

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολο-
γιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Ζωγράφου 157 73, Αθήνα



14 Ιουλίου 2003

Στοχαστικά Fractals και Εφαρμογή στην Μοντελοποίηση Φωνής

Γεώργιος Αλέξανδρος Δημάκης

Επιβλέπων Καθηγητής: Πέτρος Μαραγκός

Διπλωματική Εργασία

*Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας
Σημάτων*



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων

Copyright © Γεώργιος Αλέξανδρος Δημάκης, 2003.
Με επιφύλαξη παντός δικαιώματος.

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

Εγκρίθηκε απο τριμελή εξεταστική επιτροπή την

Γεώργιος Αλέξανδρος Δημάκης, *Στοχαστικά Fractals και Εφαρμογή στην Μοντελοποίηση Φωνής*, Διπλωματική Εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Τομέας Σημάτων, Ελέγχου και Ρομποτικής, Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων, 14 Ιουλίου 2003.

Σελίδες: 82

Περιεχόμενα

Περιεχόμενα	6
Ευχαριστίες	7
Περίληψη	9
Abstract	11
Κατάλογος Σχημάτων	14
1 Αυτοομοιότητα και Fractals	15
1.1 Ιστορικά στοιχεία	15
1.2 Ντερμιτιστική Αυτοομοιότητα	16
1.3 Στατιστική Αυτοομοιότητα	17
1.4 Power Laws	19
1.4.1 Power laws και τα ονόματά τους	20
1.5 Power laws και Αυτοομοιότητα	21
1.5.1 Αναλλοίωτο κλίμακας και συμμετρίες στην φύση	22
2 Ανελίξεις $1/f$	25
2.1 Κίνηση Brown	26
2.2 Fractional Brownian Motion	28
2.2.1 Φάσμα Ισχύος του fBm	31
3 Σύνθεση και εκτίμηση $1/f$ ανελίξεων	33
3.1 Γενικά περί Σύνθεσης	33
3.2 Σύνθεση με Fourier	34
3.3 Σύνθεση με Wavelets	36

3.3.1	Multiresolution ανάλυση	37
3.3.2	Wavelets και $1/f$	39
3.4	Σύνθεση διδιάστατων fractals	40
3.5	Εκτίμηση με wavelets	42
3.6	Εκτίμηση με variance	44
3.7	Σύνθεση και εκτίμηση με Cholesky διάσπαση του πίνακα συμμεταβλητότητας	44
3.7.1	Εκτιμητής Μέγιστης πιθανοφάνειας	45
3.8	Φασματική εκτίμηση με την μέθοδο GPH	46
3.9	Αξιολόγηση αλγορίθμων σύνθεσης	47
3.10	Αξιολόγηση αλγορίθμων εκτίμησης	48
4	Μοντελοποίηση φωνής	51
4.1	Γραμμικό μοντέλο και Μοντέλο AM-FM	51
4.2	Αποδιαμόρφωση σημάτων AM-FM	53
4.2.1	Συστοιχίες φίλτρων και αποδιαμόρφωση	54
5	Ένα μοντέλο για τυρβώδη φωνήματα	57
5.1	Εξαγωγή χαρακτηριστικών και εκτίμηση παραμέτρων	58
5.2	Έλεγχος εγκυρότητας	59
5.3	Τεχνητά σήματα συντονισμού και εκτίμηση	60
5.4	Συμπεράσματα των πειραμάτων	61
6	Θεωρητική ανάλυση του προτεινόμενου μοντέλου	69
6.1	Υπολογισμός αυτοσυσχέτισης	69
6.2	Φάσμα ισχύος	71
6.3	Σύνδεση με κατανομές α -Stable	72
7	Συμπεράσματα	75
7.1	Συμβολή της Διπλωματικής εργασίας	75
7.2	Κατευθύνσεις για μελλοντική έρευνα	76

Ευχαριστίες

Η διπλωματική αυτή εργασία ξεκίνησε σαν μελέτη των στοχαστικών Fractals στα πλαίσια του μεταπτυχιακού μαθήματος 'Χάος και Fractals' το οποίο διδάσκει ο καθηγητής Πέτρος Μαραγκός και το οποίο παρακολούθησα το 2001. Αφού μελέτησα ορισμένα χαρακτηριστικά των $1/f$ ανελίξεων ξεκινήσαμε μια προσπάθεια να εφαρμόσουμε αυτή την θεωρία στην επεξεργασία φωνής. Συνολικά εργάστηκα πάνω σε αυτά τα θέματα περισσότερο απο δύο χρόνια και αφιέρωσα σημαντική πνευματική προσπάθεια.

Ήταν τιμή και προνόμιο για εμένα να έχω σαν επιβλέπων για αυτή την προσπάθεια τον Καθηγητή Πέτρο Μαραγκό. Υπήρξε εξαιρετικός δάσκαλος και θέλω να τον ευχαριστήσω για τον πολύ χρόνο που αφιέρωσε σε εμένα.

Θέλω επίσης να ευχαριστήσω τα παιδιά στο εργαστήριο για τις συμβουλές τους, την φιλική υποστήριξη τους και για το ότι έχουν δημιουργήσει μια ευχάριστη ερευνητική ομάδα στην οποία χαιρόμουν να είμαι μέλος.

Τέλος, θέλω να ευχαριστήσω την οικογένεια μου και τους φίλους μου για την απεριόριστη υποστήριξη, σε κάθε τομέα, που μου παρέχουν όλα αυτά τα χρόνια.

Περίληψη

Στην παρούσα εργασία γίνεται μια ανάλυση των εφαρμογών των στοχαστικών fractals στην επεξεργασία σήματος. Συγκεκριμένα εξετάζεται μια κλάση απο στοχαστικά fractals που ονομάζονται $1/f$ ανελίξεις. Αφού γίνει μια θεωρητική ανάλυση των ιδιοτήτων αυτών των ανελίξεων, γίνεται μια συγκριτική μελέτη των αλγορίθμων που υπάρχουν στην βιβλιογραφία για σύνθεση και εκτίμηση τέτοιων σημάτων. Η πρωτότυπη συνεισφορά αυτής της εργασίας είναι η εφαρμογή αυτής της θεωρίας σε σήματα φωνής και η δημιουργία ενός μοντέλου για τυρβώδη φωνήματα. Το μοντέλο αυτό αναλύεται τόσο θεωρητικά όσο και πειραματικά και αποδεικνύεται η ικανότητα του να μοντελοποιεί τυρβώδεις ήχους.

Λέξεις Κλειδιά

Ψηφιακή Επεξεργασία Σημάτων, Στοχαστικά Fractals, $1/f$ στοχαστικές ανελίξεις, Στατιστική αυτοομοιότητα, Μη-γραμμική Μοντελοποίηση φωνής, Τυρβώδεις ήχοι.

Abstract

In this thesis we present an analysis of stochastic Fractals and their applications to Signal Processing. Specifically, we focus on a class of stochastic Fractals, namely $1/f$ processes. After presenting their theoretical properties, we survey the algorithms available in the literature for synthesis and estimation of such processes. The original contribution of this work is the proposition of a nonlinear model based on $1/f$ processes for turbulent speech sounds. We analyze the proposed model and establish its validity, both theoretically and experimentally.

Keywords

Signal Processing, Stochastic Fractals, $1/f$ Stochastic Processes, Statistical Self-similarity, Nonlinear Speech modeling, Turbulent sounds.

Κατάλογος Σχημάτων

1.1	Η καμπύλη του Koch παραμένει αναλλοίωτη κάτω από αφρινικούς μετασχηματισμούς	17
1.2	Στατιστική Αυτοομοιότητα. Οι Στατιστικές ιδιότητες παραμένουν αναλλοίωτες κάτω από την κατάλληλη μεγέθυνση	18
3.1	Σύνθεση με την μέθοδο του Fourier filtering για διαφορετικούς εκθέτες γ	35
3.2	Συναρτήσεις $\psi(x)$ και $\phi(x)$ για το Daubechies wavelet γένους 2	36
3.3	Προβολή σήματος φωνής σε χώρους προσέγγισης $V_0, V_{-1}, V_{-2}, V_{-3}$	38
3.4	Επαναληπτική σύνθεση $1/f$ με Wavelets	41
3.5	Σύνθεση 2d fractals και Multifractals με Wavelets	42
3.6	Εκτίμηση γ με Wavelet variance	43
3.7	Φάσμα τεχνητών σημάτων που δημιουργούνται από αλγόριθμους σύνθεσης	49
4.1	Αποδιαμόρφωση φωνήματος με χρήση Gabor φίλτρων και SESA. Από [11].	53
5.1	Δυναμική GPH εκτίμηση τεχνητού $1/f$ σήματος	61
5.2	Πειράματα με φωνήματα /s/ (άνω σειρά) και /z/ (κάτω σειρά). Στήλες: α) Σήμα φωνής $s(t)$. β) Φάσμα ισχύος του $s(t)$ και φίλτρο Gabor. γ) Στιγμιαία συχνότητα $F(t)$. δ) Διαμόρφωση φάσης $\hat{P}(t)$. ε) Φάσμα ισχύος του $\hat{P}(t)$ και εκτιμώμενη κλίση. ζ) variance των wavelet συντελεστών λεπτομέρειας	63
5.3	Πειράματα με φωνήματα /f/ (άνω σειρά) και /v/ (κάτω σειρά). Στήλες: α) Σήμα φωνής $s(t)$. β) Φάσμα ισχύος του $s(t)$ και φίλτρο Gabor. γ) Στιγμιαία συχνότητα $F(t)$. γ) Διαμόρφωση φάσης $\hat{P}(t)$. ε) Φάσμα ισχύος του $\hat{P}(t)$ και εκτιμώμενη κλίση. ζ) variance των wavelet συντελεστών λεπτομέρειας	64

5.4	Πειράματα με φωνήματα /a/ (άνω σειρά) και /b/ (κάτω σειρά). Στήλες: α) Σήμα φωνής $s(t)$. β) Φάσμα ισχύος του $s(t)$ και φίλτρο Gabor. γ) Στιγμιαία συχνότητα $F(t)$. δ) Διαμόρφωση φάσης $\hat{P}(t)$. ε) Φάσμα ισχύος του $\hat{P}(t)$ και εκτιμώμενη κλίση. ζ) variance των wavelet συντελεστών λεπτομέρειας	65
5.5	Δύο πραγματικά σήματα διαμόρφωσης φάσης και ένα τεχνητό σήμα $1/f$.	66
5.6	(α) Τεχνητός συντονισμός με $1/f$ διαμόρφωση φάσης $P(t)$. (β) Στιγμιαία συχνότητα. (γ) $P(t)$ και $\hat{P}(t)$. (δ) Φάσμα ισχύος για το $P(t)$ και $\hat{P}(t)$. .	67
6.1	Φάσματα ισχύος που ακολουθούν κατανομή a -Stable για $a = 1, 1.6, 2$. .	72

Κεφάλαιο 1

Αυτοομοιότητα και Fractals

1.1 Ιστορικά στοιχεία

«Οι μαθηματικοί του 19ου αιώνα είχαν έλλειψη φαντασίας καθώς οι ίδιες παθολογικές δομές που επινόησαν σαν αντιπαραδείγματα για την φυσική διαίσθηση αποδείχθηκε τελικά ότι εμφανίζονται εγγενείς σε πολλά αντικείμενα που παρατηρούμε καθημερινά.»

Freeman Dyson

Το έτος 1855 σημαδεύεται από τον θάνατο του Gauss ενώ την ίδια εποχή κορυφαίοι Γερμανοί μαθηματικοί ξεκινούν να θέτουν τα θεμέλια της μαθηματικής θεωρίας των fractal. Το 1855 ο Weierstrass είναι σαράντα ετών, ο Kronecker τριάντα δύο ενώ ο Cantor ένα μικρό παιδί μόλις δέκα ετών. Ο Gauss είχε αποκαλέσει τα μαθηματικά 'επιστήμη των ματιών' δίνοντας έτσι έμφαση στην γεωμετρική —διαισθητική— αντίληψη. Ο Weierstrass ήταν φανατικός αναλύστας και προτιμούσε την αυστηρή θεμελίωση του λογισμού (στην οποία είχε συνεισφέρει σημαντικά) από την γεωμετρική διαίσθηση. Για να δείξει την υπεροχή των αναλυτικών μεθόδων (σε διαισθητικούς όπως ο μεγάλος του αντίπαλος Kronecker) κατασκεύασε μια συνάρτηση (αποκαλέστηκε ένα από τα 'τέρατα των μαθηματικών') που είναι παντού συνεχής και πουθενά παραγωγίσιμη. Ο Georg Cantor προσπαθώντας να κατανοήσει το άπειρο κατασκεύασε αργότερα άλλο ένα 'τέρας', μια απείρως διασκορπισμένη σκόνη (σύνολο υπεραριθμήσιμο αλλά μέτρου μηδέν) που φέρει το όνομα του.

Το ενδιαφέρον για αυτές τις κατασκευές αρχικά εστιάστηκε στις ιδιότητες τους που παραβίαζαν την διαίσθηση των μαθηματικών και αντιμετωπίζονταν σαν μεμονομένες —ενδιαφέρουσες υπάρξεις που καλό είναι κανείς γνωρίζει και να αποφεύγει. Το πραγματικά

όμως ενδιαφέρον στοιχείο που οδηγεί σε ένα νέα μαθηματικά, είναι ο *τρόπος κατασκευής τους*. Ο Weierstrass ήθελε να κατασκευάσει μια ‘απείρως γωνιώδη’ καμπύλη και την κατασκεύαζε προσθέτοντας όλο και μικρότερα τρίγωνα. Επαναλαμβάνοντας δηλαδή την ίδια διαδικασία σε μικρότερες κλίμακες. Την ίδια αρχή ακολούθησε ο Cantor αφαιρώντας συνεχώς ευθύγραμμα τμήματα σε όλο και μικρότερες κλίμακες. Η ίδια ιδέα υπάρχει στις κατασκευές των Hilbert, Koch και Sierpinski. Το ειρωνικό στοιχείο είναι ότι αυτές οι κατασκευές της ανάλυσης οδηγούν τελικά σε μια *δαισθητική γεωμετρία των fractals*. Ο Mandelbrot, πατέρας της θεωρίας, αποκάλεσε αυτές τις κατασκευές ProtoFractals.

1.2 Ντετερμινιστική Αυτοομοιότητα

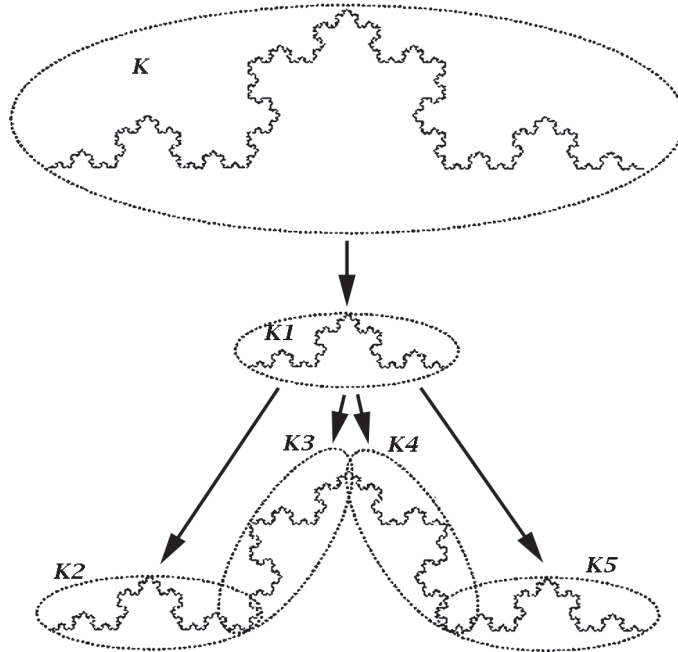
Η Αυτοομοιότητα ενός συνόλου είναι μια *Συμμετρία*. Συμμετρία είναι μια ακριβής μαθηματική έννοια και σημαίνει πως το σύνολο παραμένει *αναλλοίωτο κάτω από μια κλάση μετασχηματισμών*. Όταν αναφερόμαστε σε Αυτοομοιότητα, οι μετασχηματισμοί αυτοί ονομάζονται μετασχηματισμοί ομοιότητας (similarity transformations) και μπορεί να είναι μεγέθυνση (ή σμίκρυνση) και παράλληλη μεταφορά. Αυτό που επίσης έχει μεγάλη σημασία είναι το τι παραμένει αναλλοίωτο. Συγκεκριμένα, όταν αναφερόμαστε σε ντετερμινιστική Αυτοομοιότητα αυτό που παραμένει αναλλοίωτο είναι το ίδιο το σύνολο.

Μια γενίκευση αυτών των μετασχηματισμών είναι οι αφφινικοί (affine transformations) όπου επιτρέπουν επιπλέον περιστροφές και ανομοιόμορφη μεγέθυνση στους δυο άξονες. Το πλεονέκτημα των αφφινικών μετασχηματισμών είναι ότι αποτελούν την γενικότερη μορφή γραμμικού αντιστρέψιμου μετασχηματισμού που μπορεί να γραφεί σε μορφή πίνακα. Σύνολα που είναι αναλλοίωτα κάτω από αφφινικούς μετασχηματισμούς ονομάζονται και αυτοαφφινικά (selfaffine). Ο Mandelbrot τονίζει συχνά την διαφορά ανάμεσα στους δύο όρους, όταν όμως δεν υπάρχει πρόβλημα σύγχισης οι περισσότεροι συγγραφείς προτιμούν τον όρο *Αυτοομοιότητα* για να περιγράφουν την γενικότερη περίπτωση. Αυτό συμβαίνει κυρίως λόγω της αστοχίας¹ του όρου αφφινικός (affine) που καθιερώθηκε από τον Leonhard Euler.

Στο Σχήμα 1.2 φαίνεται η Αυτοομοιότητα της καμπύλης του Koch. Συγκεκριμένα η αρχική καμπύλη (K) προκύπτει σαν ένωση από τέσσερις μετασχηματισμένες (K2, K3, K4, K5) εκδοχές της.

Οι ιδέες αυτές της ντετερμινιστικής Αυτοομοιότητας γενικεύτηκαν σε μια πολύ όμορφη θεωρία που ξεκίνησε από τον J. Hutchinson και επεκτάθηκε από τον M. Barnsley και τους συνεργάτες του. Η θεωρία ονομάζεται IFS (Iterated Function Systems)[38] και μοντελοποιεί fractals με την χρήση αφφινικών μετασχηματισμών.

¹όπως παραδέχεται ο ίδιος ο Mandelbrot



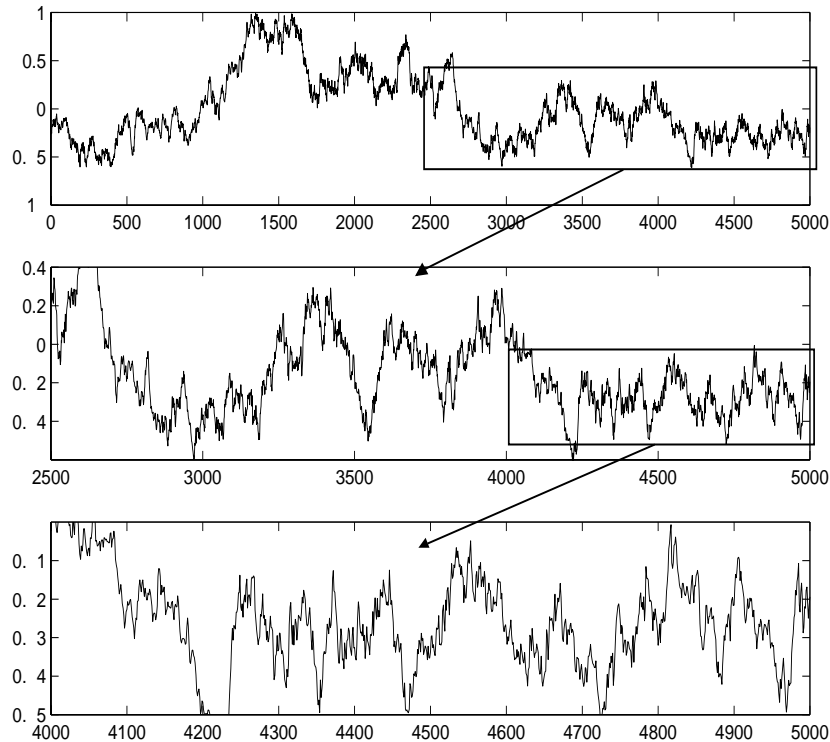
Σχήμα 1.1: Η καμπύλη του Koch παραμένει αναλλοίωτη κάτω από αφρινικούς μετασχηματισμούς

Παρόλη την ομορφιά και την απλότητα τους, τα ντερμινιστικά fractals αποτελούν πολύ περιοριστικά μοντέλα για τις περισσότερες εφαρμογές.

1.3 Στατιστική Αυτοομοιότητα

Για να μπορέσουμε να προσεγγίσουμε περισσότερο τα πραγματικά φυσικά συστήματα, πρέπει να χαλαρώσουμε τις απαιτήσεις μας. Όταν αναφερόμαστε σε Στατιστική Αυτοομοιότητα και πάλι έχουμε κάτι το οποίο παραμένει αναλλοίωτο κάτω από μετασχηματισμούς κλίμακας (στην γενικότερη περίπτωση αφρινικούς). Σε αντίθεση όμως με την ντερμινιστική εκδοχή, αυτό που παραμένει αναλλοίωτο δεν είναι το ίδιο το σύνολο αλλά οι στατιστικές ιδιότητες του συνόλου. Έτσι όταν για παράδειγμα αναφερόμαστε σε μια Gaussian στοχαστική ανέλιξη αρκεί να παραμένουν αναλλοίωτες η μέση τιμή και η αυτοσυσχέτιση. Στην γενικότερη περίπτωση πρέπει να ισχύει η ίδια σχέση για τις ροπές

υψηλότερης τάξης ή, ιδανικά, για της απο κοινού συναρτήσεις πυκνότητας πιθανότητας.



Σχήμα 1.2: Στατιστική Αυτοομοιότητα. Οι Στατιστικές ιδιότητες παραμένουν αναλλοίωτες κάτω από την κατάλληλη μεγέθυνση

Στο Σχήμα 1.3 απεικονίζεται αυτή η ιδέα της Στατιστικής Αυτοομοιότητας με μια υλοποίηση στοχαστικής ανέλιξης Wiener. Οπώς φαίνεται και στο σχήμα, όταν κάνει κανείς μεγέθυνση η ανέλιξη δεν παραμένει ίδια. Αυτό που παραμένει ίδιο είναι οι στατιστικές της ιδιότητες, η ανέλιξη δηλαδή συμπεριφέρεται με παρόμοιο στατιστικά τρόπο σε όλες τις κλίμακες. Συμβολικά, αυτό σημαίνει πως μια στοχαστική ανέλιξη είναι στατιστικά αυτοόμοια εάν ισχύει

$$x(t) \stackrel{P}{=} a^{-H} x(at). \quad (1.1)$$

Όπου $\stackrel{P}{=}$ σημαίνει πως οι δύο ανελιξεις έχουν ίδιες όλες τις στατιστικές τους ιδιότητες (π.χ. για όλες τις ροπές m τάξης). Αυτή ονομάζεται στατιστική αυτοομοιότητα με την αυστηρή έννοια (Strict-sense Self-similarity). Εάν εργαστούμε σε επίπεδο δευτέρας τάξης (μελετήσουμε δηλαδή μόνο την μέση τιμή και την αυτοσυσχέτιση) τότε μια στοχαστική

ανέλιξη $x(t)$ θα ονομάζεται στατιστικά αυτοόμοια όταν ισχύει η σχέση για την μέση τιμή:

$$E[x(t)] = a^{-H}E[x(at)] \quad (1.2)$$

και για την αυτοσυσχέτιση:

$$R_x(t, s) = E[x(t)x(s)] = a^{-2H}R_x(at, as). \quad (1.3)$$

Σε αυτή την περίπτωση η ανέλιξη ονομάζεται στατιστικά αυτοόμοια με την ευρεία έννοια (Wide sense Self-similarity).

1.4 Power Laws

Μια ποσότητα $O(x)$ ακολουθεί Power law όταν μεταβάλεται σαν

$$O(x) = Cx^a, \quad (1.4)$$

Όπου a θα ονομάζεται εκθέτης του Power law και C είναι μια σταθερά κανονικοποίησης.

Εάν λογαριθμήσουμε και τις δύο μεριές της προηγούμενης σχέσης λαμβάνουμε:

$$\log(O(x)) = \log C + \gamma \log x. \quad (1.5)$$

Το οποίο σημαίνει πως μια Power law ποσότητα μεταβάλεται σαν εύθεια γραμμή σε διπλά λογαριθμικό χαρτί (log-log plot) και αυτό αποτελεί τον δημοφιλέστερο τρόπο αναγνώρισης των Power laws σε πειραματικά δεδομένα.

Ένα παράδειγμα ποσότητας που ακολουθεί Power law είναι το φάσμα ισχύος για μια οικογένεια στοχαστικών ανελιξεων που ονομάζονται $1/f$ και αποτελούν την σημαντικότερη κλάση Στατιστικά Αυτοόμοιων ανελιξεων. Συγκεκριμένα, το (γενικευμένο) φάσμα ισχύος τους ακολουθεί μια Power law μείωση της μορφής.

$$S_x(\omega) \propto \frac{\sigma_x^2}{|\omega|^\gamma}, \quad (1.6)$$

με κάποια παράμετρο γ . Οι στοχαστικές ανελιξεις $1/f$ αποτελούν το βασικό εργαλείο μοντελοποίησης σε αυτή την εργασία και μια λεπτομερής ανάλυση των ιδιοτήτων τους δίνεται σε επόμενο κεφάλαιο.

Οι Power laws σχετίζονται άμεσα με την αυτοομοιότητα (όπως θα φανεί και παρακάτω) και εμφανίζονται πάρα πολύ συχνά στην φύση.

1.4.1 Power laws και τα ονόματά τους

Ο Zipf [7] σχεδίασε την συχνότητα εμφάνισης των λέξεων σε κείμενα σε σχέση με συχνότητα τους (τοποθέτησε δηλαδή την συχνότερη λέξη πρώτη, την δεύτερη σε συχνότητα δεύτερη κτλ). Σε διπλά λογαριθμικό χαρτί παρατήρησε πως όποιο κείμενο και αν επέλεγε εμφανιζόταν μια ευθεία γραμμή. Επίσης ισχυρίστηκε πως μπορούσε να διακρίνει ανάμεσα σε γλώσσες ή ακόμη και συγγραφείς από τους εκθέτες γ που μετρούσε. Στην συνέχεια μελέτησε το μέγεθος των πόλεων ως προς το πλήθος τους, το μέγεθος των πολέμων (σε αριθμό νεκρών) ως προς το μέγεθος τους, τον αριθμό των δημοσιεύσεων ανά επιστήμονα, την κατανομή εισοδημάτων και πολλές άλλες ποσότητες.

Ο Zipf, 'ο άνθρωπος που σχεδίαζε τα πάντα σε λογαριθμικό χαρτί', δεν ήταν όμως ο πρώτος που παρατήρησε αυτό το φαινόμενο. Η παλαιότερη αναφορά που είναι γνωστή, τουλάχιστον σε εμένα, είναι από τον Pareto[9] το 1897 όπου το ονόμασε Pareto distribution για την κατανομή εισόδηματος (και ονομάζεται ακόμη έτσι από οικονομολόγους).

Τέλος ο Benford [10] (ξανα)ανακάλυψε τους Power laws σε μια εντυπωσιακή πραγματικά εφαρμογή που ονομάζεται νόμος του Benford για τα πρώτα ψηφία. Ας υποθέσουμε ότι διαθέτουμε μια λίστα από αριθμούς που αντιπροσωπεύουν αποτελέσματα μετρήσεων από κάποια φυσική ποσότητα (π.χ. μήκη δρόμων αλλά όχι αριθμούς τηλεφώνων) και έστω ότι παρατηρούμε το πρώτο τους ψηφίο. Περιμένει κανείς πως η πιθανότητα να είναι 1 ή 2 ή 3 ... ή 9 είναι η ίδια. Στην πραγματικότητα όμως δεν είναι, αλλά ακολουθεί Power law (!). Δηλαδή η πιθανότητα να είναι μικρό (πχ 1 ή 2) είναι πολύ μεγαλύτερη από το να είναι 9. Ο νόμος του Benford είναι ένα καθαρά μαθηματικό αποτέλεσμα και έχουν προταθεί κάποιες εξηγήσεις που δεν σχετίζονται με αυτοομοιότητα.

Πρόσφατα οι Power laws έχουν τραβήξει μεγάλο ενδιαφέρον στην βιβλιογραφία της Επιστήμης Υπολογιστών και Τηλεπικοινωνιών. Συγκεκριμένα, παρατηρήθηκε πως η κατανομή του βαθμού (δηλαδή πόσα links έχει η κάθε σελίδα) (Degree) του γράφου του Web στο Internet ακολουθεί Power law. Η παρατήρηση αυτή συνεπάγεται πως ο γράφος του Web παρουσιάζει μια μορφή αυτοομοιότητας και αυτό έχει πολύ σημαντικές επιδράσεις σε αλγόριθμους Data mining που αναζητούν έγγραφα μέσα σε αυτό το δίκτυο. Επίσης σε μια πρόσφατη εργασία οι αδερφοί M. Faloutsos, P. Faloutsos και C. Faloutsos[39] έδειξαν πως και η φυσική τοπολογία του Internet ακολουθεί Power law. Τέλος η μορφή της κίνησης στο δίκτυο (αριθμός πακέτων, μέγεθος πακέτων) έχει μοντελοποιηθεί πρόσφατα με $1/f$ ανελιξεις που εμφανίζουν όπως είπαμε Power law στο φάσμα ισχύος τους και αυτό έχει τραβήξει σημαντικό ενδιαφέρον στην κοινότητα της επεξεργασίας σήματος. Τα σήματα ροής πληροφορίας σε δίκτυα εμφανίζουν αρκετές ιδιομορφίες σε σχέση με τις κλασικές $1/f$ ανελιξεις. Συγκεκριμένα, είναι αυστηρά θετικά, μοιάζουν να έχουν χρονικά χαρακτηριστικά που μεταβάλλονται με τον χρόνο και εμφανίζουν μεγάλες αποκλίσεις από

την μέση τους τιμή. Το στοιχείο της χρονικής μεταβλητότητας οδηγεί σε επεκτάσεις multifractal των $1/f$ μοντέλων που παρουσιάζουν πολύ ενδιαφέρον. Αντίστοιχα η μεγάλη διακύμανση από την μέση τιμή καθιστά τις Gaussian ανελίξεις ακατάλληλες και συνεπώς καταφεύγει κανείς σε a -Stable κατανομές που είναι ένα θέμα που θα απασχολήσει και εμάς στην θεωρητική ανάλυση που γίνεται σε επόμενο κεφάλαιο. Μια πολύ καλή αναφορά για την έρευνα στον τομέα αυτόν είναι το [30].

1.5 Power laws και Αυτοομοιότητα

Σχεδόν σε κάθε εργασία πάνω σε Power laws ή $1/f$ ανελίξεις, αναφέρεται πως αυτοί σχετίζονται με την αυτοομοιότητα. Παραδόξως όμως, είναι όμως πολύ δύσκολο να βρεί κανείς στην βιβλιογραφία μια λεπτομερή ανάλυση που να εξηγεί αυτή την σχέση. Η παρούσα ανάλυση είναι ένας συνδιασμος από τους (Sornette, Wornell) [17, 18].

Σε μια διάσταση, ας υποθέσουμε ονομάζουμε την ποσότητα που θέλουμε να παραμένει αναλλοίωτη κάτω από αλλαγές κλίμακας $O(x)$. Τότε η αλλαγή κλίμακας θα είναι $x \rightarrow \lambda x$. Τότε θα υπάρχει μια σταθερά (δηλαδή μια ποσότητα που δεν εξαρτάται από το x , $\mu(\lambda)$) τέτοια ώστε:

$$O(x) = \mu O(\lambda x), \quad (1.7)$$

Η εξίσωση (1.7) ορίζει αυτό που οι μαθηματικοί ονομάζουν *Ομογενή συνάρτηση*. Το θεμελιώδες αποτέλεσμα είναι πως η λύση της συναρτησιακής αυτής εξίσωσης είναι Power law:

$$O(x) = Cx^a, \quad a = -\frac{\ln \mu}{\ln \lambda} \quad (1.8)$$

Η εξίσωση (1.7) είναι η θεμελιώδης ιδιότητα που συνδέει τους Power Laws με την αυτοομοιότητα. Είναι σημαντικό ότι η λύση αυτής της εξίσωσης έχει αυτή την μορφή όταν έχουμε συνεχή αυτοομοιότητα για όλες τις δυνατές αλλαγές κλίμακας (δηλαδή η παράμετρος λ μπορεί να πάρει οποιαδήποτε τιμή). Σε αυτή την περίπτωση οι εξισώσεις αυτές ονομάζονται στην μαθηματική βιβλιογραφία ομογενείς (homogeneous) και δεν παρέχουν μεγάλη ποικιλία για μοντελοποίηση φαινομένων. Ας σημειωθεί ότι όλα τα κλασικά protofractals (Cantor, Sierpinsky, κτλ) χαρακτηρίζονται από διακριτή αυτοομοιότητα. Η διακριτή αυτοομοιότητα είναι ένα πολύ ενδιαφέρον θέμα που παρουσιάζει όμως πολλές μαθηματικές δυσκολίες. Πολύ πρόσφατα οι Borgnat, Flandrin, Amblard συνέδεσαν στην εργασία τους [40] την διακριτή αυτοομοιότητα με κυκλοστάσιμες Cyclostationary ανελίξεις.

1.5.1 Αναλλοίωτο κλίμακας και συμμετρίες στην φύση

Το 1905 η Emmy Amalie Noether απέδειξε το ακόλουθο θεώρημα: 'Για κάθε συνεχή συμμετρία στους νόμους της φύσης αντιστοιχεί ένας νόμος διατήρησης'. Συνεχή συμμετρία για τους νόμους της φύσης σημαίνει πως οι νόμοι της φύσης δεν αλλάζουν εάν εφαρμοστεί ένας τέτοιος μετασχηματισμός στο φυσικό σύστημα. (π.χ. Το ότι οι νόμοι της φύσης είναι ίδιοι στην Γή και στον γαλαξία της Ανδρομέδας αντιστοιχεί σε συμμετρία χωρικής μετακίνησης).

Αυτό οδηγεί στις ακόλουθες εκπληκτικές αντιστοιχίες:

1. Η συμμετρία της χωρικής μετακίνησης (Space translation) οδηγεί στην αρχή διατήρησης της Ορμής.
2. Η συμμετρία της χωρικής περιστροφής (Space rotation) οδηγεί στην αρχή διατήρησης της Στροφορμής
3. Η συμμετρία της χρονικής μετακίνησης (Time homogeneity) οδηγεί στην αρχή διατήρησης της Ενέργειας
4. Η συμμετρία της ολικής αλλαγής φάσης στην χβαντομηχανικής (Global phase shift) οδηγεί στην αρχή διατήρησης του Ηλεκτρικού φορτίου.

Στην φύση υπάρχει επίσης η συμμετρία της αλλαγής κλίμακας. Αυτό με απλά λόγια σημαίνει πως οι νόμοι της φύσης παραμένουν ίδιοι κάτω απο αλλαγές κλίμακας δηλαδή αλλαγές *Μονάδων μέτρησης*. Για παράδειγμα, τα αποτελέσματα Γάλλων πειραματιστών που χρησιμοποιούν μέτρα και κιλά δεν μπορεί να διαφέρουν απο τα αποτελέσματα Άγγλων συναδέλφων που επιμένουν να χρησιμοποιούν ίντσες και pounds. Για αυτή την περίπτωση το θεώρημα της Noether οδηγεί στην *Διαστατική ανάλυση*. Η βαθιά αυτή σχέση ανάμεσα στην Διαστατική ανάλυση και την αυτοομοιότητα έχει χρησιμοποιηθεί ήδη απο το 1941 απο τον κορυφαίο μαθηματικό A.N. Kolmogorov και έχει οδηγήσει σε ένα θεμελιακό ποσοτικό αποτέλεσμα για την Τυρβώδη Ροή που είναι γνωστό και σαν Νόμος του Kolmogorov[21]:

$$E(k, r) \propto r^{2/3} k^{-5/3} \quad (1.9)$$

Όπου $k = 2\pi/\lambda$ είναι ο κυματαριθμός σε μια πεπερασμένη περιοχή, r είναι η έκλυση ενέργειας και $E(k, r)$ είναι το φάσμα κυματαριθμού (velocity wavenumber spectrum) δηλαδή ο μετασχηματισμός Fourier των χωρικών συσχετίσεων. Ο νόμος του Kolmogorov είναι ένα Power Law για την τυρβώδη ροή υπάρχει συνεπώς κάποια σχέση ανάμεσα στην (στατιστική) αυτοομοιότητα και την τύρβη που ακόμη και σήμερα δεν έχει ξεκαθαριστεί. Ο Mandelbrot [3] και άλλοι έχουν διατυπώσει εικασίες για το ότι πολλά γεωμετρικά

χαρακτηριστικά της τυρβώδους ροής είναι fractals. Οι εικασίες όμως αυτές παραμένουν χωρίς πειραματική επιβεβαίωση και για αυτό πολλές φορές αναφέρεται πως η τυρβώδης ροή αποτελεί το τελευταίο ανοιχτό πρόβλημα της κλασσικής φυσικής. Αυτές οι ιδέες της σύνδεσης της τύρβης με τα Fractals αποτελούν το βασικό κίνητρο για το μοντέλο που προτείνεται σε αυτή την εργασία για τους τυρβώδεις ήχους.

Κεφάλαιο 2

Ανελίξεις $1/f$

Όπως αναφέρθηκε η οικογένεια των $1/f$ στοχαστικών ανελίξεων είναι η σημαντικότερη κλάση στατιστικά αυτοόμοιων ανελίξεων. Το κοινό τους χαρακτηριστικό είναι πως έχουν (γενικευμένο) φάσμα ισχύος που ακολουθεί power law

$$S_x(\omega) \propto \frac{\sigma_x^2}{|\omega|^\gamma} \quad (2.1)$$

για κάποιο εκθέτη γ .

Μια πραγματικά τεράστια οικογένεια απο φυσικά φαινόμενα εμφανίζουν αυτή την $1/f$ φασματική συμπεριφορά για μεγάλο εύρος συχνοτήτων. Μερικά παραδείγματα[17, 24, 18] είναι

- Ηλεκτρομαγνητικές διακυμάνσεις όπως για παράδειγμα η μεταβολές στην ακτινοβολία γαλαξιών , ένταση φωτεινών πηγών και μεταβολή ροής σε υπεραγωγούς.
- Θόρυβοι σε ηλεκτρονικές συσκευές όπως transistors, λυχνίες και διόδους
- Διακυμάνσεις σε συχνότητα σε ταλαντωτές όπως οι κρύσταλλοι quartz και τα ατομικά ρολόγια
- Φαινόμενα που οφείλονται σε ανθρώπινη δραστηριότητα όπως μεταβολές στην κίνηση αυτοκινήτων και δικτύων, μεταβολές συχνότητας στην μουσική.
- Τρόπος διακύμανσης της υψής σε σύννεφα, εικόνες εδάφους και ακτές.

Είναι συνεπώς πολύ ενδιαφέρον να βρεί κανείς στοχαστικά μοντέλα που να δημιουργούν τέτοιου είδους ανελίξεις ώστε να τις μελετήσει.

2.1 Κίνηση Brown

Η απλούστερη στοχαστική ανάλυση 1/f είναι ο 'τυχαίος περίπατος' ή κίνηση Brown. Η ονομασία προέρχεται από τον βοτανολόγο Brown που το 1828 παρατήρησε την ακανόνιστη κίνηση κόκκων γύρης πάνω στην επιφάνεια του νερού. Η φυσική μοντελοποίηση έγινε από τον A. Einstein το 1905 που πρότεινε πως κάθε κόκκος γύρης συγκρούεται συνεχώς με τα μόρια του νερού δημιουργώντας ένα μακροσκοπικό φαινόμενο. Αργότερα ο Perrin επιβεβαίωσε πειραματικά τις προβλέψεις του Einstein και χρησιμοποίησε το μοντέλο για να προσδιορίσει πειραματικά τον αριθμό των μορίων σε φυσικά δείγματα.

Η πρώτη ακριβής μαθηματική θεμελίωση της κίνησης Brown δόθηκε από τον κορυφαίο μαθηματικό Norbert Wiener. Συγκεκριμένα, ο Wiener όρισε αρχικά έναν διακριτό 'τυχαίο περίπατο' με τον εξής τρόπο: Έστω n ανεξάρτητες τυχαίες μεταβλητές X_i που μπορούν να πάρουν τις τιμές $+s$ ή $-s$ με πιθανότητα 1/2. Τότε

$$E[X_i] = 0, E[X_i^2] = s^2. \quad (2.2)$$

Ο τυχαίος περίπατος $X(t)$ ορίζεται σαν το συσσωρευτικό άθροισμα των X_i :

$$X(nT) = \sum_{i=1}^{i=n} X_i \quad (2.3)$$

Ισχύει προφανώς:

$$E[X(nt)] = 0, E[X(nt)^2] = ns^2 \quad (2.4)$$

Η πιθανότητα ο περίπατος να βρεθεί περάσει από το σημείο (nT, m) είναι

$$P(X(nT) = m) = \binom{n}{\frac{1}{2}(m+n)} \left(\frac{1}{2}\right)^n. \quad (2.5)$$

Όταν ο αριθμός των βημάτων n είναι πολύ μεγάλος η διωνυμική κατανομή πλησιάζει στην Gaussian:

$$P(X(nT) = m) \stackrel{n \rightarrow \infty}{\approx} \frac{1}{\sqrt{n\pi/2}} e^{-\frac{m^2}{2n}} \quad (2.6)$$

Η στοχαστική ανάλυση Wiener ορίζεται σαν το όριο του διακριτού τυχαίου περιπάτου όταν $T \rightarrow 0, n \rightarrow \infty$. Έχουμε δείξει όμως πως

$$E[Q(t)^2] = ns^2 = \frac{ts^2}{T}, \quad t = nT. \quad (2.7)$$

Συνεπώς για να έχει νόημα το όριο θα πρέπει το variance $E[Q(t)^2]$ να παραμένει πεπερασμένο συνεπώς πρέπει το s να τείνει στο μηδέν σαν \sqrt{T} :

$$s^2 = \alpha T \quad (2.8)$$

Τότε το όριο υπάρχει και είναι μια στοχαστική ανέλιξη συνεχούς χρόνου

$$B(t) = \lim_{T \rightarrow 0} X(t) \quad (2.9)$$

που ονομάζεται ανέλιξη Wiener.

Η πυκνότητα πιθανότητας της $B(t)$ είναι Gaussian με μηδενική μέση τιμή και τυπική απόκλιση at . Αυτό προκύπτει αν αντικαταστήσουμε στην σχέση (2.6) $w = ms$, $t = nT$:

$$Prob(w \leq B(t) \leq w + dw) = \frac{1}{\sqrt{2\pi at}} e^{-w^2/2at}. \quad (2.10)$$

Επίσης μπορούμε να υπολογίσουμε την αυτόσυσχέτιση της $B(t)$: Αν $t_1 < t_2$ η διαφορά $B(t_2) - B(t_1)$ είναι ανεξάρτητη από το $B(t_1)$ και συνεπώς

$$\begin{aligned} R_B(t_1, t_2) &= E[B(t_1)B(t_2)] = E[(B(t_2) - B(t_1))B(t_1) + B(t_1)^2] \\ &= E[(B(t_2) - B(t_1))E[(w_1)]] - E[B(t_1)^2], \end{aligned} \quad (2.11)$$

όμως $E[(w_1)] = 0$ άρα η σχέση (2.11) γίνεται

$$R_B(t_1, t_2) = E[B(t_1)^2] = \frac{t_1 s^2}{T} = \alpha t_1 \quad (2.12)$$

Συνεπώς γενικότερα για οποιαδήποτε t_1, t_2 ισχύει

$$R_B(t_1, t_2) = \alpha \min(t_1, t_2). \quad (2.13)$$

Οι στοχαστικές ανελιξεις που θα μελετήσουμε ορίζονται σαν στοχαστικά ολοκληρώματα της κίνησης Brown. Η στοχαστική ολοκλήρωση και η γενικότερη θεωρία των στοχαστικών διαφορικών εξισώσεων αποτελεί έναν αναπτυσσόμενο κλάδο των μαθηματικών. Στην παρούσα εργασία και για όλους τους πρακτικούς λόγους ένα στοχαστικό ολοκλήρωμα θεωρείται σαν το όριο ενός αθροίσματος που λειτουργεί πάνω σε έναν διακριτοποιημένο λευκό θόρυβο.

2.2 Fractional Brownian Motion

Το δημοφιλέστερο μοντέλο για στοχαστικές ανελιξίες $1/f$ προτάθηκε από τον B.Mandelbrot στην θεμελιώδη εργασία του [15]. Η ιδέα βασίζεται στην γενίκευση της κίνησης Brown. Η κίνηση Brown μπορεί να μοντελοποιηθεί σαν ένα στοχαστικό ολοκλήρωμα

$$B(t) = \int_0^t dB(s) = \int_0^t w(\tau)d\tau \quad (2.14)$$

Όπου $w(t)$ είναι ο λευκός gaussian θόρυβος. Η πρώτη γραφή (στοχαστική ολοκλήρωση με βάση την ανέλιξη Wiener είναι προσφιλής στην μαθηματική βιβλιογραφία ενώ η δεύτερη είναι ποιο κοντά στις τεχνικές των Ηλεκτρολόγων μηχανικών και μπορεί να νοηθεί σαν ένα γραμμικό φίλτρο που οδηγείται από λευκό θόρυβο. Συνεπώς, με βάση το ζευγάρι γενικευμένων μετασχηματισμών Fourier [17]:

$$\frac{|\tau|^{\gamma-1}}{2\Gamma(\gamma)\cos(\gamma\pi/2)} \xleftrightarrow{F} \frac{1}{|\omega|^\gamma}, \quad (2.15)$$

αν θέλουμε να έχουμε φάσμα της μορφής $1/\omega^\gamma$, θα πρέπει να επιλέξουμε φίλτρο με κρουστική απόκριση:

$$h(t) = \frac{1}{\Gamma(H+1/2)} t^{H-1/2} u(t) \quad (2.16)$$

όπου $u(t)$ είναι η βηματική συνάρτηση και $\gamma = 2H + 1$. Σε αυτή την περίπτωση η κρουστική απόκριση έχει μετασχηματισμό Laplace [17]:

$$H(s) = \frac{1}{s^{2H+1}} \quad (2.17)$$

Το πρόβλημα είναι πως αυτό το σύστημα είναι ασταθές καθώς η συνέλιξη

$$B_L(t) = h(t) * w(t) = \frac{1}{\Gamma(H+1/2)} \int_{-\infty}^t (t-\tau)^{H-1/2} w(\tau) d\tau \quad (2.18)$$

δεν συγκλίνει.

Κάτω από αυτό το πλαίσιο γίνεται πλέον φανερό για ποιόν λόγο ο Paul Lévy [22] όρισε το ακόλουθο κλασματικό ολοκλήρωμα:

$$L_H(t) = \frac{1}{\Gamma(H+1/2)} \int_0^t (t-\tau)^{H-1/2} w(\tau) d\tau, \quad (2.19)$$

ξεκινώντας την ολοκλήρωση απο το μηδέν. Η ανέλιξη αυτή αποτελεί το πρώτο κλασματικό ολοκλήρωμα που δίνει στατιστικά αυτοόμοια στοχαστική ανέλιξη, έχει όμως το πολύ βασικό μειονέκτημα ότι δεν είναι έχει κανένα στάσιμο χαρακτηριστικό. Αυτό οφείλεται στο ότι το σημείο μηδέν αποκτά μεγάλη σημασία και λειτουργεί σαν 'σημείο αναφοράς' που καταστρέφει το αναλλοίωτο κάτω απο μεταφορά που απαιτείται για να υπάρχουν στάσιμα χαρακτηριστικά.

Το 1968 οι Benoit Mandelbrot και John Van Ness δημοσίευσαν την θεμελιώδη εργασία [15] που ορίζει την κλασματική κίνηση Brown (fractional Brownian Motion, fBm):

Έστω παράμετρος H με $0 < H < 1$. Τότε ορίζουμε

$$B_H(0) = 0,$$

$$B_H(t) = \frac{1}{\Gamma(H + 1/2)} \left(\int_{-\infty}^0 (t - \tau)^{H-1/2} - (-\tau)^{H-1/2} w(\tau) d\tau \right. \\ \left. + \int_0^t (t - \tau)^{H-1/2} w(\tau) d\tau \right) \quad (2.20)$$

Μπορεί να δοθεί μια πιο συμμετρική μορφή στον παραπάνω ορισμό αν την γράψουμε σαν την συγκλίνουσα διαφορά δύο ολοκληρωμάτων που απολίνουν μεμονομένα:

$$B_H(t_2) - B_H(t_1) = \frac{1}{\Gamma(H + 1/2)} \left(\int_{-\infty}^{t_2} (t_2 - \tau)^{H-1/2} w(\tau) d\tau - \int_{-\infty}^{t_1} (t_1 - \tau)^{H-1/2} w(\tau) d\tau \right) \quad (2.21)$$

Το βασικό χαρακτηριστικό που καθιστά την ανέλιξη fBm χρήσιμη για μοντελοποίηση είναι ότι οι μεταβολές της είναι στατιστικά αυτοόμοιες και στατικές. Ισχύει δηλαδή

$$B_H(t + \tau) - B_H(t) \stackrel{P}{=} a^{-H} (B_H(t + a\tau) - B_H(t)). \quad (2.22)$$

Επίσης ισχύει ο ακόλουθος νόμος για το variance που γενικεύει τον αντίστοιχο νόμο της κίνησης Wiener:

$$\text{var}[B_H(t_2) - B_H(t_1)] = \mathbf{E}[(B_H(t_2)) - B_H(t_1)]^2 = (t_2 - t_1)^{2H} V_H \quad (2.23)$$

Όπου

$$V_H = \Gamma(H + 1/2)^{-2} \left(\int_{-\infty}^0 (1 - s^{H-1/2} - (-s)^{H-1/2})^2 ds + \frac{1}{2H} \right). \quad (2.24)$$

Με χρήση αυτού του αποτελέσματος μπορεί να υπολογιστεί η αυτοσυσχέτιση της ανέλιξης fBm:

$$R_{B_H}(t, s) = \mathbb{E}[B_H(t)B_H(s)] = \frac{\sigma_H^2}{2}(|s|^{2H} + |t|^{2H} - |t - s|^{2H}), \quad (2.25)$$

όπου η σταθερά είναι:

$$\sigma_H^2 = \text{var}(x(1)) = \Gamma(1 - 2H) \frac{\cos(\pi H)}{\pi H}. \quad (2.26)$$

Η ανέλιξη fBm είναι συνεχής (σχεδόν παντού) αλλά δεν είναι διαφορίσιμη με την mean square έννοια δηλαδή

$$\limsup_{t \rightarrow t_0} \left| \frac{B_H(t) - B_H(t_0)}{t - t_0} \right| = \infty \quad (2.27)$$

με πιθανότητα 1. Επίσης, έχει αποδειχθεί πως οι υλοποιήσεις της ανέλιξης είναι Fractals με κλασματική διάσταση Hausdorff-Besicovitch που δίνεται απο την απλή σχέση

$$D = 2 - H \quad (2.28)$$

Το πρόβλημα της μη παραγωγισιμότητας είναι σημαντικό γιατί έτσι χάνεται το βασικό πλεονέκτημα των στάσιμων μεταβολών. Ο Mandelbrot στο [15] εισάγει την ανέλιξη μεταβολών με τον ακόλουθο τρόπο: Εισάγει έναν απλό πυρήνα εξομάλυνσης της μορφής

$$\phi_\delta(t) = \begin{cases} \frac{1}{\delta} & t \in [0, \delta] \\ 0 & t \notin [0, \delta]. \end{cases} \quad (2.29)$$

Και δημιουργεί την εξομαλυμένη ανέλιξη

$$B_H^\delta(t) = B_H(t) \star \phi_\delta(t) \quad (2.30)$$

Η B_H^δ είναι παραγωγίσιμη και όταν το δ είναι πολύ μικρό τείνει στην B_H . Η k παράγωγος της B_H^δ είναι

$$(B_H^\delta(t))^{(k)} = (-1)^k \int_{-\infty}^{\infty} B_H(s) \phi^{(k)}(t - s) ds \quad (2.31)$$

Τέλος, ένα πολύ σημαντικό θεώρημα που αποδεικνύει ο Mandelbrot είναι πως κάθε (μη σταθερή) Gaussian ανέλιξη που έχει την ιδιότητα της αυτοομοιότητας και των στάσιμων μεταβολών είναι fBm. Αυτό σημαίνει πως εάν θέλουμε να παραμένουμε στα απλούστερα δυνατά στοχαστικά fractals θα πρέπει αναγκαστικά να χρησιμοποιήσουμε το μοντέλο fBm.

2.2.1 Φάσμα Ισχύος του fBm

Δεν πρέπει φυσικά να ξεχνάμε πως όλο το μοντέλο fBm δημιουργείται για να μοντελοποιήσει στοχαστικές ανελίξεις που έχουν φάσμα $1/f$. Σε πολλά βιβλία το φάσμα ισχύος μιας ανελίξης $x(n)$ (με N δείγματα) ορίζεται σαν ο μετασχηματισμός Fourier της συνάρτησης αυτοσυσχέτισης.

Το πρόβλημα εδώ είναι πως οι ανελίξεις fBm δεν είναι στάσιμες και συνεπώς δεν έχουν φάσμα ισχύος με αυτή την έννοια καθώς η συνάρτηση αυτοσυσχέτισης είναι δύο μεταβλητών.

Ένας καλύτερος ορισμός του φάσματος ισχύος είναι ο ακόλουθος

$$P_x(\omega) = \mathbb{E}\left[\frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) \exp(-j\omega n) \right|^2\right] \quad (2.32)$$

Το οποίο σημαίνει πως λαμβάνουμε το μέτρο του Fourier μετασχηματισμού (το οποίο είναι ένα σύνολο απο τυχαίες μεταβλητές) και στην συνέχεια υπολογίζουμε την μέση τιμή για κάθε συχνότητα. Το θεώρημα Wiener-Khinchin μας εξασφαλίζει πως εάν η ανελίξη που μελετάμε είναι στάσιμη (και η αυτοσυσχέτιση της φθίνει αρκετά γρήγορα) τότε ο δεύτερος αυτός ορισμός είναι ισοδύναμος με τον μετασχηματισμό Fourier της αυτοσυσχέτισης.

Ο P. Flandrin στην βασική του εργασία [31] έδειξε χρησιμοποιώντας μετασχηματισμούς Time-Frequency όπως ο μετασχηματισμός Wigner-Ville και ο μετασχηματισμός Wavelet, ότι το γενικευμένο φάσμα ισχύος ακολουθεί power law $1/f^\gamma$ και ισχύει η σχέση

$$\gamma = 2H + 1 \quad (2.33)$$

δηλαδή ο εκθέτης γ συνδέεται με απλό τρόπο με την παράμετρο H της ανελίξης fBm και απο εδώ και στο εξής θα αναφερόμαστε πάντοτε στην παράμετρο γ .

Κεφάλαιο 3

Σύνθεση και εκτίμηση $1/f$ ανελιξίων

Αυτό που επιθυμούμε να κάνουμε στην πράξη είναι να χρησιμοποιήσουμε το θεωρητικό μοντέλο fBm για την μοντελοποίηση σημάτων που λαμβάνουμε από φυσικά συστήματα. Αυτό εισάγει έναν αριθμό από νέα προβλήματα που πρέπει να λυθούν. Συγκεκριμένα χρειαζόμαστε ψηφιακούς αλγορίθμους για την σύνθεση υλοποιήσεων fBm καθώς και αλγορίθμους που να μπορούν να εκτιμήσουν τις παραμέτρους του μοντέλου μας από διακριτά δείγματα ενός σήματος το οποίο μάλιστα, στις περισσότερες περιπτώσεις περιέχει και θόρυβο.

Δεδομένου του ότι θεωρητικά τα fractals περιέχουν άπειρη λεπτομέρεια (μη πεπερασμένο συχνοτικό περιεχόμενο) είναι θεωρητικά αδύνατον να ψηφιοποιηθούν καθώς με βάση το θεώρημα του Shannon θα χρειαζόμαστε απείρως μεγάλη συχνότητα δειγματοληψίας. Συνεπώς το μόνο που μπορούμε να κάνουμε είναι να ψηφιοποιήσουμε ένα low pass filtered fractal που πρακτικά σημαίνει ότι μελετάμε προσεγγίσεις μέχρι μια πεπερασμένη κλίμακα. Το επίπεδο ανάλυσης στο οποίο θα φτάσουμε εξαρτάται από τους φυσικούς και υπολογιστικούς περιορισμούς που διαθέτουμε. Μια θεωρητική ανάλυση των ιδιοτήτων που διατηρούνται όταν από ένα fractal περνάμε στην low pass προσέγγιση του είναι ένα πολύ ενδιαφέρον θέμα που παραμένει σε μεγάλο βαθμό ανοιχτό.

3.1 Γενικά περί Σύνθεσης

Το βασικό πρόβλημα με την σύνθεση fBm είναι οι αλληλεπιδράσεις μεγάλου χρόνου (Long Range Dependence, LRD) που εισάγει η πολυωνυμική πτώση της αυτοσυσχέτισης. Αυτό πρακτικά σημαίνει πως ένα δείγμα εξαρτάται σε πολύ μεγάλο βαθμό από το πολύ

μακρινό παρελθόν του. Τα ARMA μοντέλα που μοντελοποιούνται με απλές εξισώσεις διαφορών έχουν εκθετική μείωση της αυτοσυσχέτισης και συνεπώς δεν είναι κατάλληλα για την σύνθεση στοχαστικών fractals. Έχουν προταθεί μοντέλα που εμφανίζουν LRD και βασίζονται σε επεκτάσεις του κλασσικού ARMA μοντέλου (Fractionally Integrated Auto-Regressive Moving Average, FARIMA [23]) αλλά η χρήση τους δεν έχει καθιερωθεί ακόμη και δεν θα μας απασχολήσουν σε αυτή την εργασία.

Σε γενικές γραμμές οι αλγόριθμοι σύνθεσης μπορούν να χωριστούν στις ακόλουθες κατηγορίες ανάλογα με τον τρόπο με τον οποίο λειτουργούν σε διαφορετικές αναλύσεις:

- **Αναδρομικοί αλγόριθμοι:** Εισάγεται μια προσέγγιση fractal σε μια (χαμηλή) ανάλυση και ο αλγόριθμος δίνει την προσέγγιση σε υψηλότερη ανάλυση. Παραδείγματα τέτοιων αλγορίθμων είναι ο αλγόριθμος μετακίνησης μέσου (Midpoint Displacement) [24] καθώς και η σύνθεση με wavelets[17]. Βασικό πλεονέκτημα αυτών των μεθόδων είναι η μεγάλη ταχύτητα ενώ μειονέκτημα είναι η αδυναμία να ελέγξει κανείς το ακριβές μέγεθος του υπο σύνθεση fractal καθώς και η όχι ακριβής θεωρητική σχέση τους με το μοντέλο fBm.
- **Αλγόριθμοι ενός βήματος:** Σε αυτούς τους αλγορίθμους το fractal σήμα συντίθεται κατεύθειας στην επιθυμητή ανάλυση. Αλγόριθμοι που ανοίκουν σε αυτή την κατηγορία είναι ο αλγόριθμος σύνθεσης με Cholesky διάσπαση του πίνακα συνδιακύμανσης [16] καθώς και η σύνθεση με φιλτράρισμα (Fourier Filtering) [24].

Σε αυτή την εργασία υλοποιήσαμε τις μεθόδους που θεωρούνται βέλτιστες στην βιβλιογραφία συγκεκριμένα την μέθοδο με φιλτράρισμα Fourier που δίνει το καλύτερο φασματικό περιεχόμενο, την μέθοδο με παραγοντοποίηση Cholesky που είναι πιο θεωρητικά θεμελιωμένη και την μέθοδο με wavelets που είναι ταχύτερη και ευέλικτη.

3.2 Σύνθεση με Fourier

Η μέθοδος με σύνθεση Fourier αναλύεται στο [24]. Εάν υποθέσουμε ότι διακριτοποιούμε το σήμα $B_H(t)$ σε πεπερασμένο χρόνο $t \in [0, t_{max}]$ και λάβουμε το διακριτό, πεπερασμένο σήμα $B_H[n]$, $n \in [0, N-1]$ τότε εάν δούμε τους συντελεστές του διακριτού μετασχηματισμού Fourier του:

$$B_H[n] = \frac{1}{N} \sum_{k=0}^{N-1} B_H[k] e^{j(2\pi/N)kn} \quad (3.1)$$

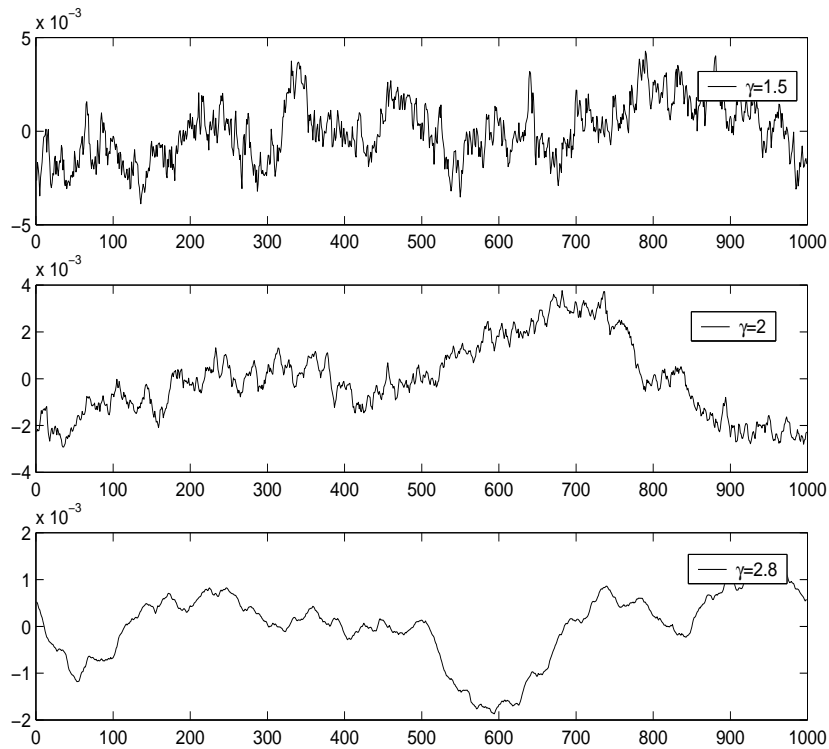
τότε για το μέτρο αυτών των τ.μ. $|B_H[k]|$ θα πρέπει να ισχύει

$$E[|B_H[k|] \propto \frac{1}{k^\gamma} \quad (3.2)$$

Αυτή η σχέση θα ισχύει όταν $0 < k < N/2$ ενώ θα πρέπει να ισχύει η ακόλουθη σχέση για τους συζυγείς μιγαδικούς συντελεστές

$$B_H[k] = B_H[N - k]^* \quad (3.3)$$

ώστε να έχουμε πραγματικό σήμα. Συνεπώς πρέπει το μέτρο $|B_H[k]|$ να είναι κανονικές τ.μ. με μέση τιμή που φθίνει σαν power law ενώ η φάση μπορεί να είναι ομοιόμορφη στο $[0, 2\pi]$. Αφού δημιουργηθούν οι συντελεστές $B_H[k]$ το σήμα $B_H[n]$ λαμβάνεται με αντίστροφο διακριτό μετασχηματισμό Fourier.



Σχήμα 3.1: Σύνθεση με την μέθοδο του Fourier filtering για διαφορετικούς εκθέτες γ

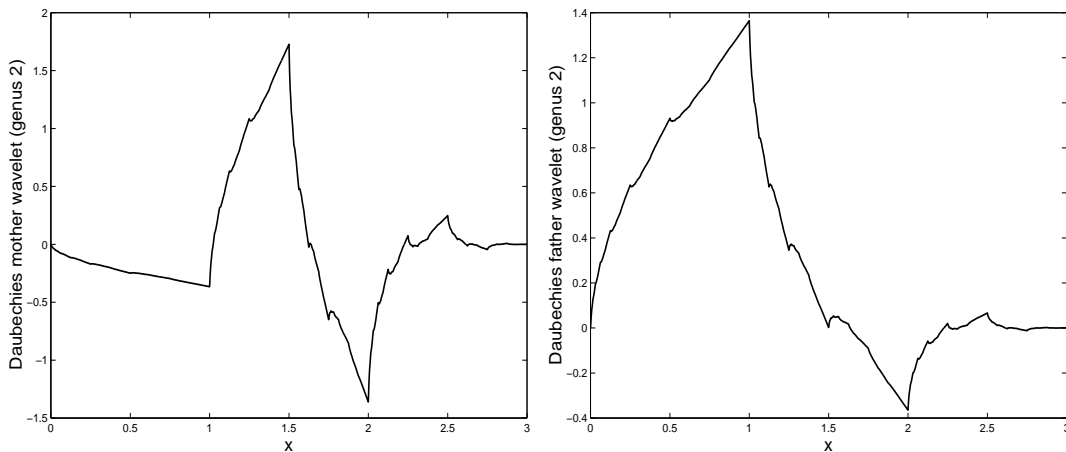
Όπως θα φανεί και στην σύγκριση με τις άλλες μεθόδους σύνθεσης, η μέθοδος φιλτραρίσματος Fourier είναι πολύ αποτελεσματική και είναι αυτή που χρησιμοποιήθηκε περισσότερο στην εργασία αυτή.

3.3 Σύνθεση με Wavelets

Ο μετασχηματισμός Wavelet είναι η προβολή ενός σήματος $x(t)$ πάνω σε μια συναρτησιακή βάση που προκύπτει από μετακινήσεις και χρονικές συμπιέσεις μιας βασικής συνάρτησης που ονομάζεται συνάρτηση $\psi(t)$ (mother wavelet). Ο χώρος προβολής είναι το σύνολο των συναρτήσεων

$$\psi_n^m(t) = 2^{m/2} \psi(2^m t - n) \quad (3.4)$$

Όπου τα m και n ονομάζονται παράμετροι κλίμακας και μετακίνησης αντίστοιχα. Χωρίς να μπούμε στις λεπτομέρειες της μαθηματικής θεωρίας των Wavelet μπορούμε να πούμε πως το σύνολο των συναρτήσεων $\psi_n^m(t)$ αποτελεί βάση για τον χώρο όταν η βασική συνάρτηση $\psi(t)$ ικανοποιεί κάποιες προϋποθέσεις[5]. Οι ιδιότητες της βάσης Wavelet καθορίζονται προφανώς από την μορφή αυτής της βασικής συνάρτησης. Μια πολύ δημοφιλής επιλογή είναι η οικογένεια από Wavelets που εισήγαγε η Daubechies[5]. Για αυτά δεν υπάρχει κλειστός τύπος που να δίνει την μορφή της συνάρτησης ψ . Παρόλα αυτά υπάρχει αλγόριθμος (Discrete Wavelet Transform, DWT) για διακριτά δεδομένα που μπορεί να υπολογίσει ταχύτατα τους συντελεστές προβολής. Επίσης τα Wavelets της Daubechies δημιουργούν ορθοκανονική βάση και έχουν μέγιστες μηδενιζόμενες ροπές για δεδομένο γένος (για δεδομένο δηλαδή μήκος του αντίστοιχου φίλτρου) που τα καθιστά κατάλληλα για την σύνθεση στοχαστικών fractals.



Σχήμα 3.2: Συναρτήσεις $\psi(x)$ και $\phi(x)$ για το Daubechies wavelet γένους 2

Η προβολή ενός σήματος $x(t)$ θα γίνεται με ολοκλήρωση:

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x_n^m \psi_n^m(t) \quad (3.5)$$

$$x_n^m = \int_{-\infty}^{\infty} x(t) \psi_n^m(t) dt. \quad (3.6)$$

3.3.1 Multiresolution ανάλυση

Ένα μεγάλο πλεονέκτημα που έχουν οι βάσεις Wavelet είναι ότι ορίζουν με φυσικό τρόπο μια ανάλυση σε πολλαπλές κλίμακες. Μια ανάλυση σε πολλαπλές κλίμακες είναι τυπικά ένα σύνολο απο διανυσματικούς χώρους (σημάτων) V_m που έχουν τις ιδιότητες

$$V_m \subset V_{m+1} \quad (3.7)$$

$$\bigcup_{m=-\infty}^{\infty} V_m = V \quad (3.8)$$

$$\bigcap_{m=-\infty}^{\infty} V_m = \emptyset \quad (3.9)$$

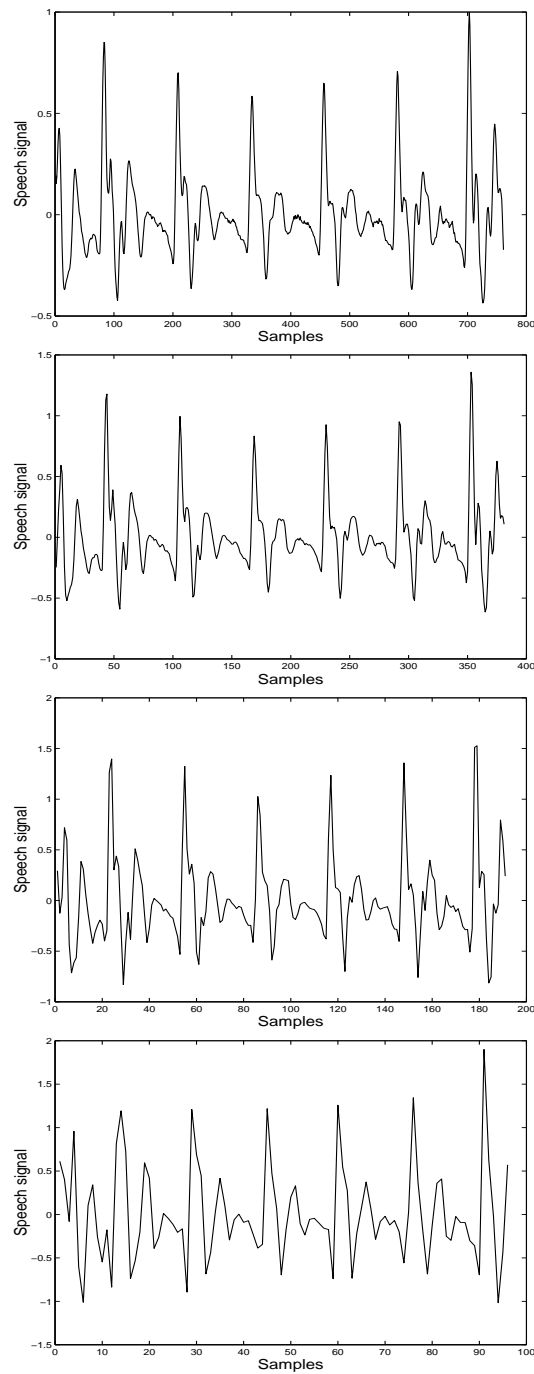
Ο χώρος V_m περιέχει τις προσεγγίσεις των σημάτων σε ανάλυση 2^m . Συνεπώς όσο μεγαλύτερη ανάλυση έχουμε τόσο πιο μεγάλη λεπτομέρεια βλέπουμε στο σήμα που μελετάμε. Μπορούμε να χωρίσουμε τον χώρο V_{m+1} στον υπόχωρο του V_m και στον κάθετο υπόχωρο O_m . Σε αυτή την περίπτωση ισχύει

$$O_m \perp V_m \quad (3.10)$$

$$O_m \oplus V_m = V_{m+1} \quad (3.11)$$

$$V = \bigoplus_{m=-\infty}^{\infty} O_m. \quad (3.12)$$

Όπου με \oplus συμβολίζουμε το ευθύ άθροισμα δύο διανυσματικών χώρων. Οι υπόχωροι V_m ονομάζονται χώροι προσέγγισης ενώ οι O_m χώροι λεπτομέρειας. Διασθητικά καταλαβαίνει κανείς πως ο χώρος O_m περιέχει την λεπτομέρεια που λείπει απο τον χώρο V_m για να 'ανέβει' στον λεπτομερέστερο χώρο προσέγγισης V_{m+1} . Μπορεί να αποδειχθεί πως ένα Mother wavelet ψ ορίζει μια Multiresolution ανάλυση. Συγκεκριμένα οι συναρτήσεις βάσης ψ_n^m για συγκεκριμένη κλίμακα m ορίζουν έναν χώρο λεπτομέρειας O_m . Οι αντίστοιχοι χώροι προσέγγισης V_m ορίζονται απο συναρτήσεις βάσης που ονομάζονται



Σχήμα 3.3: Προβολή σήματος φωνής σε χώρους προσέγγισης $V_0, V_{-1}, V_{-2}, V_{-3}$

Father wavelets ϕ_n^m και προκύπτουν σαν γραμμικός συνδιασμός των Mother wavelets. Οι συναρτήσεις ϕ_n^m ορίζονται με τον ίδιο τρόπο σαν μετακινήσεις και συμπιέσεις μιας βασικής συνάρτησης:

$$\phi_n^m(t) = 2^{m/2} \phi(2^m t - n). \quad (3.13)$$

Το ενδιαφέρον είναι πως υπάρχει ένας πολύ απλος χαρακτήρισμός που καθορίζει την συμπεριφορά που έχουν οι συντελεστές λεπτομέρειας x_n^m για ανελίξεις $1/f$.

3.3.2 Wavelets και $1/f$

Η σύνθεση Wavelets και $1/f$ στοχαστικών ανελίξεων αποτελεί σήμερα ενεργό χώρο έρευνας. Όλα τα αποτελέσματα που αναφέρονται προέκυψαν την τελευταία δεκαετία. Μια πολύ καλή αναφορά που αναλύει τα αποτελέσματα που χρησιμοποιούμε είναι [17]. Το βασικό αποτέλεσμα είναι πως μια ορθοκανονική βάση wavelet είναι πολύ κοντά στην Karhunen-Loéne βάση της ανελίξης. Αυτό σημαίνει πως οι συντελεστές wavelet x_n^m είναι ασυσχέτιστες τυχαίες μεταβλητές. Στην πραγματικότητα αυτό το αποτέλεσμα είναι προσεγγιστικό δηλαδή αυτό που έχει αποδείξει ο Wornell είναι πως η συσχέτιση των συντελεστών wavelet φθίνει εκθετικά γρήγορα με την απόσταση τους στην κλίμακα και στον χρόνο. Εάν δεχτούμε πως οι συντελεστές είναι πραγματικά ασυσχέτιστες τ.μ. τότε η ανάπτυξη σε βάσεις wavelet δημιουργεί μια νέα οικογένεια απο αλγόριθμους που μπορούν να χρησιμοποιηθούν τόσο για σύνθεση όσο και για εκτίμηση $1/f$ στοχαστικών ανελίξεων.

Όσον αφορά την σύνθεση, ο Wornell αποδεικνύει πως εάν η Wavelet βάση που χρησιμοποιείται έχει αρκετές μηδενιζόμενες ροπές τότε η ανελίξη που προκύπτει αν επιλέψουμε σαν συντελεστές κανονικές ανεξάρτητες τυχαίες μεταβλητές με μηδενική μέση τιμή και variance που ακολουθεί τον νόμο

$$\text{var}[x_n^m] = \sigma^2 2^{-\gamma m} \quad (3.14)$$

θα έχει φάσμα που θα είναι 'nearly $1/f$ ' δηλαδή θα έχει φάσμα με την ιδιότητα

$$\frac{\sigma_L^2}{|\omega|^\gamma} \leq S_x(\omega) \leq \frac{\sigma_U^2}{|\omega|^\gamma} \quad (3.15)$$

για κάποιες σταθερές σ_L και σ_U .

Το θεώρημα αυτό οδηγεί στον ακόλουθο αλγόριθμο σύνθεσης¹ που προσθέτει λεπτομέρεια επαναληπτικά: Αρχικά επιλέγεται ένα (τετριμμένο ή μή) αρχικό σήμα. Στην συνέχεια υπολογίζονται οι συντελεστές λεπτομέρειας ψ_n^m για μια κλίμακα m . Αυτοί είναι

¹Ο επαναληπτικός αλγόριθμος σύνθεσης με wavelets είναι (απο όσο γνωρίζω) προτότυπος

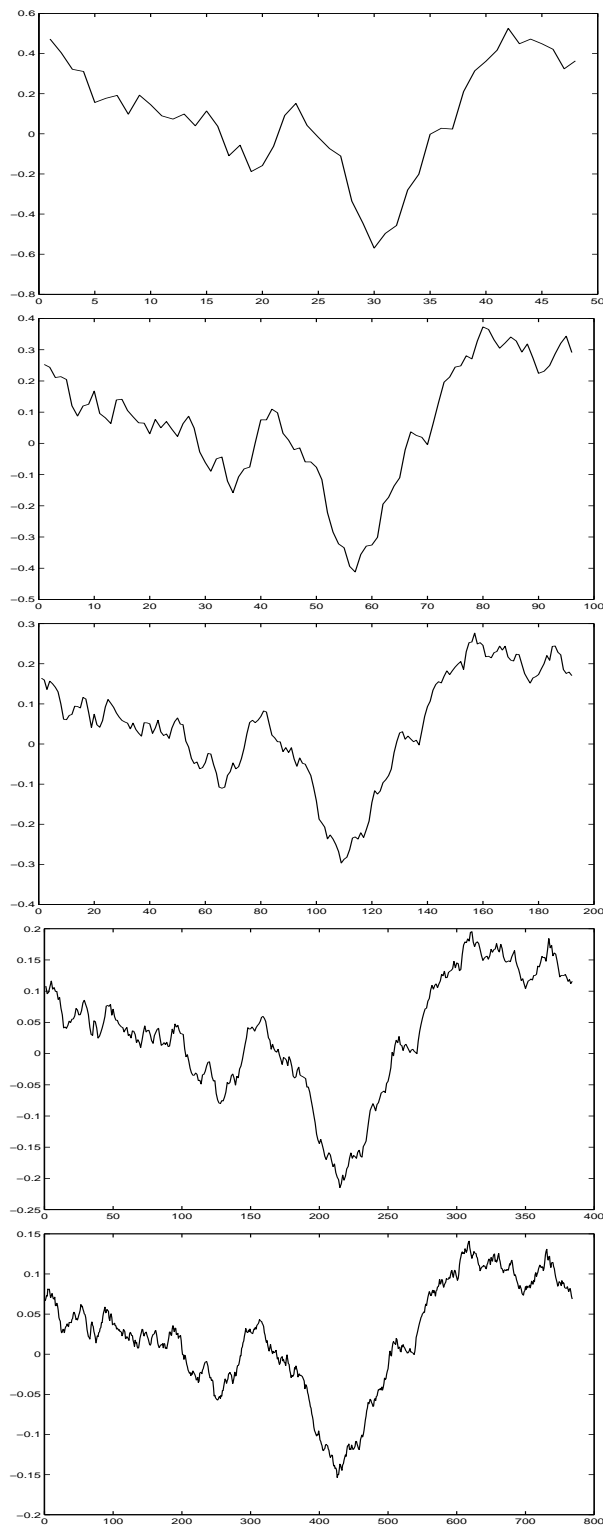
ανεξάρτητες κανονικές τυχαίες μεταβλητές με variance που δίνεται από την (3.14). Με αυτούς τους συντελεστές κατασκευάζεται ένα νέο σήμα (με διπλάσιο αριθμό δειγμάτων) που προκύπτει αν χρησιμοποιήσουμε σαν συντελεστές προσέγγισης ϕ , το αρχικό σήμα και σαν συντελεστές λεπτομέρειας ψ τις τυχαίες μεταβλητές. Η μεθοδολογία αυτή εικονίζεται στο Σχήμα 3.3.2

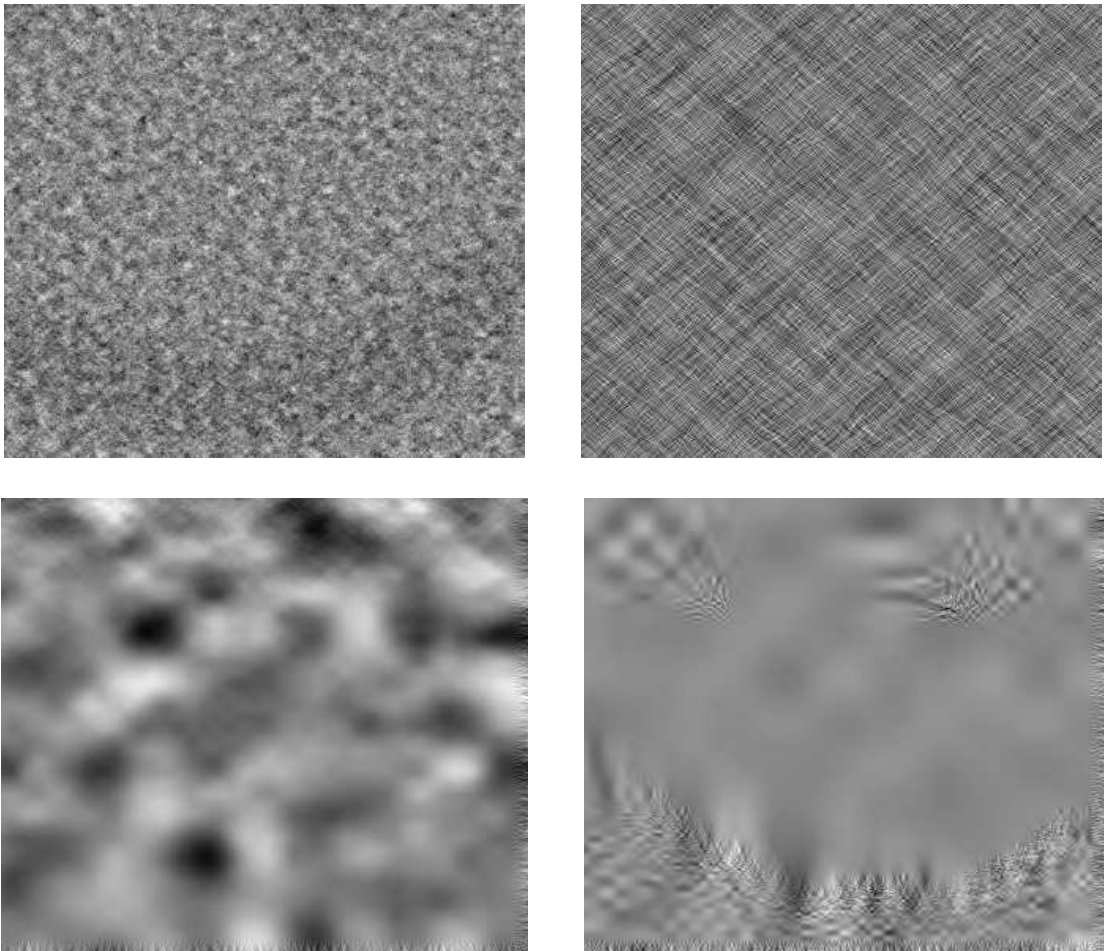
3.4 Σύνθεση διδιάστατων fractals

Η επαναληπτική μέθοδος σύνθεσης με wavelets μπορεί να γενικευτεί σε δύο διαστάσεις. Ο τρόπος γενίκευσης εξαρτάται από το είδος των διδιάστατων wavelet που θα χρησιμοποιηθούν. Στην απλούστερη περίπτωση διδιάστατων wavelets υπάρχει μια βάση προσέγγισης (όπως και στην μονοδιάστατη περίπτωση) αλλά τρεις βάσεις λεπτομέρειας που αντιστοιχούν σε οριζόντια, κατακόρυφη και διαγώνια λεπτομέρεια. Κάτω από αυτόν τον φορμαλισμό, μπορούμε να γενικεύσουμε τον νόμο του variance (3.14) σε δύο διαστάσεις. Σε αυτή την περίπτωση, από τον συντελεστή γ περνάμε σε ένα διάνυσμα $\Gamma = [\gamma_h, \gamma_v, \gamma_d]$ με 3 συνιστώσες που δίνουν το πώς μεταβάλλεται το variance των συντελεστών wavelet για τους συντελεστές οριζόντιας, διαγώνιας και κατακόρυφης λεπτομέρειας. Αυτός ο φορμαλισμός έχει πολλά πλεονεκτήματα στην σύνθεση διδιάστατων στοχαστικών fractal. Συγκεκριμένα μπορεί κανείς να επιλέξει διαφορετικό γ εκθέτη (και συνεπώς διαφορετική fractal διάσταση) για τις οριζόντιες, τις κατακόρυφες και τις διαγώνιες μεταβολές τις εικόνες. Στην εικόνα (3.4) τα σχήματα απεικονίζουν (α) διδιάστατη $1/f$ εικόνα με $\gamma_h = \gamma_v = \gamma_d = 2$ (β) $\gamma_h = \gamma_v = 3, \gamma_d = 1.5$. Επίσης ο φορμαλισμός μπορεί να επεκταθεί για την σύνθεση Multifractal εικόνων. Με τον όρο Multifractal² εννοούμε πως οι γ εκθέτες δεν είναι σταθεροί αλλά είναι χωρικές συναρτήσεις $\gamma(x, y)$. Σε αυτή την περίπτωση μπορεί να χρησιμοποιηθεί μια εικόνα-μάσκα που δίνει την τοπική τιμή των εκθετών. Οι εικόνες (γ) και (δ) είναι τέτοια Multifractals με εκθέτες που στην (γ) μειώνονται με τον κατακόρυφο άξονα Y ενώ στην (δ) περίπτωση είναι ένα μια εικόνα προσώπου που χαμογελάει.

Η επέκταση των μοντέλων $1/f$ και Multifractals σε δύο διαστάσεις καθώς και οι εφαρμογές στην Όραση Υπολογιστών (για παράδειγμα στην μοντελοποίηση υψής) είναι νέα πεδία έρευνας και πολλά παραμένουν να γίνουν σε αυτόν τον τομέα. Το ασθενέστερο σημείο αυτήν την στιγμή είναι η εκτίμηση όπου ακόμη και στην μία διάσταση παρουσιάζει σημαντικές δυσκολίες.

² Αυτός είναι ο απλούστερος ορισμός του Multifractal

Σχήμα 3.4: Επαναληπτική σύνθεση $1/f$ με Wavelets

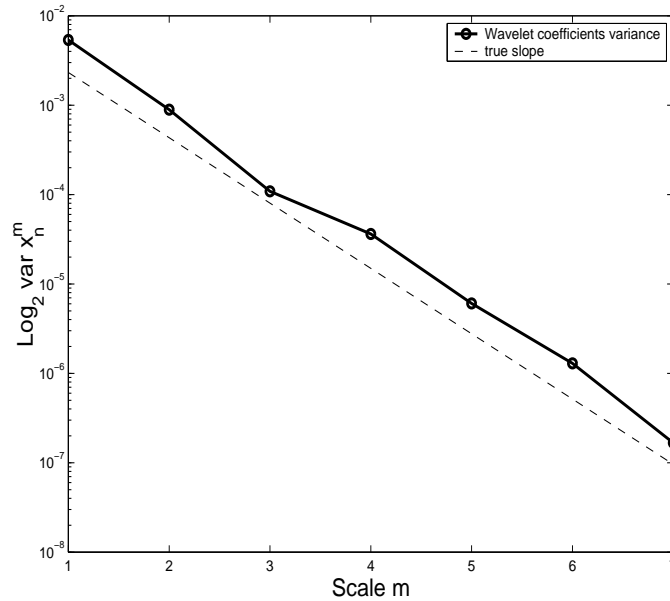


Σχήμα 3.5: Σύνθεση 2d fractals και Multifractals με Wavelets

3.5 Εκτίμηση με wavelets

Ο νόμος του variance (3.14) για τους συντελεστές λεπτομέρειας της ανάλυσης Wavelet μιας $1/f$ ανέλιξης παρέχει μια απλή μή παραμετρική μέθοδο εκτίμησης. Ο αλγόριθμος είναι ο ακόλουθος:

1. Αρχικά γίνεται Wavelet ανάλυση του σήματος, υπολογίζονται οι συντελεστές x_n^m .
2. Για κάθε κλίμακα m υπολογίζεται το Sample variance των x_n^m .



Σχήμα 3.6: Εκτίμηση γ με Wavelet variance

3. Αφού $\text{var}[x_n^m] \propto 2^{-\gamma m}$ ισχύει

$$\log_2(\text{var}[x_n^m]) = -\gamma m \quad (3.16)$$

Συνεπώς το γ μπορεί εκτιμηθεί με ελάχιστα τετράγωνα σαν κλίση ευθείας σε \log_2 -Scale plot του variance

Η μέθοδος αυτή εκτίμησης απεικονίζεται στο Σχήμα 3.5 και αποτελεί την απλούστερη μέθοδο που βασίζεται σε Wavelets.

Οι Wornell και Oppenheim στην προσπάθεια να δημιουργήσουν εκτιμητές που είναι πιο αποτελεσματικοί εισήγαγαν [25] το 1992 έναν αλγόριθμο Expectation Maximization (EM) που είναι υπολογιστικά ταχύτερος και θεωρητικά έμοιαζε ελκυστικός. Ο Brett Ninness όμως έδειξε [26] το 1998 ότι όταν $\gamma > 1$ ο αλγόριθμος συγκλίνει σε λανθασμένη εκτίμηση με πιθανότητα 1. Επίσης στην εργασία του Ninness παρουσιάζονται πειραματικά αποτελέσματα που δείχνουν πως ο EM αλγόριθμος μπορεί να έχει σημαντικές αποκλίσεις από την πραγματική τιμή και πως αυτή η εκτίμηση χειροτερεύει όσο το γ γίνεται μεγαλύτερο.

Στην παρούσα εργασία χρησιμοποιήσαμε τον αλγόριθμο εκτίμησης που βασίζεται στο Wavelet variance καθώς είναι απλός, δεν κάνει επιπλέον υποθέσεις και είναι εύκολο να ελεγχθεί το πλήθος των κλιμάκων που συμμετέχουν στην εκτίμηση.

3.6 Εκτίμηση με variance

Ο νόμος του variance (2.23) μπορεί να χρησιμοποιηθεί για να γίνει μια εκτίμηση του εκθέτη γ στο πεδίο του χρόνου. Συγκεκριμένα οι Leu και Paramarcou [27] χρησιμοποιούν το block-T Allan variance το οποίο ορίζεται για ένα σήμα $y(t)$

$$V_y(T) = \frac{1}{2T^2} (\mathbb{E}[y(t+T) - y(t)]^2 + \mathbb{E}[y(t-T) - y(t)]^2); \quad (3.17)$$

και αποδεικνύουν πως όταν το σήμα είναι fBm ισχύει για αυτό ένα power law:

$$V_{BH}(T) = cT^{\gamma-1} \quad (3.18)$$

Όπου η παράμετρος T έχει την έννοια κλίμακας στην οποία λαμβάνονται οι διαφορές. Συνεπώς απο την σχέση (3.18) μπορεί να εκτιμηθεί ο εκθέτης γ με ελάχιστα τετράγωνα στο log-log plot του Allan variance σε σχέση με την παράμετρο κλίμακας T .

Στην παρούσα εργασία υλοποιήσαμε την εκτίμηση με variance και την αξιολογήσαμε σε σχέση με τις άλλες μεθόδους για την μοντελοποίηση φωνής.

3.7 Σύνθεση και εκτίμηση με Cholesky διάσπαση του πίνακα συμμεταβλητότητας

Στην εργασία [16] υπολογίζεται η πυκνότητα πιθανότητας των διαφορών του διακριτοποιημένου fBm και με βάση αυτή δίνεται ένας αλγόριθμος σύνθεσης και ένας αλγόριθμος εκτίμησης. Το βασικό πλεονέκτημα είναι πως όλη η ανάλυση γίνεται σε διακριτό χρόνο και έτσι μπορεί κανείς να αναφέρεται στην διακριτή παράγωγο του fBm χωρίς να έχει τα προβλήματα που ανάγκασαν τον Mandelbrot να εξομαλύνει την κίνηση για να παραγωγίσει. Η ανάλυση ξεκινάει απο την διακριτοποίηση:

$$B_H[n] = B_H(nT_s). \quad (3.19)$$

Στην συνέχεια μπορεί κανείς να ορίσει τις διαφορές ενός βήματος που λειτουργούν σαν διακριτή παράγωγος:

$$X[n] = B_H[n] - B_H[n-1] \quad (3.20)$$

Όπου η διακριτή ανέλιξη $X[n]$ ονομάζεται Discrete Fractional Gaussian Noise, DFGN. Με λίγους αλγεβρικούς υπολογισμούς προκύπτει πως τα second order statistics του FGN είναι

$$\mathbb{E}[X[n]] = 0 \quad (3.21)$$

$$r[n] = \mathbb{E}[X[m]X[m+n]] = \frac{\sigma}{2} [|n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H}] \quad (3.22)$$

Το φάσμα μπορεί να υπολογιστεί απο τον μετασχηματισμό Fourier της αυτοσυσχέτισης αλλά αυτός ο υπολογισμός είναι πολύ δύσκολο να γίνει αναλυτικά. Όλα τα δείγματα του DFGN είναι απο κοινού Gaussian και με γνωστό πίνακα συμμεταβλητότητας και συνεπώς η πυκνότητα πιθανότητας προκύπτει άμεσα:

$$P(\mathbf{x}; H) = \frac{1}{(2\pi)^{N/2} |R|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{x}^T R^{-1} \mathbf{x}\right) \quad (3.23)$$

όπου $\mathbf{x} = [X[0], X[1], \dots, x[N-1]]^T$ οι τιμές του τυχαίου διανύσματος και R ο πίνακας συμμεταβλητότητας

$$R = E[\mathbf{x}\mathbf{x}^T] \quad (3.24)$$

δηλαδή με στοιχεία $R_{ij} = r[|i-j|]$.

Ο R είναι θετικά ορισμένος και συνεπώς μπορεί να εφαρμοσθεί σε αυτόν Cholesky Decomposition:

$$R = LL^T \quad (3.25)$$

Όπου L είναι κάτω τριγωνικός. Η ιδέα εδώ είναι πως εάν εφαρμόσουμε τον μετασχηματισμό

$$\mathbf{x} = L\mathbf{y} \quad (3.26)$$

τότε

$$E[\mathbf{y}\mathbf{y}^T] = E[L^{-1}\mathbf{x}\mathbf{x}^T L^{-T}] = I \quad (3.27)$$

Δηλαδή ο πίνακας συμμεταβλητότητας του \mathbf{y} είναι μοναδιαίος και συνεπώς τα στοιχεία του είναι ασυσχέτιστα. Με αυτή την παρατήρηση μπορεί να κατασκευαστεί ένας αλγόριθμος σύνθεσης που βασίζεται στην μέθοδο Cholesky: Αρχικά υπολογίζεται ένα διάνυσμα \mathbf{y} απο τυχαίες ανεξάρτητες κανονικές μεταβλητές. Για το δεδομένο γ που θέλουμε να συνθέσουμε, υπολογίζεται ο πίνακας R και στην συνέχεια απο αυτόν με την μέθοδο Cholesky ο πίνακας L . Τότε το ζητούμενο DFGN x , δίνεται απο τον μετασχηματισμό (3.26). Τέλος, το διακριτό fBm υπολογίζεται με συσσωρευτικό άθροισμα του x :

$$B_H[n] = \sum_{i=0}^{i=n} X[i] \quad (3.28)$$

3.7.1 Εκτιμητής Μέγιστης πιθανοφάνειας

Οι εκτιμητές μέγιστης πιθανοφάνειας (Maximum Likelihood estimators, MLE) θεωρούνται θεωρητικά βέλτιστοι. Συγκεκριμένα μπορεί να αποδειχθεί ότι οι εκτιμητές MLE είναι ασυμπτωτικά Unbiased και efficient ενώ ακολουθούν ασυμπτωτικά Gaussian κατανομή.

Εφόσον έχει υπολογιστεί αναλυτικά η μορφή της απο κοινού πυκνότητας πιθανότητας μπορεί να κατασκευαστεί [16] ένας MLE εκτιμητής. Συγκεκριμένα, το παρατηρούμενο δείγμα \mathbf{x} που θέλουμε να εκτιμήσουμε αντικαθίσταται στην σχέση (3.23). Τότε, η σχέση (3.23) γίνεται μια συνάρτηση μόνο της μεταβλητής H και αρκεί να βρεθεί η τιμή που την μεγιστοποιεί. Το πρόβλημα είναι πως συχνά (όπως και στην περίπτωση του fBm) δεν είναι δυνατόν να λυθεί αναλυτικά το πρόβλημα της μεγιστοποίησης. Σε αυτές τις περιπτώσεις απαιτείται η χρήση αριθμητικών μεθόδων για την εύρεση του μέγιστου (το οποίο ενδέχεται να μην υπάρχει ή να μην είναι μοναδικό).

Για την περίπτωση μας, παρόλο που αυτός ο εκτιμητής είναι ο θεωρητικά ορθότερος (ειδικά στην παρουσία θορύβου) υπάρχουν πολλά προβλήματα που τον καθιστούν ακατάλληλο για χρήση στην μοντελοποίηση φωνής. Πέρα από το μεγάλο υπολογιστικό του βάρος το βασικό του πρόβλημα είναι ότι είναι μια παραμετρική μέθοδος που βασίζεται στον μοντέλο fBm. Στην πράξη όμως, όπως θα φανεί και σε επόμενο κεφάλαιο, τα σήματα που εξάγουμε από την φωνή είναι low pass filtered 1/f ανελιζέες δηλαδή ταιριάζουν με το μοντέλο fBm για περιορισμένες κλίμακες (ή συχνότητες). Συνεπώς, για να έχουμε αποτελεσματική εκτίμηση απαιτείται η χρήση μη παραμετρικών μεθόδων στις οποίες μπορεί κανείς να ρυθμίσει το εύρος των κλιμάκων (ή συχνοτήτων) που συμμετέχουν στην εκτίμηση.

3.8 Φασματική εκτίμηση με την μέθοδο GPH

Η μέθοδος φασματικής εκτίμησης GPH προτάθηκε από τους Geweke και Porter-Hudak [28] και βασίζεται στο γεγονός πως όταν έχουμε ένα σήμα $x(t)$ 1/f με φασματικό περιεχόμενο

$$S(\omega) \propto \frac{\sigma_x^2}{|\omega|^\gamma} \quad (3.29)$$

τότε εάν λογαριθμήσουμε αυτή την σχέση λαμβάνουμε

$$\log S(\omega) \simeq \log C - \gamma \log(\omega) \quad (3.30)$$

για μια σταθερά C .

Το βασικό πρόβλημα με όλες τις μεθόδους εκτίμησης που καθιστά την μέθοδο GPH πολύ αποτελεσματική είναι πως η σχέση (3.29) δεν ισχύει για όλες τις συχνότητες και συνήθως εμφανίζεται μια εκθετική πτώση για συχνότητες μεγαλύτερες από κάποιο όριο. Συνεπώς στην πραγματικότητα η σχέση (3.30) ισχύει όταν $\omega \mapsto 0^+$. Συνεπώς, το βασικό πλεονέκτημα είναι πως η μέθοδος GPH είναι μια τοπική φασματική μέθοδος που βασίζεται μόνο στο κομμάτι του φάσματος κοντά στις χαμηλές συχνότητες και με αυτό τον τρόπο

δεν επιρρεάζεται από πιθανό low pass φιλτράρισμα που ενδεχομένως εμφανίζεται για μεγαλύτερες συχνότητες.

Για την εκτίμηση του φάσματος ισχύος χρησιμοποιείται το περιοδόγραμμα με παράθυρο.

$$S(\omega) = \left| \frac{1}{\sqrt{2\pi \sum_{t=1}^n |h_t|^2}} \sum_{t=1}^n h_t x(t) e^{it\omega} \right|^2 \quad (3.31)$$

και το παράθυρο που χρησιμοποιείται

$$h_t = (1 - \exp(2\pi it/n))^p \quad (3.32)$$

προτάθηκε από τους Hurvich και Chen [29] και είναι ειδικά σχεδιασμένο για αυτή την εκτίμηση.

Αφού γίνει η εκτίμηση του φάσματος ισχύος, ο εκθέτης γ υπολογίζεται με την μέθοδο ελαχίστων τετραγώνων από την κλίση του φάσματος σε log-log διάγραμμα όπως φαίνεται στην σχέση (3.30).

Το πρόβλημα είναι ότι η σχέση (3.30) ισχύει όπως αναφέρθηκε μόνο για περιοχή κοντά στην μηδενική συχνότητα και συνεπώς στην εκτίμηση κλίσης δεν πρέπει να χρησιμοποιούνται όλες οι συχνότητες παρα μόνο ένα υποσύνολο από αυτές που ανοίγει σε ένα εύρος συχνοτήτων M . Η επιλογή αυτού του εύρους είναι ένα δύσκολο πρόβλημα που περιλαμβάνει ένα tradeoff: Μικρότερο M οδηγεί σε εκτιμήσεις που έχουν μεγάλο variance αλλά σίγουρα δεν είναι biased από κάποια εκθετική πτώση του φάσματος σε μεγαλύτερες συχνότητες. Αντίθετα μεγαλύτερο M οδηγεί σε εκτιμήσεις με μικρότερο variance (αφού υπάρχουν περισσότερα σημεία για να κάνει κανείς παρεμβολή) αλλά υπάρχει ο κίνδυνος η εκτίμηση να είναι διαφορετική από την πραγματική τιμή λόγω αλλαγής στην συμπεριφορά του φάσματος. Αρκετές τεχνικές έχουν προταθεί ώστε να βρεθεί η βέλτιστη τιμή του εύρους συχνοτήτων M [30] αλλά βασίζονται σε επιπλέον υποθέσεις. Μια απλή προσέγγιση είναι να δημιουργήσει κανείς ένα Dynamic GPH plot (εικόνα 5.1) να εκτιμήσει δηλαδή την κλίση για πολλά διαφορετικά M και να δείξει το πως μεταβάλλεται η εκτίμηση σαν συνάρτηση του M . Η μέθοδος αυτή αποτελεί ίσως την καλύτερη μέθοδο εκτίμησης για γενικά σήματα $1/f$ και είναι αυτή που συμπεριφέρθηκε με τον καλύτερο τρόπο στα πειράματα.

3.9 Αξιολόγηση αλγορίθμων σύνθεσης

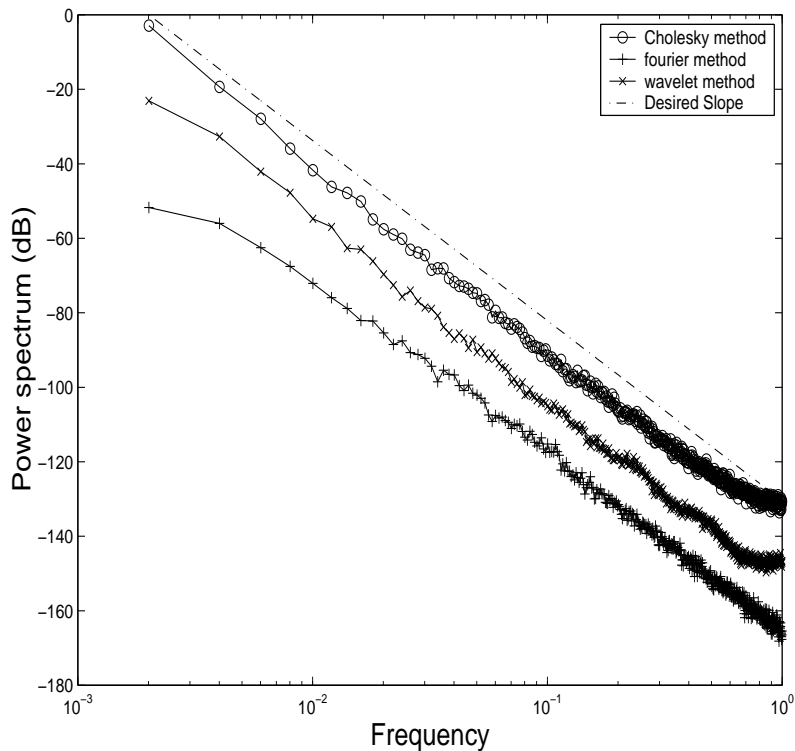
Οι μέθοδοι σύνθεσης $1/f$ ανελίξεων που υλοποιήθηκαν και αξιολογήθηκαν σε αυτή την εργασία είναι οι ακόλουθες:

- Σύνθεση με Fourier: Πλεονεκτήματα: Μεγάλη ταχύτητα, θεωρητικά ακριβής, φασματικό περιεχόμενο που ακολουθεί ακριβώς το $1/f$ νόμο.
Μειονεκτήματα: Η μέθοδος έχει το θεωρητικό πρόβλημα πως το σήμα που συντίθεται είναι περιοδικό. Αυτό μπορεί να ξεπεραστεί εάν συνθέσει κανείς ένα μεγαλύτερο σήμα από αυτό που χρειάζεται και κρατήσει ένα υποσύνολο του.
- Σύνθεση με Wavelets: Πλεονεκτήματα: Μεγάλη ταχύτητα, επαναληπτική σύνθεση. Η μέθοδος μπορεί να γενικευτεί εύκολα σε δύο διαστάσεις και σε Multifractal σήματα.
Μειονεκτήματα: Το φασματικό περιεχόμενο δεν ακολουθεί ακριβώς $1/f$ συμπεριφορά. Η επαναληπτική σύνθεση δεν επιτρέπει την ακριβή επιλογή του μήκους του υπο σύνθεση σήματος.
- Σύνθεση με Cholesky διάσπαση του πίνακα συμμεταβλητότητας: Πλεονεκτήματα: θεωρητικά ακριβής μέθοδος για το μοντέλο fBm. Μειονεκτήματα: Μεγάλο υπολογιστικό φορτίο και απαιτήσεις μνήμης, Το φασματικό περιεχόμενο δεν ακολουθεί με μεγάλη ακρίβεια $1/f$ συμπεριφορά.

Για να γίνει μια πειραματική αξιολόγηση του φασματικού περιεχομένου των σημάτων που συντίθενται με τις ακόλουθες μεθόδους έγινε το ακόλουθο πείραμα. Δημιουργήθηκαν 100 δείγματα με την κάθε μέθοδο, υπολογίστηκε το φάσμα τους με την μέθοδο του περιοδογράμματος και μετά υπολογίστηκε ο μέσος όρος για τα 100 δείγματα. Όπως φαίνεται και στην εικόνα 3.9 η μέθοδος fourier δίνει το καλύτερο φασματικό περιεχόμενο, αυτό δηλαδή που ακολουθεί καλύτερα την $1/f$ καμπύλη ενώ οι μέθοδοι wavelet και Cholesky δημιουργούν κάποιες αποκλίσεις για μικρές συχνότητες.

3.10 Αξιολόγηση αλγορίθμων εκτίμησης

Για την αξιολόγηση των αλγορίθμων εκτίμησης έγινε το ακόλουθο πείραμα: Για κάθε τιμή του εκθέτη γ δημιουργήθηκαν 100 υλοποιήσεις $1/f$ (με σύνθεση fourier) και στην συνέχεια εκτιμήθηκαν με τις τρεις μεθόδους GPH, Wavelet variance και Time-domain Variance. Τα αποτελέσματα φαίνονται στον πίνακα 3.10.



Σχήμα 3.7: Φάσμα τεχνητών σημάτων που δημιουργούνται απο αλγόριθμους σύνθεσης

γ	\overline{GPH}	Var(GPH)	$\overline{Wavelet}$	Var(Wavelet)	$\overline{T - Var}$	Var(T-Var)
1.5	1.52	0.12	1.38	0.01	1.57	0.01
2	1.98	0.13	1.82	0.02	1.92	0.01
2.4	2.39	0.12	2.25	0.02	2.23	0.01
2.8	2.87	0.14	2.70	0.03	2.55	0.01

Απο τα αποτελέσματα του πίνακα φαίνεται πως η μέθοδος που συμπεριφέρεται καλύτερα είναι η GPH. Παρόλλο που οι μέθοδοι Wavelet και Time variance έχουν σημαντικά μικρότερη διακύμανση, μοιάζουν να εμφανίζουν κάποιο bias που γίνεται εντονότερο για μεγαλύτερες τιμές του γ .

Η μέθοδος μέγιστης πιθανοφάνειας δεν χρησιμοποιήθηκε για δύο λόγους. Ό σημαντικότερος είναι οτι η μέθοδος υποθέτει ότι έχουμε δειγματοληπτημένες υλοποιήσεις fBm.

Όπως όμως αναφέρθηκε, (και αναλύεται περισσότερο σε επόμενο κεφάλαιο) στην πραγματικότητα είχαμε low pass filtered $1/f$ ανελίξεις. Σε αυτή την περίπτωση η πιθανοφάνεια δεν ήταν μια συνάρτηση με μοναδικό μέγιστο (όπως θα έπρεπε και πειραματικά παρατηρούν στο [16]) αλλά μια αύξουσα συνάρτηση του γ . Επίσης, ακόμη και για μικρά σήματα, η μέθοδος είναι εξαιρετικά επίπονη υπολογιστικά.

Οι άλλες μέθοδοι δεν είχαν πρόβλημα με το low pass filtered $1/f$ σήμα γιατί μπορεί κανείς να ρυθμίσει το πλήθος απο κλίμακες που συμμετέχουν στην εκτίμηση. Στην μέθοδο GPH αυτο σημαίνει πως ρυθμίζει κανείς το φασματικό εύρος M , στην μέθοδο Wavelet την μέγιστη και την ελάχιστη κλίμακα που συμμετέχουν στην παρεμβολή και τέλος στο variance την μέγιστη και την ελάχιστη τιμή της απόστασης T .

Κεφάλαιο 4

Μοντελοποίηση φωνής

4.1 Γραμμικό μοντέλο και Μοντέλο AM-FM

Οι συντονισμοί σε ένα σήμα φωνής (Speech Resonances) είναι οι συχνοτικές περιοχές όπου εμφανίζεται ισχυρό φασματικό περιεχόμενο. Δημιουργούνται όταν ένα τμήμα του φωνητικού σωλήνα διεγείρεται κοντά σε κάποια ιδιοσυχνότητα του και συντονίζεται. Ας σημειωθεί ότι ο φωνητικός σωλήνας έχει περιπλοκή γεωμετρία και οι ιδιοσυχνότητες του μεταβάλλονται ταχύτατα κατά την διάρκεια παραγωγής φωνής καθώς κινείται για να δημιουργήσει τους επιθυμητούς ήχους. Πρόκειται συνεπώς για ένα μη γραμμικό πρόβλημα ροής με περίπλοκες και δυναμικά μεταβαλλόμενες συντοριακές συνθήκες. Για να μπορέσει συνεπώς κανείς (με τις σημερινές γνώσεις και τεχνολογίες) να μοντελοποιήσει αυτό το φυσικό σύστημα είναι απαραίτητο να κάνει πολλές απλοποιητικές παραδοχές.

Το απλούστερο μοντέλο που χρησιμοποιήθηκε στο παρελθόν με σημαντική επιτυχία είναι το γραμμικό μοντέλο. Αυτό βασίζεται στο μοντέλο πηγής -φίλτρου (source-filter model) στο οποίο τα μη γραμμικά φαινόμενα που εμφανίζονται στον φωνητικό σωλήνα προσεγγίζονται με γραμμική ακουστική. Αυτή η προσέγγιση οδηγεί στο πολύ γνωστό μοντέλο γραμμικής πρόβλεψης, στο οποίο κάθε συντονισμός φωνής αντιστοιχεί σε έναν πόλο μιας συνάρτησης μεταφοράς.

Το γραμμικό μοντέλο έχει εφαρμοστεί στην κωδικοποίηση, σύνθεση και αναγνώριση φωνής με σχετική επιτυχία. Για να κατασκευαστούν πιο αποτελεσματικές εφαρμογές, αποκλίσεις από το γραμμικό μοντέλο μοντελοποιούνται συχνά με όρους σφάλματος.

Υπάρχουν ισχυρά θεωρητικά και πειραματικά στοιχεία [32, 33, 34] για την ύπαρξη σημαντικών μη γραμμικών αεροδυναμικών φαινομένων κατά την διάρκεια παραγωγής φωνής που δεν μπορούν να μοντελοποιηθούν με το γραμμικό μοντέλο. Μια προσέγγιση σε αυτό το πρόβλημα είναι να τροποποιηθεί κανείς το γραμμικό μοντέλο (που μπορεί να

θεωρηθεί μια προσέγγιση πρώτης τάξης) ώστε να ληφθούν υπόψη μη γραμμικά φαινόμενα. Οι μη γραμμικές ιδιότητες της φωνής μπορούν να ερευνηθούν με δυο διαφορετικές προσεγγίσεις:

- Ξεκινώντας από τις (μη γραμμικές) διαφορικές εξισώσεις Navier-Stokes να βρεθεί μια (προσεγγιστική) λύση για την ροή του αέρα μέσα στον φωνητικό σωλήνα όπως για παράδειγμα γίνεται στο [35]. Η προσέγγιση αυτή αποτελεί μια συνθετική στρατηγική που ξεκινάει από τους φυσικούς νόμους και προσεγγίζει το σύστημα παραγωγής φωνής.
- Μπορεί κανείς να ξεκινήσει αντίστροφα δημιουργώντας μη γραμμικά μοντέλα επεξεργασίας σήματος και να προσπαθήσει με αυτά να ανιχνεύσει μη γραμμικά φαινόμενα και χρήσιμη πληροφορία απευθείας από το σήμα της φωνής. Η δεύτερη αυτή μέθοδος αποτελεί μια προσέγγιση που ξεκινάει από το παρατηρήσιμο μέγεθος (το σήμα φωνής) και προσπαθεί να εξάγει πληροφορίες από αυτό.

Η προσέγγιση που υιοθετούμε σε αυτή την εργασία είναι αυτή της επεξεργασίας σήματος. Συγκεκριμένα, βασιζόμενοι στο μη γραμμικό μοντέλο διαμορφώσεων AM-FM παρατηρούμε ορισμένα σημαντικά χαρακτηριστικά για τυρβώδεις ήχους.

Η διαμόρφωση πλάτους και συχνότητας έχει χρησιμοποιηθεί εκτεταμένα σε συστήματα τηλεπικοινωνιών για να αυξήσει την αντοχή της μεταδιδόμενης πληροφορίας σε θόρυβο. Οι Maragos, Kaiser και Quatieri [12] εισήγαγαν την χρήση μοντέλων AM και FM στην μοντελοποίηση συντονισμών φωνής και πρότειναν τον αλγόριθμο διαχωρισμού ενέργειας ESA για την αποτελεσματική αποδιαμόρφωση σημάτων που περιέχουν τέτοιες διαμορφώσεις. Στο μη γραμμικό AM-FM μοντέλο, οι συντονισμοί μοντελοποιούνται σαν αποσβενόμενα ημίτονα διαμορφωμένα κατά πλάτος και συχνότητα:

$$x(t) = A(t)\cos[\phi(t)] = A(t)\cos\left[2\pi \int_0^t F(\tau)d\tau\right] \quad (4.1)$$

και το συνολικό σήμα δίνεται σαν μια υπέρθεση από τέτοια σήματα AM-FM

$$\sum_k A_k(t)\cos[\phi_k(t)], \quad (4.2)$$

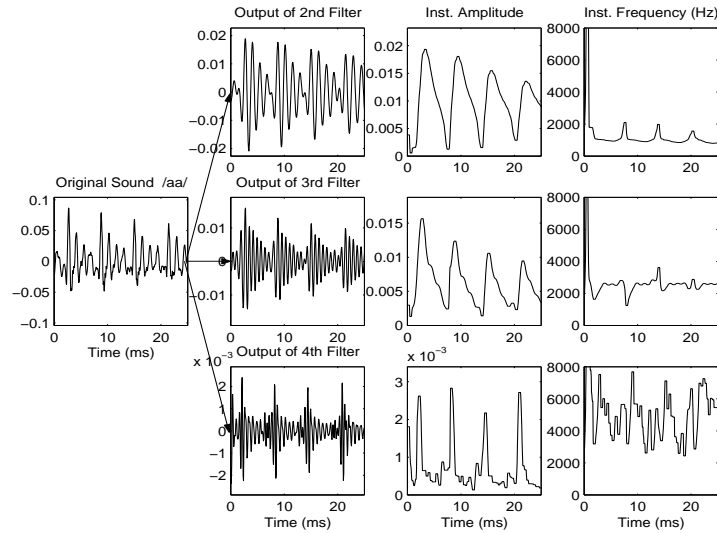
ένα για κάθε συντονισμό. Το $A(t)$ είναι το στιγμιαίο πλάτος και $F(t)$ είναι η στιγμιαία κυκλική συχνότητα του συντονισμού.

Εάν πάρουμε την μέση τιμή της στιγμιαίας συχνότητας για ένα μικρό χρονικό παράθυρο

$$F_c = (1/T) \int_0^T F(t)dt, \quad (4.3)$$

όπου T είναι συγκρίσιμο με την διάρκεια του pitch(10-30ms), τότε το F_c μπορεί να θεωρηθεί σαν συχνότητα του φέροντος γύρω απο την οποία γίνεται η διαμόρφωση.

Σε αντίθεση με το γραμμικό μοντέλο, το μοντέλο AM-FM εμπεριέχει τόσο την μέση συχνότητα του συντονισμού F_c όσο και επιπλέον πληροφορία για την απόκλιση τις στιγμιαίας συχνότητας απο την μέση της τιμή.



Σχήμα 4.1: Αποδιαμόρφωση φωνήματος με χρήση Gabor φίλτρων και SESA. Απο [11].

4.2 Αποδιαμόρφωση σημάτων AM-FM

Το πρόβλημα της αποδιαμόρφωσης συνίσταται στο πως μπορεί κανείς αν λάβει ένα απομονωμένο σήμα συντονισμού της μορφής (4.1) να ανακτήσει το στιγμιαίο πλάτος $A(t)$ και την στιγμιαία συχνότητα $F(t)$. Ο αλγόριθμος διαχωρισμού ενέργειας ESA [12] επιτυγχάνει την αποδιαμόρφωση με την χρήση του μη γραμμικού τελεστή Teager-Kaiser.

$$\Psi[x(t)] = [\dot{x}(t)]^2 - x(t)\ddot{x}(t) \quad (4.4)$$

Όπου $\dot{x} = dx/dt$. Με την χρήση αυτού του τελεστή μπορεί κανείς να υπολογίσει τα σήματα στιγμιαίου πλάτους και στιγμιαίας συχνότητας με τον ακόλουθο τρόπο:

$$\sqrt{\frac{\Psi[\dot{x}(t)]}{\Psi[x(t)]}} \approx 2\pi F(t) \quad , \quad \frac{\Psi[x(t)]}{\sqrt{\Psi[\dot{x}(t)]}} \approx |A(t)| \quad (4.5)$$

Που ονομάζεται αλγόριθμος διαχωρισμού ενέργειας ESA (Energy Separation Algorithm) [12]. Ο ESA εκτελεί AM-FM αποδιαμόρφωση με το να εντοπίζει την φυσική ενέργεια της πηγής και να την χωρίζει σε ενέργεια λόγω πλάτους ταλάντωσης και ενέργεια λόγω συχνοτικής μεταβολής.

Ο αλγόριθμος δίνει πολύ καλές εκτιμήσεις για τα σήματα $A(t)$, $F(t)$ όταν αυτά δεν μεταβάλλονται πολύ γρήγορα σε σχέση με την φέρουσα συχνότητα.

Η διακριτή μορφή του αλγορίθμου ESA ονομάζεται *DESA*. Ένα θεμελιώδες πρόβλημα στην διακριτοποίηση είναι το πως θα υπολογιστεί σε διακριτό χρόνο η τιμές των παραγώγων που χρειάζονται για τον αλγόριθμο. Η απλούστερη μέθοδος είναι να χρησιμοποιήσει κανείς διαφορές ενός βήματος (one step differences). Αυτή η προσέγγιση είναι ικανοποιητική όταν δεν υπάρχει θόρυβος. Στην αντίθετη περίπτωση μπορεί να χρησιμοποιηθεί ένα απλό smoothing του σήματος ενέργειας. Επίσης, όπως έχει δείχθει στο [13] το ζωνοπερατό φίλτράρισμα που χρησιμοποιείται για να απομονώσει κάθε συντονισμό βοηθά στην περίπτωση του θορύβου καθώς περιορίζει το εύρος του μόνο στην φασματική περιοχή του φίλτρου.

Πρόσφατα, οι Dimitriadis, Maragos [14, 11] πρότειναν πρόσφατα την χρήση Splines για τον υπολογισμό παραγώγων με περισσότερη ευρωστία σε θόρυβο (Spline-ESA).

Συνοψίζοντας, μπορούμε να πούμε πως ο αλγόριθμος ESA είναι ένας πολυ αποτελεσματικός τρόπος αποδιαμόρφωσης για τους ακόλουθους λόγους:

1. Έχει πολύ χαμηλή υπολογιστική πολυπλοκότητα ($O(N)$, όπου N είναι το μήκος του παραθύρου ανάλυσης).
2. Δημιουργεί πολύ μικρά σφάλματα στην αποδιαμόρφωση.
3. Έχει πολύ υψηλή χρονική διακριτική ικανότητα καθώς λειτουργεί σε ένα πολύ μικρό παράθυρο.
4. Έχει μια χρήσιμη και διαισθητική ερμηνεία και σχετίζεται με την ανίχνευση της φυσικής ενέργειας της ακουστικής πηγής.

Εκτεταμένες πειραματικές μελέτες [12, 36, 37] έχουν δείξει πως αυτές οι διαμορφώσεις κατά πλάτος/συχνότητα υπάρχουν σε πραγματικά σήματα φωνής και είναι αναγκαίες για την φυσικότητα της.

4.2.1 Συστοιχίες φίλτρων και αποδιαμόρφωση

Ο βασικός περιορισμός του ESA είναι πως απαιτεί το σήμα να είναι συχνοτικά εντοπισμένο (Narrowband) γύρω από την βασική συχνότητα φέροντος που αποτελεί και

την συχνότητα συντονισμού F_c . Αυτός ο περιορισμός καθιστά αναγκαία την χρήση συστοιχιών απο φίλτρα (Filter Banks) τα οποία έχουν σαν σκοπό να απομονώνουν τον κάθε συντονισμό. Συνεπώς, κάθε χρονικό κομμάτι του σήματος φωνής φιλτράρεται ταυτόχρονα απο ζωνοπερατά φίλτρα και η έξοδος του κάθε φίλτρου αποδιαμορφώνεται με την χρήση του ESA.

Η διαδικασία αυτή απεικονίζεται στην εικόνα 4.1

Κεφάλαιο 5

Ένα μοντέλο για τυρβώδη φωνήματα

Η βασική συνεισφορά αυτής της εργασίας είναι ότι προτείνει ένα μοντέλο διαμόρφωσης φάσης για τυρβώδη φωνήματα. Το μοντέλο αυτό και κάποια αρχικά πειραματικά αποτελέσματα έχουν δημοσιευτεί στο [1] ενώ η θεωρητική ανάλυση του επόμενου κεφαλαίου καθώς και περισσότερα πειραματικά στοιχεία ετοιμάζονται προς δημοσίευση[2]. Η ιδέα είναι να μοντελοποιήσει κανείς κάθε συντονισμό φωνής $R(t)$ με διαμόρφωση φάσης γύρω από μια κεντρική συχνότητα F_c :

$$R(t) = A \cos(2\pi F_c t + P(t)), \quad (5.1)$$

Όπου $P(t)$ είναι ένα σήμα διαμόρφωσης φάσης (Phase modulation) και η παράγωγος του είναι το σήμα διαμόρφωσης FM. Βασιζόμενοι σε ισχυρά πειραματικά στοιχεία (που θα παρουσιαστούν σε επόμενη ενότητα) καθώς και θεωρητικό υπόβαθρο που συνδέει τις αυτοόμοιες στοχαστικές ανελίξεις με την τυρβώδη ροή, μοντελοποιούμε το σήμα διαμόρφωσης $P(t)$ σαν μια στοχαστική ανέλιξη $1/f$. Κάτω από αυτόν τον φορμαλισμό η ανάγκη για πειραματικό έλεγχο του μοντέλου καθώς και η εκτίμηση των παραμέτρων του είναι τα πρώτα προβλήματα που καλούμαστε να επιλύσουμε. Σε αυτό το μοντέλο δεν ασχολούμαστε με την μορφή του σήματος διαμόρφωσης πλάτους $A(t)$ που φαίνεται να περιέχει πολύ λιγότερη χρήσιμη πληροφορία και στην παρακάτω ανάλυση θεωρείται σταθερό ($A(t) = A$).

5.1 Εξαγωγή χαρακτηριστικών και εκτίμηση παραμέτρων

Για να μπορέσουμε να ελέγξουμε και να εκτιμήσουμε το προτεινόμενο μοντέλο δημιουργήθηκε ο ακόλουθος αλγόριθμος που χωρίζει το πρόβλημα στα ακόλουθα υποπροβλήματα:

1. Απομόνωση του κάθε συντονισμού από το υπόλοιπο σήμα φωνής. Για να υλοποιηθεί αυτός ο φασματικός χωρισμός χρησιμοποιήθηκε ένα φίλτρο Gabor για δύο λόγους. Πρώτον τα φίλτρα Gabor εμφανίζουν ελάχιστο γινόμενο διασποράς στα πεδία χρόνου-συχνότητας και δεύτερον λόγω της συμμετρίας τους δεν εισάγουν επιπλέον διαμόρφωση συχνότητας στο φιλτραρισμένο σήμα.
2. Εξαγωγή των σημάτων AM και FM. Η αποδιαμόρφωση αυτή έγινε με χρήση του αλγορίθμου ESA και τα σήματα που εξάγονται είναι η στιγμιαία συχνότητα $F(t)$ και το στιγμιαίο πλάτος $A(t)$.
3. Ο αλγόριθμος ESA εμφανίζει υπο ορισμένες συνθήκες αστάθειες[12] που δημιουργούν έντονες κορυφές στο σήμα στιγμιαίας συχνότητας. Για τον περιορισμό αυτών των φαινομένων το σήμα $F(t)$ φιλτράρεται με ένα μη γραμμικό median φίλτρο που μειώνει τις μεγάλες ακμές.
4. Το σήμα διαμόρφωσης φάσης $\hat{P}(t)$ υπολογίζεται με το τρέχον ολοκλήρωμα της στιγμιαίας συχνότητας:

$$\hat{P}(t) = 2\pi \int_0^t (F(\tau) - F_c) d\tau, \quad (5.2)$$

όπου F_c είναι η μέση τιμή του $F(t)$ και αντιπροσωπεύει την κεντρική συχνότητα του συντονισμού.

5. Τέλος, ο εκθέτης γ υπολογίζεται από το σήμα διαμόρφωσης φάσης $\hat{P}(t)$

Το τελευταίο αυτό βήμα, είναι σίγουρα και το πιο περίπλοκο. Το πρόβλημα συνίσταται στον υπολογισμό του εκθέτη γ από μια (θορυβώδη) υλοποίηση του $\hat{P}(t)$ οπότε εφαρμόζονται όλες οι μέθοδοι εκτίμησης που αναλύονται στο προηγούμενο κεφάλαιο. Πριν όμως ασχοληθούμε με την εκτίμηση παραμέτρων θα εξετάσουμε την εγκυρότητα του μοντέλου.

5.2 Έλεγχος εγκυρότητας

Για να έλεγχο της εγκυρότητας του μοντέλου (5.1) υλοποιήθηκε ένα σύνολο απο ελέγχους που δοκιμάστικαν σε πλήθος απο φωνήματα απο την βάση δεδομένων ήχων φωνής TIMIT. Με βάση την ανάλυση του προηγούμενο κεφαλαίου είναι φανερό πως εάν το σήμα $P(T)$ είναι πραγματικά υλοποίηση μιας ανέλιξης $1/f$ θα πρέπει να έχει τις ακόλουθες ιδιότητες

1. Το φάσματος ισχύος του θα πρέπει να ακολουθεί έναν power law δηλαδή θα πρέπει να είναι ευθεία σε log-log άξονες.
2. Το variance των wavelet συντελεστών λεπτομέρειας θα πρέπει να ακολουθεί την σχέση (3.5) οπότε το Variance-Scale plot θα πρέπει να είναι ευθεία ημιλογαριθμικούς άξονες.
3. Στο πεδίο του χρόνου θα πρέπει να ισχύει η σχέση (2.23) και συνεπώς το variance θα πρέπει και αυτό να ακολουθεί power law και να είναι ευθεία σε log-log άξονες.

Στις εικόνες (5.2),(5.3) δίνεται η πορεία του αλγορίθμου εξαγωγής της στιγμιαίας φάσης και εκτίμησης κατα πόσο το σήμα διαμόρφωσης φάσης $P(t)$ είναι $1/f$ για μερικά φωνήματα. Η εκτίμηση του εκθέτη γ έχει γίνει με εκτίμηση GPH όπως περιγράφεται σε προηγούμενο κεφάλαιο.

Επίσης στο Σχήμα 5.5 δίνονται δύο σήματα διαμόρφωσης φάσης απο πραγματικά φωνήματα σε αντιπαραβολή με ένα τεχνητό σήμα $1/f$ για να φανεί και στο πεδίο του χρόνου η σχέση ανάμεσα στα πραγματικά σήματα και στο προτεινόμενο μοντέλο.

Όπως φαίνεται σε όλα τα πειράματα, η υπόθεση ότι η διαμόρφωση φάσης ακολουθεί $1/f$ συμπεριφορά επαληθεύεται και το μοντέλο μοιάζει να λειτουργεί αποτελεσματικά.

5.3 Τεχνητά σήματα συντονισμού και εκτίμηση

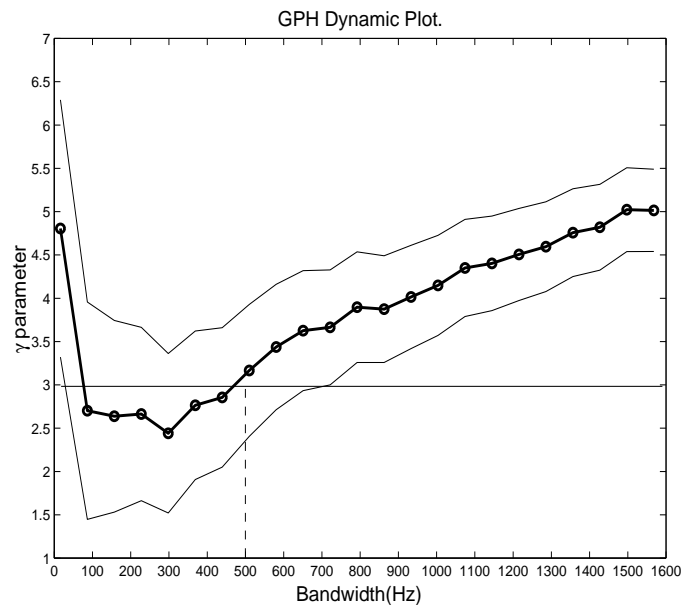
Για να ελέγξουμε την απόδοση του αλγορίθμου εξαγωγής του σήματος φάσης $P(t)$ δημιουργήσαμε τεχνητά σήματα συντονισμού με $1/f$ διαμόρφωση φάσης και συγκρίναμε το αρχικό σήμα $P(t)$ με το εκτιμώμενο $\hat{P}(t)$. Η σύνθεση του τεχνητού $1/f$ σήματος, $P(t)$ έγινε με Fourier σύνθεση όπως αυτή περιγράφεται σε προηγούμενο κεφάλαιο. Από τα αποτελέσματα αυτού του πειράματος οδηγούμαστε στα ακόλουθα συμπεράσματα τα οποία φαίνεται και στην εικόνα 5.3

- Ένα διαμορφωμένο κατα φάση $1/f$ σήμα δημιουργεί έναν ευρύ συντονισμό με αργή μείωση συχνοτήτων. Η ακριβής μορφή αυτού του φάσματος έχει υπολογιστεί αναλυτικά στην θεωρητική ανάλυση που ακολουθεί. Η παρατήρηση αυτή συμφωνεί με τους συντονισμούς μεγάλου εύρους που εμφανίζονται πειραματικά σε τυρβώδη φωνήματα.
- Το ανακατασκευασμένο σήμα διαμόρφωσης φάσης $\hat{P}(t)$ είναι μια low pass έκδοση του αρχικού σήματος $P(t)$. Αυτό είναι αποτέλεσμα του Gabor φιλτραρίσματος αλλά και της πεπερασμένης ικανότητας της διαμόρφωσης φάσης να μεταφέρει υψηλές συχνότητες.

Το πρόβλημα συνεπώς συνίσταται στο να εκτιμηθεί ο εκθέτης γ από μια low pass filtered $1/f$ ανέλιξη που έχει φάσμα όπως αυτό που φαίνεται στο (δ) της 5.3. Η μορφή αυτή του φάσματος που ακολουθεί $1/f$ νόμο για κάποιο εύρος συχνοτήτων M και στην συνέχεια εμφανίζει πτώση εμφανίζεται εκτός από το τεχνητό σήμα και στα πειράματα με πραγματικά φωνήματα (5.2) και (5.3).

Η καλύτερη μέθοδος για εκτίμηση low pass filtered $1/f$ αποδείχθηκε πειραματικά πως είναι η μέθοδος GPH με δυναμικό εύρος M . Στην εικόνα 5.1 απεικονίζεται η δυναμική GPH εκτίμηση. Στον οριζόντιο άξονα τοποθετείται η τιμή του εύρους του φάσματος που χρησιμοποιείται για την τρέχουσα εκτίμηση και στον κατακόρυφο ή εκτιμώμενη τιμή του γ για το εύρος αυτό. Η πραγματική τιμή του εκθέτη είναι $\gamma = 3$. Μπορεί να φανεί πως η εκτίμηση είναι σωστή για εύρος M μέχρι περίπου 500Hz. Για μεγαλύτερες συχνότητες η εκτίμηση ανεβαίνει σε εντελώς λανθασμένες τιμές καθώς έχουν αρχίσει να την επηρεάζει η εκθετική πτώση του φάσματος. Το αποτελεσματικό εύρος συχνοτήτων σχετίζεται με το εύρος του φίλτρου Gabor που χρησιμοποιείται για να απομονώσει τον συντονισμό. Πειραματικά στοιχεία δείχνουν πως αποτελεσματικές τιμές είναι της τάξης του $\frac{\alpha}{4}$ όπου α είναι η rms τιμή του φίλτρου Gabor.

Συνοψίζοντας, μπορούμε να πούμε πως το πρόβλημα της εκτίμησης του εκθέτη γ μιας $1/f$ ανέλιξης από μια υλοποίηση, δεν έχει βρεί ακόμη ικανοποιητική λύση στην βιβλιογραφία. Παρόλο που υπάρχει πληθώρα άρθρων πάνω στο θέμα της εκτίμησης



Σχήμα 5.1: Δυναμική GPH εκτίμηση τεχνητού $1/f$ σήματος

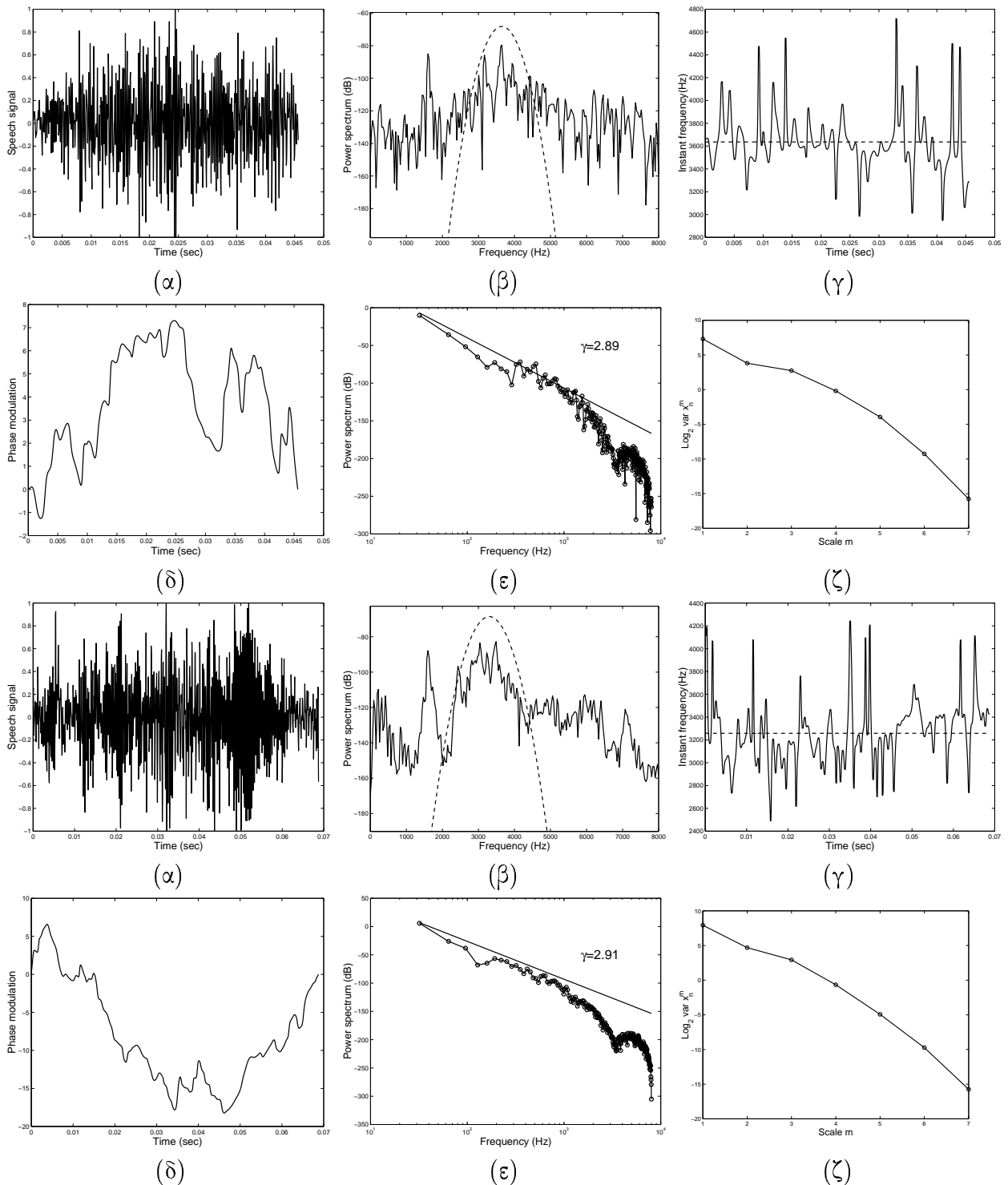
και πολλαπλές μέθοδοι έχουν προταθεί[25, 26, 27], τα σφάλματα εκτίμησης παραμένουν σημαντικά.

5.4 Συμπεράσματα των πειραμάτων

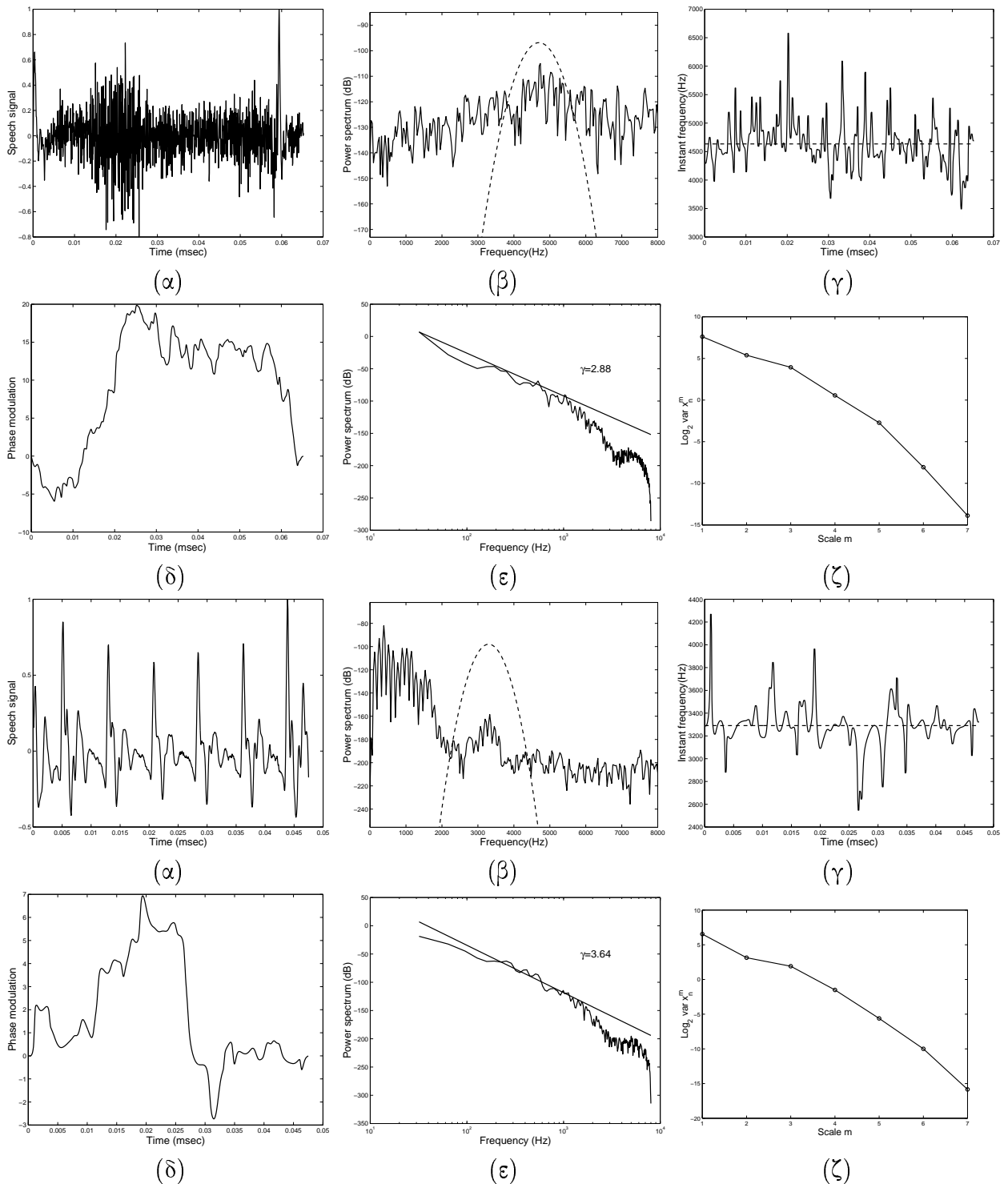
Όπως φαίνεται απο τα πειράματα, το μοντέλο διαμόρφωσης φάσης με σήματα $1/f$ είναι πολύ κοντά στα σήματα που εξάγονται πειραματικά. Σε αυτό το σημείο πρέπει να αναφερθεί πως το μοντέλο μοιάζει να είναι ακατάλληλο για άλλες κατηγορίες φωνημάτων όπως τα φωνήεντα. Αυτό οφείλεται στο ότι διαμορφωμένα κατά φάση σήματα $1/f$ δημιουργούν συντονισμούς με μεγάλο συχνοτικό εύρος (rms μεγαλύτερες των 1000Hz). Τέτοιου είδους συντονισμοί εμφανίζονται στα τυρβώδη φωνήματα (fricatives) (τόσο έμφωνα όσο και άφωνα) καθώς και (σε όχι τόσο μεγάλο βαθμό) σε φωνήματα εκρηκτικής φύσης (stops). Αντίθετα, σε φωνήεντα, το φασματικό περιεχόμενο της διαμόρφωσης φάσης φθίνει πολύ γρήγορα για να είναι αποτελεσματική μια μοντελοποίηση $1/f$. Για περιπτώσεις στενών φασματικά συντονισμών χρειάζεται μια διαφορετική μορφή σημάτων στιγμιαίας συχνότητας με μικρότερη μεταβλητότητα. Οι περιορισμοί του μοντέλου φαίνονται περισσότερο στα Σχήματα 5.4 όπου φαίνονται οι αποκλίσεις απο το power law τόσο στα φάσματα ισχύος

όσο και στους συντελεστές wavelet.

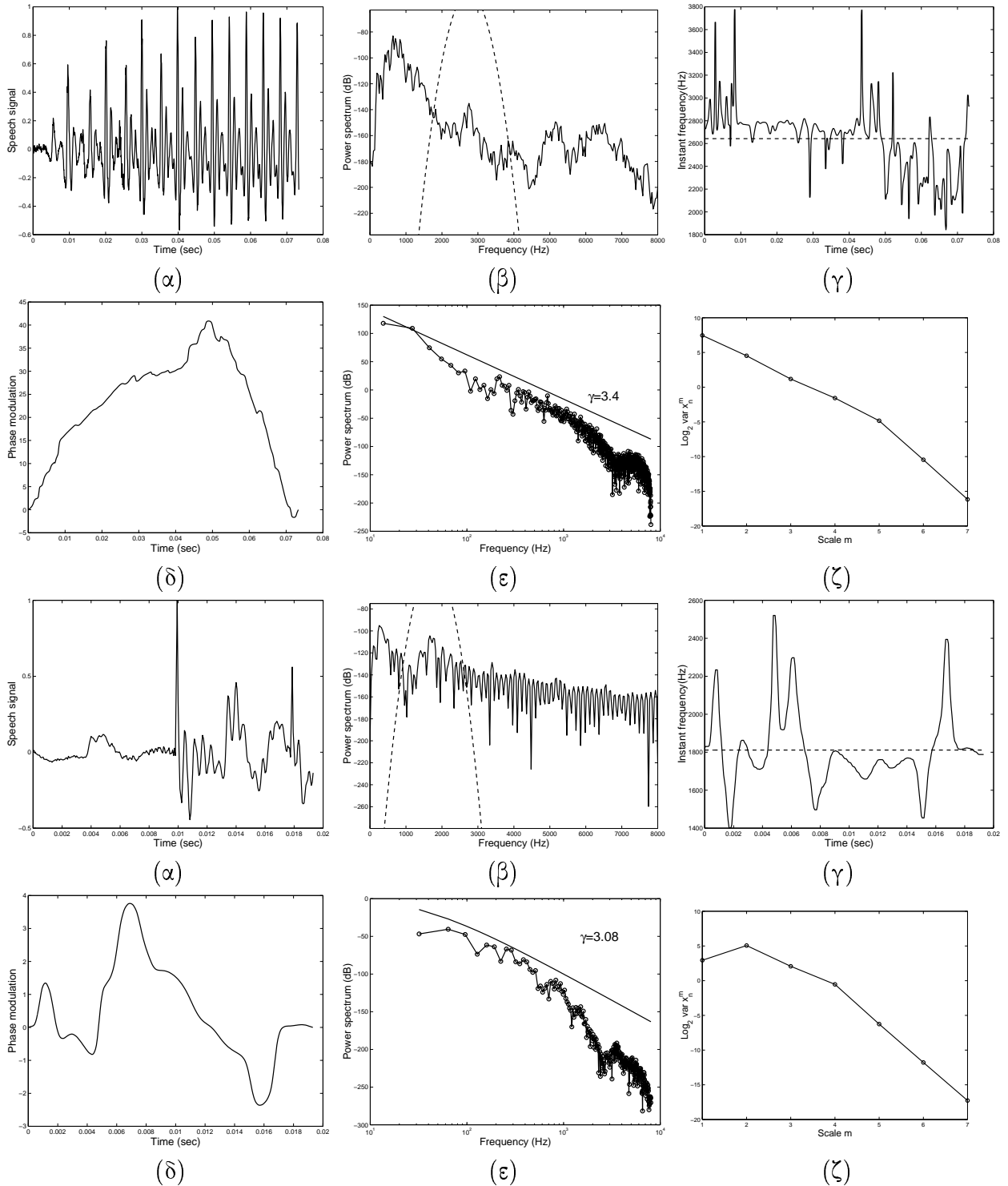
Ενδιαφέρον επίσης παρουσιάζει το εάν μπορεί κανείς να διακρίνει ανάμεσα σε έμφωνους ή άφωνους τυρβώδεις ήχους από τον εκτιμώμενο φασματικό εκθέτη γ . Μετά από εκτενέστατους πειραματισμούς καταλήξαμε στο συμπέρασμα πως οι άφωνοι τυρβώδεις ήχοι (όπως το /s/ και το /f/) εμφανίζουν, σε γενικές γραμμές, μικρότερους εκθέτες γ (και συνεπώς μεγαλύτερη Fractal διάσταση) από τους αντίστοιχους έμφωνους (/z/ και /v/) αλλά οι διαφορές είναι αρκετά μικρές ώστε να αναδειχθούν ξεκάθαρα από τα σφάλματα εκτίμησης που, όπως αναφέρθηκε, είναι σημαντικά.



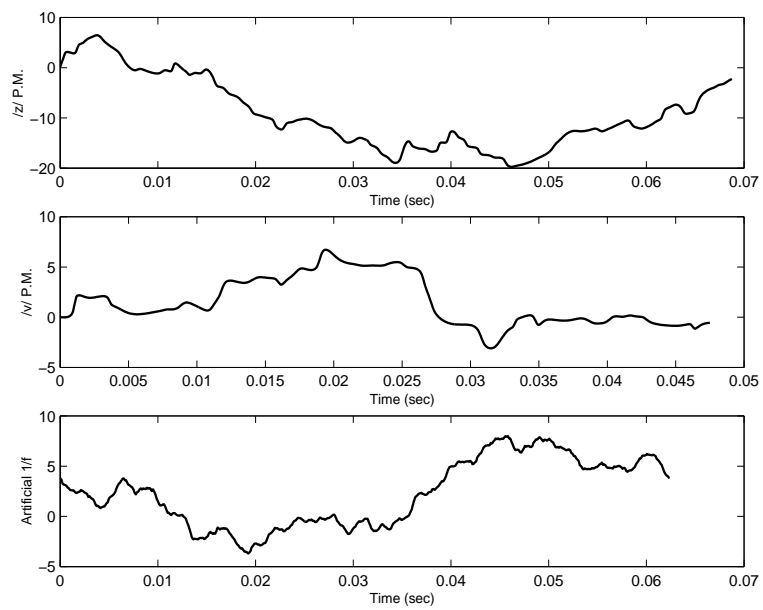
Σχήμα 5.2: Πειράματα με φωνήματα /s/ (άνω σειρά) και /z/ (κάτω σειρά). Στήλες: α) Σήμα φωνής $s(t)$. β) Φάσμα ισχύος του $s(t)$ και φίλτρο Gabor. γ) Στιγμιαία συχνότητα $F(t)$. δ) Διαμόρφωση φάσης $\hat{P}(t)$. ε) Φάσμα ισχύος του $\hat{P}(t)$ και εκτιμώμενη κλίση. ζ) variance των wavelet συντελεστών λεπτομέρειας



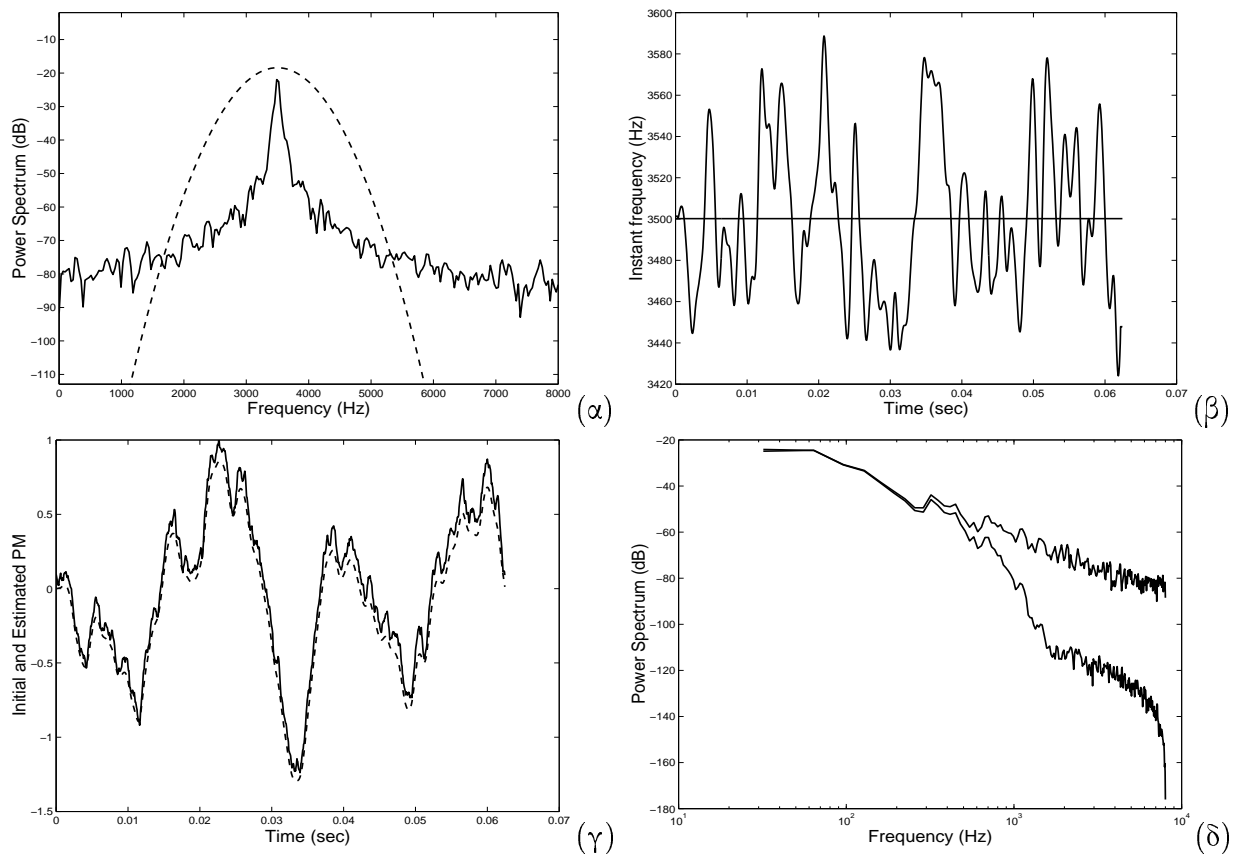
Σχήμα 5.3: Πειράματα με φωνήματα /f/ (άνω σειρά) και /v/ (κάτω σειρά). Στήλες: α) Σήμα φωνής $s(t)$. β) Φάσμα ισχύος του $s(t)$ και φίλτρο Gabor. γ) Στιγμιαία συχνότητα $F(t)$. γ) Διαμόρφωση φάσης $\hat{P}(t)$. ε) Φάσμα ισχύος του $\hat{P}(t)$ και εκτιμώμενη κλίση. ζ) variance των wavelet συντελεστών λεπτομέρειας



Σχήμα 5.4: Πειράματα με φωνήματα /a/ (άνω σειρά) και /b/ (κάτω σειρά). Στήλες: α) Σήμα φωνής $s(t)$. β) Φάσμα ισχύος του $s(t)$ και φίλτρο Gabor. γ) Στιγμιαία συχνότητα $F(t)$. δ) Διαμόρφωση φάσης $\hat{P}(t)$. ε) Φάσμα ισχύος του $\hat{P}(t)$ και εκτιμώμενη κλίση. ζ) variance των wavelet συντελεστών λεπτομέρειας



Σχήμα 5.5: Δύο πραγματικά σήματα διαμόρφωσης φάσης και ένα τεχνητό σήμα $1/f$



Σχήμα 5.6: (α) Τεχνητός συντονισμός με $1/f$ διαμόρφωση φάσης $P(t)$. (β) Στιγμαία συχνότητα. (γ) $P(t)$ και $\hat{P}(t)$. (δ) Φάσμα ισχύος για το $P(t)$ και $\hat{P}(t)$.

Κεφάλαιο 6

Θεωρητική ανάλυση του προτεινόμενου μοντέλου

6.1 Υπολογισμός αυτοσυσχέτισης

Όπως φαίνεται και στο προηγούμενο κεφάλαιο, παρατηρήσαμε πειραματικά οι συντονισμοί φωνής σε τυρβώδη φωνήματα μπορούν να μοντελοποιηθούν με διαμορφωμένες κατατάξη $1/f$ ανελίξεις. Σε αυτό το κεφάλαιο σκοπός μας είναι να αναλύσουμε θεωρητικά και να εξάγουμε τα χαρακτηριστικά που πρέπει να έχει ένας συντονισμός που διαμορφώνεται κατά τάξη από ένα στοχαστικό Fractal.

Συγκεκριμένα θα αποδείξουμε το ακόλουθο θεώρημα[2]

Θεώρημα: Έστω η στοχαστική ανελίξη

$$x(t) = A \cos(\omega_c t + \lambda P_H(t) + \phi_0) \quad (6.1)$$

όπου λ είναι μια σταθερά διαμόρφωσης, $P_H(t)$ είναι μια ανελίξη fBm και ϕ_0 είναι μια τυχαία τάξη ομοιόμορφα κατανομημένη στο $[-\pi, \pi]^1$.

Τότε ισχύουν τα παρακάτω:

1. Η ανελίξη $x(t)$ είναι Wide Sense Stationary.
2. Η αυτοσυσχέτιση της δίνεται από την σχέση

$$R_{xx}(t) = \frac{1}{2} \cos(\omega_c t) \exp\left(-\frac{1}{2} \sigma_H^2 |t|^{2H} \lambda^2\right) \quad (6.2)$$

¹H, γενικότερα, οποιαδήποτε τ.μ ώστε $E[e^{j\phi_0}] = 0$ και $E[e^{j2\phi_0}] = 0$

Απόδειξη: Έστω

$$w(t) = e^{j\lambda P_H(t)}, \quad z(t) = w(t)e^{j\omega_c t} \quad (6.3)$$

τότε ισχύει

$$\begin{aligned} R_{xx}(\tau) &= \frac{1}{2} \text{Re}[R_{zz}(\tau)] \\ R_{zz}(\tau) &= R_{ww}(\tau)e^{j\omega_c \tau} \end{aligned} \quad (6.4)$$

αφού

$$x(t) = \frac{z(t) + z^*(t)}{2} \quad (6.5)$$

και για την ανέλιξη $z(t)$ έχουμε

$$\mathbb{E}[z(t)] = \mathbb{E}[w(t)e^{j\omega_c t}] \mathbb{E}[e^{j\phi_0}] = 0. \quad (6.6)$$

και

$$R_{zz^*}(\tau) = \mathbb{E}[z(t+\tau)z^*(t)] = \mathbb{E}[w(t+\tau)w^*(t)] \mathbb{E}[e^{2j\phi_0}] = 0 \quad (6.7)$$

$$R_{zz}(\tau) = \mathbb{E}[z(t+\tau)z^*(t)] = e^{j\omega_c \tau} \mathbb{E}[e^{j\lambda(P_H(t+\tau) - P_H(t))}] \quad (6.8)$$

Όμως επειδή οι μεταβολές (increments) του fBm είναι Strict Sense Stationary η αυτο-συσχέτιση R_{zz} μπορεί να γραφεί σαν

$$R_{zz} = e^{j\omega_c \tau} \mathbb{E}[e^{j\lambda P_H(\tau)}] = e^{j\omega_c \tau} \mathbb{E}[w(t)] \quad (6.9)$$

Όμως ισχύει

$$\mathbb{E}[w(t)] = \mathbb{E}[e^{j\lambda P_H(t)}] = \Phi_{P_H}(\lambda, t) \quad (6.10)$$

Όπου $\Phi_{P_H}(\lambda, t)$ είναι η χαρακτηριστική συνάρτηση της τ.μ. $P_H(t)$. Στην περίπτωση του fBm για δεδομένο t , η $P_H(t)$ είναι μια Gaussian τ.μ. με variance (όπως αναφέρεται σε προηγούμενο κεφάλαιο)

$$\sigma_{P_H(t)}^2 = \sigma_H^2 |t|^{2H} \quad (6.11)$$

όπου

$$\sigma_H^2 = \Gamma(1 - 2H) \frac{\cos(\pi H)}{\pi H} \quad (6.12)$$

και μπορεί εύκολα να αποδειχθεί πως έχει χαρακτηριστική συνάρτηση

$$\Phi_{P_H}(w, t) = \exp\left(-\frac{1}{2} \sigma_{P_H(t)}^2 w^2\right) \quad (6.13)$$

και συνεπώς η ζητούμενη αυτοσυσχέτιση είναι

$$R_{ww}(t) = \exp(-\frac{1}{2}\sigma_{P_H(t)}^2\lambda^2) = \exp(-\frac{1}{2}\sigma_H^2|t|^{2H}\lambda^2) \quad (6.14)$$

και για το $x(t)$ είναι

$$\begin{aligned} R_{xx}(t) &= \frac{1}{2}Re[R_{zz}(t)] = \frac{1}{2}Re[e^{j\omega_c t} R_{ww}(t)] \\ R_{xx}(t) &= \frac{1}{2}\cos(\omega_c t) \exp(-\frac{1}{2}\sigma_H^2|t|^{2H}\lambda^2) \end{aligned} \quad (6.15)$$

ΟΕΔ.

6.2 Φάσμα ισχύος

Το φάσμα ισχύος της διαμορφωμένης ανέλιξης $S(\omega)$ μπορεί να υπολογιστεί με τον μετασχηματισμό Fourier(F) της αυτοσυσχέτισης $R_{xx}(t)$

$$S(\omega) = F(R_{xx}(t)) \quad (6.16)$$

Θα αποδείξουμε παρακάτω απο την σύνδεση με α -Stable κατανομές ότι δεν υπάρχει κλειστή μορφή για αυτό το αποτέλεσμα αυτού του μετασχηματισμού για οποιοδήποτε H . Υπάρχουν μόνο δυο ειδικές περιπτώσεις στις οποίες μπορούμε να έχουμε αναλυτικά αποτελέσματα. Συγκεκριμένα για $H = 0.5$ έχουμε φάσμα ισχύος:

$$\begin{aligned} S_{H=0.5}(\omega) &= \frac{1}{2}\sqrt{\frac{2}{\pi}}\sigma_h^2 \left(\frac{1}{\lambda^4\sigma_H^4 + 4(\omega - \omega_c)^2} \right. \\ &\quad \left. + \frac{1}{\lambda^4\sigma_H^4 + 4(\omega + \omega_c)^2} \right) \end{aligned} \quad (6.17)$$

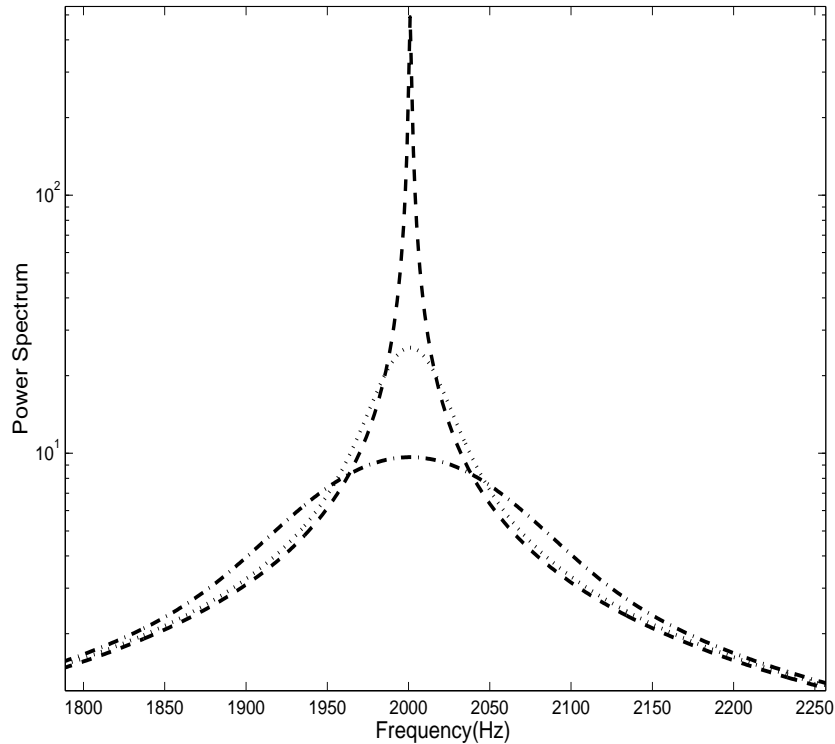
που είναι το άθροισμα απο δύο συντονισμούς τύπου Cauchy κεντραρισμένους στις συχνότητες ω_c και $-\omega_c$. Για $H = 1$ ² έχουμε

$$\begin{aligned} S_{H=1}(\omega) &= \frac{1}{2\sqrt{\lambda^2\sigma_H^2}} \left(\exp \frac{(\omega - \omega_c)^2}{2\lambda^2\sigma_H^2} \right. \\ &\quad \left. + \exp \frac{(\omega + \omega_c)^2}{2\lambda^2\sigma_H^2} \right) \end{aligned} \quad (6.18)$$

που είναι το άθροισμα δύο Gaussian τύπου συντονισμών στις συχνότητες ω_c και $-\omega_c$.

²που είναι μόνο οριακή περίπτωση για fBm

6.3 Σύνδεση με κατανομές a -Stable



Σχήμα 6.1: Φάσματα ισχύος που ακολουθούν κατανομή a -Stable για $a = 1, 1.6, 2$

Οι (συμμετρικές) a -Stable κατανομές είναι μια κατηγορία απο κατανομές με αργή μείωση (heavy-trail distributions) και άπειρο variance. Ορίζονται απο την μορφή της χαρακτηριστικής τους συνάρτησης που έχει την μορφή

$$\Phi(\theta) = \exp\left(-\left|\frac{\theta}{\sigma}\right|^\alpha\right) \quad (6.19)$$

και δεν υπάρχει κλειστή μορφή για την πυκνότητα πιθανότητας για a διαφορετικό απο 1 και 2. Μια πολύ καλή αναφορά για αυτές τις κατανομές είναι η [20]. Πρόσφατα, έχουν προταθεί μοντέλα στοχαστικών ανελίξεων [30] που εμφανίζουν a -Stable κατανομή πυκνότητας πιθανότητας σαν γενίκευση των Gaussian αυτοόμοιων στοχαστικών ανελίξεων.

Είναι πολύ ενδιαφέρον όμως να παρατηρήσει κανείς πως η χαρακτηριστική συνάρτηση (6.19) έχει την ίδια μορφή με την συνάρτηση αυτοσυσχέτισης που υπολογίσαμε θεωρητικά (6.2) αν θέσουμε $\alpha = 2H$. Ο πολλαπλασιασμός με με την συνάρτηση συνημιτόνου έχει απλα σαν αποτέλεσμα να κεντράρεται ο συντονισμός στις συχνότητες ω_c και $-\omega_c$. Αυτή η αναλογία οδηγεί άμεσα στα ακόλουθα συμπεράσματα:

- Το φάσμα ισχύος δεν μπορεί να γραφεί σε κλειστή μορφή για H διαφορετικό από 0.5 και 1. Αυτό προκύπτει άμεσα από την αναλογία με τις κατανομές α -Stable αφού η συνάρτηση αυτοσυσχέτισης αντιστοιχεί στην χαρακτηριστική συνάρτηση, το φάσμα ισχύος θα αντιστοιχεί στην συνάρτηση πυκνότητας πιθανότητας που όπως αναφέρθηκε είναι γνωστό πως δεν μπορεί να γραφεί σε κλειστή μορφή.
- Δείξαμε ότι για fBm (που είναι Gaussian) το φάσμα ισχύος της διαμορφωμένης κατά φάση ανέλιξης $x(t)$ έχει μορφή που είναι ίδια με την πυκνότητα πιθανότητας μιας α -stable κατανομής κεντραρισμένης στο κέντρο του συντονισμού ω_c .

Στο Σχήμα 6.3 απεικονίζεται η μορφή του φάσματος ισχύος για τιμές της παραμέτρου $\alpha = 1, 1.6, 2$. Η κεντρική συχνότητα συντονισμού βρίσκεται στα 2000Hz .

Αυτό το αποτέλεσμα δείχνει πως υπάρχει κάποια σχέση ανάμεσα σε διαμορφωμένες κατά φάση Gaussian αυτοόμοιες στοχαστικές ανελιξεις και α -Stable κατανομές. Μια ενδιαφέρουσα ιδιότητα των α -Stable κατανομών είναι πως παραμένουν αναλλοίωτες κάτω από συνέλιξη και ότι αποτελούν σημείο έλξης κάτω από πολλαπλές συνέλιξεις για άλλες συναρτήσεις πυκνότητας πιθανότητας. Το αποτέλεσμα αυτό είναι γνωστό σαν γενικευμένο (Levy) κεντρικό οριακό θεώρημα. Εάν συνδέσουμε αυτό το θεώρημα με τα δικά μας αποτελέσματα, το φάσμα ισχύος των $1/f$ διαμορφωμένων ανελιξεων είναι επίσης αναλλοίωτο κάτω από συνέλιξη των φασμάτων ισχύος. Αυτό είναι ισοδύναμο με πολλαπλασιασμό των συναρτήσεων αυτοσυσχέτισης και συνεπώς σημαίνει πως οι ανελιξεις που μελετάμε είναι ελκτικά σημεία για τον πολλαπλασιασμό ανεξάρτητων στοχαστικών ανελιξεων. Συγκεκριμένα, εάν ένας αριθμός από ανεξάρτητες στοχαστικές ανελιξεις πολλαπλασιαστεί στο πεδίο του χρόνου, το φάσμα ισχύος του γινομένου θα τείνει στην μορφή που υπολογίσαμε (6.16). Αυτή μοιάζει να είναι μια μαθηματική σύνδεση ανάμεσα στο προτεινόμενο μοντέλο και στα πολλαπλασιαστικά μοντέλα για τύρβη που έχουν προταθεί από την Ρωσική σχολή [21]. Η διερεύνηση αυτών των αποτελεσμάτων είναι ένα θέμα που απαιτεί περισσότερη θεωρητική και πειραματική έρευνα.

Κεφάλαιο 7

Συμπεράσματα

7.1 Συμβολή της Διπλωματικής εργασίας

Εκτιμώντας τα αποτελέσματα αυτής της εργασίας, θεωρούμε ότι συνέβαλε

- Στην ανασκόπηση των θεωρητικών ιδιοτήτων των στοχαστικών Fractals και των μοντέλων που τα παράγουν.
- Στην συγκριτική μελέτη και αξιολόγηση των αλγορίθμων που υπάρχουν στην βιβλιογραφία για σύνθεση και εκτίμηση παραμέτρων για $1/f$ ανελίξεις.
- Προτάθηκε ένα μη γραμμικό μοντέλο για τυρβώδεις ήχους και έγινε πειραματική επιβεβαίωση της καταλληλότητας του
- Τέλος, δόθηκε μια θεωρητική ανάλυση που προσδιορίζει τις στατιστικές ιδιότητες του προτεινόμενου μοντέλου.

Το μη-γραμμικό μοντέλο και τα πειραματικά αποτελέσματα δημοσιεύθηκαν σε διεθνές συνέδριο [1] και η θεωρητική ανάλυση καθώς και περισσότερα πειραματικά αποτελέσματα ετοιμάζονται για δημοσίευση σε περιοδικό [2].

7.2 Κατευθύνσεις για μελλοντική έρευνα

Γενικά, τα Power laws και τα στοχαστικά Fractals, αποτελούν ένα πολύ ενδιαφέρον πεδίο έρευνας αυτή την στιγμή. Η εμφάνιση αυτών των φαινομένων συχνότατα στην φύση δεν έχει εξηγηθεί ακόμη με ικανοποιητικό τρόπο. Επίσης, η μεγάλη τους συχνότητα εμφάνισης σημαίνει πως μπορούν να αποτελέσουν χρήσιμα εργαλεία μοντελοποίησης.

Όσον αφορά τα στοχαστικά μοντέλα για $1/f$ ανελίξεις, το fBm έχει σαν βασικό πλεονέκτημα την απλότητα του αλλά και αρκετά μειονεκτήματα. Συγκεκριμένα, δεν επιτρέπει εκθέτες μεγαλύτερους από $\gamma = 3$ το οποίο είναι σημαντικός περιορισμός. Επίσης, είναι Gaussian ανέλιξη πράγμα που δεν είναι επιθυμητό για φαινόμενα που εμφανίζουν μεγάλες διακυμάνσεις γύρω από την μέση τους τιμή (Impulsive). Έχουν προταθεί στην βιβλιογραφία διάφορες γενικεύσεις που επιλύουν αυτά τα προβλήματα, αλλά δεν έχει καθιερωθεί ακόμη κάποιο μοντέλο.

Το πρόβλημα της εκτίμησης παραμέτρων από μια (διακριτή, θορυβώδη) υλοποίηση $1/f$ ανέλιξης είναι θεμελιώδες και δεν έχει βρει ικανοποιητική λύση. Πρόκειται για μια ενεργή περιοχή έρευνας όπου έχουν εμφανιστεί τα τελευταία χρόνια πλήθος από εργασίες.

Τέλος, ένα πολύ ενδιαφέρον θέμα για μελλοντική έρευνα είναι η μελέτη της επίδρασης που έχει στα fractal η δειγματοληψία. Οι περισσότερες εργασίες βασίζονται σε συνεχή χρόνο (που απλοποιεί την ανάλυση) αλλά δεν έχει γίνει ακόμη σαφές σε τι συχνότητες πρέπει να δειγματοληπτεί κανείς ένα fractal ώστε να διατηρεί κάποιες από τις ιδιότητες του καθώς και το πόσο αυτές επιρρεάζονται σε σχέση με την δειγματοληψία.

Βιβλιογραφία

- [1] P. Maragos, A. G. Dimakis and I. Kokkinos. “Some Advances in Nonlinear Speech Modeling using Modulations, Fractals and Chaos” Proc. of International Conference on Digital Signal Processing (DSP2002).
- [2] P. Maragos and A. G. Dimakis “Phase Modulated 1/f processes with application to Turbulent sounds” (Manuscript in preparation, To be submitted for publication).
- [3] B. Mandelbrot *The Fractal Geometry of Nature*, W. H. Freeman, New York, 1983.
- [4] H. M. Teager and S. M. Teager, “Evidence for Nonlinear Sound Production Mechanisms in the Vocal Tract”, in *Speech Production and Speech Modelling*, W.J. Hardcastle and A. Marchal, Eds., NATO Adv. Study Inst. Series D, vol. 55, France, July 1989.
- [5] I. Daubechies, *Ten lectures on wavelets*, Philadelphia, PA, SIAM, 1992.
- [6] J. F. Kaiser, “Some Observations on Vocal Tract Operation from a Fluid Flow Point of View”, in *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, I. R. Titze & R. C. Scherer (Eds.), Denver Center for Performing Arts, Denver, 1983.
- [7] G.K. Zipf, “Human behaviour and the principle of Least effort”, Hafner Publishing, NY, 1949.
- [8] G.Troll, P. Graben, “Zipf’s Law is not a consequence of the central limit theorem” *Phys.Rev.* 57 1347-55 (1998).
- [9] V. Pareto, “Cours d’economie politique”, Lausanne, 1897.
- [10] R.A. Raimi, “The peculiar distribution of first digits”, *Scientific American*, 121, Dec. 1969.

-
- [11] D. Dimitriadis, P. Maragos, and A. Potamianos, "Modulation Features for Speech Recognition", *Proc. ICASSP-2002*, Orlando.
- [12] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy Separation in Signal Modulations with Application to Speech Analysis", *IEEE Trans. Signal Processing*, vol. 41, pp. 3024–3051, Oct. 1993.
- [13] A. C. Bovik, P. Maragos and T. F. Quatieri, "AM-FM Energy Detection and Separation in Noise using Multiband Energy Operators", *IEEE Trans. Signal Processing*, vol. 41, pp. 3024–3051, Dec. 1993.
- [14] D. Dimitriadis and P. Maragos, "An Improved Energy Demodulation Algorithm Using Splines", *Proc. ICASSP*, Salt Lake, 2001.
- [15] B.B. Mandelbrot and J.W. Van Ness, "Fractional Brownian motion, fractional noises and applications", *SIAM Review*, vol.10(4), p.422–437, 1968
- [16] T. Lundahl, W. Ohley, S. Kay and S. Siffert, "Fractional Brownian Motion: A maximum likelihood estimator and its applications to image texture", *IEEE Trans. Med. Imag.* vol. MI-5, pp152-161, 1986.
- [17] G. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach*, Prentice Hall, 1995.
- [18] D. Sornette, *Critical Phenomena in Natural Sciences, Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*, Springer-Verlag 2002.
- [19] A. Papoulis *Probability, Random variables and Stochastic Processes*, Third Edition, McGraw-Hill, 1991.
- [20] C.L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions*, Wiley, 1995
- [21] U. Frisch, *Turbulence: The Legacy of A.N. Kolmogorov*, Cambridge University Press, Cambridge, 1995
- [22] P. Levy, "Random functions: General theory with special reference to Laplacian random functions". *Univ. California Publ. Stat.* 1 (1953) pp 331-390
- [23] B. Pesquet-Popescu and J. Levy Vehel "Stochastic Fractal Models for Image Processing" *IEEE Signal Processing Magazine*, September 2002, Vol. 19, No5 pp 48-62

-
- [24] H.O. Peitgen, D. Saupe eds. *The Science of Fractal Images*, Spriger-Verlag 1988.
- [25] G. Wornell and A. Oppenheim, “Estimation of fractal signals from noise measurements using wavelets”, *IEEE Trans. Signal Processing*, vol 40, pp611-623, 1992.
- [26] B. Ninness, “Estimation of $1/f$ noise”, *IEEE Trans. Signal Processing*, vol 44, pp32-46, 1998.
- [27] J. Leu, A. Papamarcou, “On Estimating the Spectral Exponent of Fractional Brownian Motion”, *IEEE Trans. Info. Theory*, Vol41, No1, January 1995.
- [28] J. Geweke and S. Porter-Hudak, “The estimation and application of long memory time series models”, *J. Time Series Anal.*, vol.4, pp.221-238,1983.
- [29] C.M. Hurvich and W.W. Chen “An effective taper for potentially overdifferenced long-memory time series”, *J. Time Series Anal.*, vol.21, no. 2, pp.155-180,2000.
- [30] O. Cappe, E. Moulines, J. Pesquet, A. Petropulu, X. Yang, “Long-Range Dependence and Heavy-Tail Modeling for Teletraffic Data”, *IEEE Signal Processing Mag.*, vol.19, May 2002.
- [31] P. Flandrin “On the spectrum of fractional Brownian motion”, *IEEE Trans. Inform. Theory*, col35, pp197-199, 1989
- [32] H. M. Teager and S. M. Teager, “Evidence for Nonlinear Sound Production Mechanisms in the Vocal Tract”, in *Speech Production and Speech Modelling*, W.J. Hardcastle and A. Marchal, Eds., NATO Adv. Study Inst. Series D, vol. 55, France, July 1989.
- [33] J. F. Kaiser, “Some Observations on Vocal Tract Operation from a Fluid Flow Point of View”, in *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, I. R. Titze & R. C. Scherer (Eds.), Denver Center for Performing Arts, Denver, 1983.
- [34] A. Barney, C.H. Shadle and P.O.A.L. Davies, “Fluid Flow in a Dynamical Mechanical Model of the Vocal Folds and Tract: Part 1 & 2”, *J. Acoust. Soc. Amer.*, 105 (1): 444-466, Jan. 1999.
- [35] G. Richard, D. Sinder, H. Duncan, Q. Lin, J. Flanagan, S. Levinson, M. Krane, S. Slimon, D. Davis, “Numerical Simulation of Fluid Flow in the Vocal Tract”, *Proc. EuroSpeech*, 1995.

-
- [36] A. Potamianos and P. Maragos, “Speech Formant Frequency and Bandwidth Tracking Using Multiband Energy Demodulation”, *J. Acoust. Soc. Amer.*, 99 (6), pp.3795–3806, June 1996.
- [37] A. Potamianos and P. Maragos, “Speech Processing Applications Using an AM–FM Modulation Model”, *Speech Communication*, vol.28, pp.195–209, 1999.
- [38] Michael F. Barnsley, *Fractals Everywhere*, Academic Press, 1993.
- [39] M. Faloutsos, P. Faloutsos and C. Faloutsos, “On power-law relationships of the Internet topology” Proc. of Conference on Applications, technologies, architectures, and protocols for computer communication ACM Press, 1999.
- [40] P. Borgnat, P. Flandrin and P. Amblard Stochastic Discrete Scale Invariance, *Signal Processing Letters*, Vol9, No6, June 2002.