

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΜΟΥΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΚΑΙ ΑΚΟΥΣΤΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Χρονική συμπίεση/διαστολή ήχου μέσω φασματικής επεξεργασίας

Γουσταύος-Ιωάννης Λοΐζος (Α.Μ. 1468)

Επιβλέπων Καθηγητής
Νικόλας Βαλσαμάκης

ΡΕΘΥΜΝΟ 2016

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή Νικόλα Βαλσαμάκη, του οποίου ήταν η ιδέα της εφαρμογής και με τον οποίο αναπτύχθηκε σε στενή συνεργασία.

Περίληψη

Στόχος της εργασίας είναι η δημιουργία πρωτότυπου αλγορίθμου επεξεργασίας ήχου που ανήκει στη κατηγορία τεχνικών ανάλυσης-ανασύνθεσης επεξεργασίας ήχου. Θεμελιώδες αρχή αποτελεί ο γρήγορος μετασχηματισμός Φουριέ (Fast Fourier Transform, FFT). Η εφαρμογή που αναπτύχθηκε παρέχει τη δυνατότητα ανεξάρτητης χρονικής διαστολής/συμπίεσης κάθε συχνοτικής ζώνης που προκύπτει από τον FFT. Επιπλέον, ο χρήστης μπορεί να ορίσει κάποια χρονική καθυστέρηση στην έναρξη αναπαραγωγής κάθε συχνοτικής ζώνης, με αποτέλεσμα τη χρονική αναδιάταξη του φάσματος. Στο τέλος της διάταξης υπάρχει ένας φασματικός ισοσταθμιστής για την ανακατανομή της έντασης μεταξύ των φασματικών ζωνών. Η εφαρμογή αναπτύχθηκε στο προγραμματιστικό περιβάλλον MaxMSP.

Λέξεις κλειδιά: επεξεργασία ήχου, σύνθεση ήχου, ανάλυση/ανασύνθεση ήχου, φασματική επεξεργασία ήχου, μουσικός προγραμματισμός, MaxMSP

Abstract

Object of the study is the creation of original sound processing algorithm that belongs to the analysis-resynthesis category of sound synthesis/processing. Principle of the study is FFT (Fast Fourier Transform). The developed application has the capability of independent time stretching/compressing of every frequency band resulting from FFT. Additionally, the user can define an amount of delay before the reproduction of every frequency band, giving him the possibility of time reconstruction of the spectrum. Moreover there is a spectral equalizer for the reallocation of amplitude between frequency bands. The application developed in the programming environment MaxMSP.

Keywords: sound processing, sound analysis/resynthesis, spectral sound processing, music programming, MaxMSP

Εισαγωγή

Στόχος της εργασίας είναι η δημιουργία πρωτότυπου αλγορίθμου επεξεργασίας ήχου που ανήκει στη κατηγορία τεχνικών ανάλυσης-ανασύνθεσης επεξεργασίας ήχου. Θεμελιώδες αρχή αποτελεί ο γρήγορος μετασχηματισμός Φουριέ (Fast Fourier Transform, FFT). Η εφαρμογή που αναπτύχθηκε παρέχει τη δυνατότητα ανεξάρτητης χρονικής διαστολής/συμπίεσης κάθε συχνοτικής ζώνης που προκύπτει από τον FFT. Επιπλέον, ο χρήστης μπορεί να ορίσει κάποια χρονική καθυστέρηση στην έναρξη αναπαραγωγής κάθε συχνοτικής ζώνης, με αποτέλεσμα τη χρονική αναδιάταξη του φάσματος. Στο τέλος της διάταξης υπάρχει ένας φασματικός ισοσταθμιστής για την ανακατανομή της έντασης μεταξύ των φασματικών ζωνών. Η εφαρμογή αναπτύχθηκε στο προγραμματιστικό περιβάλλον MaxMSP.

Στο κεφάλαιο 1 γίνεται η ανάλυση κάποιων στοιχείων της θεωρίας του Μετασχηματισμού Fourier με έμφαση στην εφαρμογή τους σε ηχητικά σήματα. Αναλύεται συνοπτικά η πορεία εξέλιξης από τον θεωρητικό Μετασχηματισμό Fourier στον STFT.

Στο δεύτερο κεφάλαιο αναφέρονται οι δύο βασικότερες εφαρμογές του STFT, το φασματογράφημα και ο Phase Vocoder. Αναλύονται οι αρχές λειτουργίας και το θεωρητικό τους υπόβαθρο.

Το τρίτο κεφάλαιο περιέχει επεξήγηση των επεξεργασιών και λειτουργιών των αντικειμένων STFT της MaxMSP και συνδέονται με τη θεωρητική ανάλυση. Έπειτα αναλύεται η εφαρμογή ενός κλασσικού Phase Vocoder, με δυνατότητα χρονικής διαστολής/συμπίεσης. Στο τέλος εξηγείται η ανάπτυξη και η λειτουργία της εφαρμογής χρονικής συμπίεσης/διαστολής ήχου μέσω φασματικής επεξεργασίας.

Το κεφάλαιο 4 καταδεικνύει τον τρόπο χρήσης της εφαρμογής, με απλά λόγια, ώστε να γίνεται αντιληπτό από κάποιον που δεν γνωρίζει το θεωρητικό υπόβαθρο της εφαρμογής.

Το πέμπτο κεφάλαιο είναι μία σύντομη αποτίμηση της τελικής εφαρμογής.

Περιεχόμενα

1. Στοιχεία Θεωρίας Μετασχηματισμού Fourier	11
1.1 Μιγαδικοί Αριθμοί, καρτεσιανές και πολικές συντεταγμένες.....	11
1.2 Ημιτονοειδές σήμα ως μιγαδικό διάνυσμα.....	12
1.3 Θεώρημα Fourier και εφαρμογή στον ήχο.....	13
1.3.1 Σειρές Fourier.....	14
1.3.2 Μετασχηματισμός Fourier.....	14
1.3.3 Μετασχηματισμός Fourier διακριτού χρόνου (DTFT).....	15
1.3.4 Διακριτός Μετασχηματισμός Fourier (DFT).....	17
1.3.5 Μετασχηματισμός Fourier μικρού χρονικού διαστήματος (STFT).....	18
1.3.6 Overlap-add Ανασύνθεση (OA).....	21
1.3.7 Περιορισμοί του STFT.....	22
2. Βασικές εφαρμογές STFT στην ανάλυση/ανασύνθεση ήχου	25
2.1 Φασματογράφημα.....	25
2.2 Phase Vocoder.....	27
2.2.1 Βασικές Αρχές.....	27
2.2.2 Παράμετροι.....	31
3.Υλοποίηση της εφαρμογής χρονικής συμπίεσης/διαστολής μέσω φασματικής επεξεργασίας	32
3.1 STFT στο προγραμματιστικό περιβάλλον MaxMSP.....	32
3.2 Σχέση ταχύτητας ροής ηχητικού σήματος με παράγοντα αλληλοεπικάλυψης μέσα στο αντικείμενο rfft~.....	34
3.3 Εφαρμογή απλού Phase Vocoder στο προγραμματιστικό περιβάλλον MaxMSP.....	36
3.4 Χρονική Συμπίεση/Διαστολή ήχου μέσω φασματικής επεξεργασίας.....	41
3.4.1 Οι τρεις καταστάσεις λειτουργίας της εφαρμογής.....	44
3.4.2 Φασματικό delay.....	46
3.4.3 Φασματικό φίλτρο.....	47
3.4.4 Λειτουργία freeze.....	48
3.4.5 Έλεγχος παραμέτρων ανασύνθεσης μέσω διεπαφής.....	49

4.Παραδείγματα χρήσης της εφαρμογής.....	50
4.1 Play once.....	51
4.2 Loop mode.....	53
4.3 Loop after finished.....	54
5.Συμπεράσματα.....	55
Παράρτημα.....	56
Βιβλιογραφία.....	57

Ευρετήριο Σχημάτων

Σχήμα 1: Γραφική παράσταση ενός μιγαδικού ως σημείο με συντεταγμένες (a,b) σε δυοδιάστατο πεδίο. Αυτά τα σημεία ορίζουν το διάνυσμα r με γωνία θ από την αρχή των αξόνων.....	11
Σχήμα 2: Διάνυσμα το οποίο περιστρέφεται κυκλικά. Η υποτείνουσα συμβολίζεται με H , με A η τετμημένη και με O η τεταγμένη.....	12
Σχήμα 3: Σήμα άπειρης έκτασης.....	14
Σχήμα 4: Η ένταση του μετασχηματισμού Fourier του σήματος στο Σχήμα 1 $X[j\omega]$ σε σχέση με τη συχνότητα (το X συμβολίζει τον Μετασχηματισμό Fourier).....	15
Σχήμα 5: Το σήμα διακριτού χρόνου $x[n]$ που προέκυψε ύστερα από δειγματοληψία του σήματος $x(t)$ (Σχήμα 1).....	16
Σχήμα 6: Η ένταση σε σχέση με τη συχνότητα του antialised DTFT του σήματος $x[n]$	16
Σχήμα 7: Συνεχές συχνοτικό φάσμα του DTFT του σήματος $x[n]$	16
Σχήμα 8: Το σήμα $\tilde{x}[n]$ ύστερα από δειγματοληψία. Περιέχει αντίγραφα του σήματος $x[n]$ επ'άπειρο.....	17
Σχήμα 9: Σχεδιάγραμμα του DFT του σήματος $\tilde{x}[n]$, διακριτού φάσματος.....	18
Σχήμα 10: Χωρισμός ηχητικού σήματος σε κομμάτια μικρής διάρκειας μέσω εφαρμογής παραθύρων (πολλαπλασιασμός ήχου με περιβάλλουσες μικρής διάρκειας).....	19
Σχήμα 11: (a) Παράθυρο με σχήμα καμπύλης Gauss. (b) Φάσμα του παραθύρου(συχνότητα σε σχέση με ένταση σε dB). Η δεύτερη κορυφή στο φάσμα είναι περίπου στα -45 dB, παρόλα αυτά φαίνεται η παραμόρφωση που θα υποστεί η φασματική ανάλυση του εισερχόμενου σήματος.....	19
Σχήμα 12: Επισκόπηση του STFT.....	20
Σχήμα 13: (α) κυματομορφή προς ανάλυση. (b) Κομμάτι πολλαπλασιασμένο με παράθυρο (c) Φάσμα έντασης (εύρος 80 dB). (d) Φάσμα φάσης (εύρος $-\pi$ έως π).....	21
Σχήμα 14: Παράδειγμα Overlap-add ανασύνθεσης.....	22
Σχήμα 15: Χρονική-συχνοτική αβεβαιότητα σε STFT ανάλυση. (α) Σήμα που δόθηκε προς ανάλυση. Αποτελείται από τόνο 1000Hz που εναλλάσσεται ακαριαία σε 4000 Hz. (b) Φάσμα που προέκυψε από την ανάλυση (εύρος 500-7000Hz, λογαριθμική κλίμακα συχνότητας)..	23
Σχήμα 16: Τρία στιγμιότυπα STFT ανάλυσης ενός ήχου που ξεκινάει με συχνότητα 1Hz και καταλήγει 3Hz. Ο συγκεκριμένος STFT έχει bins ανά 1 Hz. Όταν η συχνότητα του ήχου είναι 2.5, πέφτει ακριβώς ανάμεσα στα δύο κοντινότερα bins και η ενέργεια απλώνεται σε όλο το φάσμα.....	24

Σχήμα 17: Φασματογραφήματα που παρουσιάζουν φάσμα ανθρώπινης φωνής ηχογραφημένη στα 44100 Hz με διαφορετικά μεγέθη παραθύρου. (α) Παράθυρο 32 δειγμάτων. Χρονική ευκρίνεια 0.725 ms, συχνοτική ευκρίνεια 1378 Hz. (b) Παράθυρο 1024 δειγμάτων. Χρονική ευκρίνεια 23.22 ms, συχνοτική ευκρίνεια 43.07 Hz. (c) Παράθυρο 8192 δειγμάτων. Χρονική ευκρίνεια 185.8 ms, συχνοτική ευκρίνεια 5.383 Hz.....	26
Σχήμα 18: Βασικό διάγραμμα ροής Phase Vocoder.....	28
Σχήμα 19: Αριστερά ημιτονικό σήμα ίδιας συχνότητας με το bin ανάλυσης έχει ως αποτέλεσμα μηδενική φασική μετατόπιση. Δεξιά ημιτονικό σήμα υψηλότερης συχνότητας με διαφορετικές φασικές μετατοπίσεις (για την ακρίβεια αύξουσα μετατόπιση) για κάθε διαδοχικό frame.....	29
Σχήμα 20: Φασικό τύλιγμα (phase unwrapping) (a) Τυλιγμένη φάση στο εύρος 0-2π (b) Ξετυλιγμένη φάση (απλό άθροισμα διαδοχικών διαφορών φάσης).....	30
Σχήμα 21: Απεικόνιση διαδικασίας STFT. Φαίνονται μεγέθη όπως FFT size, hop size και overlap factor (παράγοντας αλληλοεπικάλυψης). Φαίνεται ότι το μέγιστο πλήθος frames του πρώτου κομματιού ανάλυσης τα οποία αλληλεπικαλύπτονται μεταξύ τους, είναι τέσσερα. Άρα ο παράγοντας αλληλοεπικάλυψης θα είναι 4.....	31
Σχήμα 22: Το αντικείμενο [pfft~] στο προγραμματιστικό περιβάλλον MaxMSP.....	32
Σχήμα 23: Αντικείμενο [fftin~].....	32
Σχήμα 24: Αντικείμενο [fftout~].....	33
Σχήμα 25: Αντικείμενο [cartopol~].....	33
Σχήμα 26: Αντικείμενο [framedelta~].....	33
Σχήμα 27: Αντικείμενο [phasewrap~].....	33
Σχήμα 28: Αντικείμενο [frameaccum~].....	33
Σχήμα 29: Αντικείμενο [poltocar~].....	33
Σχήμα 30: Αντικείμενο [fftinfo~].....	34
Σχήμα 31: Αντικείμενο [dspstate~].....	34
Σχήμα 32: Αριστερά κεντρικό patch παραδείγματος, δεξιά μέσα στο αντικείμενο [pfft~]....	35
Σχήμα 33: Πάνω buffer κεντρικού patch παραδείγματος, κάτω buffer ο οποίος γράφτηκε μέσα στο αντικείμενο pfft~. Ο κάθετος άξονας αντιπροσωπεύει ένταση και ο οριζόντιος χρόνο σε ms.....	36
Σχήμα 34: Βασικός αλγόριθμος απλού Phase Vocoder.....	37
Σχήμα 35: Μέσα στο pfft~ του απλού Phase Vocoder.....	38

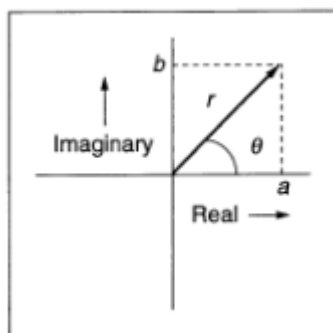
Σχήμα 36: Πάνω φαίνεται ο buffer των τιμών έντασης της ανάλυσης και κάτω ο buffer των τιμών απόκλισης φάσης. Ο οριζόντιος άξονας αντιπροσωπεύει δείγματα (οι τιμές που αναγράφονται όμως είναι χρόνος σε ms που θα χρειαστεί για να αναπαραχθούν σε παρούσα σε συχνότητα δειγματοληψίας). Ο κάθετος άξονας αντιπροσωπεύει τιμές εντάσεων για τον πάνω buffer και τιμές απόκλισης φάσης για τον κάτω.....	39
Σχήμα 37: Τρόπος εγγραφής τιμών FFT ανάλυσης σε buffer. N ισούται με μέγεθος του frame (πλήθος bins).....	40
Σχήμα 38: Διάγραμμα ροής αλγορίθμου ανεξάρτητου time stretching για κάθε bin.....	42
Σχήμα 39: Μέσα στο [pfft~] της εφαρμογής.....	43
Σχήμα 40: Αριστερά ορίζεται η ταχύτητα ανάγνωσης για κάθε bin μέσω του αντικειμένου multislider. Πάνω είναι οι υψηλές συχνότητες και κάτω οι χαμηλές. Το εύρος τιμών των ταχυτήτων ανάγνωσης είναι -1 έως 1. Στη μέση φαίνεται το επεξεργασμένο φάσμα του ήχου. Χρησιμοποιήθηκε το ηχητικό δείγμα drumloop από τα δείγματα της Max. Δεξιά φαίνεται το φάσμα του αρχικού αρχείου.....	44
Σχήμα 41: Φάσμα επεξεργασμένου ήχου με ίδιες ρυθμίσεις όπως στο Σχήμα 29 σε λειτουργία Loop.....	45
Σχήμα 42: Η ίδιες τιμές επεξεργασίας με τα δύο προηγούμενα σχήματα, στη τρίτη λειτουργία (επανάληψη αναπαραγωγής).....	46
Σχήμα 43: Ηχητικό δείγμα drumloop από τα δείγματα της Max, με γραμμικά αυξανόμενο delay από τις χαμηλές στις υψηλές συχνότητες, στη τρίτη λειτουργία (Loop after finished) της εφαρμογής. Εύρος delay 1 sec.....	47
Σχήμα 44: Χρήση φασματικού φίλτρου σε ηχητικό δείγμα drumloop, σε λειτουργία loop (after finished).....	48
Σχήμα 45 : Χρήση freeze για μεσαίο συχνοτικό εύρος. Χρησιμοποιήθηκε το ηχητικό δείγμα drumloop σε κατάσταση λειτουργίας εφαρμογής επανανάγνωσης.....	49
Σχήμα 46: Επιφάνεια εργασίας εφαρμογής.....	50
Σχήμα 47: Χρονική καθυστέρηση χαμηλομεσαίων συχνοτήτων σε ηχητικό δείγμα vibes-a1. Στο αριστερό μέρος του φασματογράφου φαίνεται το φάσμα του αρχικού ήχου.....	52
Σχήμα 48:Φασματικά επεξεργασμένο δείγμα με διαφορετική ταχύτητα αναπαραγωγής ανά συχνοτικά εύρη.....	52
Σχήμα 49: Παράδειγμα εφαρμογής σε λειτουργία loop.....	53
Σχήμα 50: Παράδειγμα χρήσης freeze σε λειτουργία loop.....	54
Πίνακας 1:Φάσεις για αθροίσματα ημιτόνων-συνημιτόνων με διαφορετικά πλάτη.....	13
Πίνακας 2: Συχνοτική ευκρίνεια σε σχέση με μήκος παραθύρου STFT ανάλυσης.....	24

1. Στοιχεία Θεωρίας Μετασχηματισμού Fourier

Η εφαρμογή που αναπτύχθηκε έχει ως πυρήνα την επεξεργασία Short time Fourier Transform (STFT) μία εφαρμογή του Discrete Fourier Transform (DFT). Οι μετασχηματισμοί αυτοί βασίζονται στο θεώρημα του Fourier το οποίο αποδεικνύει ότι κάθε περίπλοκο περιοδικό σήμα μπορεί να αναπαρασταθεί ως άθροισμα πολλών απλών σημάτων, δηλαδή ημιτόνων, διαφορετικών εντάσεων, συχνοτήτων και φάσεων. Στη συνέχεια του κεφαλαίου θα περιγράψουμε το μαθηματικό υπόβαθρο του μετασχηματισμού Fourier.

1.1 Μιγαδικοί Αριθμοί, Καρτεσιανές και πολικές συντεταγμένες

Οι πραγματικοί αριθμοί, μπορούμε να πούμε ότι, αντιστοιχούν σε σημεία μίας άπειρης, συνεχούς ευθείας (Roads,1996). Στο σύνολο αυτό δεν υπάρχει κάποιος αριθμός όπου υψωμένος στο τετράγωνο θα ισούται με -1 . Ένα υπερσύνολο των πραγματικών αριθμών, είναι οι μιγαδικοί, οι οποίοι περιέχουν ένα στοιχείο που το τετράγωνό του ισούται με -1 . $j^2=-1$. Το j καλείται η φανταστική μονάδα. Οι μιγαδικοί αριθμοί μπορούν να απεικονιστούν γραφικά σε ένα δισδιάστατο πεδίο. Μπορούν να απεικονιστούν ως σημεία συντεταγμένων (a,b) . Ο κατακόρυφος άξονας καλείται φανταστικός άξονας και ο οριζόντιος πραγματικός. Οι συντεταγμένες (a,b) είναι οι καρτεσιανές συντεταγμένες του σημείου.

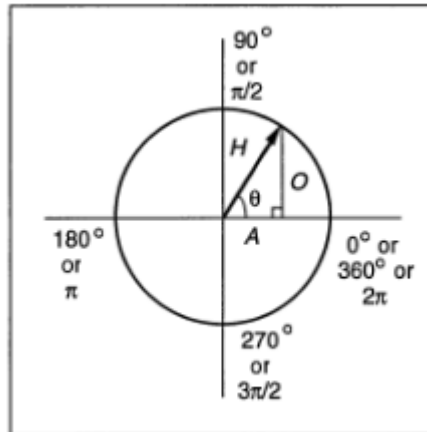


Σχήμα 1: Γραφική παράσταση ενός μιγαδικού ως σημείο με συντεταγμένες (a,b) σε δισδιάστατο πεδίο. Αυτά τα σημεία ορίζουν το διάνυσμα r με γωνία θ από την αρχή των αξόνων

Ένας άλλος τρόπος να εκφραστεί ένας μιγαδικός αριθμός είναι ως διάνυσμα. Το διάνυσμα αυτό ορίζεται από την απόσταση από την αρχή των αξόνων (r) και από τη γωνία του σε σχέση με την αρχή των αξόνων. Αυτές είναι οι πολικές συντεταγμένες του σημείου. Το r καλείται και ένταση του διανύσματος.

1.2 Ημιτονοειδές σήμα ως μιγαδικό διάνυσμα

Εάν θέλουμε να συμπεριλάβουμε και τη διάσταση του χρόνου, μπορούμε να σκεφτούμε το διάνυσμα αυτό να περιστρέφεται γύρω από την αρχή των αξόνων περνώντας από το σημείο 0 μοιρών, όσο περνάει ο χρόνος, χαράζοντας έτσι τον μοναδιαίο κύκλο, εάν υποθέσουμε ότι το μέτρο του διανύσματος ισούται με 1.



Σχήμα 2: Διάνυσμα το οποίο περιστρέφεται κυκλικά. Η υποτείνουσα συμβολίζεται με H, με A η τετμημένη και με O η τεταγμένη.

Μέσω αυτού του συστήματος μπορούμε να εκφράσουμε ένα ημιτονοειδές σήμα πλήρως. Η γωνιακή ταχύτητα του διανύσματος εκφράζεται ως ακτίνια ανά χρόνο. Η συχνότητα του ημιτονοειδούς σήματος μπορεί να εκφραστεί ως γωνιακή συχνότητα ($\omega=2\pi f$) του διανύσματος

$$x(t)=\sin(\omega t)$$

Η ένταση του σήματος είναι ανάλογη με το μέτρο του διανύσματος r, το οποίο μπορεί να προκύψει και από τις καρτεσιανές συντεταγμένες

$$r = \sqrt{a^2 + b^2} \quad (1.1)$$

Η αρχική θέση του σήματος μπορεί να οριστεί ως αρχική φάση, δηλαδή ως μετατόπιση από τη θέση 0π, κι έτσι το ημιτονοειδές σήμα μπορεί να δηλωθεί ως $x(t)=r\sin(\omega t+\phi)$. Η αρχική φάση μπορεί να προκύψει μέσω των καρτεσιανών συντεταγμένων από τον τύπο

$$\phi = \arctan\left(\frac{b_0}{a_0}\right) \quad (1.2)$$

Ένας άλλος τρόπος να αναπαρασταθεί ένα σήμα είναι με το άθροισμα ενός ημιτόνου και ενός συνημιτόνου με διαφορετικές εντάσεις το οποίο ισούται με ένα ημίτονο με μετατοπισμένη φάση.

$$r\sin(\omega t+\phi)=A\sin(\omega t)+B\cos(\omega t) \quad (1.3)$$

Η εξίσωση για την εύρεση της φάσης είναι

$$\phi = \arctan(B/A) \quad (1.4)$$

Sine	+	Cosine	Phase shift in degrees	Phase shift in terms of π	Phase shift in radians
sin	+	0 cos	0	0	0
sin	+	0.5 cos	22.5	$\pi/8$	0.3926
sin	+	cos	45	$\pi/4$	0.7853
0.5 sin	+	cos	67.5	$3\pi/8$	1.1780
0 sin	+	cos	90	$\pi/2$	1.5707
-0.5 sin	+	cos	112.5	$5\pi/8$	1.9634
-sin	+	cos	135	$3\pi/4$	2.3561
-sin	+	0.5 cos	157.5	$7\pi/8$	2.7488
-sin	+	0 cos	180	π	3.1415
-sin	+	-0.5 cos	202.5	$9\pi/8$	3.5344
-sin	+	-cos	225	$5\pi/4$	3.9269
-0.5 sin	+	-cos	247.5	$11\pi/8$	4.3196
0 sin	+	-cos	270	$3\pi/2$	4.7123
0.5 sin	+	-cos	292.5	$13\pi/8$	5.1050
sin	+	-cos	315	$7\pi/4$	5.4977
sin	+	-0.5 cos	337.5	$15\pi/8$	5.8904
sin	+	-0 cos	360	2π	6.2831

Πίνακας 1: Φάσεις για αθροίσματα ημιτόνων-συνημιτόνων με διαφορετικά πλάτη

Επίσης μέσω της σχέσης του Euler

$$e^{j\theta} = \cos(\theta) + j\sin(\theta) \quad (1.5)$$

το τριγωνομετρικό συνημίτονο μπορεί να γραφτεί ως

$$\cos(\theta) = 1/2[e^{j\theta} + e^{-j\theta}] \quad (1.6)$$

Δηλαδή οι τριγωνομετρικές εξισώσεις μπορούν να γραφούν και με εκθετική μορφή μιγαδικών, το οποίο συνηθίζεται γιατί απλοποιεί τις πράξεις που χρειάζονται για τη διαχείριση του σήματος.

1.3 Θεώρημα Fourier

Όπως ειπώθηκε αρχικά, ο αλγόριθμος επεξεργασίας αυτής της εργασίας βασίζεται στο θεώρημα του Fourier. Το θεώρημα αυτό αποδεικνύει ότι κάθε περίπλοκο περιοδικό σήμα μπορεί να αναπαρασταθεί ως άθροισμα πολλών απλών σημάτων, δηλαδή ημιτόνων, διαφορετικών εντάσεων, συχνοτήτων και φάσεων. Η εφαρμογή που αναπτύχθηκε έχει ως πυρήνα την επεξεργασία Short time Fourier Transform (STFT) μία εφαρμογή του Discrete Fourier Transform (DFT). Στη συνέχεια θα παρουσιαστεί το μαθηματικό πλαίσιο και οι βασικές ιδέες που αφορούν τις επεξεργασίες αυτές σε επίπεδο ψηφιακού σήματος ήχου.

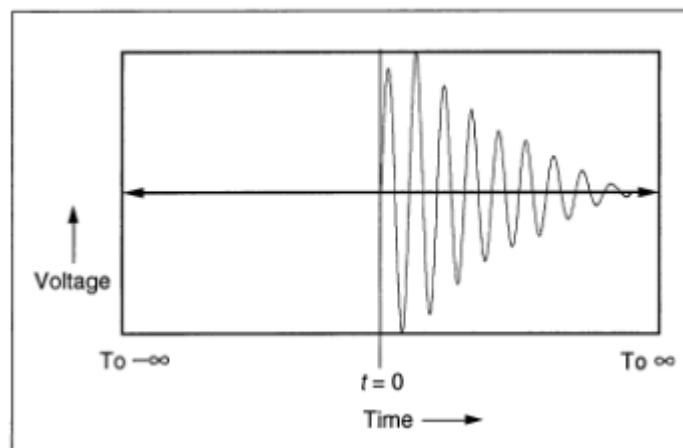
1.3.1 Σειρές Fourier

$$x(t) = C_0 + \sum_{n=1}^{\infty} C_n \cos(n\omega_0 t + \theta_n) \quad (1.7)$$

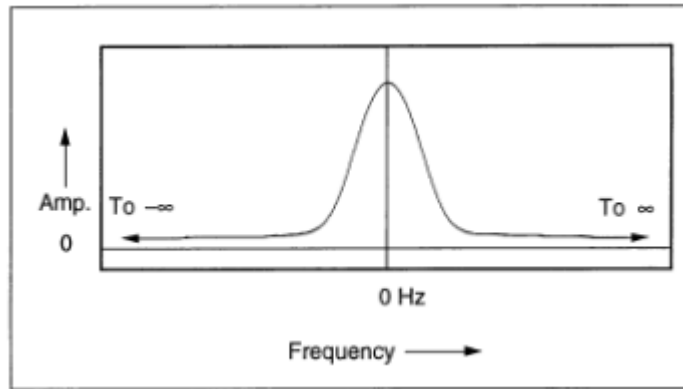
Η συνάρτηση $x(t)$ είναι ένα άθροισμα αρμονικά συσχετιζόμενων ημιτονοειδών συναρτήσεων με συχνότητα $\omega_n = n\omega_0 = 2\pi/T$ (Roads,1996). Το πρώτο ημιτονοειδές στοιχείο είναι η θεμέλιος και έχει περίοδο T . Οι σταθερές C_n και θ_n είναι η ένταση και η φάση του κάθε στοιχείου αντίστοιχα. Μέσω της συνάρτησης αυτής μπορεί να δομηθεί ένα περιοδικό σήμα.

1.3.2 Μετασχηματισμός Fourier

Για να αναλύσουμε ένα μη περιοδικό σήμα θα χρειαστεί να χρησιμοποιήσουμε τον μετασχηματισμό Fourier (M.F.). Έστω ότι έχουμε ένα συνεχές σήμα απείρου μήκους και εύρους. Το θεώρημα Fourier λέει ότι μπορεί να αναδομηθεί με άπειρα ημιτονοειδή κύματα διαφορετικών εντάσεων, συχνοτήτων και αρχικών φάσεων που θα αντιπροσωπεύουν όλες τις συχνότητες από 0 μέχρι το άπειρο, με ένα κάτοπτρο στις αρνητικές συχνότητες. Οι αρνητικές συχνότητες αυτές, δεν έχουν φυσική σημασία.



Σχήμα 3: Σήμα άπειρης έκτασης



Σχήμα 4: Η ένταση του μετασχηματισμού Fourier του σήματος στο Σχήμα 1 $X(j\omega)$ σε σχέση με τη συχνότητα (το X συμβολίζει τον Μετασχηματισμό Fourier)

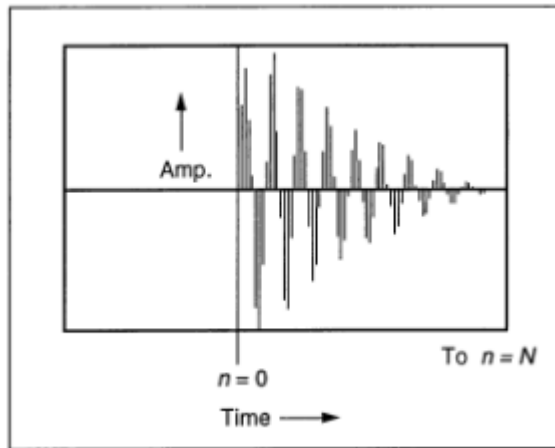
Κάθε τιμή της $X(f)$ (μετασχηματισμός Fourier της $x(t)$) είναι ένας μιγαδικός αριθμός από τον οποίο μπορούμε να εξάγουμε την ένταση και τη φάση της αντίστοιχης συχνότητας. Βέβαια το Σχήμα 2 δεν απεικονίζει πλήρως τον Μ.Φ., αλλά μόνο τις εντάσεις που η κάθε συχνότητα έχει στον ήχο. Οι αρχικές φάσεις τους παραλείπονται σε αυτή το διάγραμμα. Παρακάτω δίνεται ο τύπος του Μ.Φ.

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (1.8)$$

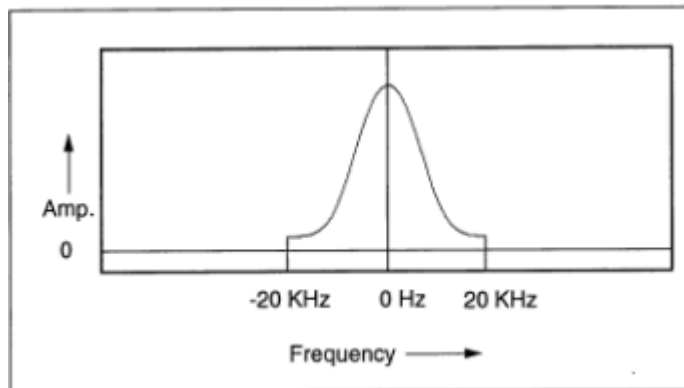
Δηλαδή ο Μ.Φ. κάθε συχνότητας είναι το ολοκλήρωμα του πολλαπλασιασμού του εισαγόμενου σήματος με το καθαρό τόνο $e^{-j\omega t}$.

1.3.3 Μετασχηματισμός Fourier διακριτού χρόνου (DTFT)

Τα ψηφιακά συστήματα μπορούν να διαχειριστούν σήματα διακριτού χρόνου (δείγματα). Έστω ότι ένα σήμα περνάει από ένα anti-aliasing φίλτρο και υπόκειται σε δειγματοληψία. Πλέον το σήμα θα έχει διακριτές τιμές έντασης $x[n]$ αλλά θα συνεχίζει να έχει συνεχές (άπειρες) τιμές συχνότητας λόγω της θεωρητικά άπειρης διάρκειάς του (άπειρα δείγματα n). Οι μετασχηματισμένες τιμές του σήματος μπορούν να σημειώνονται ως $X(e^{j\theta})$. Το σήμα θα είναι κάπως έτσι.

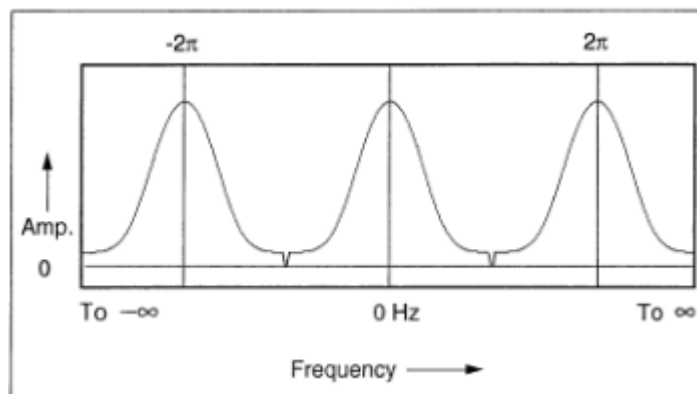


Σχήμα 5: Το σήμα διακριτού χρόνου $x[n]$ που προέκυψε ύστερα από δειγματοληψία του σήματος $x(t)$ (Σχήμα 1)



Σχήμα 6: Η ένταση σε σχέση με τη συχνότητα του antialiased DTFT του σήματος $x[n]$

Στη πραγματικότητα, όμως, οι άπειρες συχνότητες του μετασχηματισμένου σήματος, περιέχουν πολλές ίδιες συχνότητες καθώς οι τιμές για συχνότητες που διαφέρουν κατά 2π είναι ίδιες.



Σχήμα 7: Συνεχές συχνοτικό φάσμα του DTFT του σήματος $x[n]$

Φαίνεται ότι ο DTFT επαναλαμβάνεται σύμφωνα με τη σχέση

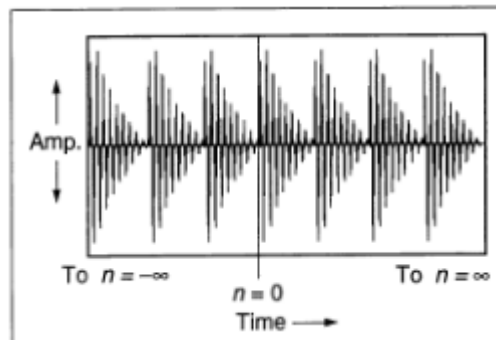
$$X(e^{j\theta_0}) = X(e^{j(\theta_0+2\pi m)}) \quad (1.9)$$

m: αυθαίρετο πολλαπλάσιο

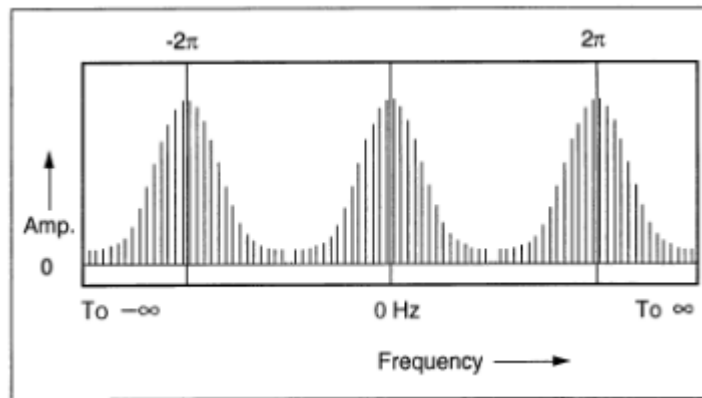
Στη πράξη οι τιμές που χρειαζόμαστε είναι από $-\pi$ έως π , οι οποίες είναι σχεδόν ίδιες με τις τιμές του MF του συνεχούς σήματος $x(t)$, $X(f)$ για εύρος από $-1/2f_s$ έως $1/2f_s$ (όπου f_s , συχνότητα δειγματοληψίας).

1.3.4 Διακριτός Μετασχηματισμός Fourier (DFT)

Τα ψηφιακά συστήματα βέβαια, δουλεύουν με διακριτές τιμές οπότε χρειάζεται μία μορφή του MF που δουλεύει με διακριτό φάσμα. Ένας τρόπος να επιτευχθεί αυτό, είναι να χρησιμοποιηθεί ένα περιοδικό σήμα. Ένα τέτοιο σήμα με περίοδο T περιέχει τη θεμέλιο συχνότητα και τις αρμονικές της. Μπορούμε να παράξουμε ένα τέτοιο σήμα αντιγράφοντας επ' άπειρο ένα σήμα πεπερασμένης διάρκειας, πίσω και μπροστά στο χρόνο. Έτσι θα παραγόταν το περιοδικό σήμα $\tilde{x}[n]$.



Σχήμα 8: Το σήμα $\tilde{x}[n]$ ύστερα από δειγματοληψία. Περιέχει αντίγραφα του σήματος $x[n]$ επ' άπειρο



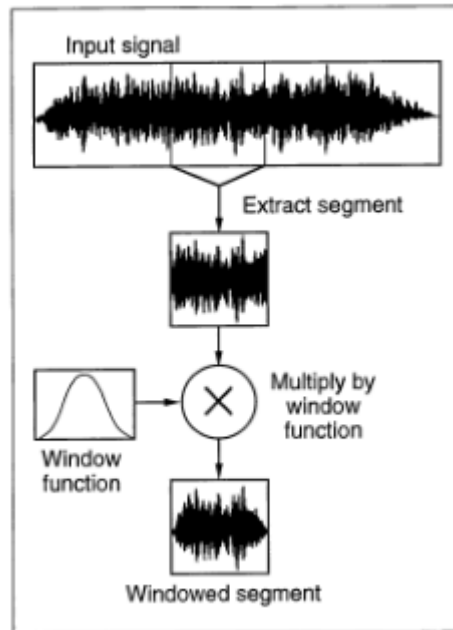
Σχήμα 9: Σχεδιάγραμμα του DFT του σήματος $\tilde{x}[n]$, διακριτού φάσματος

Φαίνεται πως το μετασχηματισμένο σήμα $\tilde{x}[n]$, δηλαδή το φάσμα του, είναι και αυτό διακριτό. Για την ακρίβεια, το φάσμα στον DFT αποτελείται από τόσες διακριτές συχνότητες όσες τα δείγματα του σήματος. Όπως η δειγματοληψία του $x(t)$ στο χρόνο έχει ως αποτέλεσμα την δημιουργία αντιγράφων στο φάσμα, έτσι η δημιουργία αντιγράφων του $x[n]$ στο χρόνο έχει ως αποτέλεσμα τη δειγματοληψία στο φάσμα. (Roads, 1996)

1.3.5 Μετασχηματισμός Fourier μικρού χρονικού διαστήματος (STFT)

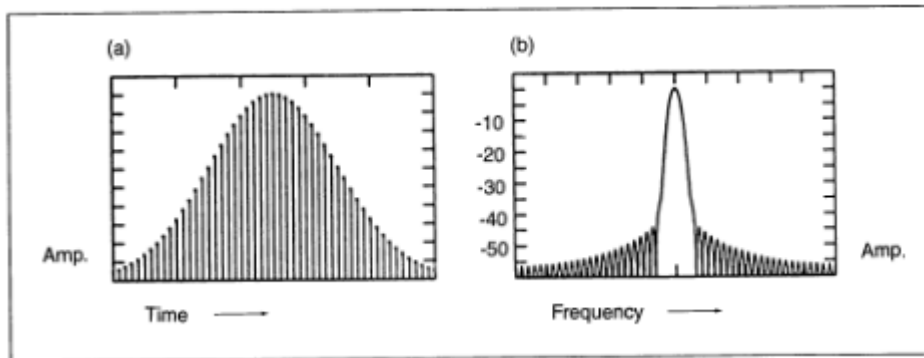
Ο Short-time Fourier transform είναι η πρακτική εφαρμογή του DFT ο οποίος δίνει τη δυνατότητα να εφαρμοσθεί ο MF σε πραγματικά σήματα, πεπερασμένης διάρκειας, που αλλάζουν συνεχώς στο χρόνο και αποτελεί τη βασική λειτουργία φασματικών επεξεργασιών και αναλύσεων των σύγχρονων λειτουργιών. Στη πραγματικότητα σχεδόν πάντα χρησιμοποιείται ο FFT (Fast Fourier Transform), ο οποίος είναι μία πιο σύντομη και υπολογιστικά πιο ελαφριά εκδοχή του DFT.

Ο STFT χωρίζει τον εισερχόμενο ήχο σε κομμάτια μικρής χρονικής διάρκειας, πολλαπλασιάζοντας τον με σύντομες περιβάλλουσες (συνήθως διάρκειας 1ms – 1s), που στην ορολογία του STFT αποκαλούνται «παράθυρα» (windows, window functions). Αυτό, βέβαια, έχει ως συνέπεια η φασματική ανάλυση να είναι προϊόν του ήχου και της περιβάλλουσας, δηλαδή να υπάρχει κάποια παραμόρφωση σε αυτή.



Σχήμα 10: Χωρισμός ηχητικού σήματος σε κομμάτια μικρής διάρκειας μέσω εφαρμογής παραθύρων (πολλαπλασιασμός ήχου με περιβάλλουσες μικρής διάρκειας)

Συνεπώς, το σχήμα του παραθύρου παίζει μεγάλο ρόλο στο τελικό φάσμα που θα προκύψει. Υπάρχουν πολλά πιθανά σχήματα παραθύρων για διαφορετικές ανάγκες. Για τις πιο πολλές μουσικές εφαρμογές του STFT μία ομαλή καμπανοειδή καμπύλη, όπως η καμπύλη του Gauss, είναι επαρκής.



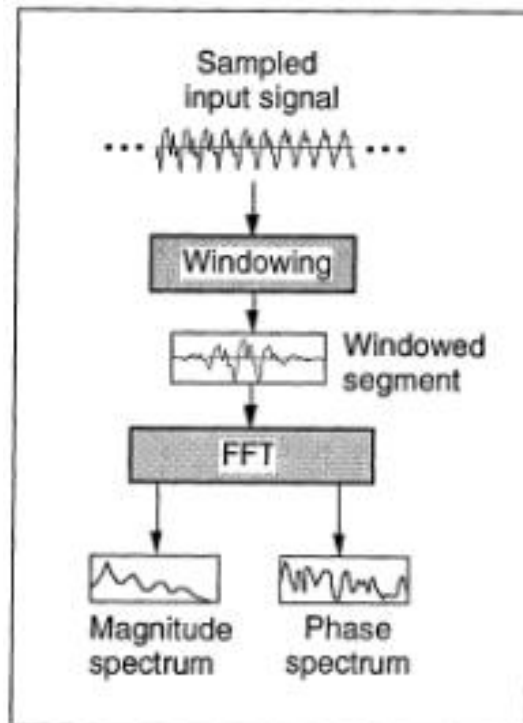
Σχήμα 11: (a) Παράθυρο με σχήμα καμπύλης Gauss. (b) Φάσμα του παραθύρου(συχνότητα σε σχέση με ένταση σε dB). Η δεύτερη κορυφή στο φάσμα είναι περίπου στα -45 dB, παρόλα αυτά φαίνεται η παραμόρφωση που θα υποστεί η φασματική ανάλυση του εισερχόμενου σήματος.

Όπως ειπώθηκε και νωρίτερα, το πλήθος δειγμάτων του ήχου που αναλύεται φασματικά είναι ίδιο με το πλήθος των κεντρικών συχνοτήτων (συνήθως αναφέρονται ως bins) που θα περιέχει η ανάλυσή του. Για παράδειγμα αν ένας ήχος χωριστεί σε κομμάτια των 100 δειγμάτων, θα προκύψουν τιμές πλάτους και φάσης για 100 bins, αρχίζοντας από 0 και φτάνοντας μέχρι τη συχνότητα δειγματοληψίας f_s , με βήμα

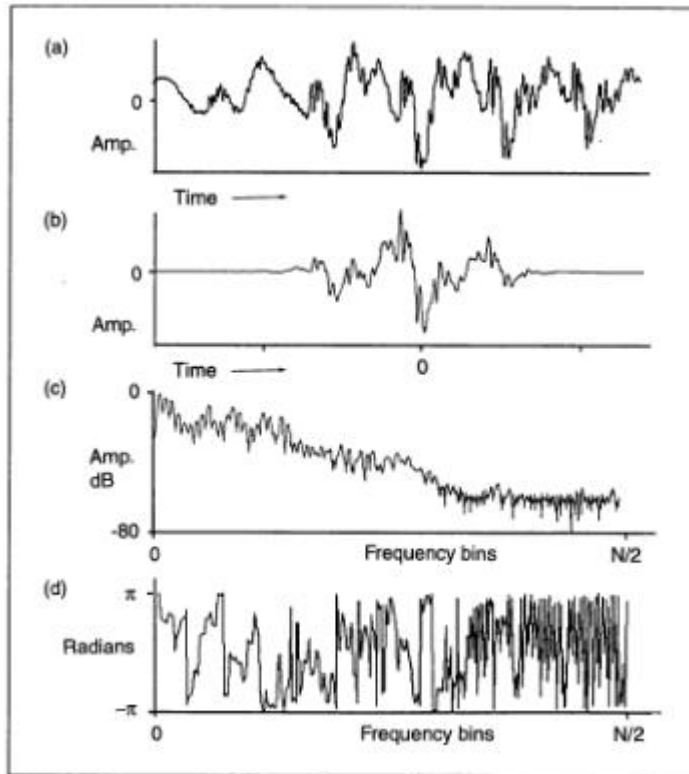
$$\frac{f_s}{\text{πλήθοςδειγμάτων}N} \quad (1.10)$$

Δηλαδή για $f_s = 40000$ το πρώτο bin ($b[0]$) θα έχει πληροφορίες για συχνότητα 0, $b[1] = 400$ Hz, $b[2] = 800$ Hz κ.ο.κ. μέχρι $b[N-1] = f_s$ δηλαδή $b[99] = 40000$. Από αυτό φαίνεται ότι χρειαζόμαστε μόνο τα μισά bin από την ανάλυση για να ικανοποιείται και το θεώρημα του Nyquist.

Οπότε ύστερα από την ανάλυση του σήματος μέσω του STFT, θα έχουμε καταλήξει σε $N/2$ πλήθος χρήσιμων bins, που το κάθε ένα θα έχει μία τιμή έντασης και μία τιμή αρχικής φάσης. Αυτές οι τιμές ύστερα μπορούν να απεικονίσουν το φάσμα και το φασικό φάσμα του ήχου σε σχέση με τη συχνότητα.



Σχήμα 12: Επισκόπηση του STFT

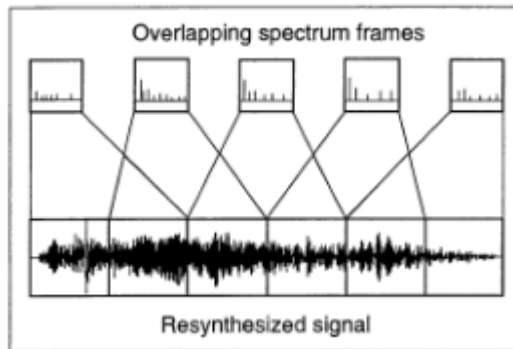


Σχήμα 13: (α) κυματομορφή προς ανάλυση. (β) Κομμάτι πολλαπλασιασμένο με παράθυρο (γ) Φάσμα έντασης (εύρος 80 dB). (δ) Φάσμα φάσης (εύρος $-\pi$ έως π).

Κάθε τέτοια ανάλυση ενός κομματιού του ήχου αποκαλείται frame (καρέ). Δηλαδή ένα frame του ήχου είναι οι $N/2$ τιμές έντασης και φάσης του ήχου των αντίστοιχων συχνοτήτων. Έτσι τελειώνει το μέρος της ανάλυσης του ήχου. Με επεξεργασία αυτών των τιμών, είτε με αλλαγή του τρόπου ανασύνθεσής τους, μπορούμε να επηρεάσουμε το φάσμα του ήχου.

1.3.6 Overlap-add Ανασύνθεση (OA)

Έχοντας τις τιμές ανάλυσης του ήχου, αυτός μπορεί να ανασυντεθεί μέσω του αντίστροφου διακριτού Μετασχηματισμού Fourier (IDFT), ο οποίος θα συνθέσει μία κυματομορφή που αντιστοιχεί στις πληροφορίες ενός frame. Στη συνέχεια αυτές οι κυματομορφές πολλαπλασιάζονται με τις ίδιες περιβάλλουσες (παράθυρα) με τις οποίες είχαν πολλαπλασιαστεί κατά το στάδιο της ανάλυσης και ύστερα προστίθονται, με αποτέλεσμα την ανασύνθεση του αρχικού ήχου. Είναι σημαντικό στη πρόσθεση αυτών των κυματομορφών, η ένταση να παραμένει σταθερή μόνιμα, αλλιώς αποτέλεσμα θα είναι η διαμόρφωση πλάτους του σήματος (AM), δηλαδή έντονη παραμόρφωση. Οπότε οι κυματομορφές αυτές, περασμένες από τα παράθυρά τους, αλληλεπικαλύπτονται, έτσι ώστε το άθροισμα των παραθύρων αυτών να είναι πάντα σταθερό.

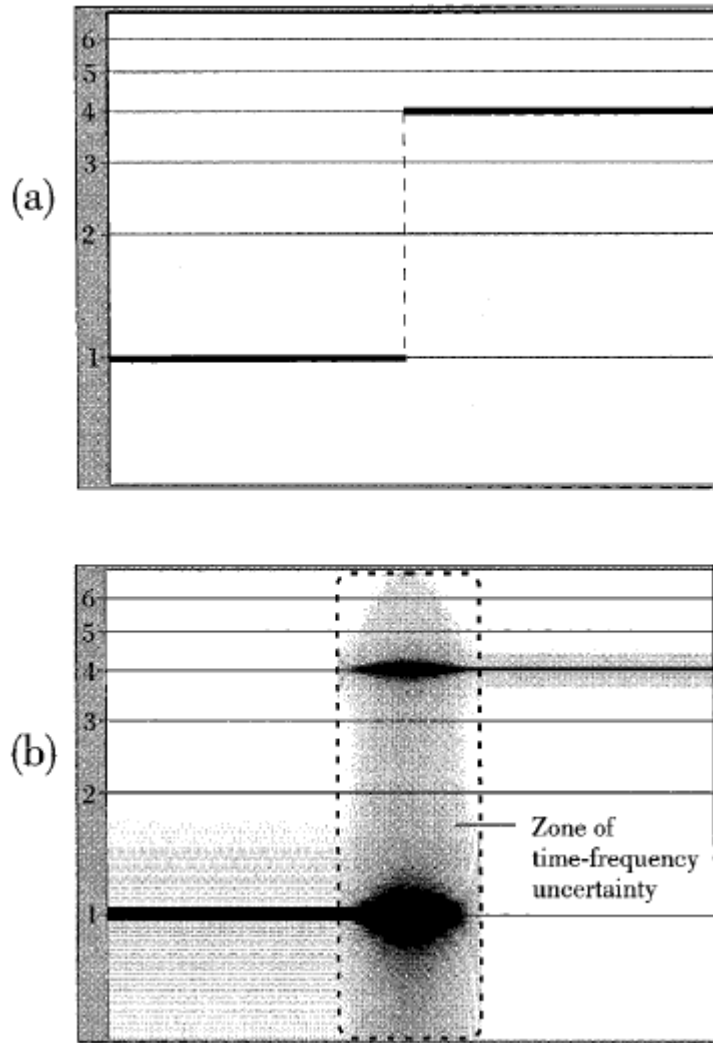


Σχήμα 14: Παράδειγμα Overlap-add ανασύνθεσης

Εάν θέλουμε να επεξεργαστούμε τον ήχο έντονα, όπως να τον επεκτείνουμε χρονικά κατά πολύ, μπορεί να παρουσιαστούν αλλοιώσεις στον ήχο, ειδικά σε ήχους με ραγδαίες αλλαγές έντασης (κρουστικούς). Αυτό μπορεί να αποφευχθεί σε μεγάλο βαθμό με περαιτέρω αλληλοεπικάλυψη των ανασυνθετημένων κομματιών που προστίθονται.

1.3.7 Περιορισμοί του STFT

Όπως αναφέρθηκε και νωρίτερα, το μέγεθος των κομματιών του ήχου που αναλύονται έχει άμεση σχέση με την συχνοτική ανάλυση του ήχου. Για την ακρίβεια για έναν STFT που δουλεύει με κομμάτια N δειγμάτων, η συχνοτική του ανάλυση θα είναι $N/2$ κεντρικών συχνοτήτων. Φαίνεται λοιπόν ότι υπάρχει μία σχέση συμβιβασμού ανάμεσα στη συχνοτική και την χρονική ακρίβεια. Για μεγάλη χρονική ακρίβεια (μικρά κομμάτια ήχου) υπάρχει μικρή συχνοτική ανάλυση (λίγες κεντρικές συχνότητες (bins)) και το αντίστροφο. Ανάλογα με τις ανάγκες ανάλυσης του κάθε ήχου επιλέγεται διαφορετικό μέγεθος παραθύρου του STFT.



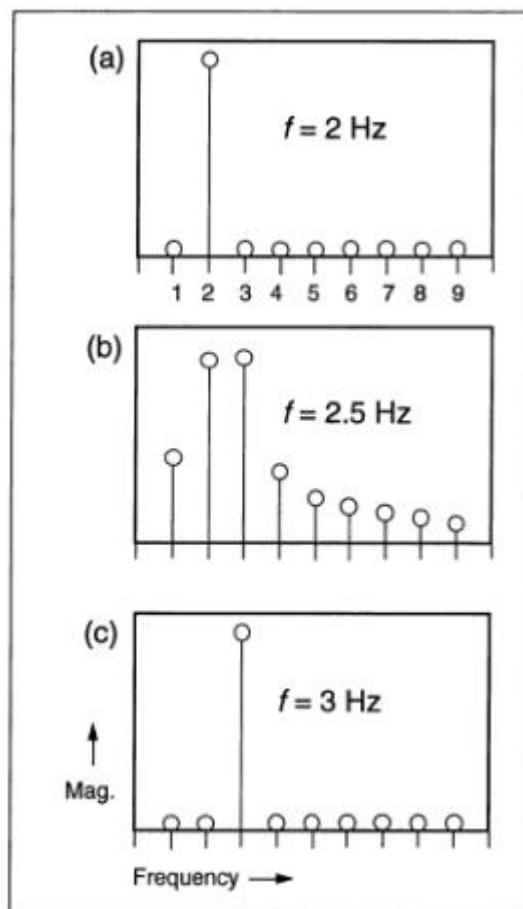
Σχήμα 15: Χρονική-συχνотική αβεβαιότητα σε STFT ανάλυση. (α) Σήμα που δώθηκε προς ανάλυση. Αποτελείται από τόνο 1000Hz που εναλλάσσεται ακαριαία σε 4000 Hz. (β) Φάσμα που προέκυψε από την ανάλυση (εύρος 500-7000Hz, λογαριθμική κλίμακα συχνότητας).

Στο σχεδιάγραμμα φαίνεται το πρόβλημα της χρονικής αβεβαιότητας. Για τα κομμάτια που αναλύθηκαν, τα οποία περιείχαν και τους δύο τόνους, η ανάλυσή τους έχει ως αποτέλεσμα μία διάχυση της ενέργειας γύρω από αυτούς καθώς και σε όλο το φάσμα. Στη συνέχεια φαίνονται διάφορες συχνотικές αναλύσεις σε σχέση με τη διάρκεια των παραθύρων σε ms.

Length of time window (in ms)	Frequency resolution (analysis bandwidth) (in Hz)
1	1000
2	500
3	330
10	100
20	50
30	33
100	10
200	5
300	3
1000 (1 sec)	1
2000	0.5
3000	0.3

Πίνακας 2: Συχνотική ευκρίνεια σε σχέση με μήκος παραθύρου STFT ανάλυσης

Γενικά για συχνότητες που βρίσκονται ανάμεσα στα bins ανάλυσης που προκύπτουν, η πλειοψηφία της ενέργειας υπάρχει στα κοντινότερα bins της συχνότητας αυτής, αλλά μικρό ποσοστό ενέργεια σκορπίζεται σε όλο το φάσμα.

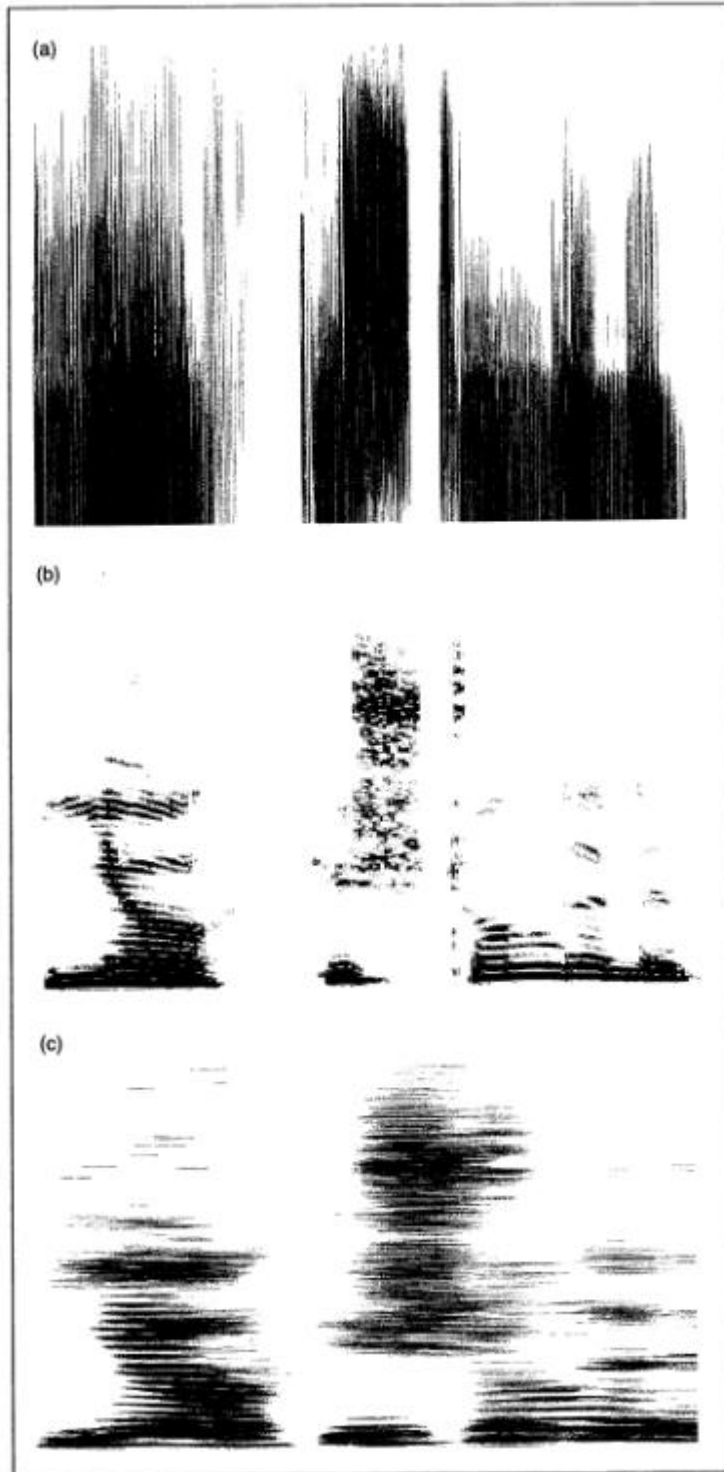


Σχήμα 16: Τρία στιγμιότυπα STFT ανάλυσης ενός ήχου που ξεκινάει με συχνότητα 1Hz και καταλήγει 3Hz. Ο συγκεκριμένος STFT έχει bins ανά 1 Hz. Όταν η συχνότητα του ήχου είναι 2.5, πέφτει ακριβώς ανάμεσα στα δύο κοντινότερα bins και η ενέργεια απλώνεται σε όλο το φάσμα.

2. Βασικές εφαρμογές STFT στην ανάλυση/ανασύνθεση ήχου

2.1 Φασματογράφημα

Η πιο συνηθισμένη εφαρμογή του STFT είναι η απεικόνιση του φάσματος, κάτι που συνήθως αποκαλείται φασματογράφημα. Αφού ο ήχος έχει αναλυθεί σε bins πληροφοριών, έχοντας σχηματίσει ένα frame, το μόνο που μένει είναι η γραφική απεικόνιση αυτού του frame ως μία στήλη τιμών εντάσεων για τις διακριτές τιμές συχνοτήτων της ανάλυσης, η οποία θα ακολουθηθεί από τη στήλη του επόμενου frame επιτρέποντας έτσι τη απεικόνιση σε πραγματικό χρόνο του σήματος. Συνήθως δίνεται η δυνατότητα ορισμού του μεγέθους του παραθύρου του STFT, για να υπάρχει η επιλογή της συχνοτικής και χρονικής ανάλυσης, ανάλογα με τον ήχο και τις ανάγκες της κάθε οπτικοποίησης. Μία ισορροπημένη τιμή παραθύρου είναι 2048 δείγματα (δηλαδή 1024 bins) για γενική χρήση. Ακραίες τιμές μεγέθους του παραθύρου συνήθως καταλήγουν σε αχνά φασματογραφήματα με περιορισμένη χρησιμότητα, λόγω της μεγάλης χρονικής είτε φασματικής αβεβαιότητας.



Σχήμα 17: Φασματογραφήματα που παρουσιάζουν φάσμα ανθρώπινης φωνής ηχογραφημένη στα 44100 Hz με διαφορετικά μεγέθη παραθύρου. (α) Παράθυρο 32 δειγμάτων. Χρονική ευκρίνεια 0.725 ms, συχνοτική ευκρίνεια 1378 Hz. (β) Παράθυρο 1024 δειγμάτων. Χρονική ευκρίνεια 23.22 ms, συχνοτική ευκρίνεια 43.07 Hz. (γ) Παράθυρο 8192 δειγμάτων. Χρονική ευκρίνεια 185.8 ms, συχνοτική ευκρίνεια 5.383 Hz.

2.2 Phase Vocoder

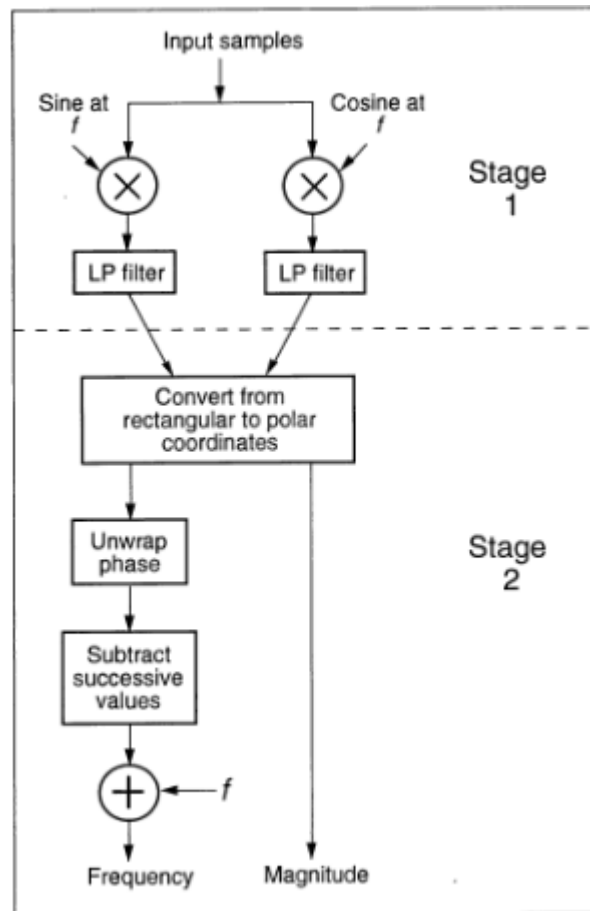
Ο όρος vocoder προέκυψε από το VOice CODER (κωδικοποιητής φωνής), που ήταν το όνομα μίας συσκευής η οποία είχε σχεδιαστεί για να μειώσει το εύρος ζώνης της φωνής με στόχο την ικανοποιητική μετάδοση της φωνής μέσω τηλεφώνου (Gordon and Strawn, 1987). Στη πράξη όμως δεν υπήρχε μείωση πληροφορίας μέσω αυτής της τεχνικής γιατί ήθελε πολλές παραμέτρους έντασης για ικανοποιητικά αποτελέσματα. Άλλο βασικό πρόβλημα ήταν το ότι δεν λάμβανε υπόψη τις τιμές φάσης, με αποτέλεσμα το ανασυνθετημένο αποτέλεσμα να είναι αρκετά φτωχό. Αργότερα αναπτύχθηκε μία εκδοχή του vocoder που χρησιμοποιούσε και τις τιμές φάσεις κατά την ανασύνθεση με πολύ καλύτερα αποτελέσματα (rphase vocoder). Όμως ακόμα ήταν αρκετά βαριά υπολογιστική διαδικασία και έτσι δεν εφαρμόζονταν για αρκετά χρόνια, μέχρι να μπορέσει να πραγματοποιηθεί με FFT στα τέλη του 1970. Έπειτα χρησιμοποιήθηκε εκτεταμένα, με αποτέλεσμα να καθιερωθεί ως ένα βασικό εργαλείο σήμερα για φασματικές επεξεργασίες ήχου.

2.2.1 Βασικές Αρχές

Υπάρχουν πολλές εκδοχές υλοποίησης του phase vocoder καθώς και πολλές επεκτάσεις του. Η βασική ιδέα, όμως, μένει η ίδια στις πιο πολλές. Αρχικά, πολλαπλασιάζει το κομμάτι του σήματος, το οποίο έχει προκύψει από πολλαπλασιασμό με το παράθυρο, ξεχωριστά με ένα ημιτονικό και ένα συνημιτονικό σήμα με συχνότητα ίση με τη κεντρική συχνότητα του συγκεκριμένου bin (Dogde, Jerse, 1997). Αυτό έχει ως αποτέλεσμα τη συχνοτική μετατόπιση των συχνοτικών συνιστωσών του εισερχόμενου ήχου κατά συν και πλην τη κεντρική συχνότητα του bin (Roads, 1996). Ύστερα και οι δύο αυτές διακλαδώσεις περνάνε από ένα χαμηλοπερατό φίλτρο ξεχωριστά που έχει ως αποτέλεσμα να αφήνει να περάσουν μόνο οι κοντινές συχνότητες του ήχου στην κεντρική συχνότητα του bin.

Π.χ. Έστω ότι έχουμε ένα κομμάτι ήχου και το πολλαπλασιάζουμε με ένα ημίτονο με συχνότητα 450 Hz (κεντρική συχνότητα bin). Έστω ότι το σήμα έχει συχνοτικό εύρος 50 – 2000 Hz. Μετά τον πολλαπλασιασμό αυτό, θα προκύψει σήμα με εύρη -400 έως 1550 Hz και 500 με 2450 Hz. Έστω ότι έπειτα το περνάμε από ένα χαμηλοπερατό φίλτρο με συχνότητα αποκοπής 25 Hz. Το άνω εύρος του σήματος αποκόβεται πλήρως και το κάτω εύρος που περνάει είναι -25 έως 25 Hz, υποθέτοντας ότι οι συχνότητες πάνω από τη συχνότητα αποκοπής αποκόπτονται πλήρως. Αυτό το εύρος συχνοτήτων που έχει περάσει αντιστοιχούν στις συχνότητες 425 -475 Hz του αρχικού σήματος. Δηλαδή το συχνοτικό εύρος -25 έως 25 Hz που περνάει από το φίλτρο έχει τη μορφή (τη κατανομή ενέργειας) άρα και τις πληροφορίες έντασης-συχνότητας (φάσμα), που έχει το εύρος 425 -475 Hz του αρχικού σήματος. Με αυτό τον τρόπο έχουμε καταφέρει να απομονώσουμε τις συχνότητες που βρίσκονται κοντά στη κεντρική συχνότητα του bin (+- 25Hz στο παράδειγμα), και να προχωρήσουμε στην ανάλυση αυτών. Το παράδειγμα αυτό εξηγεί τη διαδικασία

φιλτραρίσματος για τη μία διακλάδωση του σήματος. Αντίστοιχο αποτέλεσμα παρέχεται και από την άλλη.



Σχήμα 18: Βασικό διάγραμμα ροής Phase Vocoder

Οπότε μετά από τα χαμηλοπερατά φίλτρα έχουμε τιμές στη μορφή της σχέσης (1.3)

$$r\sin(\omega t + \phi) = A\sin(\omega t) + B\cos(\omega t)$$

Το αποτέλεσμα φαίνεται να είναι ημιτονικό σήμα σε μορφή αθροίσματος ημιτόνου και συνημιτόνου ίδιας συχνότητας. Άρα στην ουσία τα δύο αυτά φίλτρα μας παρέχουν τις καρτεσιανές συντεταγμένες (A,B ή πραγματικό και φανταστικό μέρος) που ορίζουν ένα μιγαδικό αριθμό ως σημείο, ο οποίος περιέχει τη πληροφορία συχνότητας και έντασης (μέτρου). Το μέτρο και η φάση του μιγαδικού αυτού μπορούν να προκύψουν από τις σχέσεις (1.1) και (1.4) αντίστοιχα.

$$r = \sqrt{a^2 + b^2} \quad (1.1)$$

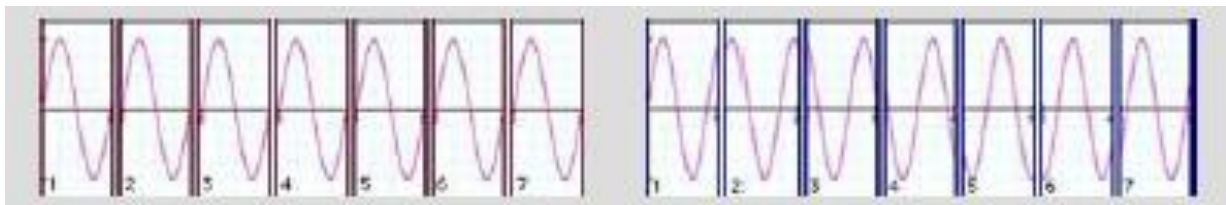
$$\phi = \arctan(B/A) \quad (1.4)$$

Έτσι γίνεται η μετατροπή από καρτεσιανές (rectangular) συντεταγμένες σε πολικές (polar), εφόσον καταλήξαμε σε τιμές μέτρου και αρχικής φάσης, οι οποίες ορίζουν έναν μιγαδικό αριθμό ως διάνυσμα.

Να θυμίσουμε σε αυτό το σημείο ότι η τιμή φάσης και η τιμή έντασης που προκύπτουν αναφέρονται για συχνότητες του αναλυόμενου κομματιού ήχου, κοντινές στη κεντρική συχνότητα του συγκεκριμένου bin για το οποίο τρέχει ο αλγόριθμος εκείνη τη στιγμή. Δηλαδή η διαδικασία αυτή θα γίνει τόσες φορές όσες το σύνολο (N) των τιμών bin της ανάλυσης, για την παραγωγή ενός frame, το οποίο θα αποτελείτε από N τιμές έντασης και φάσης, ένα ζευγάρι για κάθε bin.

Στη συνέχεια η τιμή έντασης αποθηκεύεται και η τιμή φάσης χρησιμοποιείται ώστε να γίνει μία καλύτερη προσέγγιση της αληθινής συχνότητας που βρίσκεται κοντά σε αυτό το bin του αναλυόμενου ήχου. Μέσω των διαφορών διαδοχικών τιμών φάσης μπορούμε να υπολογίσουμε την απόκλιση της πραγματικής τιμής συχνότητας του ήχου από τη κεντρική συχνότητα του bin.

Για παράδειγμα. Έστω ότι έχουμε ένα ημιτονικό σήμα με μισή συχνότητα της συχνότητας του κοντινότερου bin που προβλέπεται από την FFT ανάλυση. Τότε θα έχουμε διαφορά φάσης 0 για όλα τα ακέραια παράθυρα και π για όλα τα περιττά (Drolics, 2011). Η μόνιμη διαφορά φάσης θα είναι τότε 50% όπως και η διαφορά συχνότητας του εισερχόμενου σήματος σε σχέση με τη προβλεπόμενη κεντρική συχνότητα του συγκεκριμένου bin από τον FFT.



Σχήμα 19: Αριστερά ημιτονικό σήμα ίδιας συχνότητας με το bin ανάλυσης έχει ως αποτέλεσμα μηδενική φασική μετατόπιση. Δεξιά ημιτονικό σήμα υψηλότερης συχνότητας με διαφορετικές φασικές μετατοπίσεις (για την ακρίβεια αύξουσα μετατόπιση) για κάθε διαδοχικό frame

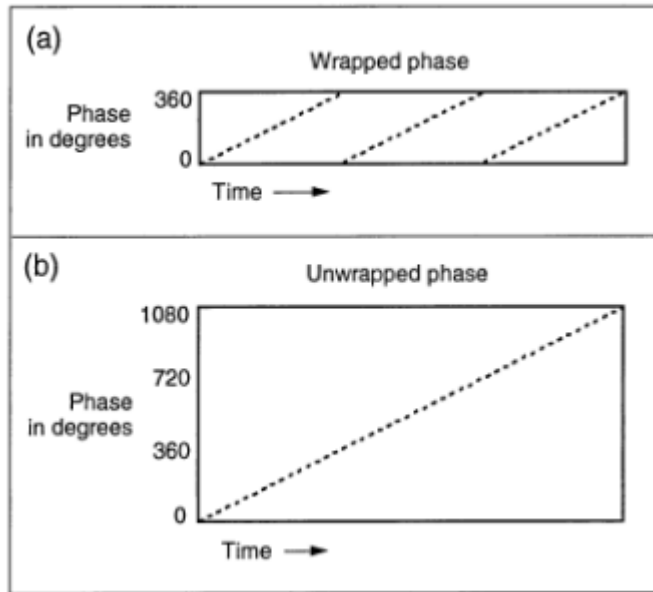
Να διευκρινιστεί ότι η διαφορά φάσης αναφέρεται σε ίδια bins για διαδοχικά frames. Επίσης η διαφορά φάσης συνήθως "τυλίγεται" (phase wrapping) στο εύρος τιμών $-\pi$ έως π .

$$\Delta\varphi_{m,k} = \varphi_{m,k} - \varphi_{m-1,k} \quad (2.1)$$

όπου m: δείκτης frame

k: δείκτης bin

Αυτή η απόκλιση συχνότητας, συχνά καλείται και στιγμιαία συχνότητα, μπορεί να προκύψει από τη διαφορά φάσης δύο διαδοχικών bins διά το χρόνο που μεσολαβεί μεταξύ τους, δηλαδή το αντίστροφο του αποτελέσματος της διαίρεσης της συχνότητας δειγματοληψίας διά το μέγεθος του FFT. Κατά την ανασύνθεση του ήχου, ακολουθεί μία διαδικασία γνωστή ως “ξετύλιγμα” φάσης (phase unwrapping) κατά την οποία προστίθονται 360 μοίρες μετά την αύξηση της διαφοράς φάσης ενός κύκλου, αντί να μηδενιστεί.



Σχήμα 20: Φασικό ξετύλιγμα (phase unwrapping) (a) Τυλιγμένη φάση στο εύρος $0-2\pi$ (b) Ξετυλιγμένη φάση (απλό άθροισμα διαδοχικών διαφορών φάσης)

Κατά την ανασύνθεση χρειαζόμαστε “απόλυτες” τιμές φάσης και όχι τη “σχετική” διαφορά φάσεων. Οπότε συνηθίζεται να χρησιμοποιείται η τρέχουσα φάση (running phase), η οποία είναι το σύνολο των διαφορών φάσεων μέχρι το frame το οποίο αναπαράγεται εκείνη τη στιγμή.

$$\varphi_R = \sum_{m=0}^{m=M} \Delta\varphi_{m,k} \quad (2.2)$$

Η τρέχουσα φάση είναι σημαντική και για την ποιοτική ανασύνθεση του ήχου με διαφορετική διάρκεια (συμπύεση-διαστολή), καθώς για αναπαραγωγή των frames με αυθαίρετη σειρά, χρειάζεται να βρίσκεται η ολική διαφορά φάσης από την αρχή και όχι η διαφορά φάσης των αναπαραγόμενων frames. Η χρονική συμπύεση-διαστολή του ήχου είναι από τις βασικές επεξεργασίες του Phase Vocoder.

2.2.2 Παράμετροι

Συνηθισμένοι παράμετροι του Phase Vocoder είναι

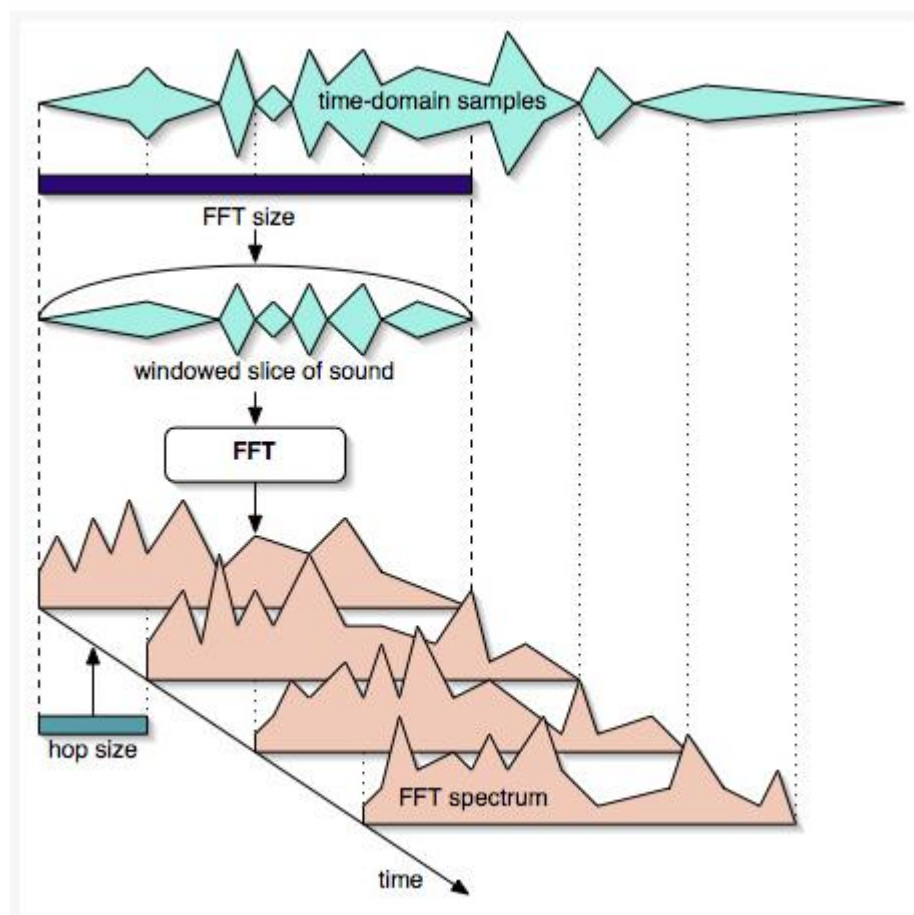
Μέγεθος FFT ή μέγεθος παραθύρου : Πλήθος δειγμάτων που κόβεται ο ήχος για αναλυθεί

Μέγεθος frame : Πλήθος bins ανάλυσης, συνήθως τα μισό από το μέγεθος FFT

Σχήμα Παραθύρου : Επιλογή σχήματος περιβάλλουσας (παραθύρου) με την οποία θα χωριστεί ο ήχος

Hop size : Το πλήθος δειγμάτων που χρειάζεται να προσπεραστούν για να αρχίσει η ανάλυση του καινούργιου κομματιού του ήχου

Βαθμός (Παράγοντας) αλληλοεπικάλυψης : Το μέγιστο πλήθος διαδοχικών παραθύρων που αλληλεπικαλύπτονται στο πρώτο αναλυόμενο κομμάτι του ήχου. Αλλιώς μέγεθος frame διά hop size



Σχήμα 21: Απεικόνιση διαδικασίας STFT. Φαίνονται μεγέθη όπως FFT size, hop size και overlap factor (παράγοντας αλληλοεπικάλυψης). Φαίνεται ότι το μέγιστο πλήθος frames του πρώτου κομματιού ανάλυσης τα οποία αλληλεπικαλύπτονται μεταξύ τους, είναι τέσσερα. Άρα ο παράγοντας αλληλοεπικάλυψης θα είναι 4.

3.Υλοποίηση της εφαρμογής χρονικής συμπίεσης/διαστολής μέσω φασματικής επεξεργασίας

Η εφαρμογή αναπτύχθηκε στο προγραμματιστικό περιβάλλον MaxMSP. Η εφαρμογή είναι μία επέκταση του Phase Vocoder. Η κεντρική ιδέα είναι η ανάπτυξη μίας εφαρμογής η οποία θα έχει τη δυνατότητα διαφορετικού ρυθμού ανάγνωσης, δηλαδή διαφορετικού βαθμού χρονικής συμπίεσης/διαστολής, για κάθε bin της φασματικής ανάλυσης του ήχου. Περαιτέρω θα δίνεται η δυνατότητα χρονικής καθυστέρησης για κάθε bin, αλλά και η δυνατότητα φιλτραρίσματος, δηλαδή εξασθένηση ή ενίσχυση της τιμής έντασης του κάθε bin.

3.1 STFT στο προγραμματιστικό περιβάλλον MaxMSP

Αρχικά, η διαδικασία φασματικής ανάλυσης μέσω STFT στο προγραμματιστικό περιβάλλον MaxMSP γίνεται με το αντικείμενο [pfft~], το οποίο παίρνει ως ορίσματα τη τιμή μεγέθους του FFT και το παράγοντα αλληλοεπικάλυψης. Το αντικείμενο απορρίπτει αυτόματα το αχρείαστο δεύτερο μισό των τιμών των bins πάνω από τη συχνότητα Nyquist, δηλαδή τις συχνότητες $f_s/2$ έως f_s . Το [pfft~] λειτουργεί ως ξεχωριστό patch με σκοπό να διαχωρίσει τις επεξεργασίες του ήχου στο πεδίο του χρόνου (ηχητικό σήμα) και στο πεδίο του φάσματος (frames). Συνήθως, λοιπόν, το ηχητικό σήμα εισάγεται σε αυτό και υπόκειται σε STFT ανάλυση, γίνεται η όποια επεξεργασία του φασματικά και στη συνέχεια ανασυντίθεται και εξέρχεται από το [pfft~] patch ως ηχητικό σήμα. Υπάρχει όμως και η δυνατότητα να εισαχθεί και να εξαχθεί ηχητικό σήμα στο [pfft~] χωρίς να υποστεί STFT ανάλυση.



Σχήμα 22: Το αντικείμενο [pfft~] στο προγραμματιστικό περιβάλλον MaxMSP

Αφού εισαχθεί το σήμα στο pfft~ αντικείμενο, μέσα στο ξεχωριστό patch πλέον, λαμβάνουμε την έξοδο του σήματος αφού έχει υποστεί STFT ανάλυση μέσω του αντικειμένου [fftin~], το οποίο έχει τρεις εξόδους: α) Πραγματική τιμή β) Φανταστική τιμή γ) Δείκτης bin. Δηλαδή λαμβάνουμε σαν έξοδο τις πραγματικές και φανταστικές τιμές του κάθε bin του κάθε frame του εισερχόμενου ήχου σε πραγματικό χρόνο.



Σχήμα 23: Αντικείμενο [fftin~]

Η αντίστοιχη έξοδος του patch αντιπροσωπεύεται από το αντικείμενο [ffftout~], το οποίο δέχεται πραγματικές και φανταστικές τιμές ως δύο ξεχωριστές εισόδους, και πραγματοποιεί τον αντίστροφο STFT, παράγοντας έτσι τη κυματομορφή ήχου που

αντιστοιχεί σε αυτές τις τιμές ανάλυσης. Το σχήμα του παραθύρου μπορεί να επιλεγεί ως όρισμα στο αντικείμενο [fftin~] και [fftout~]. Η προεπιλογή είναι παράθυρο Hanning.



Σχήμα 24: Αντικείμενο [fftout~]

Για τη μετατροπή των πραγματικών και φανταστικών τιμών σε ένταση και φάση υπάρχει το αντικείμενο [cartopol~] (Cartesian to polar coordinates, καρτεσιανές σε πολικές συντεταγμένες) το οποίο μετατρέπει τις τιμές με τον τρόπο που αναλύσαμε στο πρώτο κεφάλαιο, σχέσεις (1.3),(1.4). Συνεπώς το αντικείμενο αυτό έχει δύο εισόδους και δύο εξόδους.



Σχήμα 25: Αντικείμενο [cartopol~]

Το αντικείμενο [framedelta~] υπολογίζει την απόκλιση της φάσης (διαφορά) ίδιων bins διαδοχικών frames, σχέση (2.1).



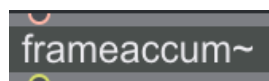
Σχήμα 26: Αντικείμενο [framedelta~]

Το αντικείμενο [phasewrap~] περιορίζει τις τιμές φάσης στο εύρος $-\pi$ έως π μέσω της διαδικασίας phase wrapping.



Σχήμα 27: Αντικείμενο [phasewrap~]

Επίσης υπάρχει αντικείμενο το οποίο υπολογίζει την τρέχουσα φάση κατά τη διαδικασία της ανασύνθεσης, το [frameaccum~], σχέση (2.2).



Σχήμα 28: Αντικείμενο [frameaccum~]

Αντίστοιχα υπάρχει αντικείμενο [poltocar~], το οποίο δέχεται τιμές έντασης-φάσης και εξάγει πραγματικές και φανταστικές τιμές, συνήθως για να τροφοδοτήσει το [fftout~].



Σχήμα 29: Αντικείμενο [poltocar~]

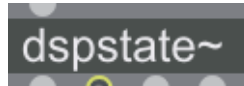
Υπάρχουν και κάποια αντικείμενα με σκοπό να παρέχουν πληροφορίες του συστήματος οι οποίες είναι απαραίτητες για τον υπολογισμό χρήσιμων τιμών. Το αντικείμενο [fftinfo~] παρέχει πληροφορίες: (α) του μεγέθους του FFT (β) του μεγέθους του frame (δηλαδή του

πλήθους bins που χρησιμοποιούνται), το οποίο ισούται και με το signal vector του patch (γ) hop size του FFT.



Σχήμα 30: Αντικείμενο [fftinfo~]

Το αντικείμενο [dspstate~] παρέχει πληροφορίες: (α) κατάστασης ηχητικού σήματος (on/off) (β) συχνότητα δειγματοληψίας.



Σχήμα 31: Αντικείμενο [dspstate~]

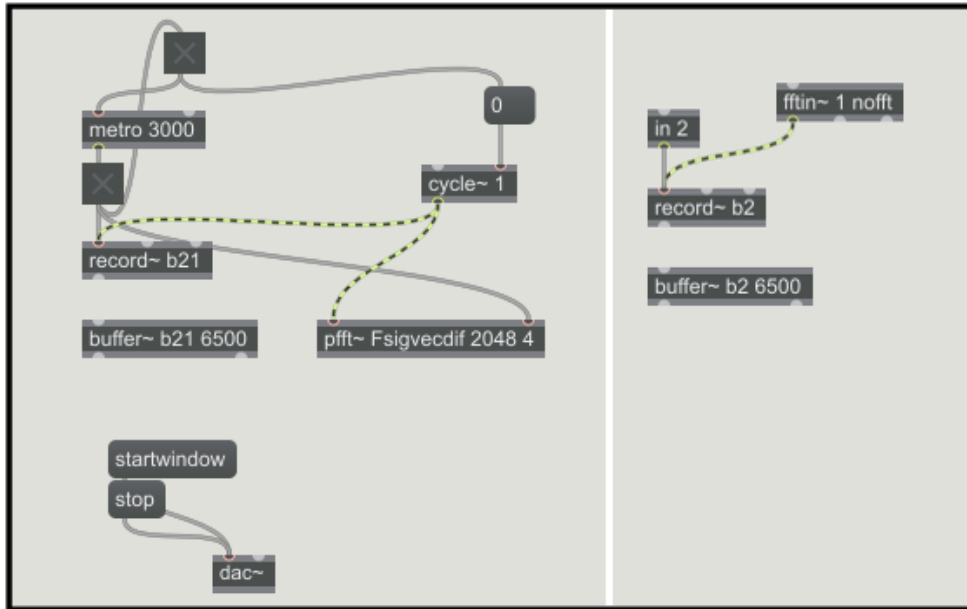
3.2 Σχέση ταχύτητας ροής ηχητικού σήματος με παράγοντα αλληλοεπικάλυψης μέσα στο αντικείμενο pfft~

Μία βασική διαφορά μέσα στο [pfft~] patch είναι ότι ο χρόνος κυλάει με διαφορετικό ρυθμό, αναλόγως τον συντελεστή αλληλοεπικάλυψης, από ότι έξω από αυτό. Δηλαδή για συντελεστή αλληλοεπικάλυψης 2, ο χρόνος θα κυλάει 2 φορές πιο γρήγορα μέσα στο [pfft~]. Το [pfft~] όμως εκτελεί μία εκδοχή του FFT, τον πραγματικό FFT αντί για τον κλασικό complex (μιγαδικό) FFT. Αυτό το επιτυγχάνει διότι τα εισερχόμενα σήματα προς ανάλυση έχουν μόνο πραγματικό μέρος. Ο πραγματικός FFT αναδιατάζει τις πραγματικές τιμές που εισέρχονται σε αυτόν σε πραγματικά και φανταστικά μέρη ενός μιγαδικού FFT με μισό μέγεθος του πραγματικού FFT. Αυτό έχει ως αποτέλεσμα την ανάλυση μόνο του πρώτου μισού του φάσματος ($0-f_s/2$), το οποίο είναι και αυτό που χρησιμοποιούμε, μέσω ενός υπολογιστικά ελαφρύτερου τρόπου. Οπότε και το μέγεθος του signal vector είναι μισό του μεγέθους του FFT.

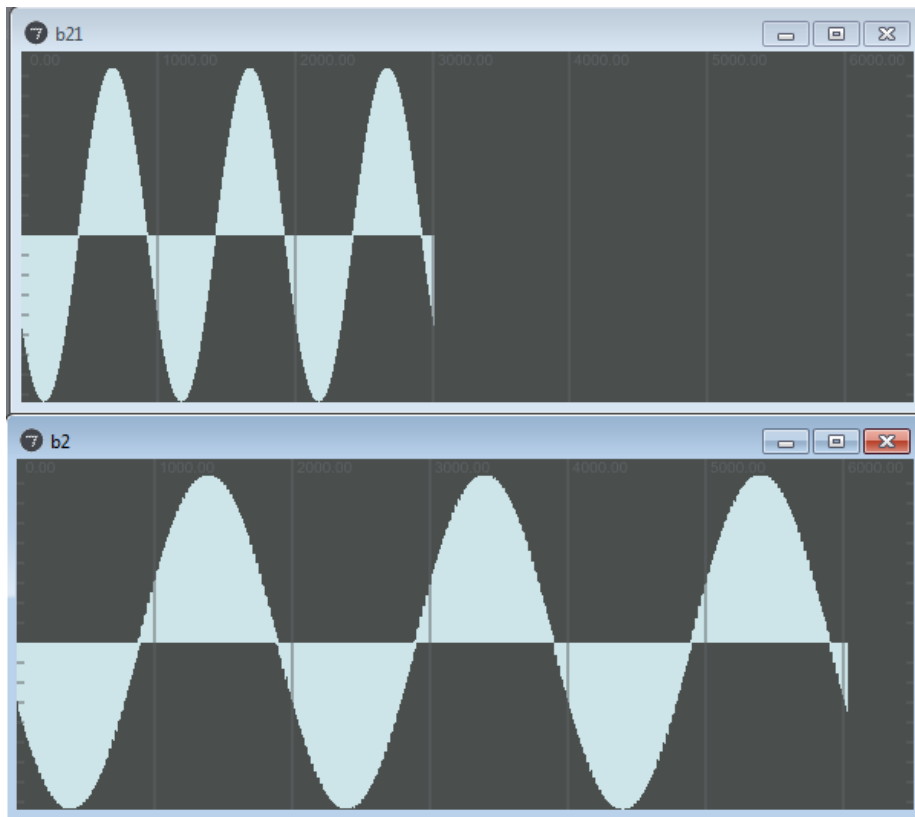
Για να υπολογίσουμε τον συντελεστή αλληλοεπικάλυψης του μιγαδικού FFT που εκτελείται, πρέπει στη πραγματικότητα να διαιρέσουμε το μισό μέγεθος του FFT (μέγεθος frame) με το hop size. Δηλαδή για ορισμό παράγοντα αλληλοεπικάλυψης 4 στο [pfft~] στη πραγματικότητα θα προκύπτει παράγοντας 2, και ο χρόνος θα τρέχει 2 φορές πιο γρήγορα μέσα στο [pfft~] από ότι έξω από αυτό. Αυτό θέλει προσοχή γιατί κατά την ανάγνωση των frames, ένας μηχανισμός ανάγνωσης μέσα στο [pfft~] θέλει τη μισή συχνότητα για να αναπαράγει τον ήχο στην αρχική του διάρκεια από ότι θέλει ο ίδιος μηχανισμός όταν δουλεύει έξω από το [pfft~], για παράγοντα αλληλοεπικάλυψης που έχει οριστεί 4 (ο πραγματικός θα είναι 2). Στην ίδια περίπτωση εάν δηλώσουμε παράγοντα αλληλοεπικάλυψης 8, ο "πραγματικός" θα είναι 4 και αν μηχανισμός ανάγνωσης δουλεύει μέσα στο [pfft~] θα πρέπει να τρέχει στο 1/4 της θεωρητικά υπολογισμένης συχνότητας για να αναπαράγει τον ήχο στην κανονική του διάρκεια, από ότι αν δουλεύει έξω από το [pfft~], που τότε δεν θα επηρεάζεται από αλλαγές στον παράγοντα αλληλοεπικάλυψης.

Για παράδειγμα, φτιάξαμε ένα patch το οποίο στέλνει ένα ημιτονικό σήμα με συχνότητα 1Hz σε ένα αντικείμενο [pfft~] (μέγεθος FFT: 2048, παράγοντας αλληλοεπικάλυψης: 4),

χωρίς να υπόκειται σε ανάλυση FFT, δηλώνοντας “noffft” στο αντικείμενο [fftin~] του [pfft~]. Μέσα στο [pfft~] υπάρχει ένας buffer με όνομα b2 και έξω από αυτό (στο κεντρικό patch) ένας με όνομα “b21”, και οι δύο χωρητικότητας 6500 ms. Ύστερα μέσω του αντικειμένου [metro] γράψαμε 3 δευτερόλεπτα του σήματος αυτού στους δύο buffers. Το αποτέλεσμα ήταν στον buffer “b21” να γραφτούν 3 δευτερόλεπτα του σήματος και στον buffer “b2” να γραφτούν 6 δευτερόλεπτα του σήματος, και τα δύο όμως να έχουν ολοκληρώσει τον ίδιο αριθμό επαναλήψεων, τρεις, που είναι το φυσιολογικό.



Σχήμα 32: Αριστερά κεντρικό patch παραδείγματος, δεξιά μέσα στο αντικείμενο [pfft~]

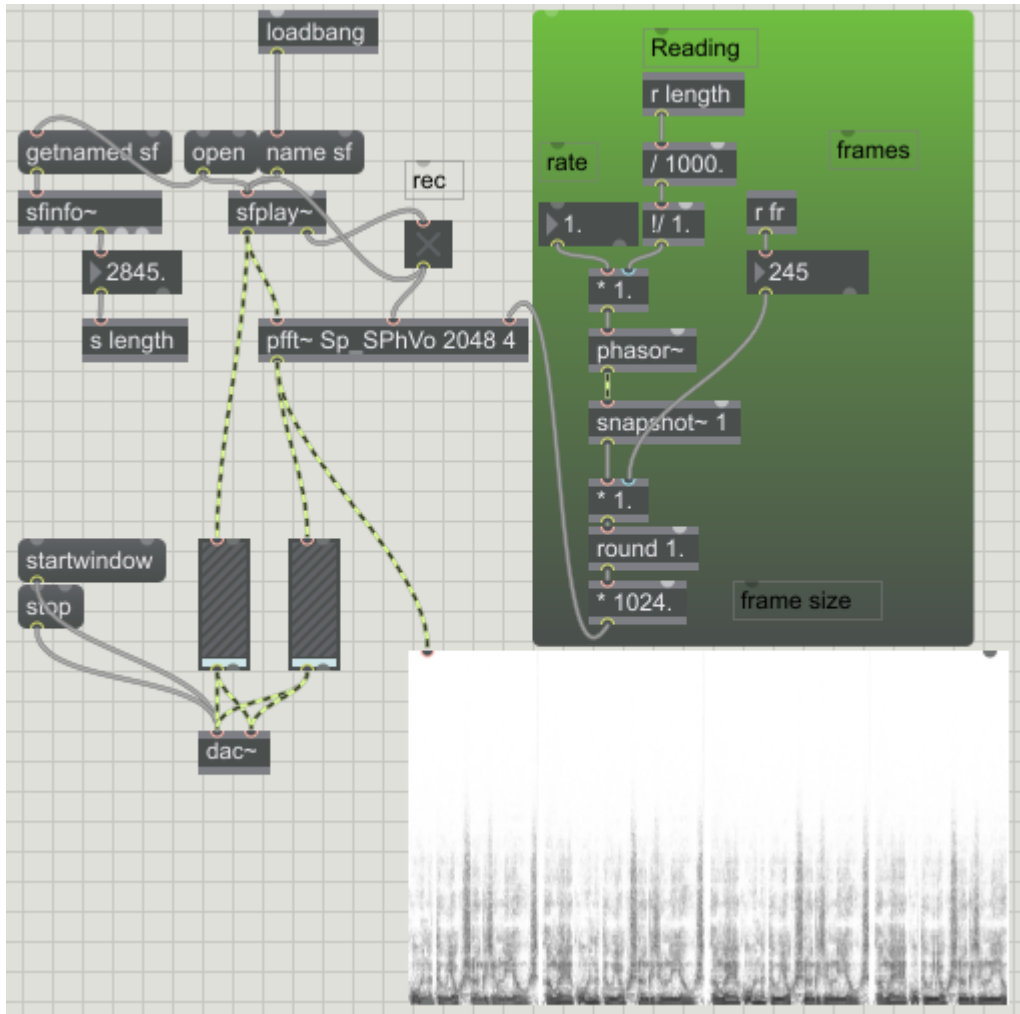


Σχήμα 33: Πάνω *buffer* κεντρικού *patch* παραδείγματος, κάτω *buffer* ο οποίος γράφτηκε μέσα στο αντικείμενο *pf~*. Ο κάθετος άξονας αντιπροσωπεύει ένταση και ο οριζόντιος χρόνο σε *ms*

Συνεπώς φαίνεται ότι ο *buffer* “b2” γραφόταν με διπλάσια συχνότητα από ότι ο “b21” που ήταν εκτός του [pf~], για αυτό και χρειάστηκε τα διπλάσια δείγματα για να καταγράψει το ίδιο σήμα.

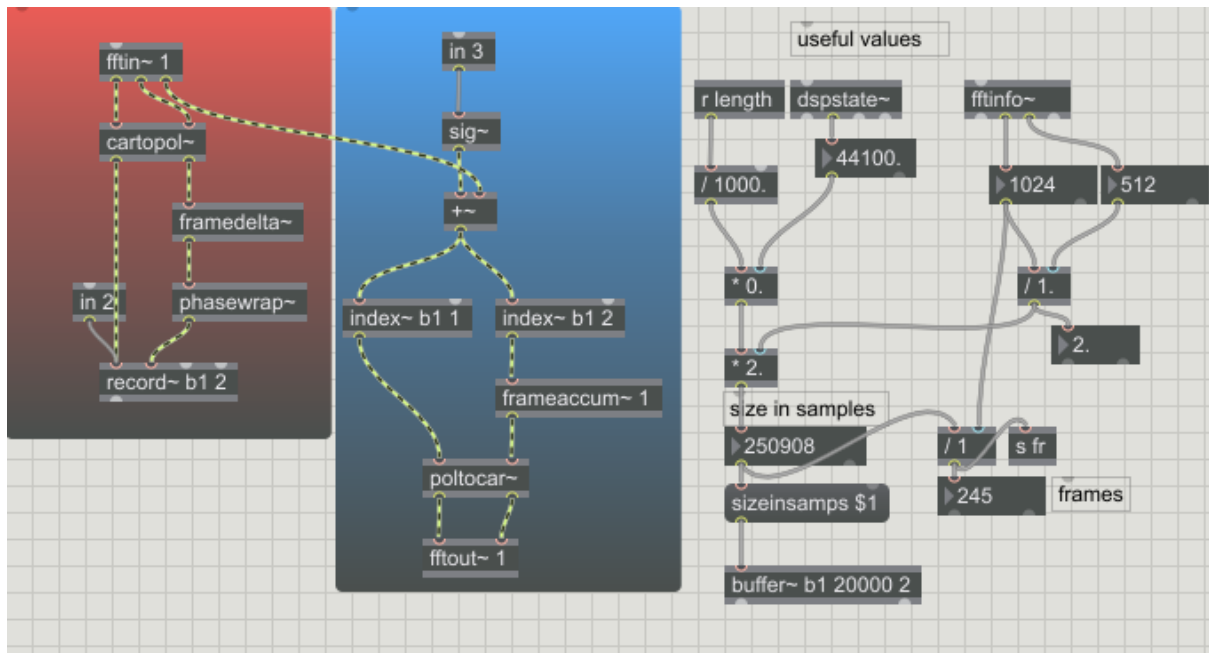
3.3 Εφαρμογή απλού Phase Vocoder στο προγραμματιστικό περιβάλλον MaxMSP

Για τη δημιουργία ενός απλού Phase Vocoder, αρκεί να αναλύσουμε φασματικά τον ήχο, να αποθηκεύσουμε αυτές τις τιμές ανάλυσης σε δύο *buffers* (τιμές έντασης-φάσης) και ύστερα να δημιουργήσουμε έναν μηχανισμό ανάγνωσης που θα δείχνει ποιο *frame* πρέπει να ανασυντεθεί κάθε χρονική στιγμή. Ακολουθεί το βασικό *patch* της εφαρμογής.



Σχήμα 34: Βασικός αλγόριθμος απλού Phase Vocoder

Ο επιλεγμένος ήχος αναπαράγεται μέσω του αντικειμένου [sfplay~]. Εφόσον επιλέξουμε αρχείο ήχου, μπορούμε να καταγράψουμε τις τιμές ανάλυσής του πατώντας τον διακόπτη με ένδειξη rec. Το αντικείμενο [sfinfo~] μας παρέχει τη διάρκεια του αρχείου που έχουμε επιλέξει σε ms. Όπως φαίνεται το αντικείμενο [pfft~] έχει μία είσοδο σήματος και δύο εισόδους μηνυμάτων.



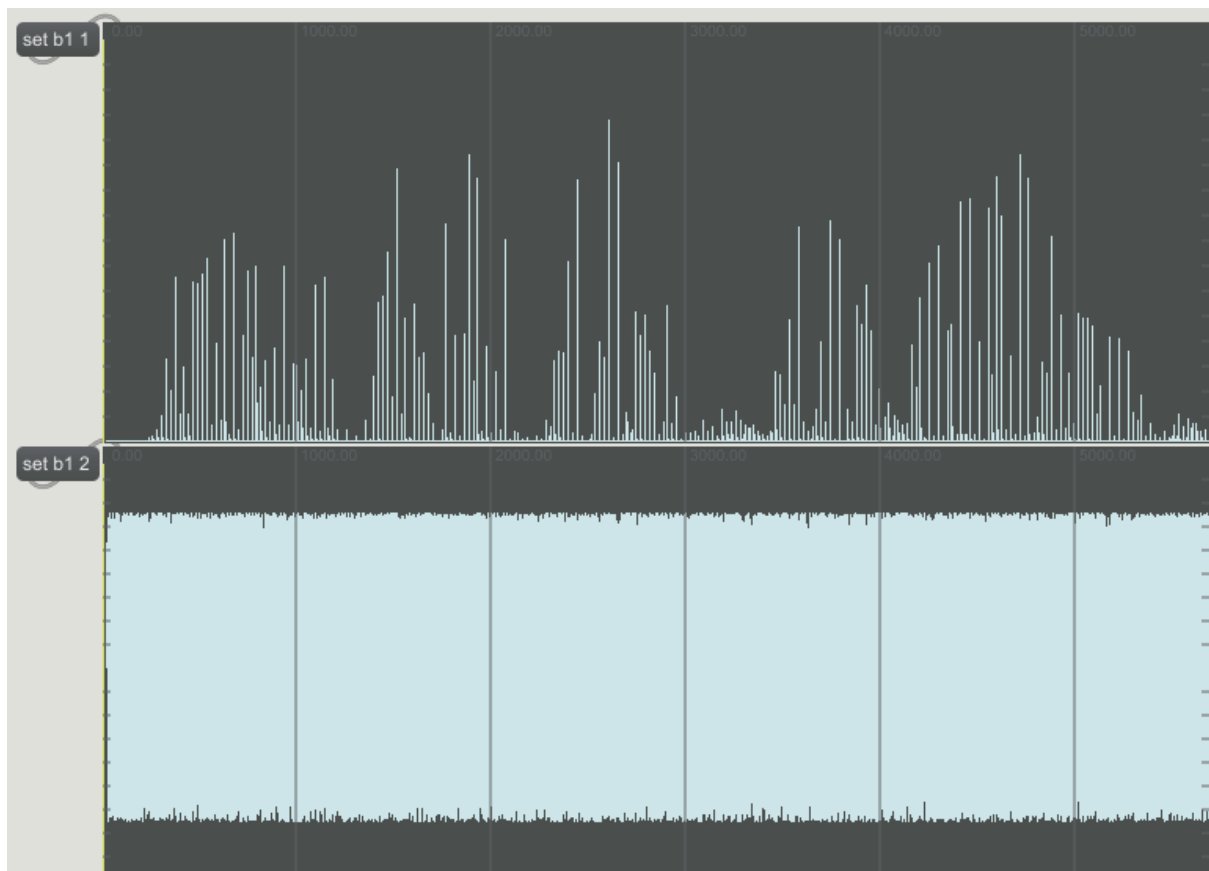
Σχήμα 35: Μέσα στο pfft~ του απλού Phase Vocoder

Το σήμα εισέρχεται στη κόκκινη περιοχή μέσω του αντικειμένου [fftin~]. Ακολουθεί την γνωστή διαδικασία, καρτεσιανές σε πολικές συντεταγμένες, υπολογισμός απόκλισης φάσης και τύλιγμα φάσης σε εύρος $-\pi$ έως π . Ύστερα οι τιμές έντασης και φάσης γράφονται μέσω του αντικειμένου [record~] σε δύο ξεχωριστά κανάλια ενός buffer (στην ουσία σε δύο buffers). Το αντικείμενο [record~] δέχεται μήνυμα έναρξης της εγγραφής από τον διακόπτη που είναι συνδεδεμένος στην δεύτερη είσοδο του [pfft~] (το αντικείμενο in 2 είναι η έξοδος της εισόδου αυτής). Ο ορισμός του buffer γίνεται στη δεξιά πλευρά του patch. Η διάρκεια του αρχείου στέλνεται από το κεντρικό patch σε ms και διαιρείται με 1000 για να μετατραπεί σε sec. Ύστερα πολλαπλασιάζεται με τη συχνότητα δειγματοληψίας για να βρεθεί η συνολική χωρητικότητα δειγμάτων που θα πρέπει να έχει ο buffer για να χωρέσει τον ήχο ακριβώς. Υπάρχουν δυο τρόποι ώστε να βρεθούν τα δείγματα ανάλυσης που απαιτούνται:

(α) εάν πάρουμε ως σύμβαση ότι ο μισός (αναδιαταγμένος μιγαδικός) FFT του πραγματικού FFT παράγει x bins για x δείγματα ανάλυσης τότε θα παράγει 44100 δείγματα για 1 ανάλυση ενός δευτερολέπτου ήχου με συχνότητα δειγματοληψίας 44100, ύστερα θα πολλαπλασιαστεί με τον παράγοντα αλληλοεπικάλυψης του μεγέθους frame, το οποίο είναι ίσο με το μέγεθος του αναδιαταγμένου μιγαδικού FFT, διά το hop size.

(β) εάν θεωρήσουμε ότι ο πραγματικός FFT σαν συνολικό αποτέλεσμα παράγει τα μισά bins σε σχέση με το πλήθος δειγμάτων που αναλύει θα έπρεπε αργότερα να πολλαπλασιάσουμε με τον κανονικό παράγοντα αλληλοεπικάλυψης, μέγεθος FFT διά hop size. Το αποτέλεσμα θα ήταν το ίδιο.

Μετά διαιρείται το μέγεθος frame διά το hop size για να βρεθεί ο διορθωμένος παράγοντας αλληλοεπικάλυψης, ο οποίος δηλώνει πόσες φορές πιο γρήγορα τρέχει το [rfft~] σε σχέση με το κεντρικό patch, όπως είπαμε και νωρίτερα. Άρα πολλαπλασιάζουμε τα δείγματα του buffer που έχουμε υπολογίσει με τον παράγοντα αυτό και ύστερα δηλώνουμε το μέγεθος του buffer σε δείγματα με το μήνυμα “sizeinsamps”. Τα ορίσματα του buffer είναι το όνομα, η προεπιλεγμένη χωρητικότητά του σε ms και το πλήθος καναλιών. Εάν διαιρέσουμε το σύνολο των δειγμάτων αυτών με το μέγεθος του frame της ανάλυσής μας, μπορούμε να βρούμε τον αριθμό των frames από τα οποία αποτελείται η ανάλυση του ήχου.

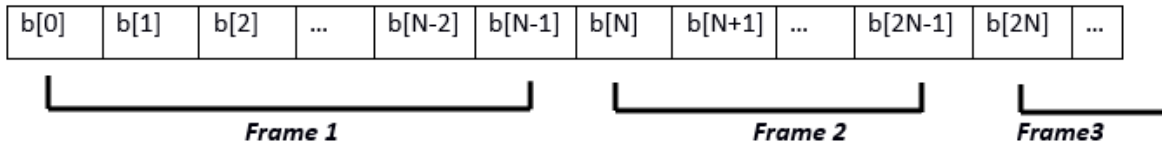


Σχήμα 36: Πάνω φαίνεται ο buffer των τιμών έντασης της ανάλυσης και κάτω ο buffer των τιμών απόκλισης φάσης. Ο οριζόντιος άξονας αντιπροσωπεύει δείγματα (οι τιμές που αναγράφονται όμως είναι χρόνος σε ms που θα χρειαστεί για να αναπαραχθούν σε παρούσα σε συχνότητα δειγματοληψίας). Ο κάθετος άξονας αντιπροσωπεύει τιμές εντάσεων για τον πάνω buffer και τιμές απόκλισης φάσης για τον κάτω.

Όπως έχει προαναφερθεί, οι τιμές φάσεις έχουν περιοριστεί στο εύρος $-\pi$ έως π (phase wrapping). Η μέγιστη τιμή έντασης εξαρτάται από το μέγεθος του FFT αλλά και από το σχήμα της συνάρτησης παραθύρου. Για hanning window (το οποίο χρησιμοποιείται) η μέγιστη τιμή έντασης που μπορεί να προκύψει ισούται με $0,25 \times$ μέγεθος FFT. Οπότε στη συγκεκριμένη εφαρμογή θα είναι $0,25 \times 2048 = 512$. Επειδή οι τιμές που προκύψαν δεν είναι τόσα μεγάλες το εύρος του buffer έντασης ορίστηκε 64 για την καλύτερη απεικόνιση

των τιμών (0-64). Το εύρος του buffer τιμών απόκλισης φάσεων είναι $-\pi-1$ έως $\pi+1$ (Σχήμα 25).

Σε αυτό το σημείο είναι χρήσιμο να αναφερθεί ότι τιμές στους buffers γράφονται ανά πακέτα frames.



Σχήμα 37: Τρόπος εγγραφής τιμών FFT ανάλυσης σε buffer. N ισούται με μέγεθος του frame (πλήθος bins)

Ο μηχανισμός ανάγνωσης φαίνεται στο πράσινο μέρος του Σχήματος 34. Αρχικά η διάρκεια του αρχείου μετατρέπεται από ms σε sec (/ 1000.) και για να βρεθεί η συχνότητα που αντιστοιχεί για περίοδο ίση με τη χρονική διάρκεια του ήχου, διαιρείται από το 1. Πολλαπλασιάζοντας αυτή τη τιμή μπορούμε να αλλάξουμε τη χρονική διάρκεια του ανασυντιθέμενου ήχου χωρίς να αλλάξει η τονικότητά του (χρονική συμπίεση/διαστολή). Δηλαδή για τιμή 1 ο ήχος θα αναπαράγεται στον αρχικό του χρόνου για τιμή 0.5 στον διπλάσιο και για τιμή 2 στο μισό. Η τιμή αυτή είναι η συχνότητα ενός φάσρα [phasor~]. Φάσρα είναι μία γεννήτρια που παράγει γραμμικά τιμές 0-1 ανά μία περίοδο. Αφού μετατρέψουμε τις τιμές σε control rate, περνώντας από δειγματοληψία την έξοδο του φάσρα ανά 1 ms, πολλαπλασιάζουμε επί το πλήθος των frames του αναλυμένου ήχου, έτσι ώστε ο φάσρα πλέον να έχει εύρος τιμών ίσο με το πλήθος των frames. Στη συνέχεια γίνεται μία στρογγυλοποίηση στη μονάδα, έτσι ώστε να προκύπτουν ακέραιες τιμές, δηλαδή τιμές που θα αντιστοιχούν ακριβώς σε ένα frame (στην αρχή του frame, στο πρώτο bin του frame). Έπειτα το αποτέλεσμα πολλαπλασιάζεται με το μέγεθος του frame, έτσι ώστε ο φάσρα να δείχνει πάντα το πρώτο bin του κάθε frame που ορίζει για ανασύνθεση, κάνοντας βήματα ίσα με το μέγεθος του frame. Δηλαδή να δείχνει τιμές $b[(m-1) \times N]$ στους buffers έντασης και φάσης, όπου m είναι το τρέχων frame και N το πλήθος των bins της ανάλυσης. Ύστερα μέσα στο [pfft~] οι τιμές μετατρέπονται σε σήμα, προστίθονται με τον δείκτη του FFT και ορίζουν ποια δείγματα των buffers θα αναγνωστούν μέσω των αντικειμένων [index~].

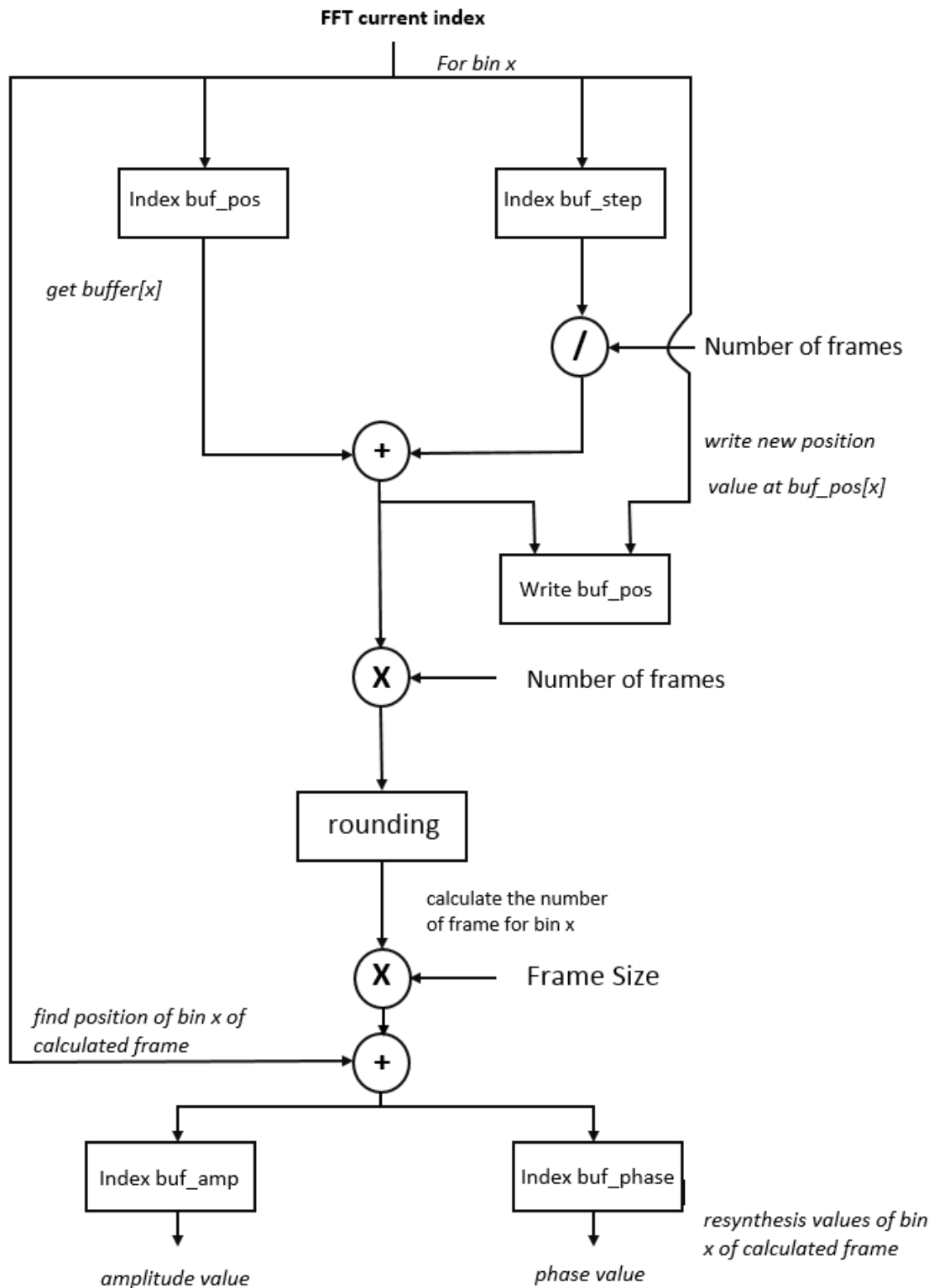
Ο δείκτης του FFT, ορίζει συνέχεια ποιο bin αναλύεται κάθε στιγμή. Δείχνει όλες τις τιμές των bins σε σειρά σε διάρκεια ενός signal vector. Για αυτό το λόγο χρησιμοποιείται και κατά την ανασύνθεση. Αυτό που γίνεται λοιπόν, είναι ότι ορίζεται η πρώτη τιμή του frame που ανασυντίθεται κάθε στιγμή από τον μηχανισμό ανάγνωσης και ύστερα σε χρόνο ενός signal vector (δηλαδή πολύ γρήγορα, σίγουρα πριν αλλάξει το frame που διαβάζεται εκείνη τη στιγμή) έχει υποδείξει στο αντικείμενο [index~] να διαβάσει όλες τις τιμές αυτού του frame και να εκτελέσει τη γνωστή διαδικασία υπολογισμού τρέχουσας φάσης, μετατροπής πολικών σε καρτεσιανών συντεταγμένων και ανασύνθεση του frame σε κομμάτι κυματομορφής μέσω του αντιστρόφου FFT, όλα σε χρόνο ενός signal vector. Οπότε μέσω αυτής της διαδικασίας μπορεί να οριστεί ποιο frame θέλουμε να συνθεθεί κάθε στιγμή

καθώς και με τι ταχύτητα διαβάζονται τα frames αυτά. Η αλλαγή της ταχύτητας ανάγνωσης των frames οδηγεί και στη κλασική επεξεργασία του time stretching.

3.4 Χρονική Συμπίεση/Διαστολή ήχου μέσω φασματικής επεξεργασίας

Η γενική ιδέα της εφαρμογής είναι η δυνατότητα διαφορετικής χρονικής συμπίεσης/διαστολής ήχου για κάθε bin της FFT ανάλυσης ανάλογα με τιμές που θα ορίζει ο χρήστης. Αναλύσαμε πως μπορούμε να εφαρμόσουμε απλό time stretching σε ήχο μέσω φασματικής επεξεργασίας. Τώρα αυτό που θέλουμε είναι κάθε bin να έχει τη δικιά του ταχύτητα ανάγνωσης και την δικιά του ανεξάρτητη θέση στο χρόνο.

Αυτό μπορεί να επιτευχθεί με τη χρήση δύο buffers, όπου ο ένας θα περιέχει τιμές θέσης του κάθε bin και ο άλλος θα περιέχει το βήμα του κάθε bin για κάθε νέο κομμάτι που ανασυντίθεται. Συνεπώς, ο καθένας θα έχει μέγεθος ίσο με τον αριθμό bins της ανάλυσης.



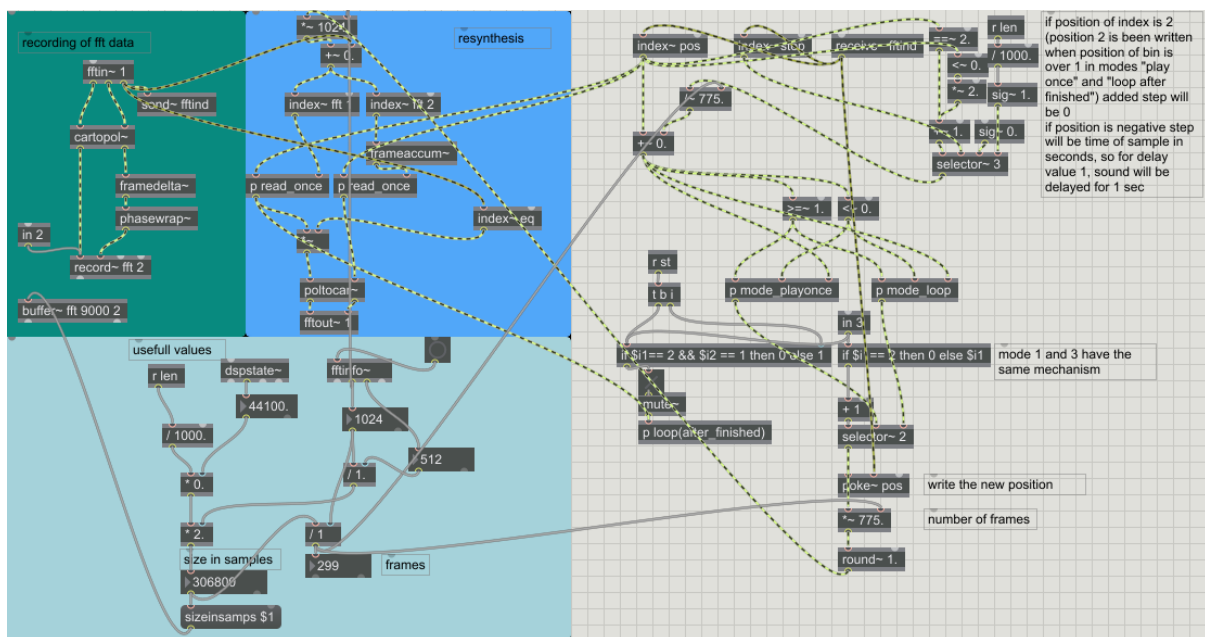
Σχήμα 38: Διάγραμμα ροής αλγορίθμου ανεξάρτητου time stretching για κάθε bin

Τα βήματα της διαδικασίας είναι τα εξής:

- 1) Ανάγνωση τιμής θέσης και βήματος για το τρέχων bin από τους αντίστοιχους buffers.

- 2) Διαίρεση τιμής βήματος διά πλήθος frames του αναλυμένου ήχου, έτσι ώστε “βήμα 1” να σημαίνει αναπαραγωγή του ήχου στην αρχική του χρονική διάρκεια.
- 3) Πρόσθεση τιμής θέσης του bin με το αποτέλεσμα της διαίρεσης της τιμής βήματος διά το πλήθος frames. Αυτή είναι η νέα τιμή θέσης του bin.
- 4) Πολλαπλασιασμός νέας τιμής θέσης επί πλήθος των frames, με σκοπό την εύρεση του frame που αναλογεί στην αντίστοιχη θέση.
- 5) Στρογγυλοποίηση τιμής ώστε να προκύψει η αρχική τιμή του frame που υπολογίστηκε.
- 6) Πρόσθεση του τρέχων bin στη τιμή για να βρεθεί η θέση του συγκεκριμένου bin στο υπολογισμένο frame.
- 7) Ανάγνωση τιμών έντασης και φάσης του bin που υπολογίστηκε.
- 8) Καταχώρηση νέας τιμής θέσης του τρέχων bin στο buffer τιμών θέσης.

Αυτή η διαδικασία επαναλαμβάνεται τόσες φορές όσες το πλήθος των bins της ανάλυσης (μέγεθος frame), με αποτέλεσμα τη σύνθεση ενός frame που αποτελείται από τιμές ανεξάρτητων frames για κάθε bin του.



Σχήμα 39: Μέσα στο [rfft~] της εφαρμογής

Όπως φαίνεται το στάδιο της ανάλυσης είναι το ίδιο με την εφαρμογή του απλού Phase Vocoder. Όλη η τροποποίηση του ήχου γίνεται κατά την διαδικασία της ανάγνωσης.

3.4.1 Οι τρεις καταστάσεις λειτουργίας της εφαρμογής

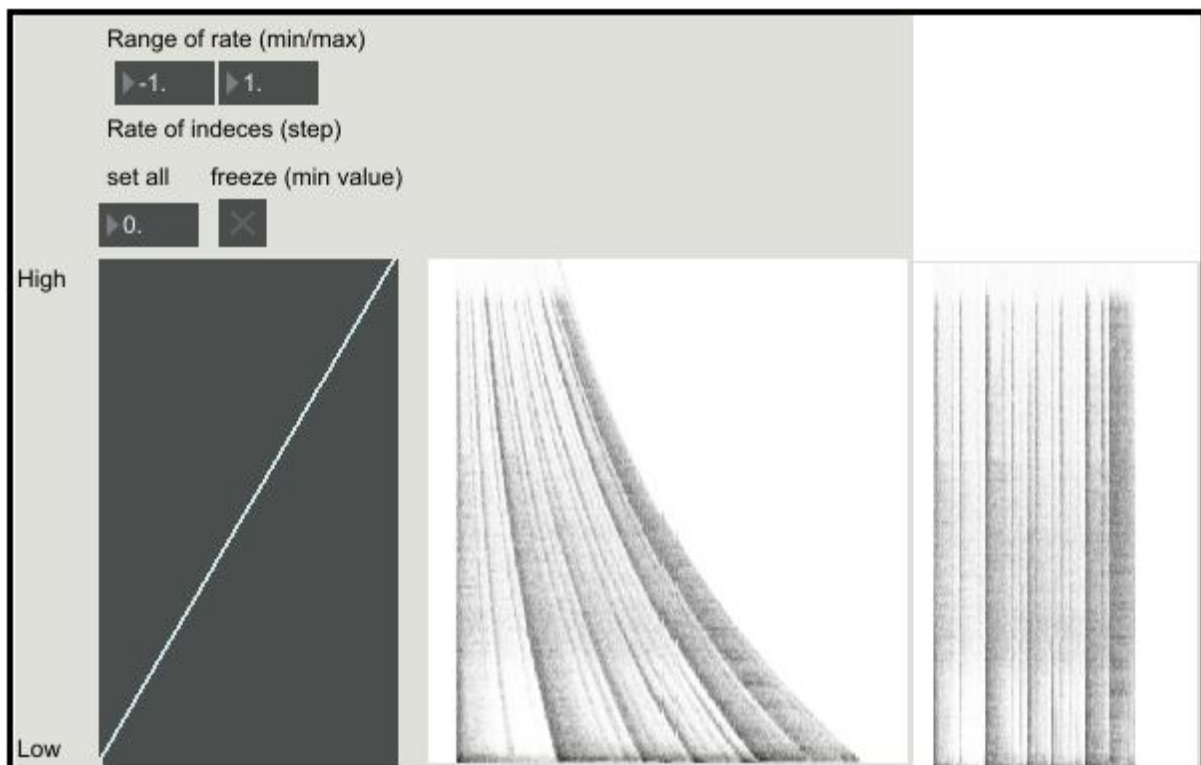
α. Μία αναπαραγωγή (Play once)

Το εύρος των τιμών θέσης των bins είναι 0 – 1. Δηλαδή 0 αντιστοιχεί στο πρώτο frame και 1 στο τελευταίο. Στο υπο-patch mode_playonce υπάρχουν διαφορετικές εξερχόμενες τιμές θέσης για συγκεκριμένα εύρη εισερχόμενων τιμών.

α) Για τιμές <1 θα εξέρχεται το εισερχόμενο σήμα

γ) για τιμές >= 1 θα εξέρχεται σήμα με τιμή 2

Η τιμή 2 δεν αντιστοιχεί σε κάποια τιμή frame της ανάλυσης. Χρησιμοποιείται για να δηλώσει ότι το συγκεκριμένο bin έχει διαβάσει μια φορά τα frames του ήχου, ή αλλιώς ότι έχει υπερβεί το τελευταίο frame της ανάλυσης του ήχου. Χρησιμοποιείται κατά την ανασύνθεση του ήχου για να δηλώσει ότι αυτό το bin έχει "τερματίσει" την ανάγνωση οπότε εξάγονται τιμές έντασης και φάσης 0 για αυτό κατά την ανασύνθεση. Επίσης για τιμή 2 το βήμα του bin είναι πάντα 0, έτσι μένει σταθερό στο χρόνο. Το αποτέλεσμα είναι να μηδενίζονται όλα τα bins που υπερβήκανε το τελευταίο frame, μέχρι να μηδενιστεί και το τελευταίο και η αναπαραγωγή του ήχου να τελειώσει με όλα τα bins παγωμένα στη θέση 2.



Σχήμα 40: Αριστερά ορίζεται η ταχύτητα ανάγνωσης για κάθε bin μέσω του αντικειμένου multislider. Πάνω είναι οι υψηλές συχνότητες και κάτω οι χαμηλές. Το εύρος τιμών των ταχυτήτων ανάγνωσης είναι -1 έως 1. Στη μέση φαίνεται το επεξεργασμένο φάσμα του ήχου. Χρησιμοποιήθηκε το ηχητικό δείγμα drumloop από τα δείγματα της Max. Δεξιά φαίνεται το φάσμα του αρχικού αρχείου.

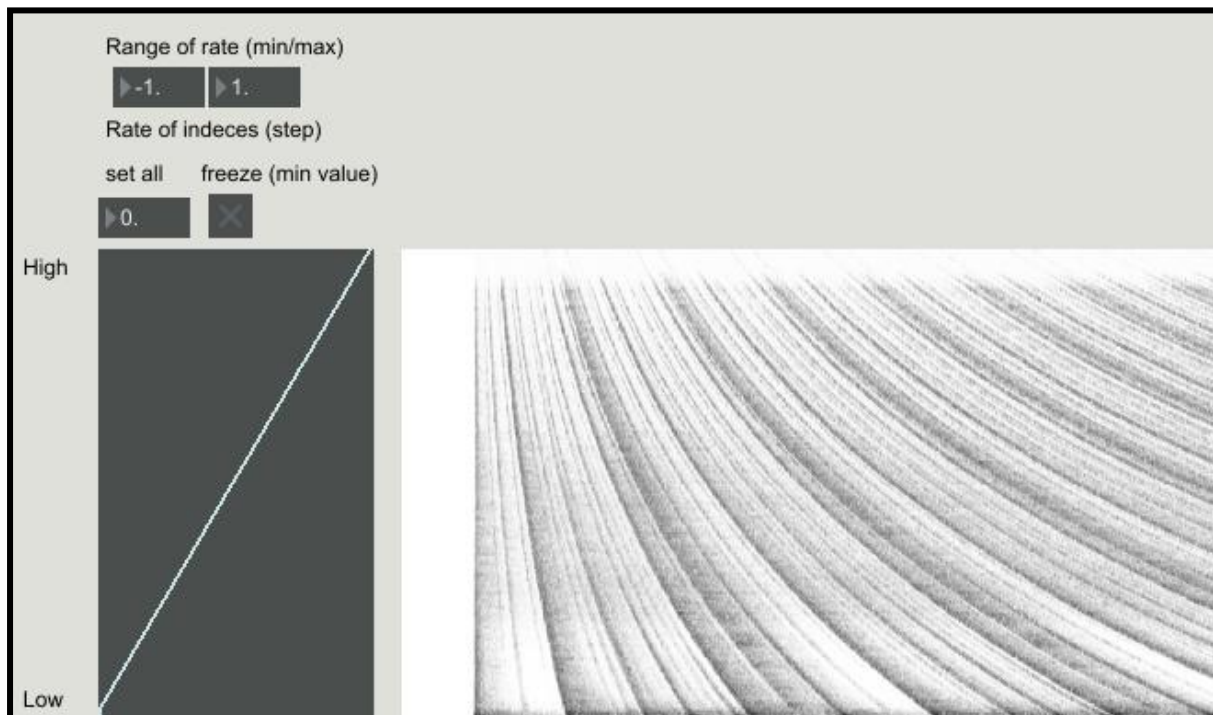
β. Επανανάγνωση των frames (Loop)

Κατά τη λειτουργία loop της εφαρμογής, κάθε bin αφού ξεπεράσει το τελευταίο frame της ανάλυσης, αρχίζει ξανά την ανάγνωση από την αρχή. Αυτό συμβαίνει στο υπο-patch mode_loop, στο οποίο διαχωρίζονται οι εξής περιπτώσεις:

α) Για τιμές 0-1 θα εξέρχεται το εισερχόμενο σήμα

β) για τιμές ≥ 1 θα εξέρχεται το εισερχόμενο σήμα -1

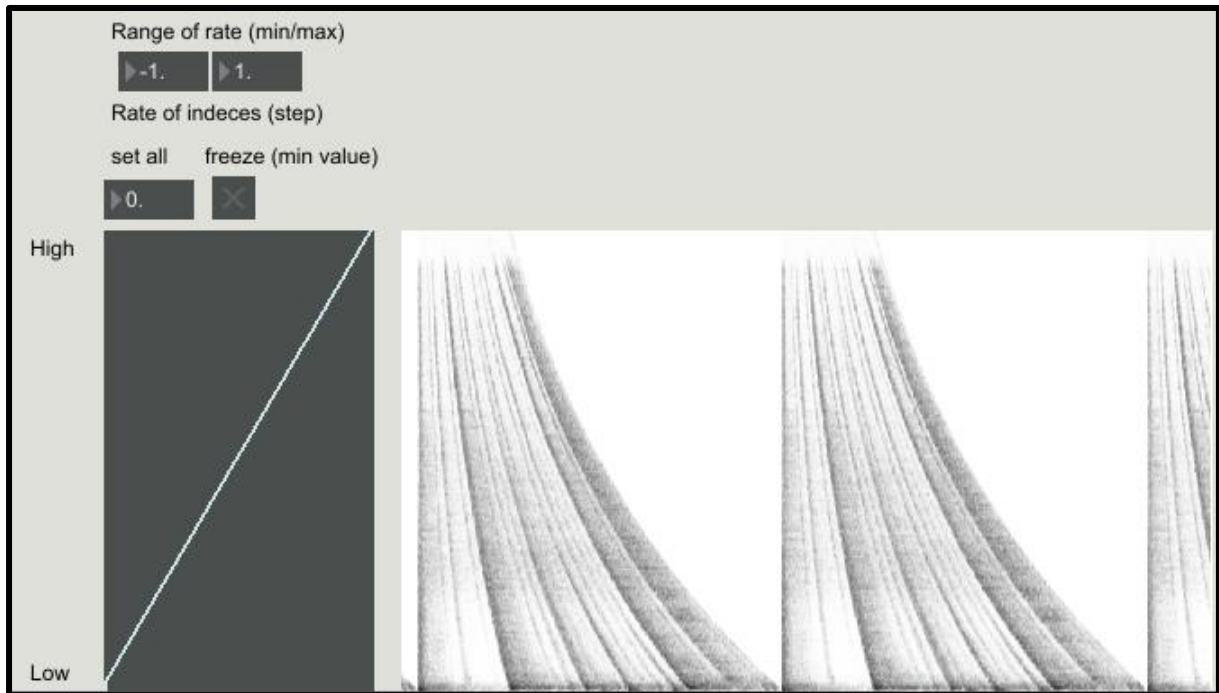
Για παράδειγμα, για βήμα bin 1 ανάλυσης ήχου 100 frames, η τιμή που θα προστίθεται σε κάθε frame ανασύνθεσης για το συγκεκριμένο bin θα είναι 0.01. Όταν το bin φτάσει στη θέση 1,01, η θέση που θα ανασυντίθεται και θα γράφεται στον buffer θέσης θα είναι 0,01, και έτσι το bin θα συνεχίζει την ανασύνθεση τιμών από τα αρχικά frames ανάλυσης του ήχου. Αυτό σημαίνει ότι το κάθε bin επαναλαμβάνεται ανεξάρτητα από το άλλο και ότι αυτή η διαδικασία συνεχίζεται έως ότου τη σταματήσει ο χρήστης. Δηλαδή ένα bin που θα αναπαράγεται με rate 3 (8 φορές πιο γρήγορα από το κανονικό) θα έχει κάνει το κύκλο των frames 64 φορές, μέχρι ένα bin με rate -3 (8 φορές πιο αργά από το κανονικό) κάνει 1 κύκλο.



Σχήμα 41: Φάσμα επεξεργασμένου ήχου με ίδιες ρυθμίσεις όπως στο Σχήμα 29 σε λειτουργία Loop

γ. Επανάληψη αναπαραγωγής (Loop after finished)

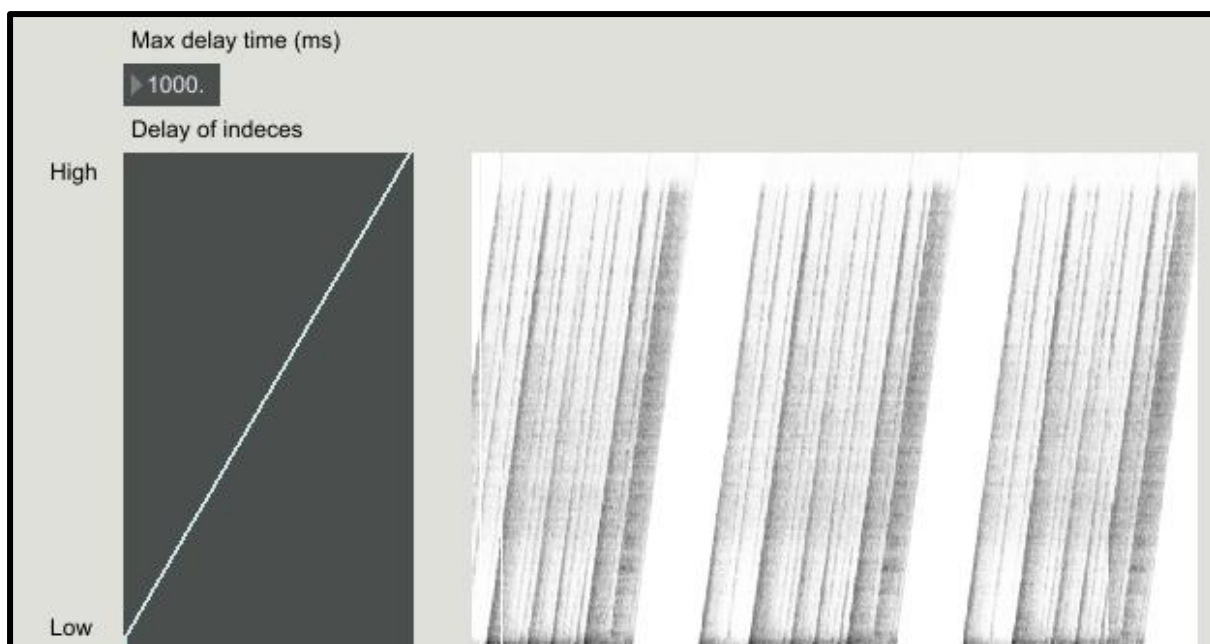
Η τρίτη λειτουργία επαναλαμβάνει τον επεξεργασμένο ήχο στην ίδια μορφή αφού τελειώσει. Μόλις και το τελευταίο bin αναπαράξει τον ήχο μία φορά, η αναπαραγωγή του ήχου ξεκινάει από την αρχή βάσει των ρυθμίσεων του χρήστη.



Σχήμα 42: Η ίδιες τιμές επεξεργασίας με τα δύο προηγούμενα σχήματα, στη τρίτη λειτουργία (επανάληψη αναπαραγωγής)

3.4.2 Φασματικό delay

Στο χρήστη δίνεται και η δυνατότητα να ορίσει κάποια αρχική καθυστέρηση στην έναρξη αναπαραγωγής του κάθε bin. Η καθυστέρηση ορίζεται σαν αρνητικές τιμές στο buffer θέσης των bins. Για αρνητικές τιμές θέσης το βήμα είναι πάντα ο χρόνος διάρκειας του αρχικός ήχου σε sec. Αυτό έχει ως αποτέλεσμα για τιμή delay 1 (θέση -1) η καθυστέρηση του συγκεκριμένου bin να είναι 1 sec. Κατά τη διαδικασία της ανασύνθεσης αρνητικές τιμές θέσης έχουν ως αποτέλεσμα το μηδενισμό έντασης και φάσης του bin.



Σχήμα 43: Ηχητικό δείγμα *drumloop* από τα δείγματα της *Max*, με γραμμικά αυξανόμενο *delay* από τις χαμηλές στις υψηλές συχνότητες, στη τρίτη λειτουργία (*Loop after finished*) της εφαρμογής. Εύρος *delay* 1 sec.

3.4.3 Φασματικό φίλτρο

Επιπλέον υπάρχει ένα φασματικό φίλτρο, το οποίο δίνει τη δυνατότητα ελέγχου έντασης των συχνοτικών στοιχείων του ήχου. Λέγοντας φασματικό φίλτρο εννοούμε ότι επιδράει στο φάσμα του ήχου μέσω συνέλιξης με τον ανασυντιθέμενο ήχο. Η συνέλιξη αντιστοιχεί σε πολλαπλασιασμό στο φασματικό πεδίο. Ο χρήστης ορίζει τιμές μέσω ενός [multislider], με πλήθος ρυθμιστικών όσο και το μέγεθος του frame. Αυτό στη συνέχεια γεμίζει έναν buffer ίδιου αριθμού θέσεων. Η κάθε τιμή αντιστοιχεί σε ένα bin. Οι τιμές του buffer καλούνται μέσω του αντικείμενου [index~], του οποίου οι τιμές ορίζονται από το δείκτη FFT του [rfft~]. Στη συνέχεια οι τιμές του buffer του φίλτρου πολλαπλασιάζονται με τις τιμές έντασης που προκύπτουν πριν την ανασύνθεση του ήχου. Εφόσον το τρέχων bin ορίζεται κοινά από το δείκτη του FFT, το κάθε bin της ανασύνθεσης πολλαπλασιάζεται με την αντίστοιχη τιμή του φασματικού φίλτρου. Με αυτό τον τρόπο επιτυγχάνεται έλεγχος της έντασης του κάθε bin.



Σχήμα 44: Χρήση φασματικού φίλτρου σε ηχητικό δείγμα *drumloop*, σε λειτουργία *loop* (after finished)

3.4.4 Λειτουργία freeze

Επιπλέον δίνεται στο χρήστη η λειτουργία freeze. Όταν ενεργοποιείται ο διακόπτης της λειτουργίας, η ελάχιστη τιμή του multislider ταχύτητας αναπαραγωγής αντικαταστίεται από την τιμή 0, με αποτέλεσμα τα bins που ορίζονται στις χαμηλότερες τιμές να παγώνουν, διαβάζοντας μόνιμα τις τιμές του ίδιου frame. Αυτό συνεπάγεται και ότι ο ήχος δεν τελειώνει ποτέ την αναπαραγωγή του. Είναι σημαντικό να αναφερθεί ότι οι τιμές βήματος ανάγνωσης αλλάζουν κατά τη διάρκεια αναπαραγωγής του ήχου αν ο χρήστης επέμβει σε αυτές (συγκεκριμένα με το που αφήσει το κλικ). Οπότε το frame που θα παγώσει θα είναι αυτό που αναπαράγεται από το συγκεκριμένο bin τη χρονική στιγμή που ο χρήστης θα του έχει ορίσει το ελάχιστο βήμα ανάγνωσης και αφήσει το κλικ.



Σχήμα 45 : Χρήση freeze για μεσαίο συχνοτικό εύρος. Χρησιμοποιήθηκε το ηχητικό δείγμα drumloop σε κατάσταση λειτουργίας εφαρμογής επανανάγνωσης

3.4.5 Έλεγχος παραμέτρων ανασύνθεσης μέσω διεπαφής

Ο χρήστης έχει τη δυνατότητα ορισμού της ταχύτητας ανάγνωσης του κάθε bin μέσω διεπαφής και συγκεκριμένα μέσω του αντικειμένου [multislider]. Το [multislider] είναι ένα αντικείμενο πολλαπλών ρυθμιστικών (sliders). Ο χρήστης έχει τη δυνατότητα ορισμού του εύρους τιμών των ρυθμιστικών αυτών. Το [multislider] πρέπει να έχει τόσο ρυθμιστικά όσο το μέγεθος του frame, όπως και ο buffer που ρυθμίζει. Οι τιμές του περνάνε από τη πράξη $y = 2^x$, όπου x είναι οι εξερχόμενες τιμές του [multislider] και y το αποτέλεσμα της πράξης. Αυτό γίνεται για περισσότερο ισοκαταναμημένο εύρος στον ορισμό των τιμών. Για παράδειγμα, με αυτή τη πράξη η τιμή 0 θα ισούται με 1 δηλαδή κανονική αναπαραγωγή του ήχου, η τιμή -1 με 0,5 (μισή ταχύτητα) και -2 με 0,25 (1/4 της ταχύτητας). Συνεπώς ο χρήστης έχει το ίδιο εύρος τιμών για να ορίσει ενδιάμεσες καταστάσεις μεταξύ της μισής και της ενός τετάρτου ταχύτητας (διαφορά 1), πράγμα που αλλιώς δε θα γινόταν (0.25-0.5, 0.5-1). Ακολουθούν κάποιες επεξεργασίες λίστας και ύστερα μέσω του αντικειμένου [poke~] οι τιμές αυτές γεμίζουν τον buffer βήματος. Ο έλεγχος των τιμών των buffers του delay και του φασματικού φίλτρου γίνεται με αντίστοιχο τρόπο. Η τροποποίηση των τιμών δεν είναι απαραίτητη για αυτούς τους ελεγκτές.

4. Παραδείγματα χρήσης της εφαρμογής

Η εφαρμογή έχει τρεις καταστάσεις λειτουργίας:

- 1.Play once
- 2.Loop
- 3.Loop after finished

Ο χρήστης μπορεί να διαλέξει μία από αυτές από το διακόπτη τριών θέσεων με τα ονόματα της κάθε μίας.



Σχήμα 46: Επιφάνεια εργασίας εφαρμογής

Αφού ενεργοποιηθεί η εφαρμογή πατώντας το κουμπί “startwindow”, πρέπει να επιλεχθεί το ηχητικό δείγμα πατώντας το κουμπί “open” που βρίσκεται πάνω από αυτό. Ύστερα πατώντας τον διπλανό διακόπτη με την ένδειξη “import after open” ο ήχος αναλύεται και καταγράφεται. Πατώντας το κουμπί space πρέπει ακούγεται ο ανασυνθετημένος ήχος στη αρχική του χρονική διάρκεια και με το αρχικό του φάσμα.

Το αντικείμενο πολλαπλών ρυθμιστικών αριστερά, με ένδειξη “Delay of bins” έχει τη δυνατότητα ορισμού καθυστέρησης αναπαραγωγής του ήχου για διαφορετικές φασματικές ζώνες. Οι χαμηλές συχνότητες ξεκινάνε από το κάτω μέρος του αντικειμένου ανεβαίνοντας προς τα πάνω. Το πρώτο ρυθμιστικό με την ένδειξη set all ελέγχει όλα τα ρυθμιστικά του αντικειμένου μαζί. Το δεύτερο ορίζει τη μέγιστη καθυστέρηση σε ms.

Μέσω του δευτέρου αντικειμένου πολλαπλών ρυθμιστικών με ένδειξη “rate of bins”, ο χρήστης έχει τη δυνατότητα να ελέγχει την ταχύτητα αναπαραγωγής κάθε συχνοτικής

ζώνης. Η σύμβαση φοράς χαμηλών-υψηλών συχνοτήτων ισχύει για όλα τα αντικείμενα. Υπάρχει πάλι ο έλεγχος όλων των ρυθμιστικών από το κουτάκι “set all” καθώς και ορισμός μέγιστης και ελάχιστης τιμής ταχύτητας αναπαραγωγής. Ο σημαίνει αναπαραγωγή στην αρχική ταχύτητα του ήχου,1 σημαίνει διπλάσια ταχύτητα-μισός χρόνος αναπαραγωγής,2 τετραπλάσια ταχύτητα κ.ο.κ. Ο διακόπτης με την ένδειξη “freeze”, όταν ενεργοποιηθεί μετατρέπει τις τιμές ελάχιστου εύρους του αντικειμένου σε 0, δηλαδή παγώνει τις συχνοτικές ζώνες αυτές. Οι τιμές του αντικείμενου “rate o bins” ανανεώνονται κατά τη διάρκεια αναπαραγωγής του ήχου εφόσον ο χρήστης τις αλλάξει, ενώ αλλαγές στις τιμές καθυστέρησης κατά τη διάρκεια της αναπαραγωγής θα ισχύουν από την επόμενη αναπαραγωγή.

Το τρίτο σε σειρά αντικείμενο είναι ένα φίλτρο το οποίο επιτρέπει στον χρήστη να εξασθενήσει συχνοτικά μέρη του ήχου. Το αντικείμενο ακολουθείται από έναν φασματογράφο.

Όταν έχει παραχθεί ένα επιθυμητό αποτέλεσμα μπορεί να εγγραφεί, πρώτα ορίζοντας τη θέση και το όνομα του αρχείου πατώντας το κουμπί “open” κάτω από την ένδειξη “set file and rec” και ύστερα ενεργοποιώντας τον διπλανό διακόπτη για να αρχίσει η εγγραφή και απενεργοποιώντας τον για να σταματήσει. Ακολουθεί επεξήγηση και παραδείγματα χρήσης των τριών καταστάσεων λειτουργίας της εφαρμογής.

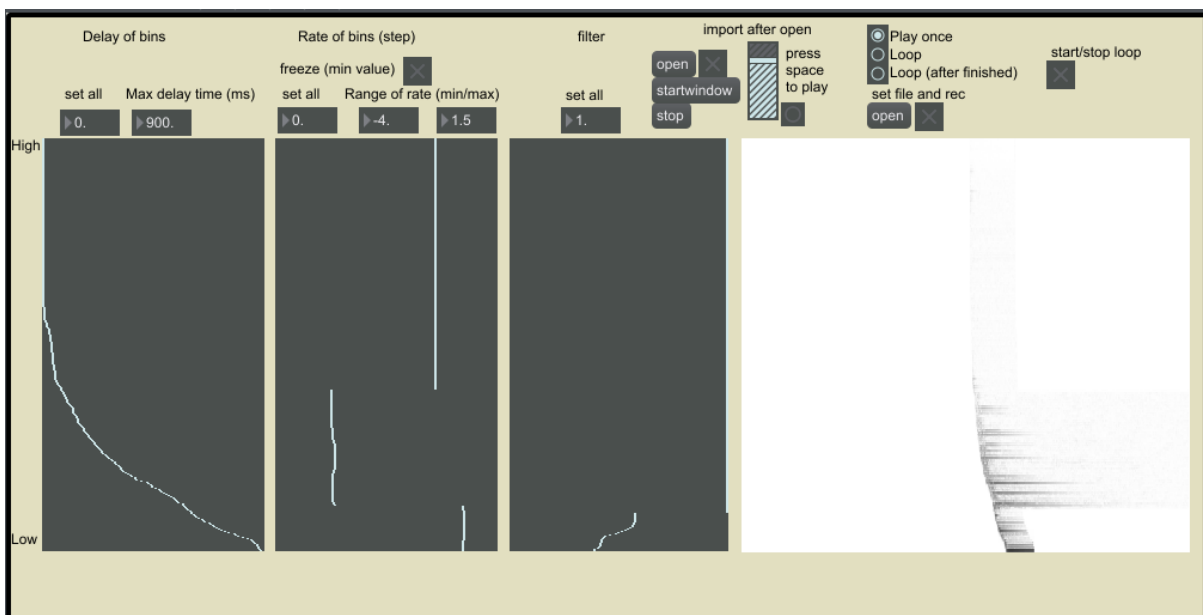
4.1 Play once

Σε αυτή τη λειτουργία ο επεξεργασμένος ήχος θα αναπαραχθεί μία φορά. Ο ήχος μπορεί να επαναληφθεί πατώντας space. Εισάγουμε το ηχητικό δείγμα vibes-a1 από τα δείγματα της max. Αρχικά θα δώσουμε μία χρονική καθυστέρηση στις χαμηλο-μεσαίες με σκοπό οι υψηλές συχνότητες να προηγούνται.



Σχήμα 47: Χρονική καθυστέρηση χαμηλομεσαίων συχνοτήτων σε ηχητικό δείγμα *vibes-a1*. Στο αριστερό μέρος του φασματογράφου φαίνεται το φάσμα του αρχικού ήχου.

Ύστερα μπορούμε να συμπιέσουμε λίγο τις χαμηλές συχνότητες και να επεκτείνουμε τις μεσαίες. Έπειτα θα εξασθενήσουμε με χρήση του φίλτρου λίγο τις χαμηλές γιατί περιέχουν πολύ ενέργεια.



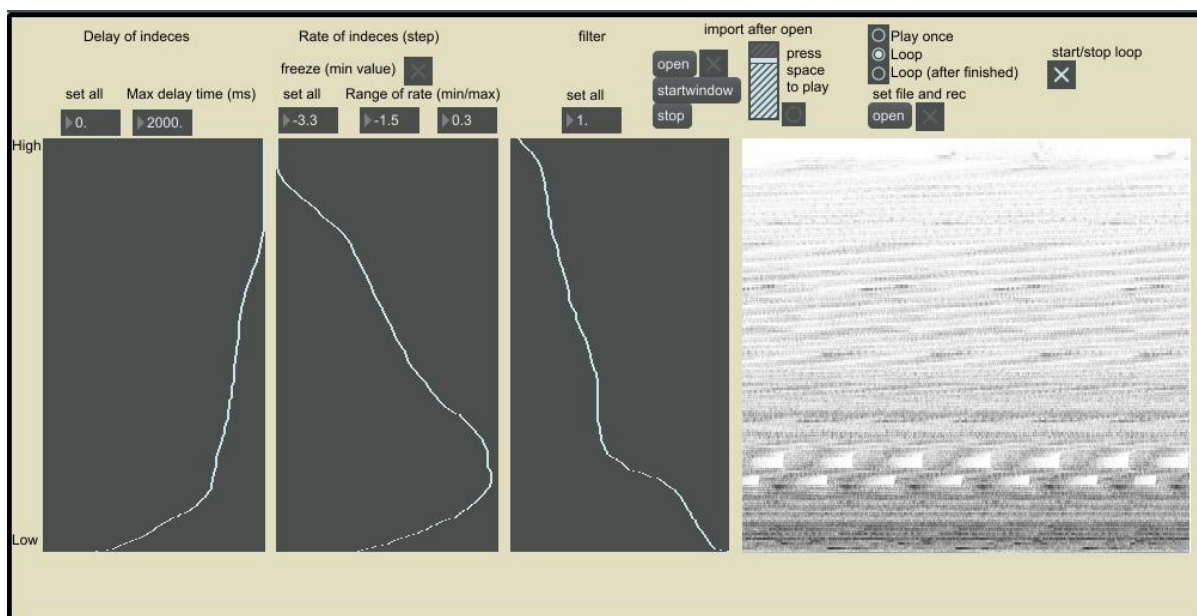
Σχήμα 48: Φασματικά επεξεργασμένο δείγμα με διαφορετική ταχύτητα αναπαραγωγής ανά συχνοτικά εύρη. (Ηχητικό δείγμα *play_once_sample.wav*)

4.2 Loop mode

Σε αυτή τη λειτουργία η κάθε συχνοτική ζώνη επαναλαμβάνεται ανεξάρτητα από την άλλη. Αυτό σημαίνει ότι μία ζώνη με μεγαλύτερη ταχύτητα αναπαραγωγής, θα έχει αναπαράξει τον ήχο παραπάνω φορές από μία άλλη με μικρότερη, μετά από κάποιο χρονικό διάστημα. Κατα κάποιο τρόπο η λειτουργία αυτή παράγει το φαινόμενο της “φασματικής πολυρυθμίας”. Αυτή η λειτουργία συνήθως παράγει πιο αφηριμένους ήχους. Με έντονη επεξεργασία ο ήχος μπορεί να χάσει την αρχική του ταυτότητα.

Με το που επιλέξουμε τη λειτουργία loop, ο διακόπτης “start/stop loop” ενεργοποιείται. Αυτό συμβαίνει επιλέγοντας και την άλλη λειτουργία loop. Όταν επιλέγεται η λειτουργία μονής αναπαραγωγής ο διακόπτης απενεργοποιείται αυτόματα. Όταν αλλάζει η λειτουργία της εφαρμογής η αναπαραγωγή του ήχου σταματάει και ξαναρχίζει. Απενεργοποιώντας τον διακόπτη “start/stop loop”, η αναπαραγωγή του ήχου σταματάει και ξαναπατώντας αρχίζει από την αρχή.

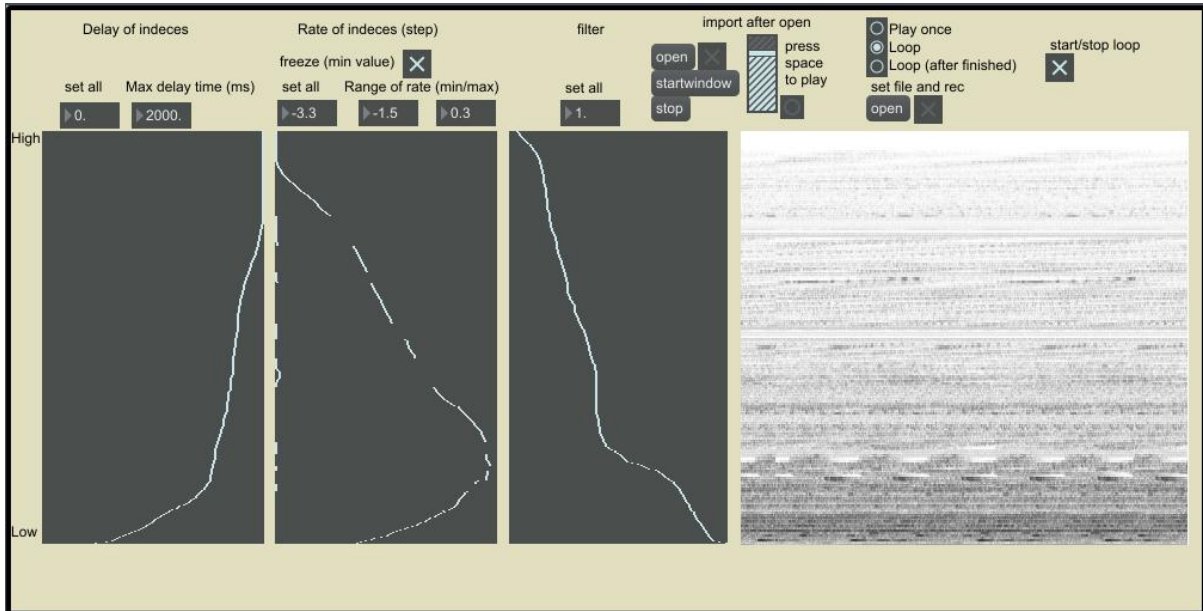
Για αρχή θα εισάγουμε το δείγμα anton-f2 της max. Θα δώσουμε μια καθυστέρηση στις μεσαίες και ψηλές. Ύστερα θα ζωγραφίσουμε μία καμπύλη με εύρος αλλαγής ταχύτητας - 3.3 έως 1. Γρήγορα ο ήχος εξελίζεται σε ένα μακρόσυρτο χαλί που προχωράει συνέχεια, με μικρές αλλαγές στις διαφορές θέσεων των ζωνών συχνοτήτων του ήχου.



Σχήμα 49: Παράδειγμα εφαρμογής σε λειτουργία loop. (Ηχητικό δείγμα loop_mode_sample.wav)

Με αυτό το αποτέλεσμα, ο χρήστης μπορεί να αλληλεπιδράσει αλλάζοντας της ταχύτητας ανάγνωσης, ενώ ο ήχος αναπαράγεται και αλλάζοντας τις σχέσεις έντασης συχνοτικά μέσω

του φίλτρου. Ενεργοποιώντας και απενεργοποιώντας το διακόπτη freeze σε συνδυασμό με αλλαγές στις ταχύτητες ανάγνωσης ο χρήστης μπορεί να παγώνει και ξεπαγώνει συχνοτικά εύρη του ήχου, δίνοντας του τη δυνατότητα να δημιουργεί συχνοτικά στάσιμα στοιχεία στον ήχο.



Σχήμα 50: Παράδειγμα χρήσης freeze σε λειτουργία loop. (Ηχητικό δείγμα loop_mode+freeze_sample.wav)

4.3 Loop after finished

Αυτή η λειτουργία υπάρχει κυρίως για χρηστικούς λόγους. Δεν έχει κάποια ιδιαίτερη επιπλέον λειτουργία σε σχέση με τη play once. Δίνει τη δυνατότητα στο χρήστη να ακούει μόνιμα το αποτέλεσμα των αλλαγών που κάνει στον ήχο, καθώς αυτός επαναλαμβάνεται όταν τελειώσει την αναπαραγωγή και η τελευταία συχνοτική ζώνη.

5. Συμπεράσματα

Η βασική ιδέα της εργασίας είναι η ανάπτυξη αλγορίθμου με δυνατότητα ανεξάρτητης μεταβολής χρονικής διάρκειας ανά συχνοτικές ζώνες ανάλυσης του FFT. Η κατηγορία σύνθεσης ήχου που ανήκει η εφαρμογή αυτή, είναι ανάλυση/ανασύνθεση ήχου και στην ουσία η εφαρμογή είναι μία επέκταση της λειτουργίας time stretching του Phase Vocoder.

Για να δημιουργηθεί η εφαρμογή, αρχικά χρειάστηκε η θεωρητική μελέτη και η αναζήτηση βιβλιογραφίας σχετικής με τον Μετασχηματισμό Fourier, και πιο συγκεκριμένα τον FFT, καθώς και γενικά μελέτη του θεωρητικού υποβάθρου τεχνικών ανάλυσης/ανασύνθεσης σύνθεσης ήχου. Μεγάλη βαρύτητα είχε η θεωρητική κατανόηση και ανάλυση της λειτουργίας του Phase Vocoder, καθώς είναι μία από τις βασικότερες εφαρμογές ανάλυσης/ανασύνθεσης ήχου και αποτελεί βάση της εφαρμογής που αναπτύχθηκε.

Στη συνέχεια ήταν αναγκαία η εξοικείωση με τις λειτουργίες ανάλυσης/ανασύνθεσης του προγραμματιστικού περιβάλλοντος MaxMSP, στο οποίο και αναπτύχθηκε η εφαρμογή. Αρχικά αναπτύχθηκε μία εφαρμογή απλού Phase Vocoder με δυνατότητα χρονικής διαστολής/συμπίεσης για όλο το φάσμα του ήχου. Ύστερα έπρεπε να σχεδιαστεί προτότυπος αλγόριθμος ο οποίος θα είχε τη δυνατότητα ανεξάρτητης χρονικής διαστολής/συμπίεσης για κάθε ζώνη ανάλυσης του FFT. Τελικά, με βάση την αρχική εφαρμογή του απλού Phase Vocoder, αναπτύχθηκε η εφαρμογή της χρονικής διαστολής/συμπίεσης μέσω φασματικής επεξεργασίας, καταφέροντας έτσι να υλοποιηθεί η αρχική ιδέα της εργασίας.

Μέσω αυτής της διαδικασίας, αποκτήθηκε ένα γενικό θεωρητικό υπόβαθρο στον τομέα των τεχνικών ανάλυσης/ανασύνθεσης ήχου καθώς και μια εικόνα των δυνατοτήτων και των εφαρμογών που υπάρχουν μέχρι σήμερα. Μέσω της ανάπτυξης της εφαρμογής αποκτήθηκε μία πρακτική, προγραμματιστική εμπειρία στο τομέα των φασματικών επεξεργασιών ηχητικού σήματος.

Η εφαρμογή που υλοποιήθηκε πραγματοποίησε την αρχική βασική ιδέα. Έφτασε σε ένα λειτουργικό στάδιο που μπορεί να χρησιμοποιηθεί για φασματική επεξεργασία εισερχομένων ηχητικών δειγμάτων από το χρήστη. Επιπλέον, η εφαρμογή έχει τη δυνατότητα αρχικής καθυστέρησης της αναπαραγωγής της κάθε ζώνης ανάλυσης του FFT, καθώς και χρήσης ενός φασματικού φίλτρου.

Ο χειρισμός της εφαρμογής παραμένει απλός. Το αρνητικό είναι ότι έχει σχετικά περιορισμένες δυνατότητες. Μία πιθανή επέκταση θα μπορούσε να είναι η αυτόματη δημιουργία τιμών ελέγχου για τα ρυθμιστικά της εφαρμογής με τις οποίες ο χρήστης θα μπορεί να αλληλεπιδράει, να επεξεργάζεται και να απορρίπτει ή να αποθηκεύει μέρη των τιμών με επιθυμητό αποτέλεσμα. Μια επιπλέον επέκταση θα μπορούσε να είναι η δυναμική αλλαγή θέσης των δεικτών ανάγνωσης των bins από παραγόμενες τιμές μέσω γεννητριών (ταλαντωτών χαμηλής συχνότητας) είται συναρτήσεων που θα ορίζει ο χρήστης, αντί της γραμμικής ανάγνωσής τους.

Παράρτημα

Το κείμενο συνοδεύεται από cd με την εφαρμογή και δείγματα ήχου από τα παραδείγματα χρήσης της.

Τα αρχεία της εφαρμογής είναι:

MainSpecStretch.maxpat

fft_stretch_buf_1.8~.maxpat

Δείγματα ήχου:

play_once_sample.wav

loop_mode_sample.wav

loop_mode+freeze_sample.wav

Η εφαρμογή αναπτύχθηκε στο προγραμματιστικό περιβάλλον Max 7.0.3 της Cycling '74.

Βιβλιογραφία

Charles, J. F. (2008), A tutorial on Spectral Sound Processing Using Max/MSP and Jitter, Computer Music Journal, Vol. 32, No. 3, Σελίδες 87-102

Cycling74, MSP Analysis Tutorial 4: Signal Processing with pfft~, https://docs.cycling74.com/max7/tutorials/14_analysischapter04, Ιούνιος 2016

Dodge C., Jerse T. A. (1997), Computer Music Synthesis, Composition, and Performance, 2nd edition, Schirmer – Thomson Learning, Broadway, New York

Droljc T. (2011), STFT Analysis Driven Sonographic Sound Processing in Real-Time using Max/MSP and Jitter, University of Hull.

Dudas R., Lippe Cort (2006), The Phase Vocoder – Part I, <https://cycling74.com/2006/11/02/the-phase-vocoder-%E2%80%93-part-i/#.V1NOBPI9670>, Ιούνιος 2016

Gordon J. and Strawn J. (1987), An Introduction to the Phase Vocoder, Stanford University, Stanford, California

Roads C. (1996), The Computer Music Tutorial, The MIT Press, Cambridge, Massachusetts.

Roads C. (2001), Microsound, The MIT Press, Cambridge, Massachusetts.

Sprenger S. M. (1999), Pitch Scaling Using The Fourier Transform, Audio DSP pages.

http://docs.happycoders.org/unsorted/computer_science/digital_signal_processing/dspdimension, Ιούνιος 2016