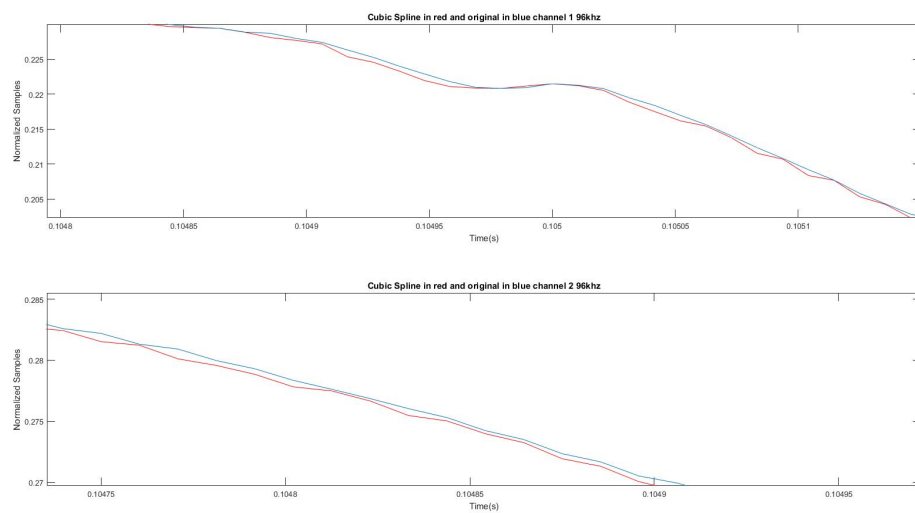


Figure 4.21: Cubic Spline and Original Comparison 96Khz



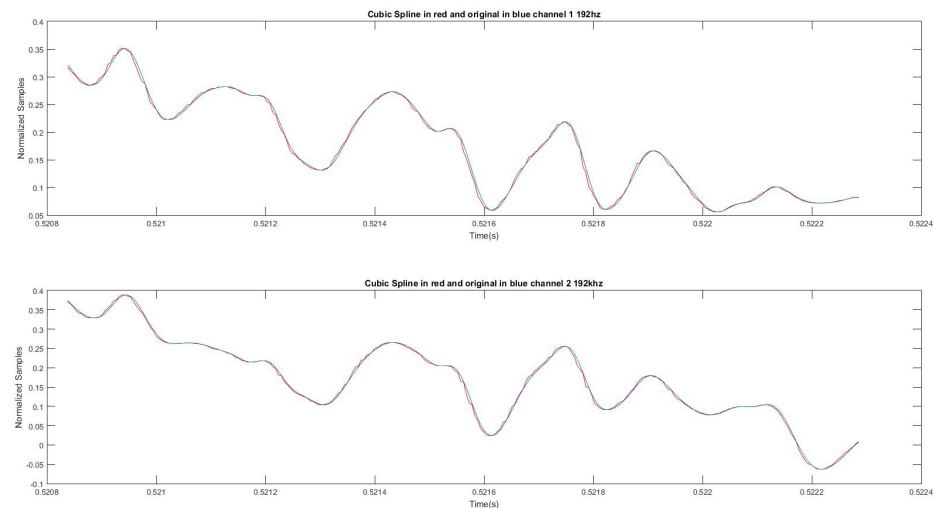


Figure 4.22: Cubic Spline and Original Comparison 192Khz

Η εφαρμογή του Resampling βλέπουμε πως μόνη της δίνει στο σήμα μας μια ομαλότητα. Το χαμηλοπερατό φίλτρο που εφαρμόζεται καταφέρνει να δημιουργήσει καμπύλες και ειδικά όσο αυξάνεται ο ρυθμός δειγματοληψίας αυτό φαίνεται όλο και πιο έντονα. Ομως όσο αυξάνουμε τον ρυθμό, τόσο πιο πολλές φορές κάνουμε interpolation ανάμεσα σε δυο διαδοχικά samples και αυτό έχει ως αποτέλεσμα να 'χαλάνε' οι καμπύλες που δημιουργούνται από το resample interpolation. Μπορούμε να διακρίνουμε τη διαφορά από το πρώτο resample στα 48khz και βλέπουμε σημαντικές διαφορές ειδικά από τα 96khz όπου εμφανίζουμε το στιγμιότυπο ολόκληρο και με zoom. Τέλος στα 192khz παρατηρούμε ακόμα πιο απότομες αλλαγές από την μέθοδο μας, ενώ το απλό resample αποδίδει αρκετά καλό απόψη καμπυλότητας. Φτάνουμε λοιπόν στο συμπέρασμα πως η μέθοδος μας αποδίδει καλύτερα όσον αφορά τις ακμές και τη καμπυλότητα, στο αυθεντικό σήμα στα 44.100HZ. Όταν την εφαρμόσουμε σε σήμα που έχει υποστεί ήδη Resampling άρα interpolation δεν αποδίδει το ίδιο καλά, ενώ η εφαρμογή χαμηλοπερατού φίλτρου στα μηδενικά που εφαρμόζει το Resample τα πηγαίνει εξαιρετικά, ειδικά όσο αυξάνουμε το ρυθμό στις ακμές των κυματομορφών.

## Κεφάλαιο 5

# Πειραματική επαλήθευση μοντέλου

Εξετάσαμε την συμπεριφορά του μοντέλου οπτικά μελετώντας τις κυματομορφές και το σπεκτρογράμμο στο κεφάλαιο 4. Σε αυτό το κεφάλαιο θα εξετάσουμε πειραματικά την ακουστική συμπεριφορά των μεθόδων μας.

Δυστυχώς λόγω περιορισμένων πόρων δεν ήμασταν σε θέση να πραγματοποιήσουμε ένα trial test στα πρότυπα των πειραμάτων που αναφέρθηκαν στο κεφάλαιο 3. Θα χρειαζόμασταν ένα καλό setup ηχείων η ακουστικών συνδεδεμένα σε έναν ενισχυτή και ένα DAC μαζί με τον player μας (θα μπορούσε να είναι ο υπολογιστής) μαζί με έναν ABX Comparator για την ακεραιότητα των αποτελεσμάτων, όπως επίσης και τουλάχιστον 20 ακροατές στους οποίους θα υποβάλαμε την διαδικασία με πολλαπλές επαναλήψεις.

### 5.1 Ακουστικά Tests

Για κάθε sample που χρησιμοποιήσαμε στο Matlab, περνάμε τις μεθόδους μας και μέσω της συνάρτησης audiowrite γράφουμε το interpolated sample και το ακούμε μέσω του player στον υπολογιστή μας. Η ακρόαση έγινε και σε 5.1 ηχεία (Logitech Z-5500) με Speaker Fill από τα Windows για Surround ήχο και σε 2.1 ακουστικά με noise cancellation.

#### 5.1.1 Interpolation Tests

Εφαρμόζουμε interpolation και ακούμε τα καινούρια αρχεία .wav που δημιουργούμε:

Αρχίζουμε από το mono drum sample διάρκειας 6 δευτερολέπτων. Παρόλο τις επαναλήψεις τόσο με τα ηχεία όσο και με τα ακουστικά δεν καταφέραμε να εντοπίσουμε κάποια αξιοσημείωτη διαφορά μεταξύ του αυθεντικού δείγματος και της μεθόδου με το Cubic Spline, όμως παρατηρήσαμε πως και τα δύο είναι ελάχιστα πιο "ζεστά" και "γεμάτα" από τη μέθοδο του Linear Interpolation.

Συνεχίζουμε μελετώντας το επόμενο δείγμα που χρησιμοποιήσαμε στα πειράματά μας, πάλι ένα drum sample διάρκειας 4 δευτερολέπτων αυτή τη φορά σε stereo mode. Αυτή τη φορά ήμασταν σε θέση να ξεχωρίσουμε το κομμάτι που έχει υποστεί interpolation από το κανονικό.

Με τα ακουστικά, μια μικρή λεπτομέρεια στο πως ακούγεται η μπότα (kick) μας έδωσε τη δυνατότητα να διακρίνουμε τις διαφορές έστω και μετά από τις πολλές επαναλήψεις χάνονταν η αίσθηση. Και σε αυτή τη περίπτωση η Linear μέθοδος ακούστηκε κατώτερη από τις άλλες, ειδικά στο snare το οποίο μας αφήνει μια "στεγνή" αίσθηση.

Στα ηχεία είδαμε και την μεγαλύτερη διαφορά μεταξύ της μεθόδου και του αυθεντικού κομματιού. Διαφορά τέτοια που μπορούμε να πούμε πως και το απα-

ιδεοτο αυτή δε θα είχε πρόβλημα στο να τη διακρίνει. Η Cubic μέθοδος έπαιζε surround ήχο με τη βοήθεια του speaker fill ,πράγμα που το αρχικό δείγμα αδυνατούσε να κάνει. Συγκεκριμένα, ακούσαμε πεντακάθαρα ένα closed hat απο τα περιφερειακά ηχεία, το οποίο ήταν μη ακουστό στο αυθεντικό δείγμα, αλλά παιζόταν μέσω των 2.1 ηχείων.

Αυτή η λεπτομέρεια δίνει στο κομμάτι μια πιο "ζεστή" αίσθηση και το κάνει να φαίνεται πιο γεμάτο. Αυτός είναι και ο σκοπός άλλωστε του speaker fill , να δημιουργήσει μια ψευδο-surround αίσθηση στον ακροατή όταν ακούει κομμάτια με 2 κανάλια. Πρακτικά δεν μπορούμε να εξηγήσουμε πως το Interpolation οδήγησε στο να βοηθήσει την surround εμπειρία γιατί δεν γνωρίζουμε πως ακριβώς λειτουργεί το speaker fill των windows αλλά ήταν μια ενδιαφέρουσα προσθήκη. Βέβαια σε συστήματα με 2.1 ηχεία, είτε υπολογιστή είτε συνδεδεμένα με ενισχυτή, αυτή η λεπτομέρεια δεν θα ήταν αισθητή.

Αυτή τη φορά θα εξετάσουμε το πλήρες δείγμα που δημιουργήσαμε για αυτήν τη διπλωματική, που περιέχει drums, bass, vocals που διαρκεί 24 δευτερόλεπτα. Και σε αυτή τη περίπτωση, με τη μέθοδο του τυφλού random shuffle , καταφέραμε να διακρίνουμε πότε το σήμα έχει υποστεί Cubic Spline interpolation με λιγότερη επιτυχία βέβαια, όχι με κάτι συγκεκριμένο όπως με το προηγούμενο δείγμα (είναι και αρκετά μεγαλύτερο) αλλά λόγω της αίσθησης που μας αφήνει ακούγοντας το, ίσως να έπαιξε ρόλο ο ψυχοακουστικός παράγοντας.

Στα ηχεία εντοπίζουμε και αυτή τη φορά σαφή βελτίωση στον surround ήχο. Αυτή τη φορά το κομμάτι χωρίς interpolation ακούγεται στα περιφερειακά ηχεία, πάλι ένα κομμάτι απο τα drums (hihat, closedhat, snares, claps) αλλά όταν το ακούμε με interpolation αυτό το σημείο ακούγεται πιο έντονα και καθαρά. Η linear μέθοδος συνεχίζει να είναι κατώτερη με τα ίδια προβλήματα που περιγράψαμε παραπάνω και ειδικά σε αυτό το sample λόγω των οργάνων που έχουν χρησιμοποιηθεί φαίνεται ξεκάθαρα. Είναι σα να έχουμε εφαρμόσει ένα reverb filter με κλειστά (ίσως και με αρνητικές τιμές) όλα τα room και wet effects , μια αντίστροφη λειτουργία δηλαδή απο τον σύνηθες ρόλο του φίλτρου αυτού στη μουσική παραγωγή.

Τελος, εξετάζουμε και ένα πλήρες κομμάτι. Όπως αναφέρθηκε στο κεφάλαιο 3, όταν προσπαθούμε να ξεχωρίσουμε έναν ήχο απο έναν άλλο στο θέμα της ποιότητας η του sampling rate , το ιδανικό είναι να χρησιμοποιούμε δείγματα 15-20 δευτερολέπτων. Παρόλο αυτο, συνεχίζουμε το πείραμα για ένα ολόκληρο κομμάτι, άλλωστε εκεί θέλουμε να χρησιμοποιείται και η μέθοδος μας.

Με τα ακουστικά δε καταφέραμε να ξεχωρίσουμε τα κομμάτια, αν και δοκίμασαμε να τα παίζουμε στο ίδιο δημείο αρκετές φορές. Έτσι δεν μπορούμε να καταλήξουμε σε κάποιο συμπέρασμα με σιγουριά. Ούτε με τα ηχεία καταφέραμε να διακρίνουμε κάποια διαφορά, ούτε στον surround ήχο, που εύκολα διακρίναμε στα προηγούμενα πειράματα. Επίσης ο παράγοντας του μεγάλου κομματιού που αναφέραμε έπαιξε ρόλο καθώς η μνήμη μας δεν κατάφερε, αν υποθέσουμε πως υπήρχαν, να συγκρατήσει τις διαφορές των κομματιών.

Η μέθοδος με το Linear Interpolation αν και αναγνωρίστηκε σχετικά εύκολα, τα πήγε ομολογουμένως καλύτερα απο ότι περιμέναμε βάση αυτών που είχαμε ακούσει στο προηγούμενα πειράματα.

### 5.1.2 Resample Tests

Εφαρμόζουμε την ίδια διαδικασία με παραπάνω αφού έχουμε εφαρμόσει re-sample και interpolation στον ήχο μας. Επιλέγουμε να μελετήσουμε τους ίδιους ρυθμούς και με τα προηγούμενα πειράματα δηλαδή στα 48, 66, 15, 96, 192 khz.

Η Linear μέθοδος που μας απογοήτευσε ακουστικά στα 44.1khz συνεχίζει να αποδίδει στα ίδια χαμηλά επίπεδα για όλους τους ρυθμούς δειγματοληψίας. Παρόλο αυτά μας άφησε την αίσθηση πως όσο ανεβαίνει ο ρυθμός, τόσο πιο πολύ "γεμίζει" ο ήχος, πράγμα που φαίνεται ιδιαίτερα στη σύγκριση του χαμηλότερου ρυθμού με το μεγαλύτερο.

Το Resample στο αυθεντικό κομμάτι ακουστικά δεν μπορούμε να εντοπίσουμε κάποιες έντονες διαφορές αν και η αίσθηση που αποκομίζουμε ανεβάζοντας το ρυθμό είναι πως ακούγεται καλύτερα και πιο ευχάριστα, πράγμα που όμως μπορεί να οφείλεται στον ψυχοακουστικό παράγοντα.

Με τη μέθοδο Cubic ενώ η εφαρμογή της στα 44.1kHz αλλάζει από την αρχική, με το Resample και ύστερα δεν μπορούμε να παρατηρήσουμε κάποια ουσιαστική διαφορά με το αυθεντικό κομμάτι στους αντίστοιχους ρυθμούς δειγματοληψίας. Φαίνεται πως η εφαρμογή των φίλτρων της μεθόδου Resample υπερσχύει αυτής του Cubic Spline Interpolation. Η αίσθηση πως όσο ανεβάζουμε το ρυθμό τόσο καλύτερα μας ακούγεται υπάρχει και σε αυτή τη περίπτωση χωρίς όμως να μπορούμε να το υποστηρίξουμε αυτό με πάσα βεβαιότητα.

Δοκιμάζουμε να κάνουμε Resample με το Cubic Spline interpolated σήμα, δηλαδή πρώτα η μέθοδος μας για το Interpolation και μετά το πέρασμα από το χαμηλοπερατό φίλτρο. Εδώ παρατηρούμε πως αυξάνοντας τη συχνότητα δειγματοληψίας το σήμα μας ακούγεται πιο κοντά με το πρώιμο Cubic σήμα, είναι σα να λέμε πως η μια μέθοδος ακυρώνει την άλλη και αυτή που θα εφαρμοστεί πρώτη είναι αυτή που στο τέλος φτάνει και στα αυτιά μας.

## 5.2 Πείραμα με θόρυβο

Βασιζόμενοι στον ισχυρισμό πως το 'απαίδευτο' αυτί δεν μπορεί να ξεχωρίσει διαφορές στην ποιότητα ειδικά όταν ακούει όλη την πληροφορία του ήχου, αποφασίσαμε να προσεθέσουμε φυσικό θόρυβο ανάμεσα στα samples και να εξετάσουμε κατά πόσο μπορούμε να ανακτήσουμε την αρχική ποιότητα, έστω και με ψυχοακουστικό τρόπο εφαρμόζοντας τις μεθόδους μας και interpolation.

Συγκεκριμένα σχεδιάσαμε το παρακάτω σύστημα:

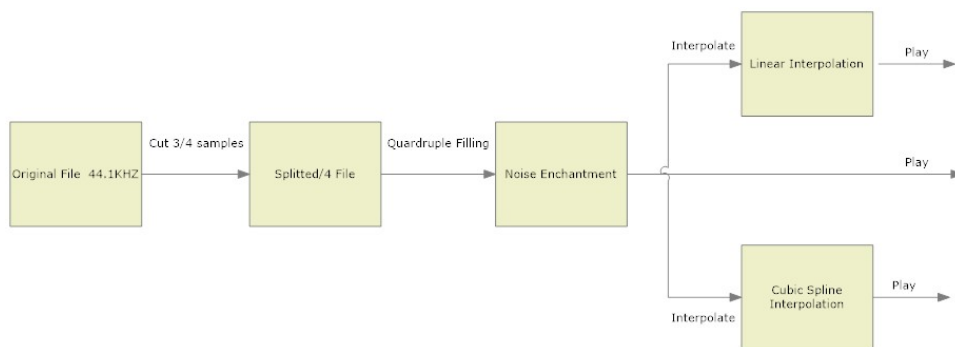


Figure 5.1: Το σύστημα εφαρμογής θορύβου και στη συνέχεια Interpolation στο αρχικό 'καθαρό' σήμα

### Οπτικό Πείραμα

Στο αρχικό σήμα, ανα τέσσερα samples κρατάμε το πρώτο και κόβουμε τα υπόλοιπα. Στη συνέχεια στη θέση των κομμένων προσθέτουμε τη τιμή του sample που κρατήσαμε αντι για αυτά. Με αυτό το τρόπο δημιουργούμε ένα φυσικό θόρυβο στο ηχητικό σήμα όπου κατά την αναπαραγωγή μοιάζει σαν να έχουν κάποιο πρόβλημα τα ηχεία μας.

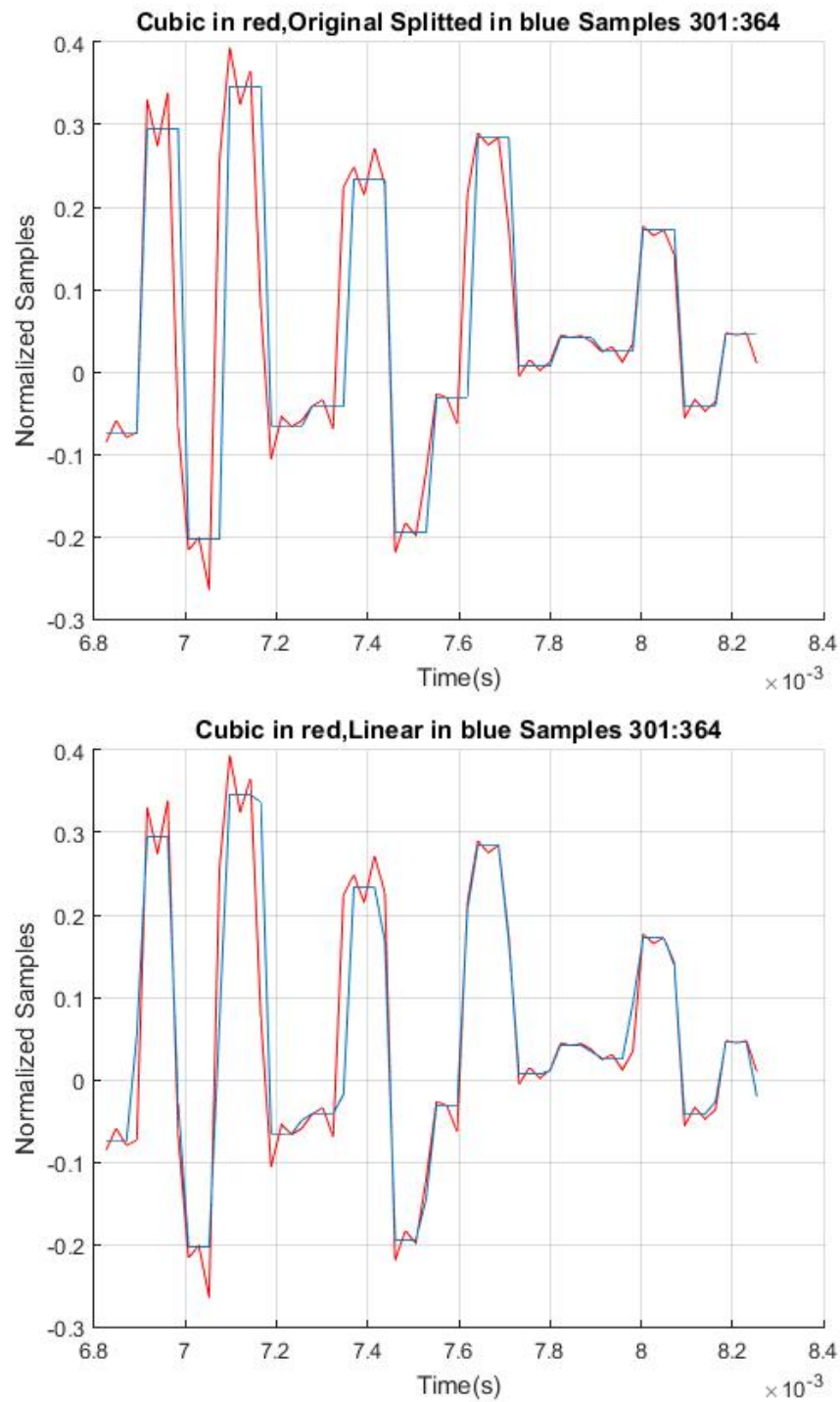


Figure 5.2: Η συμπεριφορά των μεθόδων σε σχέση με το αρχικό σήμα με θόρυβο

Παρατηρούμε πως με το θόρυβο η κυματομορφή μοιάζει σαν μια μορφή τετραγωνικού παλμού ενώ με την εφαρμογή interpolation μοιάζει περισσότερο σε κυματομορφή που απεικονίζει ένα ηχητικό σήμα.

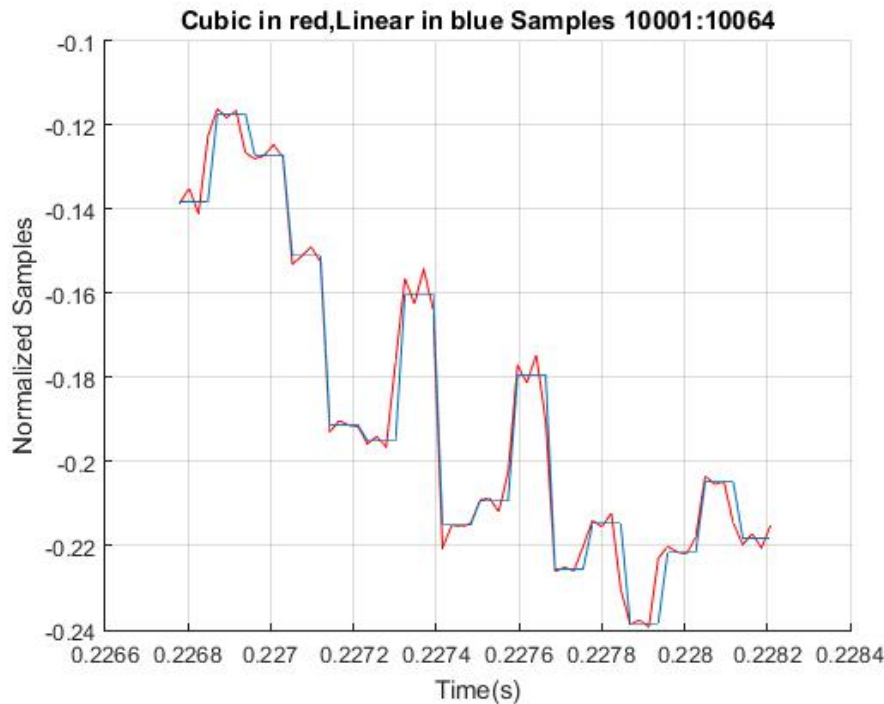


Figure 5.3: Επόμενο στιγμιότυπο των μεθόδων μας μετά την εφαρμογή θορύβου.

### Ακουστικό Πείραμα

Ακούμε τα κομμάτια με τον θόρυβο που δημιουργήσαμε, με ή χωρίς interpolation με τη διαδικασία που ακολουθήσαμε στην ενότητα 5.1. Παρατηρούμε αρχικά πως η μέθοδος Linear όπως περιμέναμε έχει τη χειρότερη ακουστική αίσθηση στα αυτιά μας. Η αμέσως όμως χειρότερη αίσθηση σε αυτό το πείραμα έρχεται από την Cubic, αφού δεν φαίνεται να καθαρίζει το κομμάτι αλλά και αυτή τονίζει ως ένα βαθμό τους θορύβους μειώνοντας την ακουστική εμπειρία. Το κομμάτι χωρίς interpolation ακούγεται πιο φυσιολογικά, δεδομένου του θορύβου που έχει εφαρμοστεί.

Ένα άλλο υποπείραμα που κάναμε σε αυτό το στάδιο είναι να κόψουμε και να κρατήσουμε από το κομμάτι το  $\frac{1}{4}$  και αντί να το γεμίσουμε με θόρυβο, να εφαρμόσουμε τη τεχνική resample επάνω στο κομμένο σήμα κατά 4 φορές πάνω στην τιμή του και γράψαμε το αποτέλεσμα στα 44100kHz. Εδώ παρουσιάστηκε και η χειρότερη από όλες εκδοχή του σήματος. Εφαρμόζοντας τώρα Cubic Spline Interpolation σε αυτό το σήμα βελτιώσαμε σε πολύ μεγάλο βαθμό την ακουστική του που συνεχίζει να είναι κατώτερος όμως όχι μόνο του σήματος με θόρυβο χωρίς interpolation αλλά και αυτού με το interpolation στο enchanted κομμένο σήμα.

### 5.3 Μελέτη του SNR

Στην ενότητα 4.2 μελετήσαμε το σเปκτρογράμμο των σημάτων πριν και μετά το interpolation και δείξαμε οπτικά πως η μέθοδος Cubic Spline κάνει το σήμα να φαίνεται πιο καθαρό όσον αφορά τη φασματική πυκνότητα ισχύος. Σε αυτήν την ενότητα συνεχίζουμε το πείραμα με τον θόρυβο, μετρώντας ποσοτικά το SNR (Signal-to-noise ratio) του αρχικού ανέπαφου σήματος, με αυτά στα οποία έχει εφαρμοστεί θόρυβος και interpolation.

Η τιμή του SNR μας δείχνει την ισχύ του σήματος σε σχέση με το θόρυβο του περιβάλλοντος. Το μετράμε σε Decibels και θέλουμε η τιμή του να είναι όσο δυνατόν μεγαλύτερη ώστε να έχουμε ένα πιο καθαρό σήμα. Στο ηχητικό απόσπα-

σμα που εξετάσαμε έχει εφαρμοστεί dither στην έξοδο για να μειωθεί τυχόν θόρυβος κβάντισης(Quantization Noise) στο αρχικό σήμα, το οποίο θέλουμε να θεωρούμε ως τέλειο.

Το μοντέλο που ακολουθούμε για τους υπολογισμούς μας είναι πως ως καθαρό σήμα παίρνουμε πάντα το original δειγματοληπτημένο κομμάτι χωρίς την εφαρμογή κανενός φίλτρου και σαν noise παίρνουμε με τη διαφορά του με το σήμα στο οποίο έχει περαστεί θόρυβος σύμφωνα με τη μέθοδο της ενότητας 5.2. Υπολογίζουμε τις τιμές τόσο με τις συναρτήσεις snr του Matlab όσο και τις θεωρητικές σύμφωνα με τους μαθηματικούς τύπους. Οι θεωρητικές τιμές που προέκυψαν είναι ακριβώς οι ίδιες με τις τιμές από τις συναρτήσεις γι αυτό δεν τις συμπεριλάβαμε στη μελέτη.

Ετσι λοιπόν προκύπτει ο παρακάτω πίνακας για το SNR:

No Interpolation		Cubic Spline Interpolation		Linear Interpolation	
Channel 1	Channel 2	Channel 1	Channel 2	Channel 1	Channel 2
12.4965dB	12.4261dB	13.6141dB	13.3747dB	14.1526dB	14.0712dB

Table 5.1: Πίνακας των SNR μετά την εφαρμογή θορύβου

Το SNR στα σήματα που έχουν υποστεί Interpolation μετά την εφαρμογή θορύβου είναι μεγαλύτερο κατά ποσοστά από 5% έως 13%. Η μέθοδος linear σημειώνει τα καλύτερα ratios στο SNR πράγμα που μας κάνει εντύπωση καθώς η ακουστική της απόδοση ήταν αρκετά χαμηλή. Τα σήματα με το θόρυβο και το interpolation μπορεί να τα πηγαίνουν αρκετά καλύτερα από άποψη SNR όμως κάτι τέτοιο δεν διαπιστώθηκε κατά το ακουστικό πείραμα. Φτάνουμε στο συμπέρασμα λοιπόν πως το SNR δε μπορεί να χαρακτηρίσει ποσοτικά τη ποιότητα ενός μουσικού κομματιού.

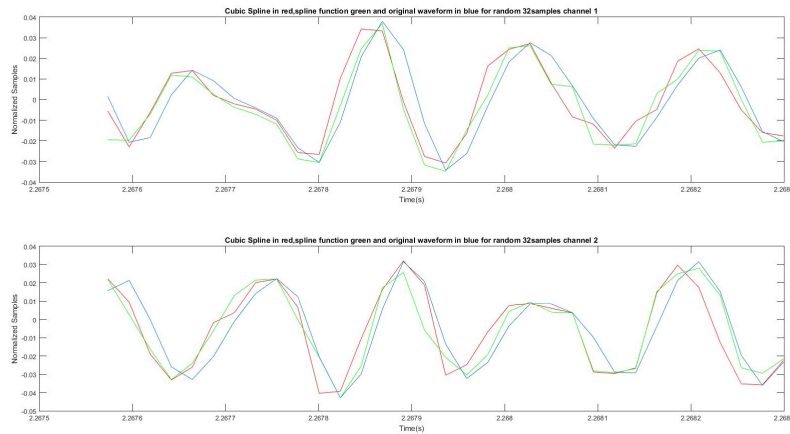
## 5.4 Επαλήθευση λειτουργίας μεθόδου

Το μεγαλύτερο ίσως κομμάτι της διπλωματικής, αφιερώθηκε στην υλοποίηση της μεθόδου για το Cubic Spline Interpolation. Από τη μελέτη, την επιλογή του καλύτερου μαθηματικού προτύπου για το Matlab και για τον ήχο ως την υλοποίηση σε κώδικα και την εφαρμογή πάνω σε αρχεία .wav, υπήρξαν αρκετές αμφιβολίες για το αν δουλεύει και πως δουλεύει ακριβώς η μέθοδος.

Για την επιβεβαίωση της λειτουργίας συγκρίναμε την μέθοδο μας με την έτοιμη συνάρτηση του Matlab spline η οποία δέχεται τις παραμέτρους x, y, χρόνος και samples στη περίπτωση μας, καθώς και το σημείο xx ο οποίος είναι το σημείο στο οποίο εφαρμόζεται το interpolation (παρεμβάλουσα).

Ο λόγος για τον οποίο δεν χρησιμοποιήσαμε εξαρχής την συνάρτηση spline για να εξετάσουμε τη συμπεριφορά του Cubic Spline interpolation στον ήχο, είναι πως θέλαμε μια πλήρη παραμετροποιήσιμη μέθοδο ώστε να τη μελετήσουμε ενδελεχώς σε όλα της τα βήματα και να έχουμε τη δυνατότητα να αλλάξουμε ό,τι επιθυμούμε και ενώ θα μπορούσαμε να μελετήσουμε το σπεκτρογράμμο και τις κυματομορφές χωρίς πρόβλημα, δε θα μπορούσαμε να μεταφέρουμε την μέθοδο σε μελλοντική σχεδίαση ενός ενσωματωμένου συστήματος και τη μεταφορά της σε κάποιο board.

Αρχικά τρέχουμε τις δύο συναρτήσεις για το Spline αυτόνομα, με διαφορετικά σημεία για τα interpolants δηλαδή για ένα stereo sample:



Παρατηρούμε πως η δικιά μας μέθοδος συμπεριφέρεται καλύτερα απο την έτοιμη συνάρτηση του Matlab οσον αφορά την ομαλότητα.

Αυτη τη φορά θα τρέξουμε την συνάρτηση spline μέσα στην μέθοδο μας ώστε να έχουμε ακριβώς τα ίδια σημεία για Interpolants για ενα mono sample :

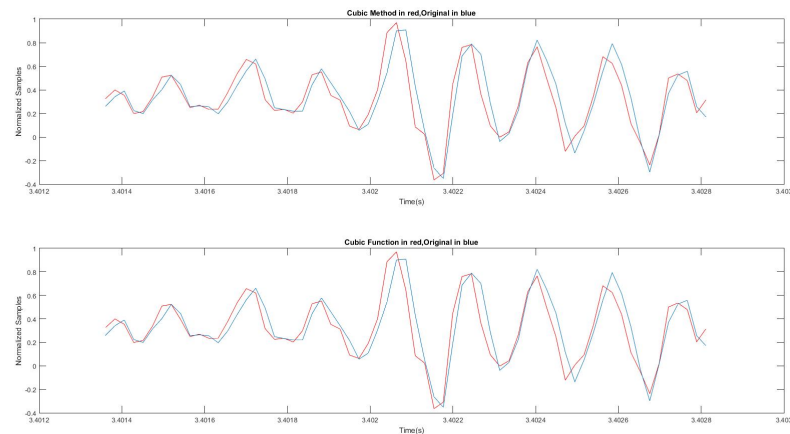


Figure 5.4: Spline Methods Comparison

Βλεπουμε λοιπόν πως έχουμε ακριβώς τα ίδια αποτελέσματα στις κυματομορφές για τις ίδιες μεθόδους,άρα επαληθεύουμε πως η μέθοδος μας λειτουργεί σωστά,τουλάχιστον απο μαθηματικής άποψης.Δε τρέξαμε τις δυο μεθόδους στην ίδια κυματομορφή όπως πριν γιατί είναι ακριβώς οι ίδιες και θα είχαμε επικάλυψη της μια και της άλλης.

## 5.5 Επαλήθευση μοντέλου Matlab για λειτουργία σε Hardware

Η Matlab μας έδωσε τη δυνατότητα να εκμεταλευτούμε όλη την ισχύ του υπολογιστικού μας συστήματος έτσι υπολογίσαμε σε ελάχιστο χρονικό διάστημα πίνακες με χιλιάδες εως εκατομμύρια δείγματα.Ετσι εφαρμόσαμε την μέθοδο Cubic Spline σε όλα τα δείγματα μαζί και γράψαμε το αποτέλεσμα σε καινούριους πίνακες.Αυτή η υλοποίηση σε μια fpga θα ήταν μη συμβατή απο άποψη τόσο πόρων αλλα και σχεδίασης.Είναι αδύνατο να εισάγουμε,να επεξεργαστούμε και να αποθηκεύσουμε μερικά εκατομμύρια δείγματα(για ένα μέσο κομμάτι τριών λεπτών σε τυπική συχνότητα δειγματοληψίας χρειαζόμαστε 8 εκατομμύρια δε-



ίγματα για κάθε κανάλι) σε ένα board,πόσο μάλλον όταν κάθε δείγμα είναι σε μορφή floating point και θα χρειαστούμε όλα τα διαθέσιμα blocks κάθε φορά πράγμα που σημαίνει πως το ρολόι της μονάδας δεν θα μπορεί να τρέχει στην μέγιστη δυνατή τιμή όπως θα θέλαμε.

Για να είναι λειτουργική λοιπόν η σχεδίαση μας,τροποποιούμε το σύστημα ώστε να εισάγουμε και να επεξεργαζόμαστε 4 δείγματα κάθε φορά βγάζοντας τους συντελεστές για κάθε ένα.Ετσι λοιπόν η ιδέα μας είναι να γίνονται streaming τα δείγματα 4-4 κάθε φορά ώστε να μπορεί να αναπεξέλθει μια σύγχρονη fpga με τους πόρους που διαθέτει και να βγάξει ως έξοδο τον νέο ήχο που έχει υποστεί το interpolation.Στο μαθηματικό μοντέλο ο αριθμός  $n$  που μας έδειχνε το πλήθος των δειγμάτων γίνεται 4 αντι για το μήκος του πίνακα που χρησιμοποιούσαμε ως τώρα στο Matlab και οι συντελεστές για το interpolation υπολογίζονται ανάλογα.

Παμε να δούμε κατα πόσο επηρεάζει το αποτέλεσμα του interpolation η τροποποίηση αυτή στη μέθοδο μας συγκρίνοντας τα αποτελέσματα πριν και μετά.Θεωρητικά για την κάθε τετράδα δειγμάτων τουλάχιστον δύο θα πέφτουνε μαζί η αρκετά κοντά,το δευτερο και το τρίτο,με το πρώτο και το τέταρτο να διαφέρουν καθώς οι τιμές που παίρνει το  $s_i$  επηρεάζει σε πολύ μεγάλο βαθμό τους συντελεστές b,c,d που δημιουργούν το spline.Τρέξαμε στο Matlab για κάθε κανάλι την ρουτίνα που υπολογίζει το Cubic πολλές φορές εισάγοντας τα δείγματα ανα τέσσερα μέχρι να τα επεξεργαστούμε όλα όσα έχει ο πίνακας και γράψαμε και εμφανίσαμε τα αποτελέσματα,συγκρίνοντας την μέθοδο όπου εισάγουμε όλο το πίνακα μαζί και το αρχικό σήμα:

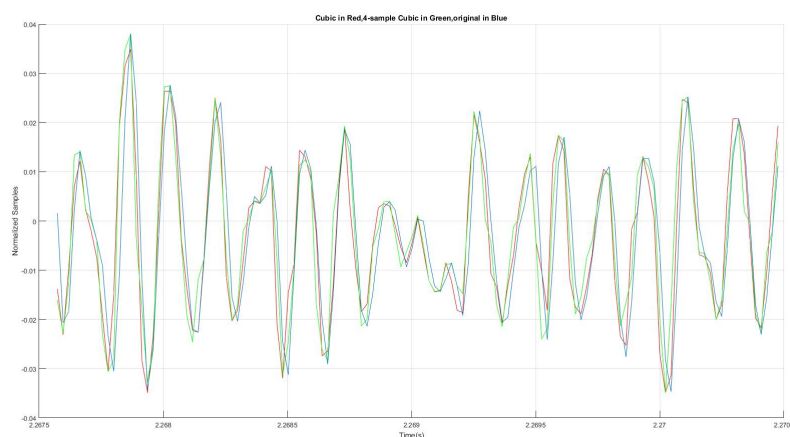


Figure 5.5: Οι 2 μέθοδοι με κόκκινο και πράσινο και η αυθεντική με μπλε.

Παρατηρούμε πως η μέθοδος τροποποιημένη για 4 δείγματα τη φορά,που φαίνεται με πράσινο χρώμα στο στιγμιότυπο,λειτουργεί αρκετά στο μοντέλο που περιμέναμε θεωρητικά καθώς αρκετά σημεία συμπίπτουν ακριβώς με τη προηγούμενη μέθοδο.Σε σύγκριση με το αρχικό σήμα του συμπεριφέρεται όπως θα περιμέναμε απο μια μέθοδο Cubic Spline Interpolation,όπως αυτή μελετήθηκε στο κεφάλαιο 4.Γενικώς κατα τα πειράματα είδαμε την streaming μέθοδο να έχει μια πιο έντονη μεταξύ διαδοχικών σημείων interpolation πράγμα άλλωστε που περιμέναμε γιατί το τελικό κομμάτι,όπως και κάθε στιγμιότυπο κυματομορφών που παρουσιάσαμε είναι μια συλλογή διαδοχικών εφαρμογών του Cubic Spline Interpolation.

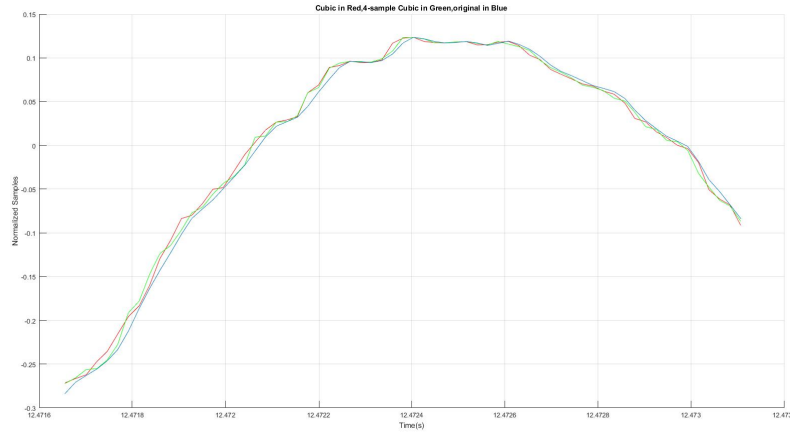


Figure 5.6: Η μέθοδος με τα 4 δείγματα με το πράσινο δείχνει εδώ πιο έντονη συμπεριφορά στο τρόπο που πάει να δημιουργήσει της καμπύλες

Εμφανίζουμε και το σπεκτρογράμμο για την "streaming" μέθοδο και παρατηρούμε πως η φασματική πυκνότητα ισχύος και ο καθαρισμός του σήματος είναι ακριβώς τα ίδια όπως τα περιγράψαμε παραπάνω.

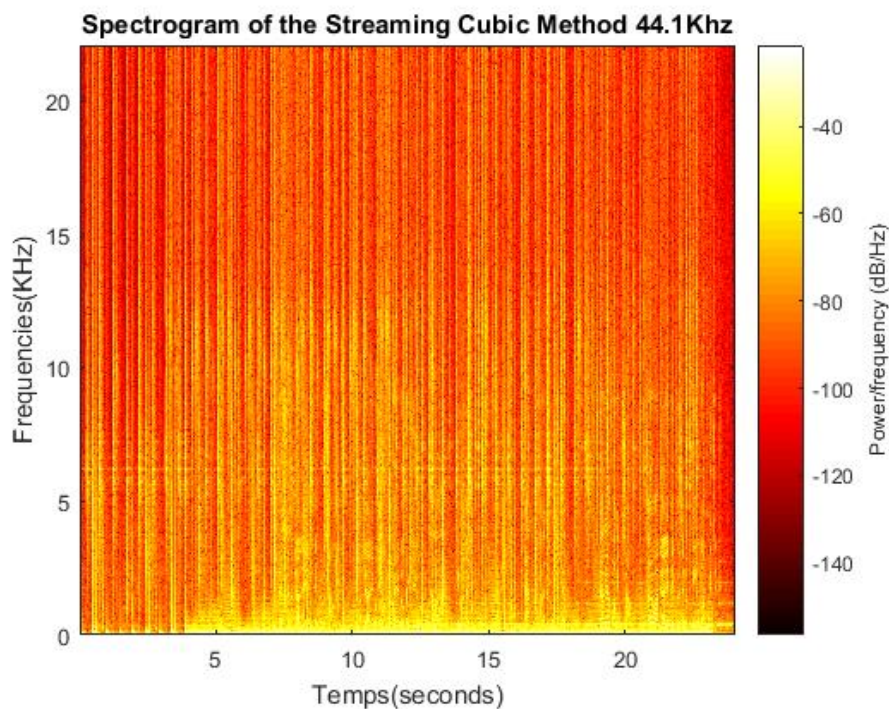


Figure 5.7: Το σπεκτρογράμμο της μεθόδου κινείται ακριβώς στα πλαίσια που περιγράψαμε στην ενότητα 4.2.

Στη συνέχεια γράφουμε το σήμα σε ένα αρχείο .wav. Ακουστικά, σε σύγκριση με τη προηγούμενη μορφή της μεθόδου δε μπορέσαμε να διακρίνουμε κάποια διαφορά, είτε με headphones είτε με τα ηχεία. Λαμβάνοντας υπόψιν όλα τα παραπάνω δεδομένα, μπορούμε να πούμε πως η "streaming" μέθοδος είναι λειτουργική οσον αφορά το interpolation στο αρχικό σήμα και μπορούμε να την χρησιμοποιήσουμε στο σχεδιασμό ενός ενσωματωμένου συστήματος σε hardware.

## Κεφάλαιο 6

# Σχεδίαση ενσωματωμένου συστήματος

### 6.1 Εισαγωγή στα Εργαλεία

Για τη σχεδίαση ενός ενσωματωμένου συστήματος βασισμένο στις μεθόδους μας για το interpolation, χρησιμοποιήσαμε τα εργαλεία Simulink της Matlab και το System Generator for DSP της Xilinx .

Το Simulink είναι μια πλατφόρμα σχεδιασμού μοντέλων τύπου block diagram και παρέχεται από τη Matlab μαζί με το κυρίως πρόγραμμα, ανάλογα την έκδοση που έχει επιλέξει ο χρήστης. Το interface του περιλαμβάνει ένα περιβάλλον σχεδιασμού με γραφικά blocks καθώς και μια βιβλιοθήκη η οποία περιλαμβάνει αυτά τα blocks από όπου και ο χρήστης επιλέγει ο,τι χρειάζεται για την μοντελοποίηση του. Υπάρχει πλήρης επικοινωνία με το workspace της Matlab έτσι ο χρήστης μπορεί να διαβάσει και να γράφει στοιχεία από και προς αυτήν.

Το System Generator for DSP είναι ένα εργαλείο σχεδίασης συστημάτων για ψηφιακή επεξεργασία σήματος το οποίο λειτουργεί μέσω του Simulink. Συγκεκριμένα παρέχει στις βιβλιοθήκες του μια πληθώρα από DSP blocks, από τα βασικά αθροιστές, καταχωρητές, πολλαπλασιαστές, πολυπλέκτες ως πιο πολύπλοκα όπως μνήμες, φίλτρα, FFT καθώς και τη δυνατότητα (όπως άλλωστε και το Simulink από μόνο του) να δημιουργήσει ο σχεδιαστής δικά του block τα οποία εκτελούν διάφορες λειτουργίες εφαρμόζοντας τους αλγορίθμους που έχει υλοποιήσει.

Ο συνδυασμός αυτών των εργαλείων μας βοηθάει στο να εξομοιώσουμε πλήρως το σύστημα που έχουν σχεδιάσει, να ελέγξουμε τη λειτουργία του θέτοντας όρια και παραμέτρους ανάλογα βέβαια και με τις δυνατότητες της fpga που σκοπεύουμε να χρησιμοποιήσουμε για το download της σχεδίασης. Για να συμβεί αυτό πρέπει να εξάγουμε τον κώδικα περιγραφής υλικού (VHDL) και αυτό μπορεί να γίνει μέσω του HDL Coder που περιλαμβάνεται στη Matlab και συνεργάζεται με το System Generator καθώς και το πρόγραμμα ISE της Xilinx μέσα από το οποίο ανάλογα την FPGA που έχουμε επιλέξει γίνεται το place and route όπου προσαρμόζεται η netlist στην αρχιτεκτονική του board και στη συνέχεια δημιουργείται το .bit file και η σχεδίαση είναι έτοιμη για το download.

### 6.2 Σχεδίαση του συστήματος

#### 6.2.1 Σχεδιασμός συστήματος *Linear Interpolation*

Σχεδιάζουμε αρχικά τη μέθοδο για το Linear Interpolation. Η μέθοδος αυτή όπως αποδείχθηκε μέσα από τη μελέτη και τα πειράματά μας να απέτυχε στην εφαρμογή της με ήχο άρα δεν θα είχε κανένα νόημα να την υλοποιήσουμε σε μια fpga, όμως την σχεδιάσαμε πρώτη για να εξοικειωθούμε με τα εργαλεία και τη προσομοίωση ώστε να υλοποιήσουμε και την πιο πολύπλοκη μέθοδο για το Cu-

bic Spline. Το Simulink μας δίνει τη δυνατότητα να διαβάσουμε δεδομένα από το Workspace της Matlab κατευθείαν στο μοντέλο μας. Για να εισάγουμε πίνακα στο μοντέλο και να χρησιμοποιήσουμε το System Generator χρησιμοποιούμε τα block **Fom Workspace** για να διαβάσουμε το πίνακα και το **Gateway In** για να ορίσουμε πως από εδώ ξεκινάει η δουλειά της fpga. Ο πίνακας αυτός πρέπει να είναι 2 διαστάσεων με τη πρώτη διάσταση να έχουμε τον χρόνο  $t$  και στη δεύτερη το σήμα μας, υποχρεωτικά καθώς μας περιορίζει το ίδιο το πρόγραμμα αφού θέλουμε να εισάγουμε matrix. Ένα πρόβλημα που εμφανίστηκε είναι τα κανάλια του ήχου, καθώς είναι ποιότητας stereo συνεπώς όταν γίνεται δειγματοληψία έχουμε ένα δισδιάστατο πίνακα με κάθε διάσταση να αντιπροσωπεύει ένα κανάλι του ηχητικού σήματος. Στη μοντελοποίηση σε Matlab καλούσαμε τη συνάρτηση που εφαρμόζει το interpolation ξεχωριστά για κάθε κανάλι και γράφαμε τον ήχο σε ένα νέο δισδιάστατο πίνακα. Για τη μοντελοποίηση για hardware έχουμε 2 επιλογές, το να εισάγουμε δύο πίνακες, έναν για κάθε κανάλι με το σύστημα να τρέχει ουσιαστικά δύο φορές ταυτόχρονα, θα χρειάζεται δηλαδή τους διπλάσιους πόρους, η να περνάμε ενα-ενα τα κανάλια με τον ήχο και να περιμένουμε να τελειώσει για το πρώτο, πράγμα που ίσως μας κοστίζει σε χρόνο εκτέλεσης και σε αυτή τη περίπτωση δεν θα έχουμε live streaming μετάδοση του ήχου. Με τη πιθανή χρησιμοποίηση του Altec Codec '97, μπορούμε να εισάγουμε σε fpga δικάναλο ήχο, άρα μπορούμε να βασίσουμε τη σχεδίαση μας σε αυτό και να διαλέξουμε την πρώτη επιλογή ώστε θεωρητικά να μην υπάρχει καθόλου καθυστέρηση στην μετάδοση του ήχου από το σύστημα. Έτσι λοιπόν το σύστημα θα τρέχει ταυτόχρονα 2 φορές για κάθε κανάλι με ακριβώς τον ίδιο τρόπο. Το μοντέλο που σχεδιάσαμε έχει μια αρκετά απλή μορφή. Κάνει streaming τα δείγματα 2-2 κάθε φορά και με τους κατάλληλους αθροιστές, πολλαπλασιαστές και διαιρέτες δημιουργείται το μοντέλο μας. Επιλέξαμε το 2by2 streaming λόγω της φύσης και της απλότητας του μοντέλου, αφού όλη η πληροφορία που χρειαζόμαστε για κάθε interpolant βρίσκεται μεταξύ δύο διαδοχικών samples.

# Linear Interpolation Module

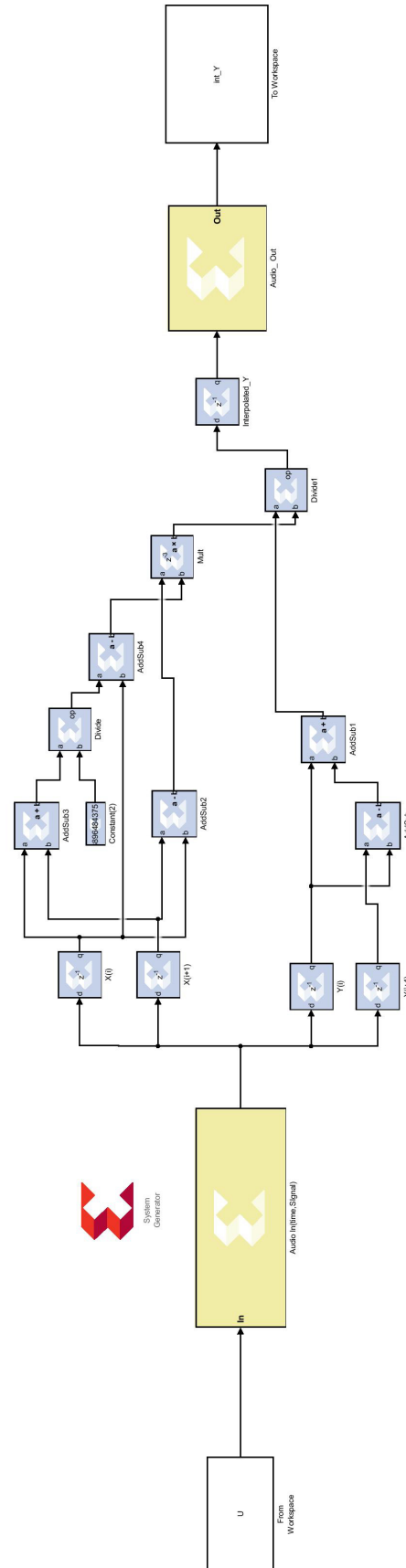


Figure 6.1: Για κάθε δύο διαδοχικά δείγματα που διαβάζουμε για κάθε κανάλι, γίνονται οι πράξεις και σε πραγματικό χρόνο δίνουμε στην έξοδο το interpolatedσήμα.



Το βασικό πρόβλημα που αντιμετωπίσαμε στην μεταφορά της σχεδίασης μας απο το περιβάλλον Matlab σε ένα περιβάλλον για hardware ήταν στο πως θα γίνεται η μετάδοση του ήχου και που θα αποθηκεύεται αυτός. Στην software υλοποίηση αποθήκευαμε όλα τα δεδομένα σε πίνακες, πριν και μετά το interpolation. Έτσι σαν είσοδο στην μέθοδο δίναμε τον πίνακα με τα δείγματα, επεξεργαζόμασταν όλα τα στοιχεία σύμφωνα με το μαθηματικό μοντέλο και επιστρέφαμε στην έξοδο τον πίνακα interpolated και στη συνέχεια γράφαμε τον ήχο σε ένα νέο αρχείο .wav. Στη hardware σχεδίαση κάτι τέτοιο είναι αδύνατο απο άποψη πόρων αλλά και ευχρηστίας, καθώς θα έπρεπε να επεξεργαστούμε και να αποθηκεύσουμε εκατομμύρια δείγματα για κάθε κανάλι. Έτσι λοιπόν καθώς μας έρχεται ο ήχος μέσω του Altec Codec, παίρνουμε απο 4 δείγματα (απο κάθε κανάλι), εφαρμόζουμε interpolation και στέλνουμε στην έξοδο το αποτέλεσμα το οποίο παίζει στα ηχεία μας. Συνεχίζουμε για κάθε τετράδα δειγμάτων μέχρι να τελειώσουμε το κομμάτι. Έτσι λοιπόν σπάμε σε κομμάτια την μέθοδο εφαρμόζοντας την τοπικά πολλές φορές μέχρι να σταματήσουν να μας έρχονται τα bitstream. Δοκιμάσαμε την παραλλαγή αυτήν και στο περιβάλλον της Matlab τρέχοντας όλα τα πειράματα και δεν παρατηρήσαμε καμία ουσιαστική διαφορά τόσο οπτικά όσο και ακουστικά που θα δημιουργούσε κάποιο λειτουργικό πρόβλημα.

Την τετράδα των δειγμάτων που δεχόμαστε την αποθηκεύουμε σε μια FIFO (ουρά). Χρησιμοποιήσαμε ένα μπλοκ 64k depth με ακρίβεια 16 bit για να μπορούμε να καλύψουμε όλο την ανάλυση του κάθε δείγματος. Αν αντι για τετράδες παίρναμε μεγαλύτερο αριθμό δειγμάτων θα χρειαζόμασταν μεγαλύτερη FIFO. Υπόψιν πως κάθε δείγμα μπαίνει στο σύστημα σε ένα κύκλο ρολογιού, άρα αφού για μια τετράδα δουλεύει είμαστε ικανοποιημένοι γιατί προσπαθούμε να κρατήσουμε αυτόν τον αριθμό χαμηλά. Το σύστημα για να είναι λειτουργικό θα πρέπει να γραφτεί ένα Control Unit το οποίο θα καθορίζει τα enables σε μνήμες, πολυπλέκτες και στην ουρά και θα βγάζει ποιο δείγμα θα περνάει μέσα στο σύστημα κάθε φορά. Το Unit αυτό θα είναι αρκετά πολύπλοκο με τους ελέγχους καθώς πρέπει να είμαστε σίγουροι πως έχουμε το σωστό sample σε κάθε περίπτωση.

Όσον αφορά τους χρονισμούς, χρησιμοποιήσαμε δεκάδες ανθροιστές, διαιρέτες, πολυπλέκτες και πολλαπλασιαστές (καθώς και negators) όπου ο καθένας θα χρειάζεται τουλάχιστον μια ακμή του ρολογιού για να κάνει τη πράξη. Στο κομμάτι που φτιάχνουμε τους συντελεστές  $\sigma$  θα χρειαστούμε λόγω της μεθόδου back substitution όλα τα δείγματα μέχρι εκείνο το σημείο επεξεργασμένα ώστε να ξεκινήσουμε απο το τελευταίο. Για την αποθήκευση αυτών των δειγμάτων θα χρειαστούμε μνήμες ram. Στην σχεδίαση μας προτιμήσαμε την dual port ram για την αποθήκευση των συντελεστών  $\alpha$  και  $\beta$ . Στην υλοποίηση μπορούμε να χρησιμοποιήσουμε μια μνήμη cache ώστε να μπορούμε να βρίσκουμε με μεγαλύτερη ταχύτητα το δείγμα που χρειαζόμαστε.

Στο παρακάτω διάγραμμα δείχνουμε ποια είναι η διαδρομή του ήχου απο τη στιγμή που διαβάζεται απο την πηγή μέσω του Altec Codec και περνάει 4-4 τα δείγματα για κάθε κανάλι στις 2 FIFO ώστε να βγουν οι συντελεστές που χρειαζόμαστε για το Spline. Στη συνέχεια με αυτές τις τιμές υπολογίζουμε τις τιμές του Interpolation και για κάθε κανάλι τις στέλνουμε πίσω στο Altec Codec ώστε αυτό να τις αναπαράγει στα ηχεία μας.

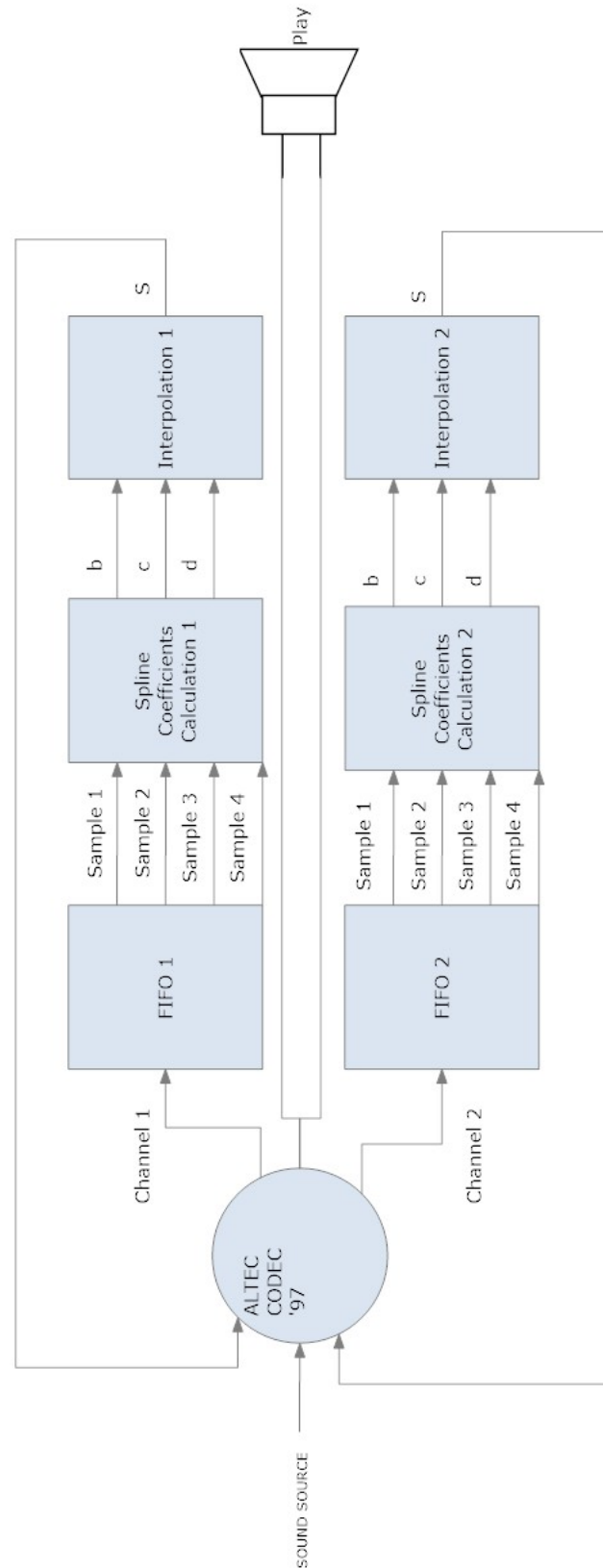


Figure 6.3: The Path of the Sound inside the System

Τα παρακάτω σχεδιαγράμματα είναι η σχεδίαση κάθε Module ξεχωριστά στο Simulink για το ενιαίο σύστημα του Cubic Spline Interpolation:



# HDelta Module

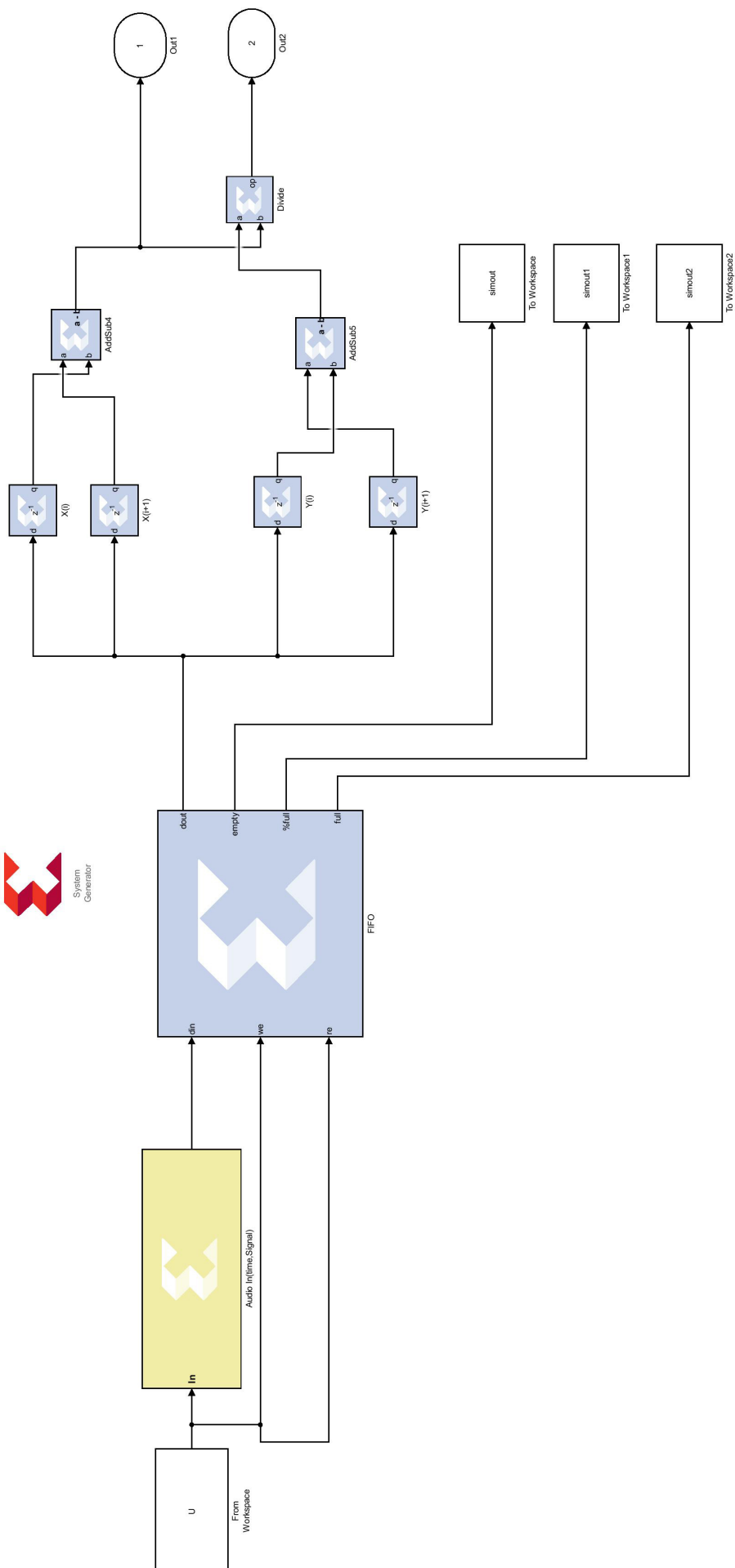


Figure 6.4: Για κάθε κανάλι αποθηκεύουμε τα δείγματα ανα τετράδα στην ουρά και στη συνέχεια βγάζουμε τις αντίστοιχες τιμές για τα  $H$  και  $\Delta$  έλτα τα οποία περνάμε στο Module Sigma.

# Sigma Module

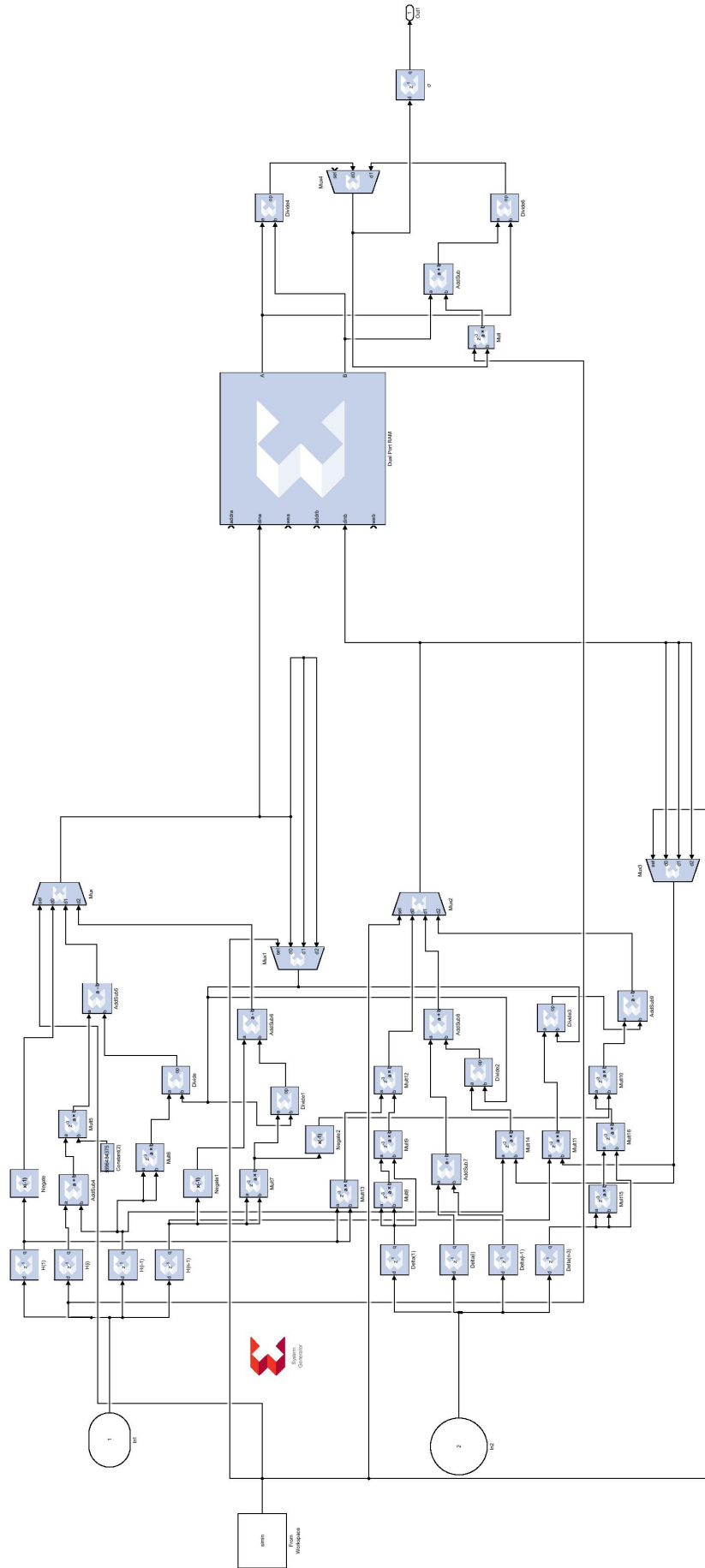


Figure 6.5: Δεχόμαστε τις τιμές  $H, \Delta$  και παράγουμε τους συντελεστές  $\alpha, \beta$ . Αποθηκεύουμε στη μνήμη τις τιμές τους και στη συνέχεια κάνουμε back substitution και υπολογίζουμε τις τιμές του  $\sigma$  τις οποίες δίνουμε δυο δυο στους καταχωρητές στο επόμενο module.

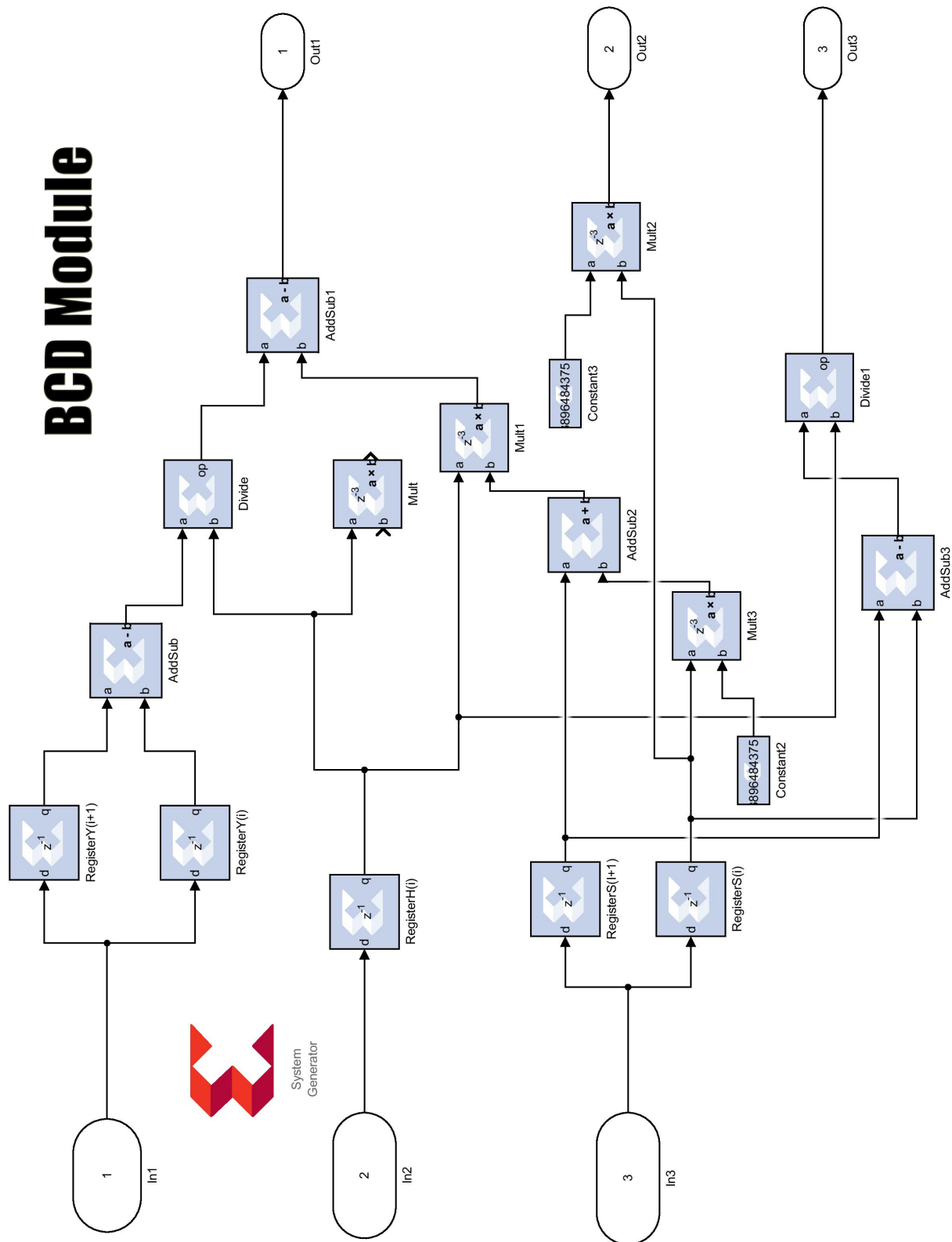


Figure 6.6: Με τις τιμές των  $\sigma, H, Y$  υπολογίζουμε τους συντελεστές  $B, C, D$  και δίνουμε το αποτέλεσμα στο επόμενο Module.

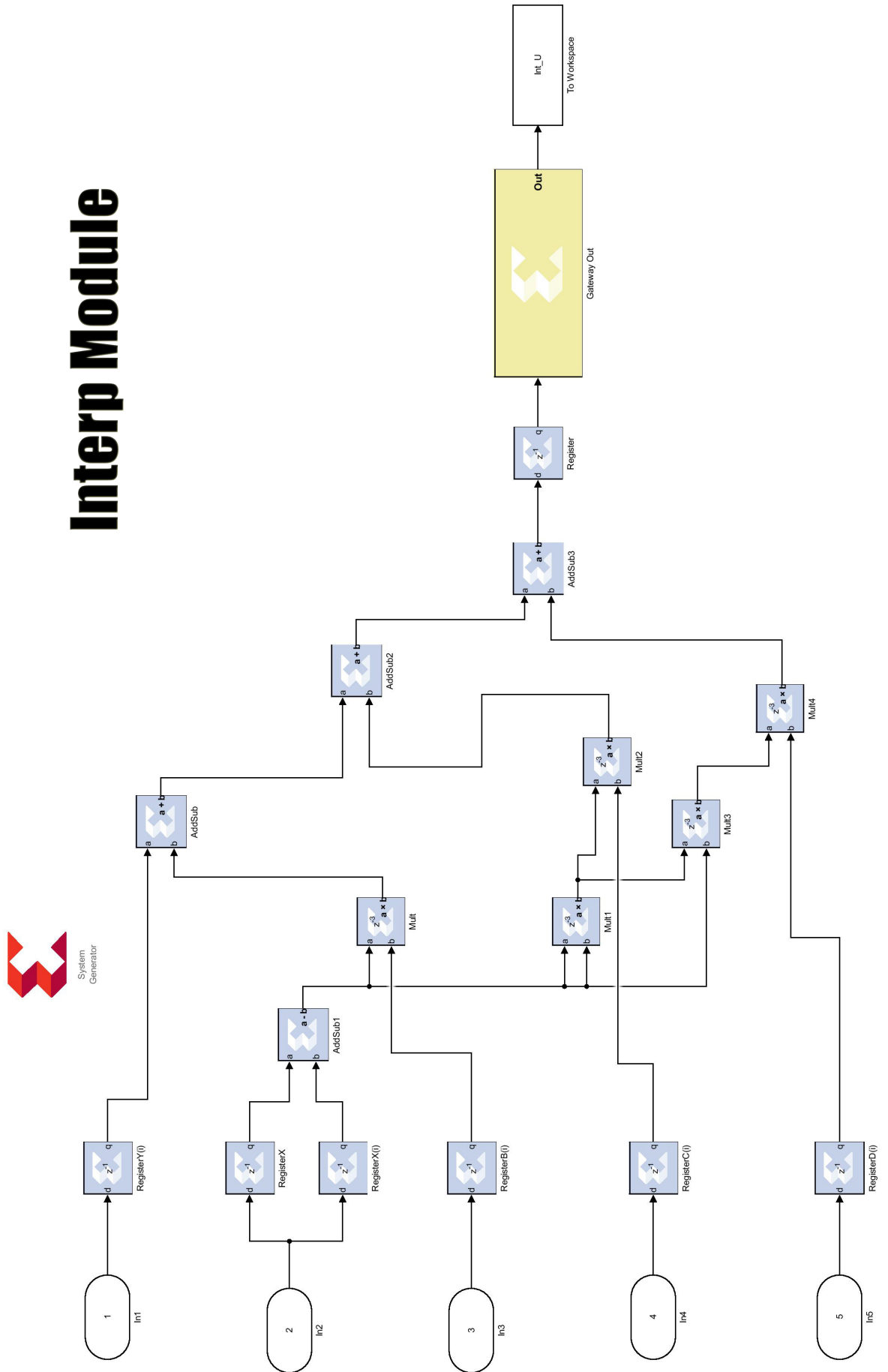


Figure 6.7: Αυτό είναι και το τελικό Module όπου υπολογίζουμε με τους συντελεστές τις τελικές τιμές των Interpolated δειγμάτων και βγάζουμε στην έξοδο το αποτέλεσμα.

## Κεφάλαιο 7

# Συμπεράσματα και μελλοντική εργασία

### 7.1 Ανασκόπηση

Σε αυτή τη διπλωματική ασχοληθήκαμε με την δημιουργία μιας μεθόδου Interpolation ,που βασίζεται στο μαθηματικό μοντέλο της Cubic Spline με γνώμονα την εφαρμογή της σε ηχητικά μουσικά σήματα βασικής ποιότητας cd.

Αρχικά ασχοληθήκαμε με τη γενική μελέτη και κατανόηση των βασικών αρχών του ψηφιακού ήχου μέσω της διαθέσιμης βιβλιογραφίας αλλά και του ίντερνετ,μέσα απο άρθρα και videos απο ψηφιακά και μη περιοδικά και forums σχετικά με το αντικείμενο.Εχοντας μια εμπειρία απο την ηχογράφηση και παραγωγή ψηφιακής μουσικής,επιλέξαμε με ποιες έννοιες θα ασχοληθούμε και θα αναδειξουμε κατα την εκπόνηση της εργασίας.Η ιδέα για τη μελέτη του Cubic Spline Interpolation πάνω στον ήχο υπήρχε εξ αρχής και μέσω της βιβλιογραφίας εξοικειωθήκαμε με το πως λειτουργεί και διαλέξαμε το μοντέλο και τη μέθοδο εκείνη που θα μας βολευε κατα την μοντελοποίηση σε Matlab .

Πριν αρχίσουμε να υλοποιούμε τη μέθοδο,μελετήσαμε τη συμπεριφορά και τη λειτουργία του ανθρώπινου αυτιού στο πως αντιλαμβάνεται τον ήχο και το ρόλο της ψυχοακουστικής σε αυτό.Στόχος μας ήταν να δούμε αν η αύξηση του ρυθμού δειγματοληψίας μπορεί να ωφελήσει τη ποιότητα του ήχου.Αναζητήσαμε διάφορες σχετικές έρευνες στη βιβλιογραφία,με σημαντικότερη αυτήν τον Mayer-Moran που υποστήριξαν πως το ανθρώπινο αυτί δεν αντιλαμβάνεται τον high-fidelity ήχο.Κατα τη διάρκεια της εκπόνησης της διπλωματικής,παρακολουθώντας την βιβλιογραφία και συγκεκριμένα τον Ιούνιο του 2016,δημοσιεύτηκε μια σχετική Meta-Analysis όπου αναλύθηκαν όλα τα σχετικά papers (συμπεριλαμβανομένου και του Mayer) με την σύγκριση normal και high-fidelity ήχου πειραματικά σε ακροατές όπου έδειξε πως έστω με μικρό ποσοστο οι ακροατές μπορούσαν να αντιληφθούν τον υψηλής ποιότητας ήχο,ειδικά οταν το πείραμα γινόταν σε εκπαιδευμένα αυτιά,το ποσοστό ανέβαινε κατακόρυφα.

Μέσω του Matlab υλοποιήσαμε ένα σύστημα που να διαβάζει τον ψηφιακό ήχο σε πίνακες και να μπορεί να κάνει resample σε αυτόν και υλοποιήσαμε βασιζόμενοι στα μαθηματικά μοντέλα τις Interpolation μεθόδους μας και συγκεκριμένα το Linear και το Cubic Spline Interpolation.Στη συνέχεια πειραματικά,ερευνήσαμε τις αλλαγές στις κυματομορφές και το σπεκτρόγραμμα επιλέγοντας μια γκάμα συχνοτήτων δειγματοληψίας εκτός της κανονικής,τα 48,66.15,96,192 KHZ.Είδαμε την μέθοδο του Cubic Spline να τα πηγαίνει εξαιρετικά όσον αφορά το σπεκτρόγραμμα σε όλο το φάσμα των συχνοτήτων σε σύγκριση τόσο με το αρχικό σήμα όσο και με την Linear Interpolation.Στις κυματομορφές η Cubic ήταν καλύτερη στα 44.1KHZ αλλά όσο αυξάναμε τον ρυθμό τόσο βλέπαμε πως πέφτει η απόδοσή της σε σχέση με το κανονικό σήμα.

Συνεχίσαμε τα πειράματα μας,αύτη τη φορά με ακουστικά test όπου για διαφορετικά ηχητικά δείγματα μελετήσαμε τη συμπεριφορά τους τόσο με την

εφαρμογή Interpolation ,όσο και resample αλλά και τα 2 μαζί. Παρατηρήσαμε πως η Cubic Spline δουλεύει πάρα πολύ καλά με το Speaker Fill των Windows όπου σε ορισμένα δείγματα και ειδικά στο κομμάτι που είχαν ψηφιακά drums παρατηρήθηκε έντονη διαφορά σε σχέση με το αυθεντικό κομμάτι. Η ακουστικότητα της Linear μεθόδου ήταν πάντα χειρότερη και μπορούσε να διακριθεί. Για παραπάνω συχνότητες όπου δηλαδή είχε εφαρμοστεί resample δε παρατηρήσαμε έντονες διαφορές παρα μόνο πως ακουστικά υπερίσχυε η Interpolation μέθοδος που εφαρμόζαμε πρώτα δηλαδή ή το resample ή το Cubic Spline Interpolation. Παρόλα αυτά, για ψυχοακουστικούς ίσως λόγους, είχαμε πάντα την εντύπωση πως τα 192 khz ακούγονται καλύτερα.

Στη συνέχεια, μέσω του Simulink της Matlab και του System Generator for DSP της Xilinx ,μοντελοποιήσαμε τις μεθόδους μας για πιθανή υλοποίηση ως ένα ενσωματωμένο σύστημα που θα επεξεργάζεται τον ήχο σε μια fpga. Αφού μελετήσαμε τις δυνατότητες που έχουν σήμερα οι fpga's και με ποιον τρόπο μπορούν αυτές να διαχειρίζονται τόσο τον δικάναλο ήχο ποιότητας 44.1khz, σχεδιάσαμε το πως θα μπορούσε να είναι ένα τέτοιο σύστημα ώστε να δούμε αν τελικά αξίζει μια τέτοια υλοποίηση. Παραμετροποιήσαμε τις μεθόδους μας ώστε να μπορούν να εφαρμοστούν σε hardware χωρίς μεγάλα προβλήματα, κατανοώντας τους περιορισμούς μιας τέτοιας εφαρμογής, κάνοντας streaming τα δείγματα ανα τετράδα και εφαρμόζοντας interpolation σε πραγματικό χρόνο.

## 7.2 Συμπεράσματα

Μέσω αυτής της διπλωματικής εργασίας μπορέσαμε να κατανοήσουμε τις σύγχρονες τάσεις και τη δημόσια συζήτηση στους χώρους της μουσικής τεχνολογίας όσον αφορά τη ποιότητα του ήχου και το αν μας καλύπτει τελικά η όχι η standard cd ποιότητα στα 44.1KHZ 16 bit. Είδαμε πως βασικό ρόλο στην ακουστική εμπειρία παίζει η ποιότητα του εξοπλισμού αναπαραγωγής εκτός από το είδος των αρχείων. Ένα εκπαιδευμένο αυτί μπορεί να αντιλαμβάνεται καλύτερα μεγαλύτερες συχνότητες δειγματοληψίας άρα ο ακροατής πριν επιλέξει τον τρόπο με τον οποίο θα ακούει τη μουσική του πρέπει να σκεφτεί τι είδους ακουστική εμπειρία μπορεί και θέλει να βιώσει, πριν επενδύσει τα χρήματά του σε κάποιο ακριβό setup και αρχεία.

Το ανθρώπινο αυτί μπορεί να περιορίζεται ακουστικά όσον αφορά τις συχνότητες που μπορεί να αντιληφθεί, αλλά όπως απέδειξε το Meta-Analysis του Dr Joshua Reiss , μπορεί να διακρίνει την ακουστική διαφορά υψηλών συχνοτήτων δειγματοληψίας, με τα UltraSounds και την ψυχοακουστική να παίζουν σημαντικό ρόλο.

Με την εφαρμογή Interpolation στο σήμα, στόχος μας ήταν να βελτιώσουμε την ποιότητα τόσο οπτικά όσο ακουστικά. Αν και οπτικά μπορούμε να πούμε πως είχαμε θετικά αποτελέσματα και ήμασταν σε θέση να το αποδείξουμε, ακουστικά η κρίση μας (όπως και κάθε μεμονωμένου ακροατή) είναι σίγουρα υποκειμενική και για να αποδείξουμε την πιθανή ακουστική ανωτερότητα της μεθόδου θα πρέπει να γίνει μέσω από ένα πείραμα στα πλαίσια των πειραμάτων που αναφέρθηκαν στο κεφάλαιο 3. Το μόνο που μπορούμε να πούμε με απόλυτη βεβαιότητα είναι η πολύ καλή συμπεριφορά της μεθόδου σε 5.1 ηχεία με speaker fill και τη δημιουργία μιας ψευδο-surround αίσθησης. Αυτό βέβαια δεν έχει καμία επιρροή σε stereo οικιακά και μη συστήματα και σε ακουστικά, ίσως να μην έχει και σε συστήματα (λειτουργικά εκτός Windows ή home cinema) που χρησιμοποιούν διαφορετική μέθοδο για το surround .

Χρησιμοποιώντας το resample της Matlab και τη μέθοδο μας μαζί, είδαμε πως το ένα interpolation εξουδετερώνει το άλλο ακουστικά και υπερισχύει αυτό που εφαρμόζεται πρώτο. Άρα δεν μπορούμε να εφαρμόσουμε τη μέθοδο μας με ασφάλεια σε μεγαλύτερους ρυθμούς δειγματοληψίας με την υπάρχουσα τεχνική. Θα πρέπει η να διαβάσουμε αρχεία ήχου high-fidelity από το mastering

τους και στη συνέχεια να περνάμε το interpolation ή να φτιάξουμε μια καινούρια resample τεχνική που δε βασίζεται στο χαμηλοπερατό φίλτρο. Αρα λοιπόν η ασφαλής χρήση της μεθόδου είναι πάνω στον κανονικό δικάναλο ήχο όπως αυτός διαβάζεται η πληροφορία από το αρχείο.

Με τη μοντελοποίηση σε Simulink και Xilinx εξετάσαμε τη προοπτική ενσωμάτωσης της μεθόδου σε ένα board fpga. Το γεγονός πως για να λειτουργήσει η μέθοδος χρειαζόμαστε να επεξεργαστούμε τα δεδομένα του ήχου τα οποία είναι σε μορφή floating point ώστε να έχουμε μεγαλύτερη ακρίβεια την οποία έχουμε ανάγκη, θα πρέπει να επιλέξουμε ένα τελευταίας κυκλοφορίας board το οποίο υποστηρίζει πλήρως το format. Ολοκληρώνοντας τη μοντελοποίηση, είχαμε στα χέρια μας ένα μοντέλο που κοστίζει αρκετά από άποψης πόρων, τόσο σε hardware όσο και σε χρόνο. Γίνεται streaming ο ήχος, λίγα-λίγα τα samples κάθε φορά ακολουθώντας το μαθηματικό μοντέλο για τη δημιουργία των συντελεστών που χρειαζόμαστε για το interpolation. Πέρα λοιπόν από τους δεκάδες αθροιστές, πολλαπλασιαστές, διαιρέτες και καταχωρητές θα χρειαζόμαστε μια dual port ram και μια FIFO που αποθηκεύουμε κάθε τετράδα samples και όλα αυτά θα πρέπει να τρέχουνε δύο φορές ταυτόχρονα, μια για κάθε κανάλι. Επίσης θα χρειαζόμαστε ένα πολύ καλογραμμένο control κομμάτι γραμμένο σε vhd1 για να διαχειριζόμαστε ποια samples θα πηγαίνουν που, τα enables των πολυπλεκτών και μνημών και τους χρόνους που θα γίνεται αυτό. Λαμβάνοντας όλα αυτά υπόψιν καταλήγουμε πως η μεταφορά της μεθόδου σε hardware λόγω πολυπλοκότητας και πόρων δεν είναι μια καλή ιδέα. Αντίθετα η software υλοποίηση σύμφωνα με τη Matlab μοντελοποίηση σε μία γλώσσα όπως η C++, θα είναι πιο εύχρηστη στο ευρύ κοινό.

### 7.3 Μελλοντική εργασία

Ολοκληρώνοντας την εργασία μείναμε με αρκετές σχέψεις τις οποίες δεν καταφέραμε να υλοποιήσουμε είτε λόγω εξοπλισμού και πόρων, είτε λόγω αντικειμένου και γνώσεων όπου θα ξεφεύγαμε από το πλαίσιο μιας διπλωματικής.

Επειδή έχουμε να κάνουμε με ποιότητα ήχου, κάτι τελείως υποκειμενικό για το κάθε άτομο όπως έχουμε αναφέρει και βλέπουμε πως ακόμα και σήμερα δεν υπάρχει ξεκάθαρη εικόνα (ίσως το Meta-Analysis το αλλάζει αυτό) για το αν το άτομο μπορεί να αντιληφθεί την high-fidelity ποιότητα, μια ιδέα για την αξιολόγηση της μεθόδου είναι η εκπόνηση κάποιων trial blind tests σε ένα μεγάλο εύρος ακροατών με διαφορετικά χαρακτηριστικά, όπως φύλο, ηλικία, εκπαιδευμένο ή όχι, αντί, στα πλαίσια των πειραμάτων που αναφέρθηκαν στο κεφάλαιο 3 ώστε να έχουμε μια γενική εικόνα για το πως αντιλαμβάνεται ο ακροατής ακουστικά τον ήχο με ή χωρίς τη μέθοδο.

Η μέθοδος για το Cubic Spline Interpolation μπορεί να φτιαχτεί σε ένα περιβάλλον εφαρμογής στο styl γνωστών Converters που κυκλοφορούν στην αγορά. Εκεί ο χρήστης θα μπορεί να βάζει αρχεία ήχου και να τα μετατρέπει σε .wav που έχουν υποστεί Interpolation. Το αν θα υποστηρίζει μόνο .wav αρχεία ή θα μπορεί να γίνεται μετατροπή από .mp3 ή .flac σε .wav και μετά Interpolation είναι στο χέρι του προγραμματιστή, όπως επίσης και το αν θα μπορεί να διαβάζει ένα cd από τον οπτικό δίσκο και να το μετατρέπει.

Το γεγονός πως είδαμε διαφορές σε surround περιβάλλον, έχει ενδιαφέρον καθώς μπορεί να μελετηθεί η εφαρμογή της μεθόδου ως φίλτρο σε audio codecs που αναπαράγουν τον ήχο σε αρχεία video ώστε να εχμεταλευτούμε και να αναδείξουμε το χαρακτηριστικό της μεθόδου που εντοπίσαμε στον surround ήχο. Ένα τέτοιο φίλτρο με κάποιες ίσως παραλλαγές θα μπορούσε να χρησιμοποιείται και από τους μηχανικούς και παραγωγούς ήχου κατά τη διάρκεια του mastering μιας ηχογράφησης, όπου θα μπορούν να τονίζουν μέρη των drums όπως τα hats και τα snares.

Ως μελλοντική δουλειά θα μπορούσαμε να αναφέρουμε την βελτίωση του

μοντέλου για το ενσωματωμένο σύστημα και το download σε μια fpga που θα μπορεί να την σηκώσει, με την εφαρμογή pipeline ανάμεσα στα units γράφοντας ένα control unit που θα στηρίζει την σχεδίαση μας.



# Bibliography

## Books and Papers

- [1] Lynn Blair. *Data Interpolation and Its Effects on Digital Sound Quality*. McMurry University, 2008. [Url Link](#)
- [2] E.Brad Meyer,David R.Moran. *Audibility of a CD-Standard A/DA/A Loop Inserted into High-Resolution Audio Playback*. Boston Audio Society, Lincoln, MA, USA, 2007. [Url Link](#)
- [3] Pras, Amandine; Guastavino, Catherine. *Sampling Rate Discrimination: 44.1 kHz vs. 88.2 kHz*. McGill University, Montreal, Quebec, Canada , 2010. [Url Link](#)
- [4] J. Robert Stuart. *Coding High Quality Digital Audio*. Meridian Audio Ltd, Stonehill, Stukeley Meadows, Huntingdon, PE18 6ED, United Kingdom, 1998. [Url Link](#)
- [5] Dan Lavry. *Sampling Theory For Digital Audio*. Lavry Engineering, Inc., 2004. [Url Link](#)
- [6] Reiss, Joshua D. *A Meta-Analysis of High Resolution Audio Perceptual Evaluation*. Queen Mary University of London, London, UK, 2016. [Url Link](#)
- [7] C. Burgel, R. Bartholomaus, W. Fiesel, J. Hilpert, A. Holzer, K. Linzmeier, M. Weishart. *Beyond CD-Quality: Advanced Audio Coding (AAC)for High Resolution Audio with 24 bit Resolution and 96 kHz Sampling Frequency*. Fraunhofer Institute for Integrated Circuits IIS-AErlangen, D-91058, Germany , 2001. [Url Link](#)
- [8] Francis Rumsey, Tim McCormick. *Sound and Recording: Applications and Theory*. CRC Press , 2014. [Url Link](#)
- [9] Gary D. Knott. *Interpolating Cubic Splines*. Springer Science and Business Media, 2000. [Url Link](#)
- [10] GEORGE E. FORSYTHE,MICHAEL A. MALCOLM,CLEVE B. MOLER *Computer Methods For Mathematical Computations* . PRENTICE-HALL, INC. ENGLEWOOD CLIFFS , N. J. 0763 2 [Url Link](#)
- [11] Andreas Spanias, Ted Painter, Venkatraman Atti *Audio Signal Processing and Coding*. John Wiley and Sons,2006. [Url Link](#)
- [12] Ivan Jeleu Tashev *Sound Capture and Processing: Practical Approaches*. John Wiley and Sons,2009. [Url Link](#)

**Web pages**

- [13] Monty. *24/192 Music Downloads ...and why they make no sense*. Red Hat Inc. and Xiph.Org , 2012. Url Link
- [14] JEFF MILROD. *Signal processing FPGAs with hard floating-point – No compromise*. Url Link
- [15] Cubic Splines and Matlab <http://people.whitman.edu/~hundledr/courses/M467/CubicSplines/> Url Link
- [16] Hydrogen Audio Forum
- [17] AvForums
- [18] Matlab Forums
- [19] James Larson audioholics.com Total Harmonic Distortion (THD): Is It a Good Indicator of Sound Quality?

**Wikipedia**

- [20] Spline interpolation
- [21] Linear interpolation
- [22] Interpolation
- [23] Digital audio
- [24] Nyquist–Shannon sampling theorem
- [25] Pulse-code modulation
- [26] Audio bit depth
- [27] Upsampling
- [28] Ultrasound
- [29] Psychoacoustics
- [30] Audio system measurements
- [31] Signal to noise ratio

**Youtube**

- [32] Production Bytes Digital Audio 101 - Bit Depth, Sampling Rate, Interpolation  
Digital Audio 102 - PCM, Bit-Rate, Quantisation, Dithering, Nyquists Sampling Theorum
- [33] Oscar Veliz Cubic Splines
- [34] Jacob Bishop 5.2.7-Curve Fitting: Spline Interpolation
- [35] sachin sharma sachin sharma
- [36] Hisham Mohiuddin Cubic Spline Interpolation