

**Τμήμα  
Μηχανικών  
Πληροφορικής τ.ε.**  
Τεχνολογικό Εκπαιδευτικό Ίδρυμα  
Δυτικής Ελλάδας

## **ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Υλοποίηση σταδίων κωδικοποίησης ηχητικού σήματος  
πλήρους φάσματος στο προγραμματιστικό περιβάλλον  
Matlab**

---

**Επιβλέπων καθηγητής: Μιχαήλ Παρασκευάς**

**Εκπονήθηκε από τους:**

- 1. Δημήτρης Πέρρος AM:1264**
- 2. Alexey Bukreyev AM:0910**

Αντίρριο Σεπτέμβριος 2015

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή

Αντίρριο, Ημερομηνία

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Ονοματεπώνυμο, Υπογραφή
2. Ονοματεπώνυμο, Υπογραφή
3. Ονοματεπώνυμο, Υπογραφή

# Ευχαριστίες

---

Αρχικά θέλουμε να ευχαριστήσουμε τους καθηγητές μας που μας στήριξαν όλα αυτά τα χρόνια.

Επίσης θέλουμε να ευχαριστήσουμε τον καθηγητή κ. Μιχαήλ Παρασκευά ιδιαίτερος για τη στήριξή του και για την επίβλεψη της παρούσας πτυχιακής εργασίας.

Τέλος, θέλουμε να ευχαριστήσω τις οικογένειές μας για την στήριξή τους όλα αυτά τα χρόνια που μας βοήθησαν για την ολοκλήρωση των σπουδών μας.

## Περιεχόμενα

Ευχαριστίες .....	3
1. ΚΕΦΑΛΑΙΟ 1ο : Ήχος.....	6
1.1. Θεωρία του Ήχου .....	6
1.2. Βασικές Αρχές Θεωρίας Ήχου .....	8
1.3. Χαρακτηριστικά Ψηφιακού Ήχου. ....	11
2. ΚΕΦΑΛΑΙΟ 2: Διαδικασία Ψηφιοποίηση Ήχου .....	15
2.1. Στάδια Ψηφιοποίησης Ηχητικών Αρχείων. ....	15
2.2. Δειγματοληψία.....	19
2.3. Προβλήματα Δειγματοληψίας .....	22
2.4. Κβαντοποίηση .....	23
3. ΚΕΦΑΛΑΙΟ 3 « Κωδικοποίηση Ψηφιακού Σήματος » .....	26
3.1. Μέθοδοι κωδικοποίηση.....	26
4. ΚΕΦΑΛΑΙΟ 4 « Συμπίεση » .....	34
4.1. Συμπίεση Ψηφιακού Ήχου .....	34
4.2. Πρότυπο MPEG.....	37
4.3. MP3 .....	40
4.4. Άλλα πρότυπα .....	47
4.5. Τύποι Αρχείων Ψηφιακού Ήχου.....	51
5. ΚΕΦΑΛΑΙΟ 5ο : ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ.....	53
Παράρτημα Α.....	57
Παράρτημα Β.....	65
Βιβλιογραφία.....	89

## ΛΙΣΤΑ ΣΧΗΜΑΤΩΝ

Σχήμα 1.2.1 Ύψος κυματομορφής.....	8
Σχήμα 1.2.2 Δύο κυματομορφές με διαφορετικό ύψος .....	9
Σχήμα 1.2.3 Περίοδος μιας κυματογράφος .....	9
Σχήμα 1.2.4 Συχνότητα κυματομορφής .....	10
Σχήμα 1.2.5 Τόνος .....	10
Σχήμα 1.2.6 Είδη κυματομορφών .....	10
Σχήμα 1.3.1 Χαρακτηριστικά του Ψηφιοποιημένου ήχου.....	14

Σχήμα 2.1.1 Διαδικασία Ψηφιοποίησης Ηχητικού Σήματος.....	17
Σχήμα 2.1.2 Διαδικασία Ψηφιοποίησης Ηχητικού Σήματος.....	18
Σχήμα 2.2.1 :(α) Αναλογικό σήμα, (β) Ψηφιακό με χαμηλή συχνότητα, (γ)Ψηφιακό σήμα με υψηλή συχνότητα.....	20
Σχήμα 2.3.1 Διαδικασία μετατροπής αναλογικού σε ψηφιακή μορφή και αντίστροφα .....	22
Σχήμα 2.4.1 Ψαλιδισμός (Clipping). .....	26
Σχήμα 3.1.1 Κωδικοποίηση PCM.....	28
Σχήμα 3.1.2 Χρήση PCM Κωδικοποιητή.....	28
Σχήμα 3.1.3 Παράδειγμα PCM Κωδικοποίηση.....	29
Σχήμα 3.1.4 Κωδικοποιητής ADPCM.....	32
Σχήμα 3.1.5 Αποκωδικοποιητής ADPCM.....	32
Σχήμα 3.1.6 : Διάγραμμα Κωδικοποιήσεων Ψηφιακού Ήχου. ....	33
Σχήμα 4.1.1 Ψυχοακουστικό Μοντέλο Ακοής.....	35
Σχήμα 4.1.2 Φαινόμενο Ηχητικής Σκίασης. ....	36
Σχήμα 4.2.1 Λόγος Συμπίεσης.....	39
Σχήμα 4.3.1 Φαινόμενο επικάλυψης. Ένας δυνατός ήχος, μεταβάλλεται τοπικά στην καμπύλη του κατωφλιού ακουστικότητας επικαλύπτοντας ένα γειτονικό αδύναμο ήχο.....	42
Σχήμα 4.3.2 Διάγραμμα βαθμίδων κωδικοποιητή MP3. Μαθηματική περιγραφή του μετασχηματισμού MDCT.....	43
Σχήμα 4.3.3 Διαδικασία Ripping και Encoding.....	45
Σχήμα 4.3.4 Διάγραμμα Βαθμίδων ενός MP3.....	46
Σχήμα 4.4.1 Διάγραμμα βαθμίδων του κωδικοποιητή Ogg Vorbis. ....	49
Σχήμα 4.5.1 Τύποι Αρχείων Ήχου.....	53
Σχήμα 5.1 : Αντιληπτική ανάλυση/ σύνθεση ακουστικού σήματος .....	53

## **1. ΚΕΦΑΛΑΙΟ 1 : ΉΧΟΣ**

### **1.1. ΘΕΩΡΙΑ ΤΟΥ ΗΧΟΥ**

Αυτό που εμείς οι άνθρωποι αντιλαμβανόμαστε ως «ήχο», δεν είναι παρά μεταβολές της πίεσης του αέρα, ικανότητα την οποία απέκτησε το είδος μας ώστε να μπορεί να εντάσσεται καλύτερα στο περιβάλλον του. Ο ήχος που παράγεται από μια πηγή και συλλαμβάνεται από το αυτί μας, αποτελεί ένα από τα πιο εντυπωσιακά στοιχεία των πολυμεσικών εφαρμογών, καθώς μπορεί να προσφέρει ακουστική απόλαυση, να εντυπωσιάσει με διαφορετικά ηχητικά εφέ και να ξεκουράζει σαν ηχητική υπόκρουση.

Είναι γνωστό το φιλοσοφικό ερώτημα, «αν ένα δέντρο πέσει στο δάσος και δεν υπάρχει κανείς να το ακούσει, θα υπάρξει ήχος ». Στο ερώτημα αυτό, η έννοια ήχος αποδίδεται με δύο ερμηνείες: (α) Αναφέρεται στην χαρακτηριστική διαταραχή που προκαλεί ένα σώμα σε κίνηση, με επακόλουθο ο ήχος να είναι μια οργανωμένη κίνηση των μορίων ενός μέσου(π.χ νερό, αέρα, κάποιο στερεό σώμα) που προκαλείται από την δόνηση ενός σώματος σε αυτό το μέσο. (β) Η άλλη έννοια, είναι άμεσα συνδεδεμένη στη ανθρώπινη αίσθηση της ακοής ως ακουστική αίσθηση που γίνεται αντιληπτή μέσω του αυτιού και οφείλεται σε μεταβολές της πίεσης, της θέσης και της φοράς. Συμπερασματικά, ο ήχος δημιουργείται, αφού ένα σώμα δονείται γρήγορα ώστε να προκαλέσει την διάδοση ενός ακουστικού κύματος στο μέσο όπου συμβαίνει η δόνηση.

Από φυσική λοιπόν άποψη, ένας ήχος παράγεται από μεταβολές της πίεσης που μεταδίδονται σε ένα μέσο που μπορεί να συμπιεστεί, ενώ παρόλο που ο μηχανισμός

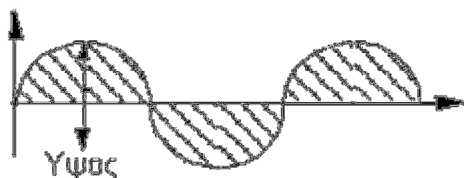
διάδοσης του ήχου διαφέρει από περίπτωση σε περίπτωση, οι βασικές αρχές μετάδοσης και παραγωγή του είναι οι ίδιες. Πιο συγκεκριμένα, τα μόρια όλων των φυσικών σωμάτων οφείλουν να κρατούν σταθερές αποστάσεις από τα γειτονικά τους μόρια. Κατά συνέπεια, αν για κάποιο λόγο τα μόρια σε μια περιοχή συμπιεστούν θα επιδιώξουν να επανέλθουν στη αρχική τους μορφή συμπιέζοντας με την σειρά τους τα μόρια των γειτονικών τους περιοχών. Με αυτό τον τρόπο, δημιουργούνται μεταβολές πίεσης που μεταδίδονται με μια ορισμένη ταχύτητα σε ένα μέσο, ούτως ώστε αυτές οι μεταβολές να αποτελούν ένα ηχητικό κύμα.

Παράλληλα, κατά την διάδοση του ηχητικού κύματος σε ένα μέσο, δεν πραγματοποιείται μεταφορά των μορίων του μέσου αλλά διαταραχής από το ένα μόριο στο διπλανό του, με αποτέλεσμα ο ήχος να χαρακτηρίζεται ως κύμα με το οποίο μεταφέρεται ενέργεια και όχι ύλη. Για παράδειγμα, με το πέρασμα ενός αεροπλάνου μπορεί να σπάσουν τα τζάμια κτηρίου, τα οποία βρίσκονται σε κοντινή απόσταση λόγω του ισχυρού ηχητικού κύματος που διαμορφώνεται.

Όπως τυπώθηκε παραπάνω, ο ήχος μεταδίδεται εξαιτίας των μεταβολών της πίεσης του αέρα ενώ οι μεταβολές αυτές, δηλαδή τα πυκνώματα και τα αραιώματα του αέρα διαδίδονται σε όλες τις κατευθύνσεις στο γύρο χώρο. Για παράδειγμα, αν κτυπήσουμε την χορδή μιας κιθάρας, τότε η χορδή θα αρχίσει να ταλαντώνεται συμπιέζοντας τα μόρια του αέρα που βρίσκονται πάνω της, μέχρι που να συναντήσουν το τύμπανο του αυτιού μας διεγείροντας ένα ολόκληρο μηχανισμό με αποτέλεσμα το συναίσθημα της ακοής. Η ικανότητα του ανθρώπου να αντιλαμβάνεται τις μεταβολές αυτής της πίεσης και να αποκωδικοποιεί το περιεχόμενο τους, στηρίζεται στα τρία πιο κάτω στάδια: (α) η κατεύθυνση των κυμάτων του ήχου στο ακουστικό μέρος του αυτιού, (β) ο εντοπισμός της διακύμανσης του και (γ) η μετάφραση της διακύμανσης σε ηλεκτρικά σήματα όπου και γίνονται κατανοητά από τον εγκέφαλο. Παράλληλα, ο ανθρώπινος εγκέφαλος έχει την ικανότητα να συγκρίνει την χρονική διαφορά με την οποία ένας ήχος φτάνει στο αυτί, οπότε και συμπεράνει την απόσταση από την οποία προέρχεται.

## 1.2. ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΘΕΩΡΙΑΣ ΉΧΟΥ

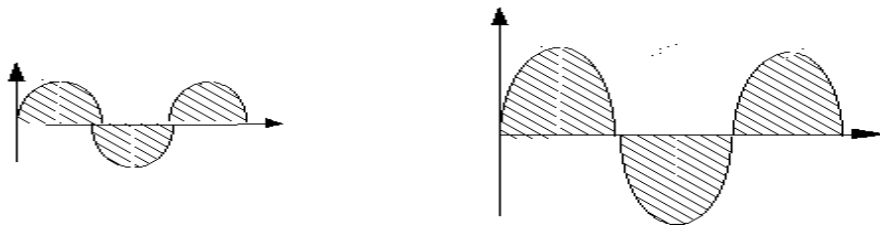
Όπως είδη έχει αναφερθεί, ο ήχος από την φύση του είναι μια αναλογική ποσότητα και για να γίνει αντικείμενο επεξεργασίας θα πρέπει να ψηφιοποιηθεί ενώ για την κατανόηση της μεθόδου εισαγωγής, οφείλουμε να θέσουμε τις βασικές αρχές θεωρίας του. Ο ήχος, είναι η αίσθηση που αντιλαμβανόμαστε όταν τα ακουστικά νεύρα ενεργοποιούνται από μόρια που δονούνται στον αέρα, π.χ όταν κτυπάει ένα κουδούνι αυτό δονείται. Η κίνηση αυτή διοχετεύεται στα περιβάλλοντα μόρια του αέρα, όπου και δονούν τα γειτονικά τους μόρια προκαλώντας με αυτή την κίνηση δόνηση στο τύμπανο. Εντούτοις, ο ήχος παρουσιάζεται και ως **κυματομορφή**, το μέγεθος της οποίας αντιπροσωπεύει την ευρύτητα του ήχου όπως φαίνεται στο σχήμα 1.2.1.



Σχ.1.2.1: Ύψος κυματομορφής.

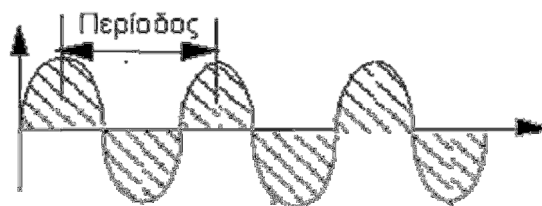


Επίσης, το **ύψος** αναδεικνύει την σχετική ισχύει ή δύναμη του ήχου. Στο σχήμα 1.2.2, ο ήχος στα δεξιά είναι πολύ πιο δυνατός σε σχέση με αυτό στα αριστερά.



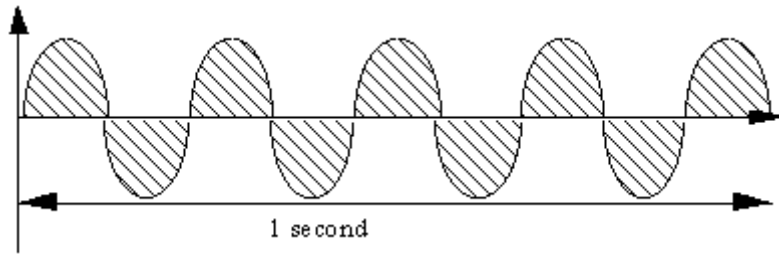
Σχ.1.2.2: Δύο κυματομορφές με διαφορετικό ύψος

Η απόσταση μεταξύ δύο συνεχόμενων κορυφών ονομάζεται **περίοδος** (period) και φαίνεται στο σχήμα 1.2.3.



Σχ1.2.3: Περίοδος μιας κυματογράφος.

Ο αριθμός των κορυφών που διαμορφώνονται σε χρονικό διάστημα ενός δευτερολέπτου λέγεται **συχνότητα**, η οποία μετράται σε Hertz (Hz) ή kilohertz (KHz).



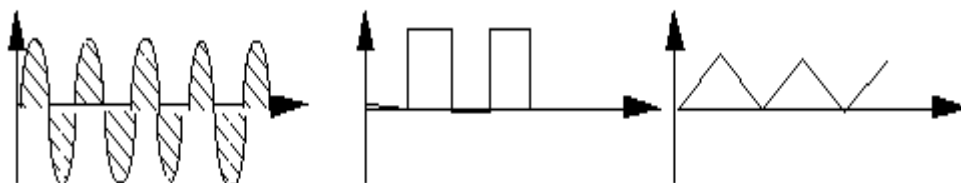
Σχ.1.2.4: Συχνότητα κυματομορφής

Ο **τόνος** (pitch) ενός ήχου χαρακτηρίζεται ως η ψυχολογική εντύπωση της υψηλότητας ή της χαμηλότητας του, με επακόλουθο όσο πιο υψηλός είναι ο τόπος τόσο πιο υψηλή είναι η συχνότητα του. Στο παρακάτω σχήμα 1.2.5 αναδύεται η σχέση μεταξύ του τόνου και της συχνότητας.



Σχ.1.2.5: Τόνος.

Στην σχήμα 1.2.6. τίθενται τρία βασικά χαρακτηριστικά δείγματα κυματομορφών ήχου, ημιτονοειδή, τετραγωνική και πριονωτή. Η ημιτονοειδής κυματομορφή έχει την ικανότητα να ακούγεται ευχάριστα, η τετραγωνική πιο σκληρά ενώ οι περίπλοκοι κυματισμοί ακούγονται φυσικοί. Συμπερασματικά, όσο πιο σταθερός και ομαλός είναι ο κυματισμός τόσο πιο σαφής και καθαρός είναι ο ήχος. Όσο πιο πολλές γωνίες διαμορφώνει ο κυματισμός τόσο πιο σκληρά ακούγεται ο ήχος.



Σχ.1.2.6: Είδη κυματομορφών.

### **1.3. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΨΗΦΙΑΚΟΥ ΉΧΟΥ.**

Η ιδιαιτερότητα κάθε ήχου οφείλεται σε ένα σύνολο από ηχητικά μεγέθη, άλλα από τα οποία είναι ανεξάρτητα από την προσωπική αντίληψη του ακροατή και αποκαλούνται αντικείμενα, ενώ άλλα είναι άμεσα συνδεδεμένα με τον ακροατή και λέγονται υποκειμενικά χαρακτηριστικά του ήχου. Τέτοια αντικειμενικά χαρακτηριστικά, είναι η συχνότητα και η ένταση και ως υποκειμενικά η ακουστικότητα, το ύψος και η χροιά.

**Συχνότητα:** Αν θέλουμε να περιγράψουμε τον ήχο που παράγουν δύο πνευστά μουσικά όργανα όπως η τρομπέτα και η τούμπα, θα παρατηρήσουμε ότι παρόλο που και τα δύο είναι παρόμοια όργανα, η τρομπέτα παράγει πιο υψηλό ήχο από την τούμπα. Το ύψος του ήχου είναι ένα υποκειμενικό γνώρισμα που σχετίζεται αμοιβαία με ένα αντικειμενικό χαρακτηριστικό, τη συχνότητα. Η συχνότητα έχει να κάνει με το πόσο γρήγορα ή αργά πάλλεται το σώμα που δημιουργεί τον ήχο, για παράδειγμα, όταν η χορδή μιας κιθάρας πάλλεται 100 φορές το δευτερόλεπτο, τότε προκαλείται ο αντίστοιχος αριθμός πυκνωμάτων ούτως ώστε, να γίνεται αντιληπτός ο συγκεκριμένος ήχος. Συμπερασματικά, η

συχνότητα ορίζει τον αριθμό των κύκλων που εκτελεί η συνάρτηση ανά δευτερόλεπτο και μετριέται σε κύκλους ανά δευτερόλεπτο ή Hertz (Hz).

Κάθε ήχος για να γίνει αντιληπτός από το ανθρώπινο αυτί, θα πρέπει να έχει συχνότητα μεταξύ 20 και 20.000Hz. Οι πιο χρήσιμες συχνότητες βρίσκονται κάτω από 10 kHz, (π.χ ομιλία, μουσική, διάφοροι θόρυβοι). Εντούτοις, υπάρχουν ήχοι με συχνότητα μεγαλύτερη των 20 kHz οι οποίοι ονομάζονται και υπέρηχοι, ενώ αυτοί με μικρότερη συχνότητα των 20 Hz λέγονται υπόηχοι. Αν και οι δύο κατηγορίες δεν έχουν επίδραση από το ανθρώπινο αυτί, μπορεί να έχουν επίδραση στη υγεία του.

**Ένταση:** Το δεύτερο βασικό γνώρισμα του ήχου είναι η ένταση, στενά συνδεδεμένη με την ισχύ του ηχητικού σήματος που διεγείρει το αυτί μας. Η κύρια αιτία που οι ήχοι έχουν διαφορετικές εντάσεις είναι ότι πιέζουν με διαφορετική δύναμη το τύμπανο του αυτιού μας, δηλαδή το πόσο έντονες είναι οι αναταράξεις που προκαλεί το σώμα που παράγει τον ήχο. Όσο πιο μεγάλη ισχύ διαμορφώνουν τα ηχητικά κύματα ολοένα και περισσότερη δύναμη εξασκούν στο μηχανισμό του αυτιού μας. Για παράδειγμα, κτυπώντας απλά την χορδή μιας κιθάρας τότε αυτή πάλλεται με μια συχνότητα, δημιουργώντας αναταράξεις που όμως δεν είναι έντονες. Αντίθετα αν κτυπήσουμε την ίδια χορδή με δύναμη παρατηρείται πως πάλλεται με την ίδια συχνότητα αλλά με πολύ εντονότερες διαταραχές. Άρα, ένταση ορίζεται από το πλάτος της δόνησης με αποτέλεσμα όσο μεγαλύτερο είναι το πλάτος τόσο ισχυρότερος είναι ο ήχος.

Στην φυσική, ως ένταση ορίζεται το ποσό της ηχητικής ενέργειας στην μονάδα του χρόνου και εκφράζεται σε  $\text{Watt/m}^2$ . Οι τιμές των ηχητικών εντάσεων που γίνονται αντιληπτές από τον άνθρωπο καλύπτουν την περιοχή από 10<sup>-12</sup> έως 10  $\text{Watt/m}^2$ . Σύμφωνα με αυτό παρουσιάζει η χρήση της γραμμικής κλίμακας στην περιγραφή της ηχητικής έντασης, ενώ ο τρόπος που γίνονται συνειδητές οι ακουστικές εντάσεις είναι λογαριθμικός και όχι γραμμικός (δηλ. η διαφορά μεταξύ των υποκειμενικών εντάσεων δύο ήχων δεν εξαρτάται από την απόλυτη τιμή της διαφοράς αυτής αλλά από το λόγο των ισχύων του). Αυτό σημαίνει ότι η αύξηση της έντασης από 2μ  $\text{Watt/m}^2$  σε 4  $\text{Watt/m}^2$ , αντιστοιχεί όπως και στην περίπτωση 5μ  $\text{Watt/m}^2$  σε 10μ  $\text{Watt/m}^2$ .

Κατά συνέπεια, για τους πιο πάνω λόγους, η σχετική τιμή μεταξύ 2 ηχητικών κυμάτων δεν μετριέται σε  $\text{Watt/m}^2$  αλλά σε bels ή πιο συχνά σε decibels( $\text{dB}$ =μονάδα μέτρηση της έντασης ενός ήχου και  $1\text{dB}$  αντιστοιχεί σε  $0,1$  bel). Για να συγκρίνουμε το την ισχύ δύο ηχητικών κυμάτων υπολογίζουμε το λόγο των ισχύων τους. Η λογαριθμική κλίμακα  $\text{dB}$  παρουσιάζει της σχέση της ηχητικής ισχύος με μια συγκεκριμένη πηγή, σε σχέση με την χαμηλότερη ηχητική ισχύ που μπορεί να γίνει αντιληπτή από το ανθρώπινο αυτί ( $10^{-12} \text{Watt/m}^2$ ).

$$\text{Ένταση ήχου} = 10 \log(P/P_0) \text{ dB}^1$$

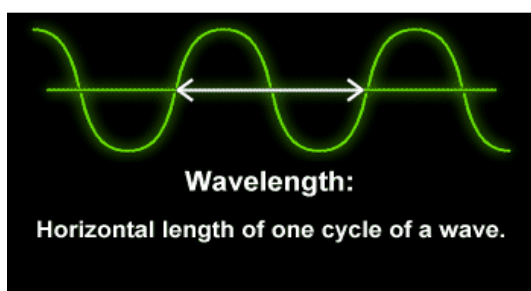
Η τιμή  $P$ = τιμή μέτρησης

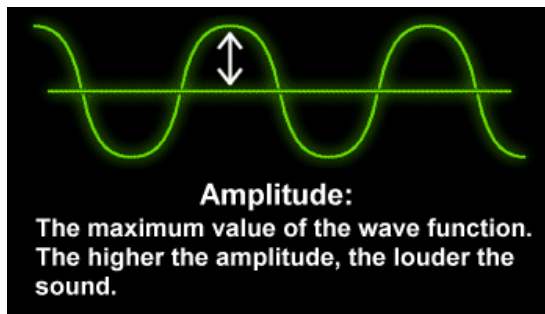
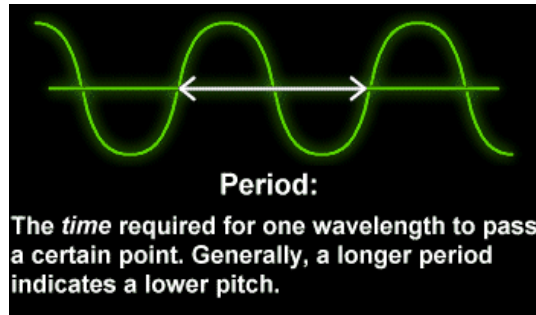
Η τιμή  $P_0$ =τιμή αναφοράς σε μονάδες  $\text{Watt/m}^2$

Παράδειγμα: Για  $P=10^{-12}\text{Watt/m}^2$  έχουμε

$$\text{Ένταση} = 10 \log(10^{-12}/10^{-12}) = 10 \log 1 = 0 \text{ db}$$

Δηλαδή, ο ήχος έντασης  $0 \text{ db}$  αντιστοιχεί στο κατώφλι ακουσικότητας ενώ ο ήχος έντασης  $140 \text{ db}$  αντιστοιχεί στο όριο πόνου. Συμπερασματικά, μικρότερη ένταση σε  $\text{db}$  αντιπροσωπεύει μεγάλη αύξηση ηχητικής ισχύος ενώ θεωρητικά, η αύξηση κατά  $3 \text{ db}$  διαμορφώνει διπλασιασμός της ηχητικής ισχύος. Έτσι, όταν ένας ήχος είναι  $10$  φορές πιο ισχυρός από την σχεδόν απόλυτη ησυχία έχει ένταση  $10\text{db}$  , $100$  φορές πιο ισχυρός έχει ένταση  $20\text{db}$  και  $1000$  φορές έχει ένταση  $30\text{db}$ .





**Σχ.1.3.1: Χαρακτηριστικά του Ψηφιοποιημένου ήχου**

**Ακουστικότητα:** Η ακουστικότητα αποτελεί ένα από τα πιο υποκειμενικά γνωρίσματα του ήχου τα οποία κάθε ακροατής αντιλαμβάνεται με διαφορετικό τρόπο. Ήχος με ίδια ένταση θα έχει μικρότερη ακουστικότητα για ένα άτομο κε προβληματική ακοή. Με επακόλουθο η ακουστικότητα να έχει σχέση με την ένταση του ήχου. Σε σχέση με την ακουστικότητα οι ήχοι διακρίνονται σε ασθενείς με ισχυρούς, ενώ μονάδα μέτρησης της είναι το Phon, δηλαδή ήχος μόλις ακούγεται έχει ακουστικότητα 1 Phon ενώ με ακουστικότητα 130 Phon προκαλεί πόνο στο αυτί.

**Ύψος:** Ένα άλλο υποκειμενικό χαρακτηριστικό αποτελεί το ύψος το οποίο συνδέεται άμεσα με τη συχνότητα το ήχου. Η σχέση αυτή διακρίνει τους ήχους σε δύο μορφές, πρώτον σε οξείς με μεγάλη συχνότητα, όπως τους ήχους ενός βιολιού και κατά δεύτερον σε βαρείς με μικρή συχνότητα, όπως τους ήχους που παράγει το μπάσο τύμπανο. Και στις δύο περιπτώσεις η συχνότητα εξαρτάται από το πάχος και το μήκος της χορδής, με αποτέλεσμα όσο πιο μεγάλη σε μήκος και διάμετρο είναι η χορδή ,τόσο μικρότερη συχνότητα

παράγεται, ενώ όσο πιο πολύ μειώνεται το πάχος και το μήκος περισσότερο αυξάνεται η συχνότητα .

**Χροιά:** Ως τρίτο βασικό υποκειμενικό γνώρισμα είναι η χροιά, χάρη στην οποία ο ήχος ξεχωρίζει ακόμη και αν τα υπόλοιπα χαρακτηριστικά του είναι τα ίδια. Η χροιά αναφέρεται στους σύνθετους ήχους και εξαρτάται από τους απλούς ήχους που αποτελούν το σύνθετο. Κάθε περιοδικός ήχος μπορεί να αναλυθεί σε ένα άθροισμα συνημιτονικών συναρτήσεων με διάφορα πλάτη και φάσεις και με συχνότητες ακέραια πολλαπλάσιες μιας θεμελιώδους συχνότητας (fundamental frequency). Οι συχνότητες αυτές αποκαλούνται αρμονικές (harmonic frequency) και το πλήθος και το σχετικό τους πλάτος είναι σε μεγάλο βαθμό υπεύθυνες για το υποκειμενικό αίσθημα της χροιάς ενός ήχου. Σε αρκετές περιπτώσεις η αρμονική συχνότητα και το μεγαλύτερο πλάτος καθορίζει και το ύψος του ήχου.

## **2. ΚΕΦΑΛΑΙΟ 2: ΔΙΑΔΙΚΑΣΙΑ ΨΗΦΙΟΠΟΙΗΣΗ ΉΧΟΥ**

### **2.1. ΣΤΑΔΙΑ ΨΗΦΙΟΠΟΙΗΣΗΣ ΗΧΗΤΙΚΩΝ ΑΡΧΕΙΩΝ.**

Ο συνδυασμός των υπολογιστών και της ψηφιακής τεχνολογίας κατά την δεκαετία του 90`, επέκτεινε στην κυριολεξία όλους τους φραγμούς που επί δεκαετίας υπήρχαν στο χώρο της μουσικής απόλαυσης. Η επιθυμία και το προαιώνιο όνειρο των φιλοσόφων μετατρέποντας τα αισθητά (ήχους – εικόνα- λέξεις) σε νοητά, υλοποιείται μέσω της

ψηφιακής «επανάστασης». Τέρμα πια στους στείρους διαχωρισμούς μεταξύ επαγγελματικών και ερασιτεχνικών στούντιο και αυτό γιατί πρόκειται ούτε λίγο ούτε πολύ για ένα θαύμα, είτε μας αρέσει είτε όχι η ψηφιοποίηση του κόσμου αποτελεί κατόρθωμα του πνεύματος.

Ο ήχος είναι μια διακύμανση πιέσεων και το μικρόφωνο αποτελεί το όργανο εκείνο που μετατρέπει ένα ηχητικό σήμα σε διακύμανση ηλεκτρονικής τάσης, αυτή η ηλεκτρονική τάση είναι το αναλογικό ηλεκτρονικό σήμα. Για να γίνει δυνατή η επεξεργασία του από τον Η/Υ απαραίτητη προϋπόθεση η ψηφιοποίηση του, μετατρέποντας το αναλογικό (analogy) σε ψηφιακό (digital) με την χρήση των μετατροπών, ADC (Analog-to-Digital-Converters) και DAC (Digital-to-Analog- Converters DACs) και αντίστροφα<sup>2</sup>. Έτσι, η διαδικασία αυτή συμπεριλαμβάνει τρεις βασικές λειτουργίες: την «δειγματοληψία» του αρχικού σήματος, τον «κβαντισμό» των τιμών του σήματος διακριτών χρόνου που προκύπτει από την δειγματοληψία και τέλος η «κωδικοποίηση».

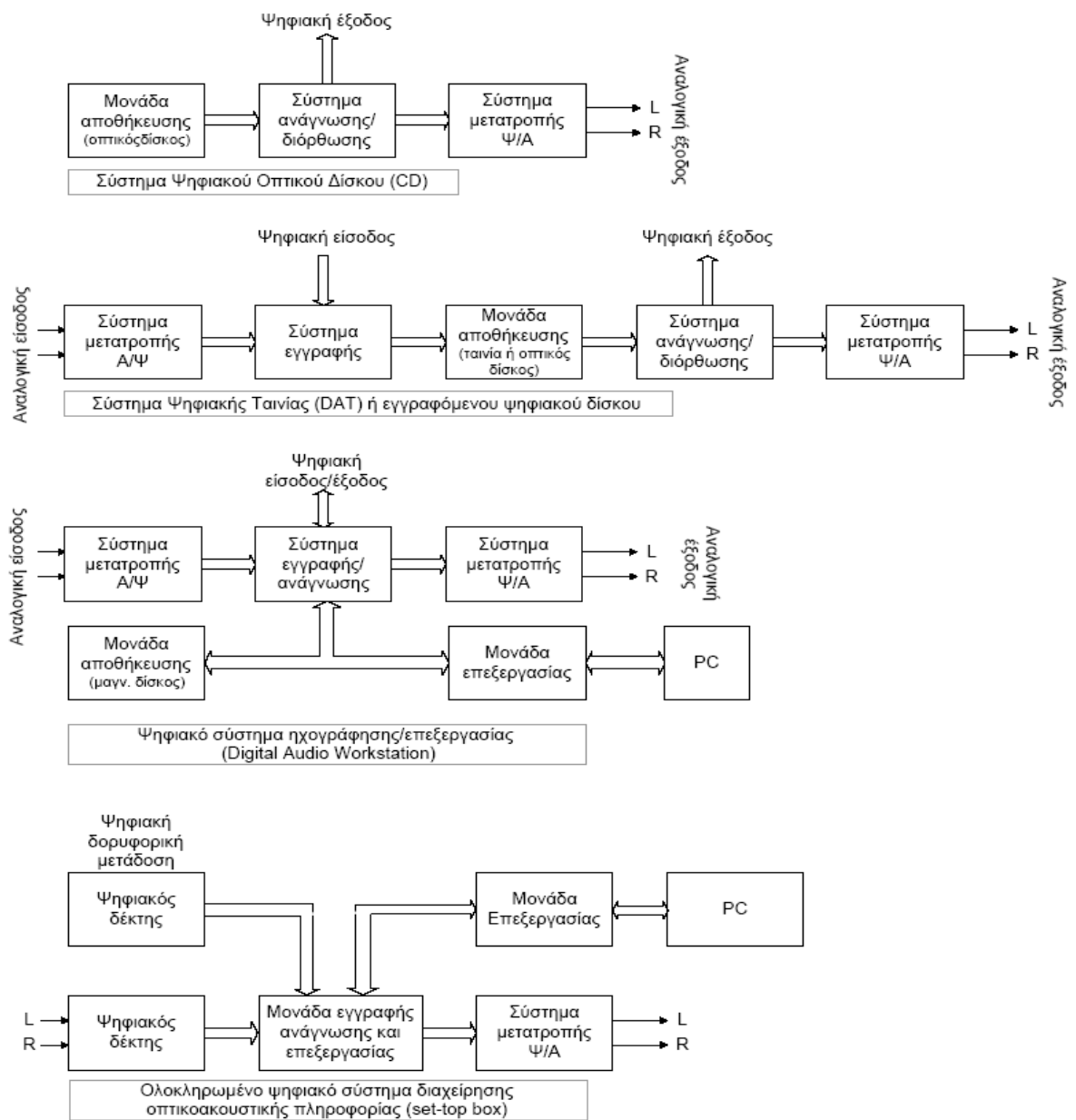
Πιο συγκεκριμένα, κατά την **δειγματοληψία** παίρνονται δείγματα του πλάτους του αναλογικού σήματος σε τακτά χρονικά διαστήματα με συνεπακόλουθο να δημιουργείται ένα στιγμιότυπο του σήματος της συγκεκριμένης τιμής, σύνολο το οποίου τα δείγματα διαμορφώνουν το ψηφιοποιημένο σήμα. Αντίθετα, στον **κβαντισμο** οι διαδοχικές τιμές της στάθμης (πλάτους) του σήματος διακριτού χρόνου  $s_d(nT_s)$  μετατρέπεται σε διακριτές (ψηφιακές) τιμές, πραγματοποιείται δηλαδή μια απεικόνιση της μορφής  $s(nT_s) = Q\{s_d(nT_s)\}$  όπου  $Q\{s_d(nT_s)\}$ <sup>3</sup> είναι η κβαντισμένη τιμή. Κατά την παραπάνω μετατροπή το τελικό σήμα είναι διακριτό τόσο ως προς το χρόνο όσο και ως προς το πλάτος που καλείται ψηφιακό.

Όπως έχει ήδη τυπωθεί, το αναλογικό σήμα ακουστικής πίεσης μετατρέπεται σε ψηφιακό με την χρήση υποσυστημάτων εφαρμογής Α/Ψ (Analog-to-Digital-Converters) και το αντίστροφο, υλοποιείται με την χρήση Ψ/Α μετατροπών (Digital-to-Analog- Converters DACs) όπου οι κβαντισμένες τιμές μετατρέπονται σε αναλογικές<sup>4</sup>. Σε τέτοια υποσυστήματα



οι διαδικασίες κβαντισμού και δειγματοληψίας πραγματοποιούνται ταυτόχρονα, ενώ η υλοποίηση των μετατροπών και στις δύο φάσεις, παρουσιάζουν τόσο θεωρητικά όσο και πρακτικά προβλήματα, ενώ για την σταθερή μετατροπή καλό θα είναι να ικανοποιούνται συγκεκριμένες συνθήκες δειγματοληψίας και κβαντισμού στις οποίες αναφορά γίνεται πιο κάτω.

#### **Σχ.2.1.1: Διαδικασία Ψηφιοποίησης Ηχητικού Σήματος**

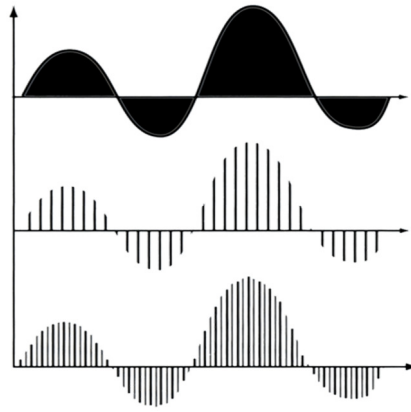


Σχ.2.1.2: Διαδικασία Ψηφιοποίησης Ηχητικού Σήματος

## 2.2. ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Όπως προαναφέρθηκε, η δειγματοληψία είναι η διαδικασία που παράγει την σειρά των δειγμάτων από τα οποία δημιουργείται η ψηφιακή πληροφορία, δηλαδή η διαδικασία κατά την οποία ένα αναλογικό σήμα  $s_c(t)$  (σήμα του οποίου η ανεξάρτητη χρονική μεταβλητή  $t$ , παίρνει συνεχείς τιμές) μετατρέπεται σε σήμα διακριτού χρόνου (σήμα το οποίο ορίζεται σε διακριτές χρονικές στιγμές), πραγματοποιείται δηλαδή μια απεικόνιση της μορφής  $s_c(t) \rightarrow s_d(nT_s)$  όπου  $n=1,2,\dots$  είναι ο αριθμός του δείγματος και  $T_s(\text{sec})$  η περίοδος δειγματοληψίας.<sup>5</sup>

Η επιλογή της περιόδου δειγματοληψίας, καθορίζει και το μέτρο της ποιότητας του ψηφιακού ήχου με επακόλουθο να αποτελεί καθοριστικό παράγοντα ορθής μετατροπής ενός σήματος συνεχούς χρόνου σε διακριτού χρόνου δείγματος. Η συχνότητα δειγματοληψίας (sampling frequency), δηλώνει τον αριθμό των δειγμάτων πλάτους του αναλογικού σήματος καθώς και το πόσες φορές το δευτερόλεπτο παίρνουμε δείγματα από το αναλογικό σήμα, το οποίο μετριέται σε χιλιάδες κύκλους ανά δευτερόλεπτο ή kHz. Όσο μεγαλύτερη είναι η συχνότητα ψηφιοποίησης τόσο καλύτερη ποιότητα αναπαραγωγής έχουμε άλλα ταυτόχρονα, με ριζική αύξηση του χώρου αποθήκευση (σχ.2.2).<sup>6</sup>



Σχ.2.2.1:(α) Αναλογικό σήμα, (β) Ψηφιακό με χαμηλή συχνότητα, (γ)Ψηφιακό σήμα με υψηλή συχνότητα

Απαραίτητη διαδικασία για την μετατροπή ενός αναλογικού ηχητικού σήματος σε ψηφιακό είναι η δειγματοληψία, όπως αναδύεται στο **θεώρημα του Nyquist** από το οποίο απορρέει ότι συχνότητα δειγματοληψίας θα πρέπει να είναι μεγαλύτερη ή ίση του διπλάσιου της μέγιστης συχνότητας του αρχικού αναλογικού σήματος,  $f_s > 2f$  ή  $T_s \leq T/2$ .

Για να ισχύει το παραπάνω πρέπει το αρχικό σήμα να αναζωπυρωθεί στο μισό της συχνότητας δειγματοληψίας περνώντας από ένα ιδανικό βαθυπερατού φίλτρου (βλ. παρακάτω).

Οι συχνότητες δειγματοληψίας που χρησιμοποιούνται σήμερα για την ψηφιακή εγγραφή μουσικής, κυμαίνονται στα 22 με 44kHz επιτρέποντας να περάσει ένα φάσμα ακουστικών συχνοτήτων από 11 έως 22kHz. Κατά συνέπεια, ο ήχος που γίνεται αντιληπτός από το ανθρώπινο αυτί δεν ξεπερνά τα 18 με 22kHz και γι' αυτό θα πρέπει να δειγματολογούμε με συχνότητα περίπου 44kHz(CD) με επακόλουθο ένα πιστό ψηφιακό αντίγραφο του αναλογικού σήματος. Ο A/D παρέχει σε κάθε δείγμα μια τιμή ακέραιου αριθμού που ισοδυναμεί με το μέγεθος του δείγματος . Η ανάλυση τιμών εξαρτάται από την ανάλυση των δειγμάτων (sampling resolution), όπου εάν ο A/D είναι 8 bits, οι τιμές μπορεί να κυμαίνονται από 0 έως 255. Συνήθως κατά την διαδικασία της ηχογράφηση επιλέγουμε την μέγιστη ανάλυση δειγμάτων που υποστηρίζεται από το hardware. Ενώ μια καλή επιλογή είναι τα 8 bits και τα οποία χρησιμοποιούνται ευρέως . Αντίθετα, για ποιότητα CD ο ψηφιακός ήχος κωδικοποιείται ως 44,1kHz στηριζόμενος στο 16bit σύστημα δηλ. το αρχικό κύμα τεμαχίζεται 44.100 φορές το δευτερόλεπτο και ένα μέσο εύρος επίπεδο εφαρμόζεται σε κάθε δείγμα (16 bit σημαίνει ότι 65.536 διαφορετικές τιμές μπορούν να οριστούν ή να κβαντοποιηθούν σε κάθε δείγμα ).

Κανάλια	Δείγμα	Δειγματοληψία	Χρόνος Δείγματος	Μέγεθος Αρχείου
Mono	8bits	11kHz	10 sec	110000 bytes
Stereo	8bits	11kHz	10 sec	220000 bytes
Mono	8bits	22kHz	10 sec	220000 bytes
Stereo	8bits	22kHz	10 sec	440000 bytes
Mono	16bits	44,1kHz	10 sec	882000 bytes
Stereo	16bits	44,1kHz	10 sec	1764000 bytes

Στον παραπάνω πίνακα παρατηρείται ο ρυθμός δειγματοληψίας (sampling rate) και ο αριθμός των δειγμάτων που συλλέγονται ανά sec. Όπως είναι φυσικό, το μέγεθος του αρχείου στο οποίο αποθηκεύεται το ψηφιοποιημένο σήμα ολοένα και αυξάνεται με το μέγεθος του δείγματος (sampling size) και το ρυθμό δειγματοληψίας (sampling rate) , με αποτέλεσμα η καλύτερη απόδοση να επιταχύνεται με δειγματοληψία 44,1 kHz με μέγεθος δείγματος 16bits σε 2 κανάλια ήχου (ποιότητα μουσικών CD).

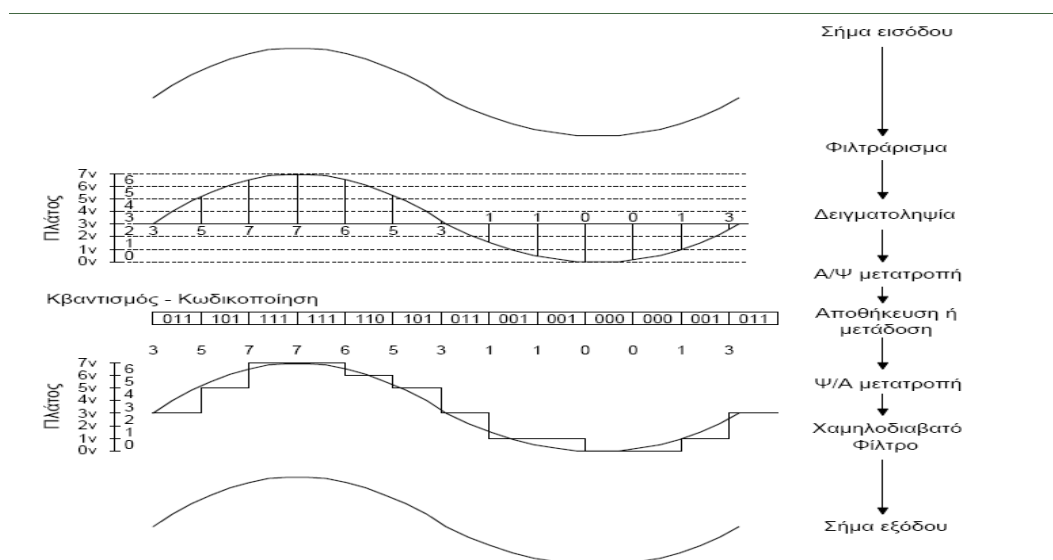
Συμπερασματικά, για να υπολογίσουμε το μέγεθος ενός ψηφιακού αρχείου το οποίο εξαρτάται από την χρονική διάρκεια και την ποιότητα του ήχου, πολλαπλασιάζουμε την συχνότητα δειγματοληψίας σε Hz με το μέγεθος του δείγματος σε bit και με την διάρκεια του ήχου σε δευτερόλεπτα :

**Χωρητικότητα (bits) = Συχνότητα (Hz) . Μέγεθος δείγματος (bits) . Διάρκεια (sec)**

### 2.3. ΠΡΟΒΛΗΜΑΤΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Ο επιστήμονας H.Nyquist, πρώτος ανακάλυψε ότι η μέγιστη συχνότητα αναλογικού σήματος που μπορεί να αποδοθεί χωρίς αλλοίωση είναι το μισό της συχνότητας δειγματοληψίας. Συχνότητες μεγαλύτερες από τα μισά της συχνότητας ψηφιοποίησης εμφανίζονται λανθασμένες σαν χαμηλές συχνότητες, με επακόλουθο το φαινόμενο αυτού του ελλιπούς ρυθμού δειγματοληψίας να ονομάζεται αναδίπλωση (aliasing). Έτσι για να μην εμφανιστούν συχνοτικά ειδώλια που θα παραμορφώσουν το σήμα μετά την δειγματοληψία οι ψηφιοποιητές φιλτράρουν όλες τις συχνότητες που είναι μεγαλύτερες από το μισό της συχνότητας δειγματοληψίας.

Κατά συνέπεια, το ηλεκτρονικό σήμα μετά το στάδιο της προενίσχυσης τροφοδοτείται σε ένα ιδανικό βαθυπερατό φίλτρο (low-pass), με στόχο την αποκοπή των συχνοτήτων μεγαλύτερων από την μέγιστη συχνότητα δειγματοληψίας για αποφυγή του φαινομένου της αναδίπλωσης. Στη συνέχεια, κατά το στάδιο της δειγματοληψίας τροποποιείται το αναλογικό ηλεκτρικό σήμα της εισόδου σε δυαδικούς αριθμούς και αποθηκεύεται στη μνήμη του Η/Υ, με αποτέλεσμα να ολοκληρώνεται ο μετασχηματισμός από το αναλογικό (analog) στο ψηφιακό (digital) (σχ.2.3)<sup>7</sup>.



Σχ.2.3.1 : Διαδικασία μετατροπής αναλογικού σε ψηφιακή μορφή και αντίστροφα

Η απόδοση ενός ψηφιακού ηχητικού σήματος ακολουθεί μια αντίστροφη διαδρομή βημάτων από αυτή της καταγραφής, δηλαδή η τροφοδότηση του σήματος υλοποιείται από τη μνήμη του Η/Υ σε ένα μετατροπέα ψηφιακού σε αναλογικού σήματος. Με βάση τον τρόπο λειτουργίας και την συχνότητα δειγματοληψίας, ο μετατροπέας παράγει ένα ηλεκτρικό σήμα που αποτελεί μια προσεγγιστική μορφή του ηλεκτρικού σήματος το οποίο ενισχύεται και στη συνέχεια τροφοδοτείται στη είσοδο του ηχείου, αποδίδοντας το σήμα.

Στο σημείο αυτό, ανακύπτει το ερώτημα αν η εισαγωγή τέτοιων φίλτρων (filter), διαμορφώνουν νέα προβλήματα κατά την ψηφιοποίηση του ήχου. Απάντηση είναι, πώς δεν υπάρχει ένα τέλειο βαθυπερατό φίλτρο που να μπορεί να αποκόψει τέλεια συχνότητες πάνω από ένα επιθυμητό όριο, ενώ η χρήση τέτοιων φίλτρων διαμορφώνει μικρές παραμορφώσεις στο σήμα στο οποίο και εκφράζεται. Ομοίως, όπως έχει ήδη τυπωθεί κατά την διαδικασία της Ψ/Α μετατροπής εξαιτίας των κβαντισμένων τιμών πλάτος του αναπαραγόμενου σήματος, διαμορφώνονται συχνότητες υψηλότερες της μέγιστης συχνότητας όπου και φιλτράρονται με ένα παρόμοιο φίλτρο εξομάλυνσης (smoothing filter)<sup>8</sup>.

## 2.4. ΚΒΑΝΤΟΠΟΙΗΣΗ

Το επόμενο εφαρμοσμένο βήμα με μεγάλη συμβολή στην ποιότητα του ήχου είναι αυτό της **κβάντωσης**, προκειμένου να αναπαρασταθεί κάθε δείγμα με την μορφή μιας δυαδικής σειράς από bits, καθώς πρέπει στο συνεχώς μεταβαλλόμενο πλάτος της τάσης του αναλογικού σήματος να τεθεί μια διακριτή τιμή. Σημαντικό είναι να τον τονίσουμε ότι η δειγματοληψία και η κβαντοποίηση είναι συμπληρωματικές διαδικασίες και ο συνδυασμός

αυτών των δύο διαδικασιών, καλείται ψηφιοποίηση. Η κβαντοποίηση παίζει καθοριστικό ρόλο στο σχεδιασμό του συστήματος του ψηφιακού ήχου, ενώ ανάλογα με τις στάθμες κβάντισης διαμορφώνεται και το format του ψηφιακού ήχου.

Βασική παράμετρος της φάσης αυτής είναι το μέγεθος του δείγματος (sampling size), με επακόλουθο όσα περισσότερα bits χρησιμοποιούνται με μεγαλύτερη ακρίβεια περιγραφής. Για παράδειγμα, αν έχουμε 8 bit, τότε μπορούν χρησιμοποιηθούν  $2^8=256$  διαφορετικές τιμές, ενώ για 16bit μπορούν να χρησιμοποιηθούν  $2^{16}=65.536$  διαφορετικές τιμές, άρα μεγαλύτερη πιστότητα και ακρίβεια.

Κατά την πεπερασμένη διακριτικότητα της ψηφιακής αναπαράστασης, εισάγεται ο *θόρυβος* ένα σημαντικό πρόβλημα τόσο κατά την ηχογράφηση όσο και κατά την αναπαραγωγή, γνωστός και ως **θόρυβος κβάντοποίησης**. Το φαινόμενο αυτό προέρχεται από το γεγονός ότι τα δείγματα του αναλογικού σήματος που θα χρησιμοποιηθούν δεν είναι κατά ανάγκη ακέραιοι και συνεπώς η μετατροπή τους σε ψηφιακά δείγματα συνεπάγεται στην στρογγυλοποίηση του δείγματος. Πρέπει, όμως να τονισθεί ότι στη περίπτωση αυτή το ψηφιοποιημένο σήμα απαρτίζεται από το άθροισμα του αναλογικού σήματος καθώς και από ένα σήμα θορύβου το οποίο και μετριέται σε αρνητικά decibel (db= μονάδα μέτρησης έντασης του ήχου) .Το σφάλμα αυτό εκφράζεται από το λόγο σήματος προς θόρυβο, (Signal to Noise Ratio, SNR)<sup>9</sup> **S/E (dB)=6.02n+176** όπου το n είναι ο αριθμός των bits της λέξης στο format του ψηφιακού ήχου. Όσο μεγαλύτερος είναι ο αριθμός των db τόσο μικρότερος είναι ο θόρυβος, ενώ τα περισσότερα συστήματα σήμερα κυμαίνονται στα 90db SNR, όπου η παραμόρφωση μετριέται επί τις εκατό του αρχικού σήματος Η ευκρίνεια του κβαντισμού και του λόγου σήματος προς θόρυβο παρουσιάζεται στον παρακάτω πίνακα όπου ενδεικτικές τιμές SNR αντιστοιχούν σε διαφορετικές τάξεις μετατροπής.

Τάξη N(bits)	SNR (db)
8	49.8



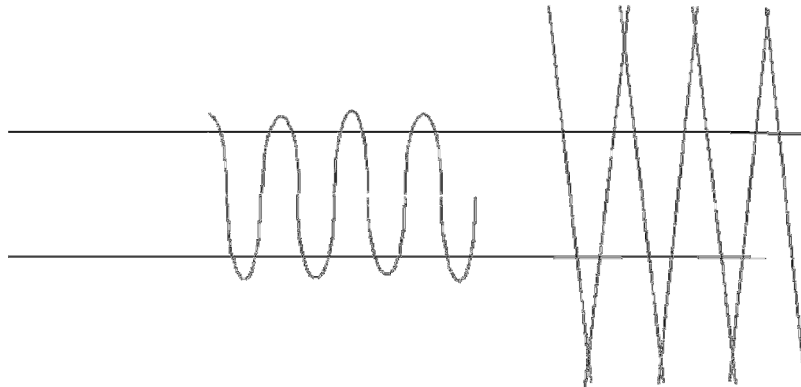
12	73.8
16	97.8
18	109.8

Πίνακας Ευκρίνειας Κβαντισμού και SNR.

Συνοψίζοντας τα πιο πάνω, ανάγεται το συμπέρασμα ότι ο θόρυβος που παράγουν τα ψηφιακά ηχητικά συστήματα, εμφανίζεται κατά την διαδικασία της αναπαραγωγής του κβαντισμένου ηχητικού σήματος, έτσι ώστε να διαμορφώνεται ένας ανεπιθύμητος χαρακτήρας.

Στην περίπτωση των 16bit συστημάτων, επιβάλλεται η χρήση του **dither** κατά την Α/Ψ, μέθοδος κατά την οποία ενσωματώνεται μικρή ποσότητα αναλογικού θορύβου πριν από το στάδιο της δειγματοληψίας, εξασθενώντας πολλά από τα προβλήματα που συζητήθηκαν παραπάνω ούτως ώστε να εμφανίζεται ένα ικανοποιητικό αποτέλεσμα. Σε γενικές γραμμές, η προσθήκη του dither τροποποιεί την μορφή του σφάλματος και το καθιστά ανεξάρτητο από το σήμα εισόδου, αποκόπτει κάθε είδους αρμονικής παραμόρφωσης η οποία εμφανίζεται κατά την μετατροπή σήματος μικρού πλάτους και τέλος βελτιώνει την διακριτή ικανότητα του κβαντιστή αυξάνοντας κατά υποκειμενικό τρόπο την δυναμική του περιοχή.

Ένας άλλος σημαντικός παράγοντας που επηρεάζει την ποιότητα του ήχου, είναι ο **ψαλιδισμός** (clipping) του σήματος, όπου κατά την εγγραφή του σήματος η μέγιστη στάθμη του σήματος εισόδου υπερβαίνει την μέγιστη στάθμη κβαντισμού διαμορφώνοντας κάποια διαστρέβλωση. Στο παρακάτω σχήμα 2.4, απεικονίζεται ένα ψαλιδισμένο σήμα με κοψίματα στην κορυφή και στη βάση με επακόλουθο ο ψαλιδισμός να μειώνεται αφού μειωθεί το μέγεθος του σήματος.



Σχ2.4.1: Ψαλιδισμός (Clipping).

### 3. ΚΕΦΑΛΑΙΟ 3 : ΚΩΔΙΚΟΠΟΙΗΣΗ ΨΗΦΙΑΚΟΥ ΣΗΜΑΤΟΣ

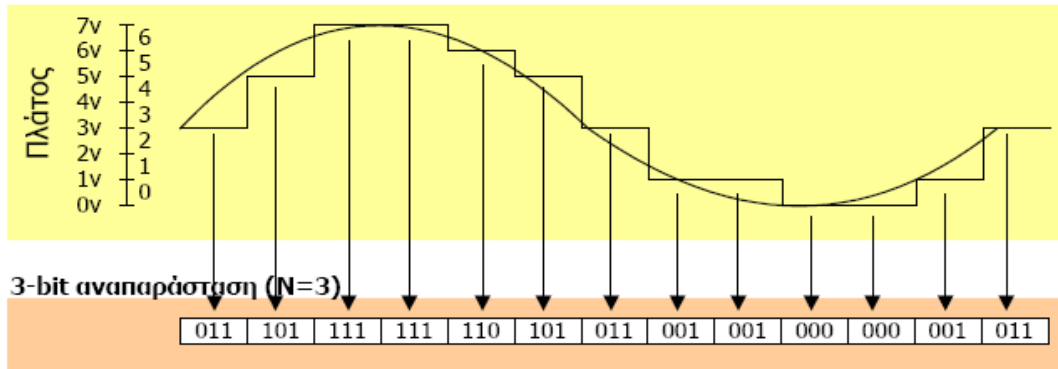
#### 3.1. ΜΕΘΟΔΟΙ ΚΩΔΙΚΟΠΟΙΗΣΗ

Ο συνδυασμός της δειγματοληψίας και της κβαντοποίησης ως συμπληρωματικές διαδικασίες αποκαλείται ψηφιοποίηση, ενώ για την εκμετάλλευση των πλεονεκτημάτων τους απαραίτητη προϋπόθεση η χρησιμοποίηση μιας διαδικασίας **κωδικοποίησης** (encoding process) για την μετατροπή του διακριτού συνόλου των τιμών των δειγμάτων σε μια πιο κατάλληλη μορφή. Κατά το στάδιο αυτό, η ένταση του κάθε δείγματος αντιστοιχεί σε ένα σταθερό δυαδικό αριθμό 0,1 και διατίθεται για την αποθήκευση κάθε δείγματος και καλείται εύρος δείγματος (τιμές που μπορεί να πάρει το κάθε δείγμα).

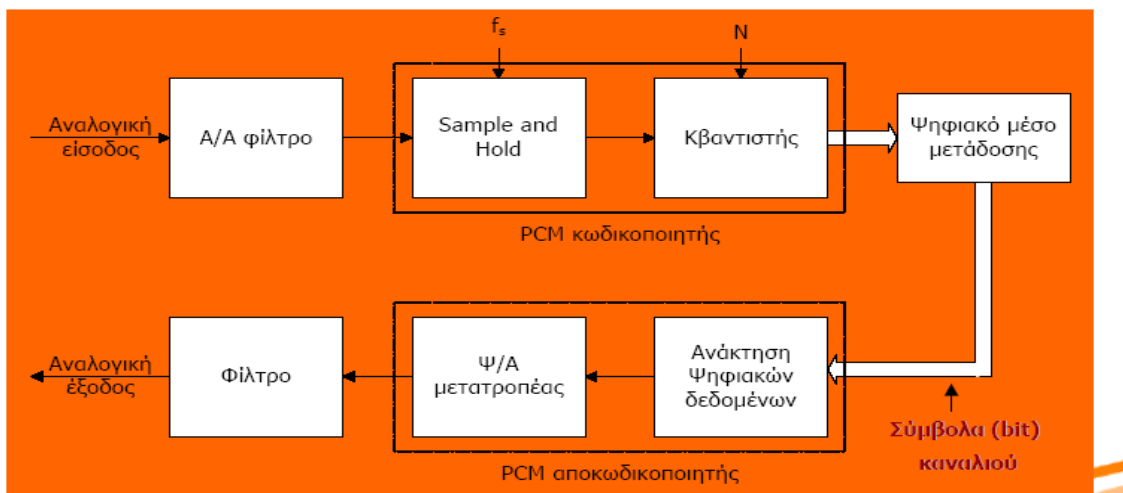
Για το ψηφιακό ήχο οι δυνατότητες ποικίλουν από 8bits ( $2^8=256$  στάθμες) έως και 16bits ( $2^{16}=65536$  στάθμες) ,με επακόλουθο η χρήση ενός δυαδικού συστήματος να αναδύει καλύτερα αποτελέσματα σε συνάρτηση με την επίδραση του θορύβου σε ένα μέσο μετάδοσης και αυτό γιατί ένα δυαδικό σύμβολο, αντέχει σε μια σχετική υψηλή στάθμη θορύβου καθώς εύκολα μπορεί να αναγεννηθεί.

❖ **PCM** (Pulse Code Modulation).

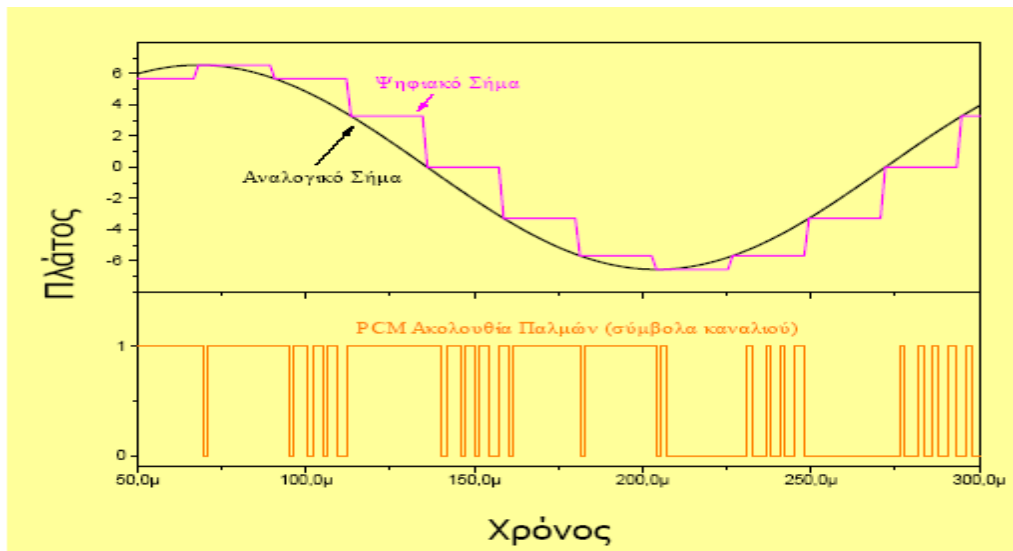
Μια από τις πιο απλές και ευρέως μεθόδους κωδικοποίησης ψηφιακού ήχου είναι η παλμοκωδική κωδικοποίηση PCM. Στην μέθοδο αυτή κάθε δείγμα αναπαριστάται με ένα σύνολο παλμών που αντιστοιχούν στον δυαδικό κώδικα και στην τιμή του δείγματος με επακόλουθο να αποθηκεύεται ένα προς ένα τα δείγματα σε ψηφιακή μορφή χρησιμοποιώντας γραμμική κωδικοποίηση. Όπως είναι αναμενόμενο, η πιστότητα του σήματος που προκύπτει είναι συνάρτηση του δυαδικού κώδικα, π.χ. ένα δυναμικό εύρος 128 τιμών θα έχει ως αποτέλεσμα την απάλειψη ήχων στο ψηφιοποιημένο σήμα με ένταση ίση ή μικρότερη από το  $1/128$  της έντασης του δυνατότερου ήχου που μπορεί να αναπαρασταθεί από το σήμα. Στο σημείο αυτό, ανακύπτει το ερώτημα αν η πιστότητα σε ένταση σε ασθενή ηχητικά σήματα εισάγει πρόβλημα. Προφανώς εισάγει και για αυτό το λόγο έχουν αναπτυχθεί άλλες τεχνικές οι οποίες χρησιμοποιούν λογαριθμική κωδικοποίηση PCM όπου και αναλύονται στις επόμενες παραγράφους. Τέτοιες τεχνικές, χρησιμοποιούν περισσότερα δυαδικά ψηφία για την αναπαράσταση ασθενών ήχων και λιγότερα για τους ήχους μεγάλης έντασης στους οποίους η ευαισθησία της ανθρώπινης ακοής σε σχετικές διαφορές έντασης είναι μειωμένη (βλ. προβλήματα δειγματοληψίας) .



Σχ 3.1.1: Κωδικοποίηση PCM



Σχ. 3.1.2 : Χρήση PCM Κωδικοποίηση



Σχ. 3.1.3: Παράδειγμα PCM Κωδικοποίησης.

#### ❖ Mu-Law PCM και A-Law PCM.

Η διαφορά της λογαριθμικής κωδικοποίησης ήχου σε σχέση με την γραμμική έγκειται στο γεγονός ότι κατά την λογαριθμική αντιστοίχιση, αντιστοιχούνται ολοένα και περισσότερες στάθμες σε χαμηλές συχνότητες και λιγότερες στις υψηλές, ενώ ταυτόχρονα πραγματοποιείται τόσο καλύτερη αναπαράσταση όσο και καλύτερη συμπίεση του σήματος με αυτό της παλμοκωδικής, η οποία δεν πραγματοποιεί καμία συμπίεση ούτως ώστε να προκύπτει ένα αρχείο ήχου κωδικοποιημένο χωρίς καμία απώλεια. Πιο συγκεκριμένα, 8bits σε συνδυασμό με λογαριθμική κωδικοποίηση καλύπτει το ίδιο εύρο τιμών 14bits και παλμοκωδική κωδικοποίηση PCM, επομένως πρόκειται για μια συμπίεση της τάξης του 1,75 προς 1. Κατά συνέπεια, στην κατηγορία αυτή της λογαριθμικής κωδικοποίησης ήχου εντάσσονται οι Mu-Law PCM και A-Law PCM, δύο μέθοδοι που έχουν τυποποιηθεί από το ITU-T (International Telecommunication Union-Telecommunication Standardization Sector) στο πρότυπο G711, Pulse Code Modulation of noise Frequencies το οποίο και περιγράφει λεπτομερές τα δύο είδη κωδικοποίησης. Συγκεκριμένα, η Mu-Law υλοποιείται σε ISDN (Integrated Services Digital Network) δίκτυα της Ιαπωνίας και Β.Αμερικής ενώ η A-Law σε δίκτυα των υπόλοιπων χωρών.

❖ **DPCM** (Differential Pulse Code Modulation).

Σε σχέση με την PCM κωδικοποίηση ψηφιακού σήματος, η διαφορική παλμοκωδική κωδικοποίηση δεν κωδικοποιεί το κάθε δείγμα ανεξάρτητα από τα υπόλοιπα αλλά σε συνάρτηση με τα γειτονικά δείγματα, αποθηκεύοντας τις διαφορές μεταξύ των διαδοχικών τιμών και όχι τις απόλυτες τιμές των δειγμάτων. Δηλαδή για την χρονική στιγμή  $t$  θα κωδικοποιηθεί η διαφορά των δειγμάτων  $\delta t - \delta t_{-1}$  όπου ως προβλεπόμενη τιμή έχει χρησιμοποιηθεί η τιμή του δείγματος  $\delta t_{-1}$  την χρονική στιγμή  $t-1$ . Η βασική έννοια DPCM που κωδικοποιεί μια διαφορά, στηρίζεται στο γεγονός ότι τα περισσότερα σήματα πηγής παρουσιάζουν σημαντικό συσχετισμό μεταξύ των διαδοχικών δειγμάτων που κωδικοποιούν, ούτως ώστε οι τιμές των δειγμάτων να διαμορφώνουν χαμηλό ποσοστό δυαδικών ψηφίων. Για την υλοποίηση της βασική έννοιας θα πρέπει να προβλέψουμε την τρέχουσα αξία των δειγμάτων η οποία βασίζεται στα προηγούμενα δείγματα καθώς και να κωδικοποιήσουμε την αξία του δείγματος και την προσληφθείσα αξία.

Μια ειδική μορφή της διαφορικής παλμοκωδικής κωδικοποίησης είναι η DM (Delta Modulation), ως απλουστευμένη DPCM κατά την οποία η διαφορά της προβλεπόμενης και της τρέχουσας τιμής του δείγματος κωδικοποιείται με ένα μόνο bit παίρνοντας τιμές σύμφωνα με την **αρχή DM**: Η παραγωγή DM είναι **0** εάν το τρέχον δείγμα έχει μικρότερο εύρος από το προηγούμενο και **1** εάν το τρέχον δείγμα έχει εύρος μεγαλύτερο από το αμέσως προηγούμενο, με επακόλουθο το DM να κωδικοποιεί την κατεύθυνση των διαφορών στο εύρος σήματος αντί την αξία της διαφοράς DPCM. Αυτό σημαίνει ότι κάθε δείγμα μπορεί να είναι είτε μεγαλύτερο είτε μικρότερο κατά ένα κβάντο από το προηγούμενο του με αποτέλεσμα ο περιορισμός να οδηγεί σε μεγάλη οικονομία αλλά αν το σήμα αλλάζει γρήγορα θα υπάρχει μεγάλη απώλεια πληροφορίας.

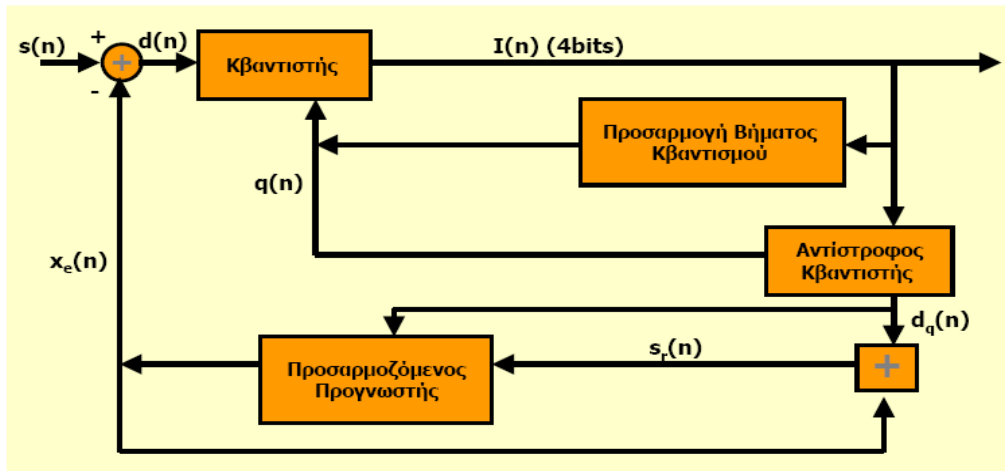
❖ **ADPCM** (Adaptive Differential Pulse Code Modulation).

Η διαφορική παλμοκωδική κωδικοποίηση, αποτελεί μια επέκταση της DPCM μεθόδου κατά την οποία προβλέπεται η τιμή ενός δείγματος με βάση την τιμή του γειτονικού δείγματος

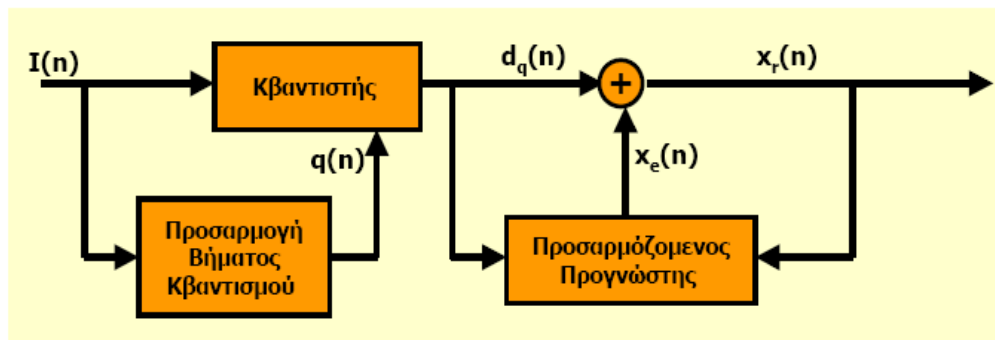
και αυτό επειδή τα γειτονικά δείγματα πολύ πιθανόν να είναι όμοια αν όχι και ίδια. Συνεπώς, η ADPCM κωδικοποίησης, υπολογίζει την διαφορά μεταξύ της πραγματικής τιμής του δείγματος και της τιμής που είχε προβλεφθεί και κωδικοποιείται με συνεπακόλουθο, η διαφορά της τιμής να είναι σχετικά ελάχιστη (4bits) απαιτώντας λιγότερα bits. Στο σημείο αυτό, ένα σημαντικό μειονέκτημα αναδύεται ανάμεσα στις δύο μεθόδους κωδικοποίησης ADPCM και DPCM και αυτό εξαιτίας της εξάρτησης του μεγέθους των διαφορών μεταξύ διαδοχικών δειγμάτων ενός ηχητικού σήματος σε σχέση με την συχνότητα.

Όπως είναι αντιληπτό, οι διαφορές μεταξύ διαδοχικών δειγμάτων σε σχέση με ήχους χαμηλών συχνοτήτων, είναι σημαντικά μικρότερες από τις αντίστοιχες διαφορές σε ήχους υψηλών συχνοτήτων, με αποτέλεσμα η κωδικοποίηση αυτών των διαφορών να εντάσσεται στο συχνοτικό περιεχόμενο του σήματος της κάθε χρονικής στιγμής, γεγονός που δεν πραγματοποιείται στις παραπάνω μεθόδους. Αντίθετα, οι μέθοδοι *κωδικοποίηση υποζώνης* (subband coding) ,κατηγοριοποιούν το ηχητικό σήμα σε δύο ή περισσότερες ζώνες συχνοτήτων και συμπιέζουν κάθε μια από αυτές ξεχωριστά με σκοπό μια τέτοια κατηγοριοποίηση να εκμεταλλεύεται τα χαρακτηριστικά της ανθρώπινης ακοής που παρουσιάζει την μέγιστη ευαισθησία στο εύρος συχνοτήτων μεταξύ 2700-3200Hz με την ευαισθησία να ελαττώνεται όσο απομακρυνόμαστε από την ζώνη αυτή.

Συνοψίζοντας τα πιο πάνω, ανάγεται το συμπέρασμα, πως σε μια τέτοια μέθοδο κωδικοποίησης υποζώνης, τα αποτελέσματα της δεν γίνονται αντιληπτά από την ανθρώπινη ακοή και αυτό επιτυγχάνεται, γιατί έχει την ικανότητα τόσο να συμπιέζει ήχους με συχνοτικό περιεχόμενο που απέχουν από την παραπάνω ζώνη όσο και να εφαρμόζει την ελάχιστη συμπίεση που ανήκει στην ζώνη αυτή. Σε μια τέτοια κατηγορία, ανήκουν τα πρότυπα MPEG audio, Dolby AC-2 και AC-3 και το RealAudio για τα οποία εκτεταμένη αναφορά γίνεται στο παρακάτω κεφάλαιο.



Σχ.3.1.4: Κωδικοποιητής ADPCM

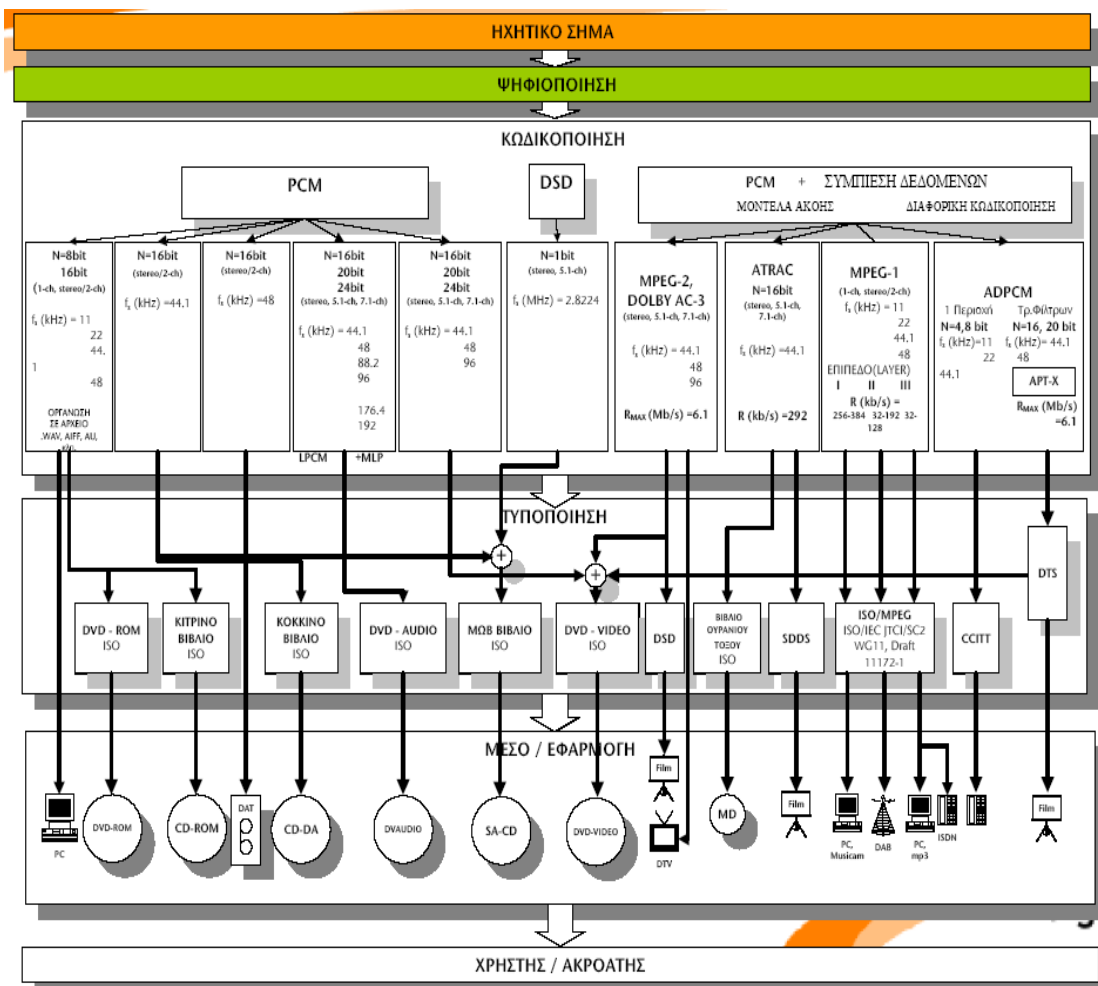


Σχ.3.1.5: Αποκωδικοποιητής ADPCM

❖ **LPC** (Linear Predictive Coding)

Μια από τις νέες τεχνολογίες κωδικοποίησης ήχου που έχουν συνταχθεί για την ανθρώπινη ομιλία και πραγματοποιούν σημαντικούς βαθμούς συμπίεσης είναι η γραμμική προβλεπόμενη κωδικοποίηση LPC, με σκοπό ο κωδικοποιητής αυτός να συγκρίνει τα σήματα φωνής που παραλαμβάνει με βάση ένα αναλυτικό μοντέλο φωνής που έχει αποθηκευμένο. Τα χαρακτηριστικά που ταιριάζουν καλύτερα στο αναλυτικό μοντέλο μεταδίδονται, ενώ ο αποκωδικοποιητής χρησιμοποιεί τα χαρακτηριστικά αυτά για να ανασυνθέσει τα φωνητικά σήματα. Παρόλα αυτά, μειονέκτημα της μεθόδου αυτής εντοπίζονται στην αδυναμία της να επεξεργαστεί άλλο σήμα εκτός από την ομιλία, η οποία δημιουργήθηκε για να εξυπηρετήσει την μετάδοση ομιλίας στην κινητή τηλεφωνία.





Σχ. 3.1.6: Διάγραμμα Κωδικοποιήσεων Ψηφιακού Ήχου

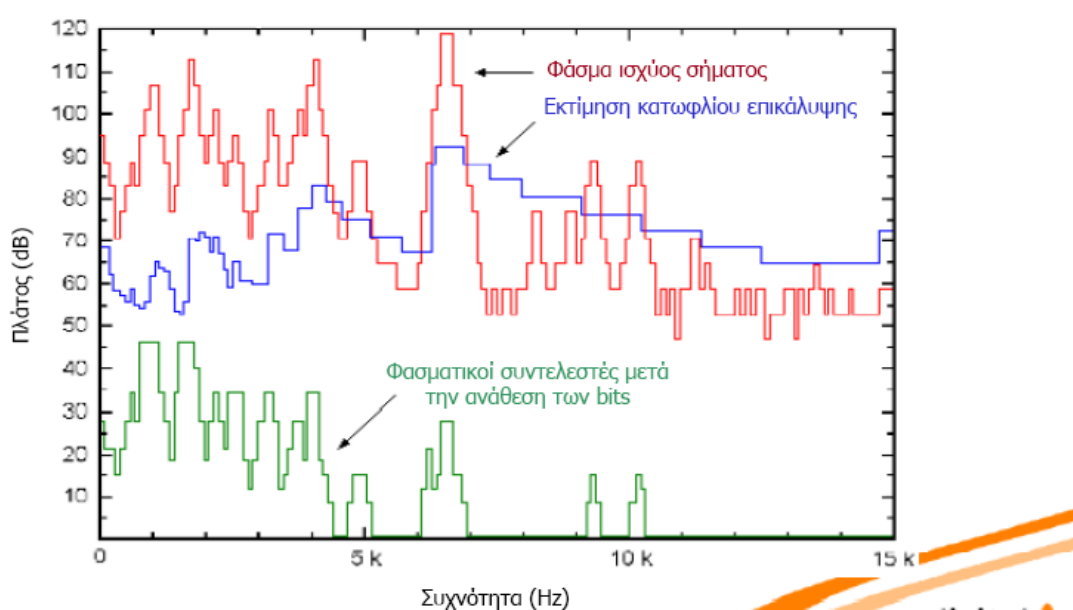
## 4. ΚΕΦΑΛΑΙΟ 4 : ΣΥΜΠΙΕΣΗ

### 4.1. ΣΥΜΠΙΕΣΗ ΨΗΦΙΑΚΟΥ ΉΧΟΥ

Με δεδομένη την ολοένα και αυξανόμενη χρήση της ψηφιακής τεχνολογίας η οποία αντικαθιστά την αναλογική που παραδοσιακά χρησιμοποιείται, εδώ και δεκαετίες στις τηλεπικοινωνίες, την ηχογράφηση και αναπαραγωγή της μουσικής αλλά και σε άλλους τομείς έχει αρχίσει παράλληλα να γίνεται έρευνα προς την κατεύθυνση της συμπίεσης ψηφιακής πληροφορίας και στην περίπτωση μας του ψηφιακού ηχητικού σήματος με στόχο τη οικονομία εύρους φάσματος. Σύμφωνα με αυτά που έχουν αναφερθεί στα παραπάνω κεφάλαια, παρατηρείται ότι κατά την διαδικασία της ψηφιοποίησης ηχητικών δεδομένων, παράγονται αρχεία μεγάλων μεγεθών τα οποία δύσκολο να μεταφερθούν ή να υποστούν επεξεργασία αφού προϋποθέτουν πολύ χρόνο, μεγάλους αποθηκευτικούς χώρους και πολύ μνήμη.

Το τεχνικό πρόβλημα του χώρου και της ανεπάρκειας στην ταχύτητα μεταφοράς έρχονται να λύσουν οι τεχνολογίες συμπίεσης του ηχητικού σήματος. Οι τεχνολογίες αυτές, χρησιμοποιούν διάφορες τεχνικές μείωσης του όγκο και της ροής των δεδομένων που απαιτούνται για την κωδικοποίηση των ηχητικών σημάτων και διακρίνονται σε δύο κατηγορίες, τις **μη απωλεστικές** (lossless) και τις **απωλεστικές** (lossy). Οι μη απωλεστικές, έχουν το ιδιαίτερο χαρακτηριστικό ότι η διαδικασία συμπίεσης δεν αλλοιώνει καθόλου την πληροφορία δηλαδή μετά την αποσυμπίεση η πληροφορία επανέρχεται στην ακριβώς στην αρχική της μορφή. Συνήθως, οι μη απωλεστικοί αλγόριθμοι συμπίεσης ψηφιακού ήχου, εφαρμόζονται σε περιπτώσεις που δεν υπάρχει κανένα περιθώριο απωλειών. Αντίθετα, οι απωλεστικές τεχνικές αλλοιώνουν τα δεδομένα, απορρίπτοντας τα μέρη εκείνα που δεν

είναι χρήσιμα με βάση κάποιο συγκεκριμένο κριτήριο (irrelevancy) . Είναι φανερό, ότι σε τέτοιες περιπτώσεις το σημασιολογικό περιεχόμενο ουσιαστικά δεν μεταβάλλεται αλλά υπεισέρχεται η έννοια της μείωσης της ποιότητας καθώς οι απωλεστικές τεχνικές είναι περισσότερο αποτελεσματικές επιτυγχάνοντας συμπίεση που φθάνει το 24:1 με καλή ποιότητα ενώ οι μη απωλεστικές σπανίως ξεπερνούν το 4:1. Εντούτοις, στις σύγχρονες μεθόδους συμπίεσης εφαρμόζονται διάφορα ψυχοακουστικά μοντέλα αντίληψης βάση των οποίων απορρίπτονται οι ήχοι που έτσι και αλλιώς δεν μπορούν να γίνουν κατανοητοί από το ανθρώπινο αυτί.



Σχ. 4.1.1: Ψυχοακουστικό Μοντέλο Ακοής

Οι σημαντικότεροι μέθοδοι συμπίεσης ψηφιακών ηχητικών αρχείων στηρίζονται στο φαινόμενο της ηχητικής σκίασης (auditory masking), που αποσκοπεί στην ακουστικότητα των ήχων. Πιο συγκεκριμένα, όταν υπάρχει ένας ήχος συγκεκριμένης συχνότητας και έντασης, άλλοι ήχοι σε κοντινές συχνότητες δεν γίνονται αντιληπτοί από το ανθρώπινο αυτί. Παράλληλα, κατά την συμπίεσης των ψηφιακών ηχητικών δεδομένων απορρίπτονται οι συχνότητες που το ανθρώπινο αυτί δεν μπορεί να ξεχωρίσει μέσα σε ένα καθορισμένο διάστημα συχνοτήτων. Αν και χάνονται ορισμένες ηχητικές πληροφορίες εντούτοις η ποιότητα του ήχου παραμένει σε υψηλά επίπεδα (Σχ.4.2).

#### Σχ. 4.1.2 : Φαινόμενο Ηχητικής Σκίασης.

Το φαινόμενο της ηχητικής σκίασης, είναι άμεσα συνδεδεμένο τόσο με το πρότυπο AC-3 Dolby Digital (H.P.A) όσο και με το Mpeg των οποίων οι διαφορές τους στηρίζονται στο τρόπο υλοποίησης τους. Εντούτοις και τα δύο πρότυπα συμπίεσης ηχητικών δεδομένων, δουλεύουν χωρίζοντας το φάσμα των ακουστικών συχνοτήτων σε υπομπάντες χρησιμοποιώντας περίπλοκους αλγόριθμους και ψυχοακουστικά μέσα για να απορρίψουν τις μη ακουστικές από το ανθρώπινο αυτί συχνοτήτες. Ο ανταγωνισμός μεταξύ των προτύπων AC-3 και Mpeg, φαίνεται να ευνοεί το μοντέλο AC-3 το οποίο στην πορεία θα αναδειχθεί ως το διεθνές πρότυπο για την συμπίεσης ψηφιακών ηχητικών δεδομένων.

Συμπερασματικά, όλες οι διαδικασίες συμπίεσης και αποσυμπίεσης υλοποιούνται μέσω κατάλληλων προγραμμάτων, όπου ο ρυθμός μετάδοσης των ψηφιακών δεδομένων είναι πολύ σημαντικός καθορίζοντας την ποιότητα του ήχου καθώς εξαρτάται τόσο από την συχνότητα δειγματοληψίας όσο και το μέγεθος του δείγματος. Για παράδειγμα, ο ήχος του

μουσικού CD χρησιμοποιεί συχνότητες δειγματοληψίας 44.1 kHz με μέγεθος δείγματος 16bits παράγοντας ένα ρυθμό δεδομένων 1.4 Mbit/sec, ενώ αν μειωθεί η συχνότητα δειγματοληψίας τότε θα χαθούν οι υψηλές συχνότητες του ήχου. Επομένως, οι περισσότεροι σύγχρονοι αλγόριθμοι συμπίεσης που χρησιμοποιούν ψυχοακουστικό μοντέλο, στηρίζονται στη μεταβολή των bit για να επιτύχουν την μεγαλύτερη συμπίεση διατηρώντας υψηλή την ποιότητα του αναπαραγόμενου ήχου.

## 4.2. ΠΡΟΤΥΠΟ MPEG

Οι προσπάθειες για μετάδοση ψηφιακού DAB (Digital Audio Broadcasting), ξεκίνησε στην Ευρώπη από το 1987 με το πρόγραμμα Eureka, στο οποίο συμμετείχαν ερευνητές τόσο από το Ινστιτούτο Fraunhofer με επικεφαλής τους Karlheinz Brandenburg όσο και τα μέλη της ομάδας Moving Picture Expert Group (MPEG). Η MPEG, ως μια συνεργασία Πανεπιστημίων, ερευνητών Ινστιτούτων και εταιριών λειτουργεί στα πλαίσια του Διεθνούς Οργανισμού Τυποποίησης γνωστή σαν ISO/IEC με στόχο την ανάπτυξη διεθνών πρότυπων για την συμπίεση και αποσυμπίεση, την επεξεργασία και την κωδικοποιημένη αντιπροσώπευση της κίνησης των εικόνων, του ήχου και τους συνδυασμούς τους.

Το όνομα MPEG, έχει επικρατήσει όμως, να αναφέρεται και στη οικογένεια των τυποποιήσεων (standards) που δημιουργήθηκαν από την ομάδα Mpeg και χρησιμοποιούνται για την μετάδοση οπτικών και ηχητικών δεδομένων σε ψηφιακή συμπιεσμένη μορφή. Στην οικογένεια Mpeg, εντάσσονται τα standards Mpeg-1 που αφορά την συμπίεση ήχου και εικόνας, το Mpeg-2 για την εφαρμογή στην ψηφιακή τηλεόραση και το Mpeg-4 ως standard για εφαρμογές επικοινωνίας πολυμέσων. Επίσης, υπάρχει στα σχέδια τους και το Mpeg-7, με στόχο την αναπαράσταση περιεχομένου (content

representation) για την αναζήτηση πληροφοριών σε εφαρμογές. Στο σημείο αυτό, οφείλουμε να επισημάνουμε ότι τα δύο τελευταία στάδια του Mpeg βρίσκονται σε υπανάπτυξη και δεν έχουν γίνει ακόμη στάνταρ ενώ το Mpeg 1 και 2 έχουν τεθεί ήδη σε εκτεταμένη εφαρμογή. Ενδιάμεσα, για αρκετό χρονικό διάστημα σε εξέλιξη υπήρξε το Mpeg 3 αλλά εγκαταλείφθηκε και ενσωματώθηκε ένα μέρος του στο Mpeg 2.

Πιο συγκεκριμένα, το πρότυπο που έγκειται το ψηφιακό ήχο είναι το Mpeg-1 Audio, ως το πρώτο διεθνές πρότυπο για την ψηφιακή συμπίεση ήχου υψηλής πιστότητας που δεν αποτελεί ένα αλγόριθμο συμπίεσης αλλά μια οικογένεια τριών διαφορετικών τεχνικών κωδικοποίησης και συμπίεσης. Και τα τρία αυτά στάδια στηρίζονται στην ίδια αρχή, δηλαδή η συμπίεση ολοκληρώνεται με το συνδυασμό ενός είδος κωδικοποίησης μετασχηματισμού και sub-band division ενώ οι διαφορές του αναδύονται στο τελικό στάδιο της κβαντοποίησης. Παράλληλα, το πρότυπο Mpeg-1 Audio προβλέπει ένα ή δύο ηχητικά κανάλια χρησιμοποιώντας 16bits για την κωδικοποίησης των δειγμάτων, ενώ η συχνότητα δειγματοληψίας του ήχου μπορεί να είναι 32kHz, 44kHz ή 48 kHz.

Όπως προαναφέρθηκε, το πρότυπο Mpeg-1 Audio στηριζόμενο στην κωδικοποίηση ψηφιακού ήχου διακρίνεται σε Mpeg-1 Audio Layer I, II και III (ή MP3). Συγκεκριμένα, το Mpeg-1 Audio Layer I χρησιμοποιήθηκε στο σύστημα συμπίεσης ψηφιακής κασέτας DCC της Philips προσφέροντας συμπίεση 4:1. Ως αποτέλεσμα, η ηχητική ποιότητα είναι μέτρια ενώ το bandwidth που απαιτείται είναι αυξημένο 192 ή 256 kbps ανά κανάλι. Το Mpeg-1 Audio Layer II (ή Mp2), χρησιμοποιήθηκε στο ψηφιακό ραδιόφωνο όπου ο αλγόριθμος αυτή της κατηγορίας έχει βελτιποιηθεί για ένα εύρος ζώνης 96 ή 128 kbps ανά μονοφωνικό κανάλι, ενώ ως αποτέλεσμα η ποιότητα είναι εφάμιλλη του CD (6:1...8:1 με 256...142 kbps για στερεοφωνικό ήχο) . Αντίθετα, το Mpeg-1 Audio Layer III ή αλλιώς Mp3 (βλ.παρακάτω), έχοντας καλύτερη απόδοση από τα παραπάνω παρουσιάζει συμπίεση περίπου 12:1 με ποιότητα που πλησιάζει αυτή των CD και ρυθμό μετάδοσης δεδομένων στα 64 kbps. Σήμερα, με βάση αυτό το πρότυπο είναι κωδικοποιημένα τα αρχεία ήχου Mp3 και τα οποία παρουσιάζονται τόσο για την μεταφορά όσο και για την φόρτωση μέσω διαδικτύου όσο και για την αναπαραγωγή ή ανάκληση από το σκληρό δίσκο. Στον πιο κάτω πίνακα 4.2.1, παρουσιάζεται ο λόγος συμπίεσης σε συνάρτηση με το ρυθμό μετάδοσης των δεδομένων που υποστηρίζει το κάθε στρώμα του προτύπου Mpeg-1 Audio.

Στρώμα	Λόγος Συμπίεσης
Mpeg-1 Audio I	4:1 με 384 kbps για στερεοφωνικό ήχο
Mpeg-1 Audio II	6:1 με 256-192 kbps για στερεοφωνικό ήχο
Mpeg-1 Audio III	12:1 με 128-112 kbps για στερεοφωνικό ήχο

Σχ.4.2.1: Λόγος Συμπίεσης

Όπως έχει αναφερθεί παραπάνω, το πρότυπο Mpeg 1 μπορεί να κωδικοποιήσει μόνο δυο κανάλια ήχου ενώ για το δίκτυο που δεν διαθέτει μεγάλο εύρος ζώνης χρησιμοποιούνται κυρίως τεχνολογίες Mpeg 2. Το Mpeg 2 για τον ήχο, επεκτείνει την κωδικοποίηση μονοφωνικού και στερεοφωνικού ήχου του Mpeg 1, χωρίζεται σε τρία επίπεδα ανάλογα με το bit rate (ρυθμό δειγματοληψίας) που πρόκειται να υπάρξει ενώ κάθε επίπεδο δεν είναι καλύτερο από το άλλο, απλώς πιο περίπλοκο. Όσο προχωράμε στα επίπεδα τόσο πιο περίπλοκος γίνεται ο κωδικοποιητής και τόσο καλύτερη εκμετάλλευση του ρυθμού δειγματοληψίας πραγματοποιείται. Ο παρακάτω πίνακας, αναδύει αποτελέσματα τεστ σύγκρισης των τριών επιπέδων με κλίμακα από το 1 ως το 5.

Επίπεδο	Bit rate	Συμπίεση	Ελάχιστη καθυστέρηση	Ποιότητα 64 kbit
I	192 kbit	4:1	19ms	---
II	128 kbit	6:1	35ms	2.1 ως 2.6
III	64 kbit	12:1	59ms	3.6 ως 3.8

Αυτή την στιγμή το αγαπημένο της βιομηχανίας είναι το επίπεδο II αφού όταν σχεδίαζαν τους εξοπλισμούς τους δεν είχαν ακόμη οριστικοποιηθεί το επίπεδο III. Όμως με

ολοένα και αυξανόμενους ρυθμούς το επίπεδο III κερδίζει θέση με πρωταρχικό χώρο το διαδίκτυο. Για δομημένη ποιότητα ήχου Mpeg Layer III απαιτεί μικρότερο bit rate ή αλλιώς για δοσμένο bit rate πετυχαίνει υψηλότερη ποιότητα ήχου. Συμπερασματικά, το πρότυπο προσφέρει δειγματοληψία ελαττωμένη κατά το ήμισυ ( 16 kHz, 22kHz και 24kHz), παρέχοντας βελτιωμένη ποιότητα για ρυθμούς μετάδοσης σε 64 Kbit/sec.

### 4.3.MP3

Το MP3 καταγράφηκε στην Ιστορία, ως μια από τις πιο θορυβώδες καινοτομίες την ψηφιακής μουσικής. Προκάλεσε χαρά στους μουσικόφιλους και πανικό στις δισκογραφικές εταιρίες, ταρακούνησε θεσμούς και αμφισβήτησε τα δεδομένα της εποχής τα οποία τελικά άλλαξαν άρδην. Η ψηφιακή μουσική είναι άρρηκτα συνδεδεμένη με το MP3, πράγμα που διαπιστώνει κανείς από το γεγονός ότι οι νέες συσκευές κατακλύζουν ριζικά και ραγδαία την αγορά.

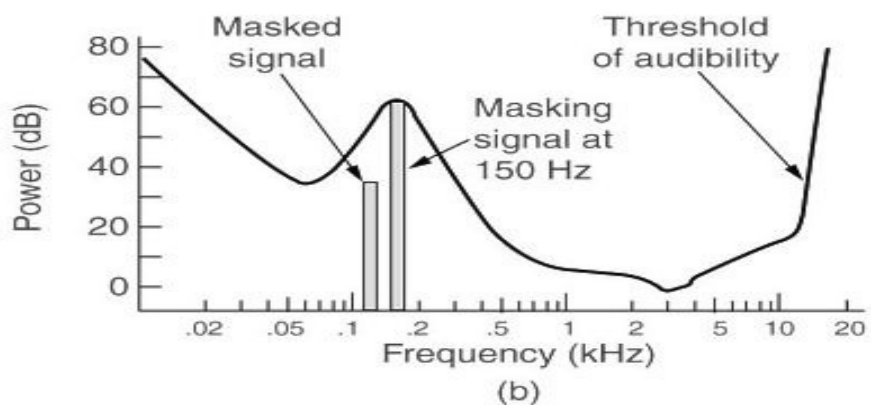
Όπως αναφέρθηκε, το Mpeg Layer III ή MP3 όπως είναι παγκόσμια γνωστό έχει αναπτυχθεί από τους ερευνητές του Ινστιτούτου Fraunhofer IIS με επικεφαλής, τους Karlheinz Branderburg και Dieter ως το πιο διαδεδομένο πρότυπο συμπίεσης ψηφιακού ήχου με στόχο την μεγαλύτερη συμπίεση και καλύτερη ποιότητα ήχου. Συγκεκριμένα, η συμπίεση των αρχείων ήχου μπορεί να φτάσει έως και 85%, χωρίς την ύπαρξη αισθητής διαφοράς στην ποιότητα απόδοσης ενώ αν η συμπίεση γίνει με ρυθμό μετάδοσης άνω των



128 Kbit/sec η διάφορά από το πρωτότυπο CD δεν είναι αντιληπτή. Αποτέλεσμα, η ύπαρξη μουσικών αρχείων 4-6 λεπτών που κανονικά θα καταλάμβαναν 40-70MB να καταλαμβάνουν χώρο μόνο 3-7MB.

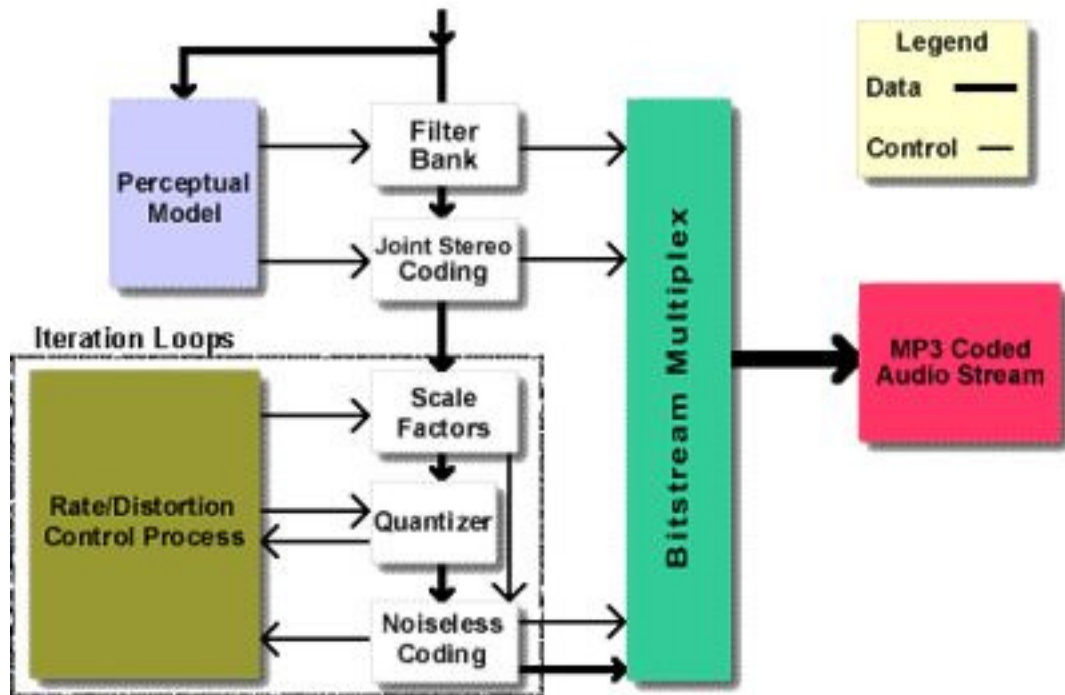
Οι codes MP3, χρησιμοποιούν ως βασικό μοντέλο ακοής αυτό που βασίζεται στις **καμπύλες κατώφλιού ακουστικότητας** (Minimal Audition Threshold) , δηλαδή την ελάχιστη ένταση που πρέπει να έχει ένας ήχος δια να τον ακούμε. Όσο μικρότερη είναι η ένταση, τόσο μικρότερη η ευαισθησία της ακοής και τόσο μεγαλύτερη στάθμη απαιτείται για να γίνει ακουστή μία συγκεκριμένη συχνότητα. Εμπειρικά αποτελέσματα, έχουν δείξει ότι το ανθρώπινο αυτί έχει μια περιορισμένη και εξαρτώμενη από την συχνότητα διακριτικότητας με επακόλουθο το κατώφλι ακουστικότητας να εξαρτάται από την ένταση του σήματος μέσα σε ένα περιορισμένο εύρος ζώνης γειτονικό αυτό της συχνότητας. Για κάθε συχνότητα του ακουστικού φάσματος το κατώφλι ακουστικότητας είναι διαφορετικό.

Ωστόσο, πολύ μεγαλύτερο ενδιαφέρον παρουσιάζει και το **φαινόμενο την επικάλυψης** (masking), με δεδομένους δυο γειτονικούς ήχους ο ισχυρότερος αλλοιώνει τοπικά την καμπύλη ακουστικότητας επικαλύπτοντας τον ασθενέστερο ο οποίος δεν γίνεται αντιληπτός από το ανθρώπινο αυτί. Έτσι το μοντέλο που χρησιμοποιείται στον codec MP3 υλοποιεί αυτό, δηλαδή υπολογίζει σε κάθε στιγμή το φασματικό περιεχόμενο του σήματος αποφασίζοντας ποιοι ήχοι επικαλύπτονται από ισχυρότερους με αποτέλεσμα να μην κωδικοποιηθούν. Σε κάθε χρονική στιγμή, ο codec MP3 έχει στη διάθεση του ένα αριθμό ψηφίων ο οποίος εξαρτάται από το βαθμό συμπίεσης που του έχει ζητηθεί, ενώ όσο εξελίσσεται η συμπίεση δεν χρησιμοποιούνται όλα τα ψηφία, ιδιαίτερα αν η στιγμιαία μορφή του σήματος είναι εύκολα συμπεσιμένη με συνεπακόλουθο την συντήρηση μιας δεξαμενής ψηφίων που απαιτούν μεγαλύτερη ακρίβεια στην κωδικοποίηση.



Σχ.4.3.1 : Φαινόμενο επικάλυψης. Ένας δυνατός ήχος, μεταβάλλεται τοπικά στην καμπύλη του κατωφλιού ακουστικότητας επικαλύπτοντας ένα γειτονικό αδύναμο ήχο.

Στο σημείο αυτό, οφείλουμε να αναφέρουμε ότι για καλύτερη συμπίεση δεδομένων το MP3 για την κωδικοποίηση των κβαντισμένων δειγμάτων χρησιμοποιεί την **εντροπική κωδικοποίηση (entropy encoding)**, η οποία κωδικοποιεί τα ψηφία που προκύπτουν από τα προηγούμενα στάδια. Η εντροπική κωδικοποίηση αποκαλείται και ως Huffman Coding. Συμπερασματικά, το ψυχοακουστικό μοντέλο για την συμπίεση ψηφιακού ήχου στην περίπτωση του MP3 τρέχει από το πεδίο του χρόνου σε αυτό της συχνότητας. Για υλοποίηση του περάσματος αυτού, χρησιμοποιείται μια πολύπλοκη μαθηματική διαδικασία κατά την οποία το υπό κωδικοποίησης μέρος του σήματος, υπόκειται σε επεξεργασία από μια τράπεζα φίλτρων (filter bank) όπου και χωρίζει το φάσμα σε 32 περιοχές και ακολουθεί ο μετασχηματισμός MDCT (Modified Discrete Cosine Transform), με στόχο την καλύτερη φασματική διακριτικότητα. Η διαδικασία αυτή, στηρίζεται στην αρχή του μετασχηματισμού Fourier όπου και αναλύει ένα σήμα εξελισσόμενο στο χρόνο σε μια σειρά συνιστωσών στο πεδίο της συχνότητας.



$$f_j = \sum_{k=0}^{2n-1} x_k \cos \left[ \frac{\pi}{n} \left( j + \frac{1}{2} \right) \left( k + \frac{1}{2} + \frac{n}{2} \right) \right]$$

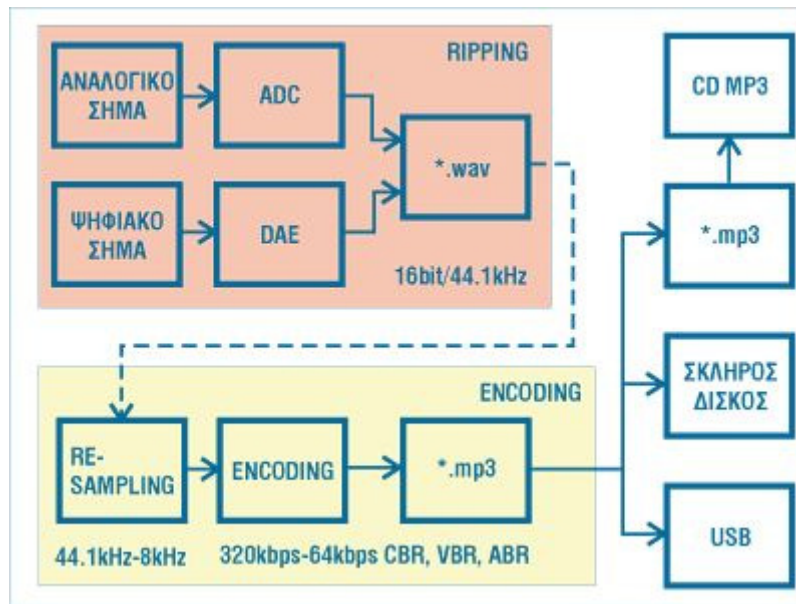
Σχ.4.3.2 : Διάγραμμα βαθμίδων κωδικοποιητή MP3. Μαθηματική περιγραφή του μετασχηματισμού MDCT

### Δημιουργία και Αναπαραγωγή Αρχείου MP3 :

Πρώτου εστιάσουμε το ενδιαφέρον μας στην δημιουργία MP3, οφείλουμε να αναφέρουμε ότι σημαντικό χαρακτηριστικό τους είναι πως δεν έχουν όλα την ίδια συμπίεση, με αποτέλεσμα το ποσοστό συμπίεσης του κάθε αρχείου να καθορίζεται από

εμάς. Όσο μεγαλύτερο είναι το ποσοστό συμπίεσης τόσο μικρότερο θα είναι το αρχείο, ενώ όσο περισσότερο συμπιέζεται ο ήχος τόσο περισσότερη πληροφορία χάνεται, γεγονός που επηρεάζει την ποιότητα του ήχου. Το μέγεθος και η ποιότητα του αρχείου ήχου ρυθμίζεται από το ροή μετάδοσης των δεδομένων.

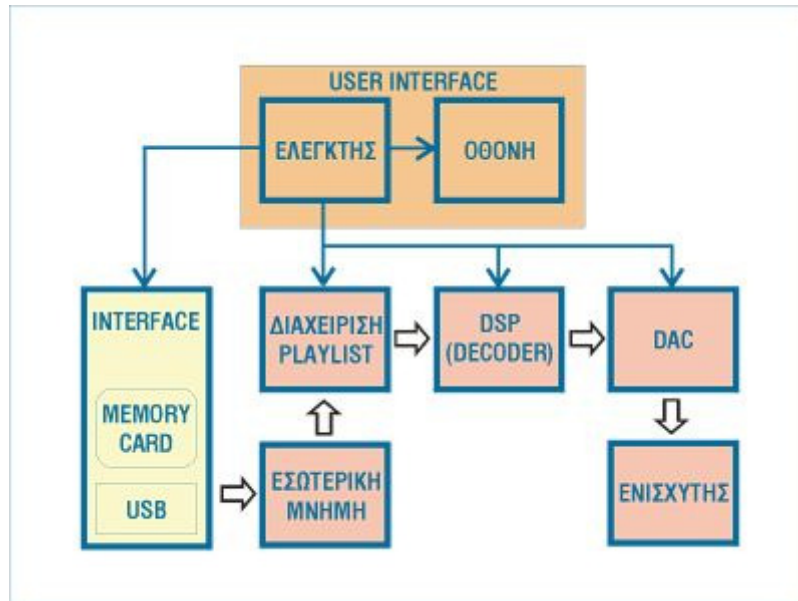
Συγκεκριμένα, την **δημιουργία ενός αρχείου MP3** υποδηλώνει η διαδικασία **ripping**, επιτρέποντας την δημιουργία ενός αρχείου wav που αποθηκεύεται στον υπολογιστή, από το πρωτογενές υλικό. Αν το υλικό αυτό είναι αναλογικό, θα πρέπει πρώτα να περάσει από έναν μετατροπέα A/D, αντίθετα αν το υλικό είναι αποθηκευμένο σε CD, τότε μπορούμε να χρησιμοποιήσουμε την διαδικασία **Digital Audio Extraction (DAE)** καθώς και να μεταφέρουμε το ψηφιακό περιεχόμενο του δίσκου απευθείας σε αρχείο wav. Το αρχείο wav που προκύπτει από τον ripper έχει προδιαγραφές που εξαρτώνται από την διαδικασία **CD Quality**, δηλαδή συχνότητα δειγματοληψίας 44.1kHz, και μήκος λέξης 16bit. (χωρίς να αποκλείονται και άλλες εκδοχές , όπως τα 48kHz ή τα 24bit). Αυτό το αρχείο, είναι η πηγή των δεδομένων που τοποθετείται στην είσοδο του encoder το οποίο πραγματοποιεί **resampling** (με βάση τις οδηγίες που του δίνουμε), κωδικοποιώντας το σήμα με βάση το μοντέλο της απωλεστικής συμπίεσης με αποτέλεσμα την δημιουργία ενός **αρχείο mp3** (Σχ.4.3.4)<sup>10</sup>. Το αρχείο αυτό, μπορεί να αποθηκευτεί τοπικά και να αναπαράγεται μέσω του σχετικού player, της κάρτας ήχου και των ηχείων του υπολογιστή, να μεταφερθεί σε κάποια εξωτερική συσκευή μέσω USB, να μετατραπεί και πάλι σε wav με απώτερο στόχο την εγγραφή του σε CD που είναι συμβατό με απλά CD players ή να εγγραφεί απ' ευθείας σε CD με στόχο να χρησιμοποιηθεί από συσκευές που είναι συμβατές με CD MP3.



Σχ. 4.3.3: Διαδικασία Ripping και Encoding

Η διαδικασία **αναπαραγωγή αρχείων MP3** παραγματοποιείται όταν αυτά βρίσκονται αποθηκευμένα στον υπολογιστή. Καταρχήν τα αρχεία με κατάληξη mp3 μέσω των "File Associations", αναδύουν με διπλό κλικ **το media player** **ούτος ώστε να μην χρειάζεται** άλλη προσπάθεια από την πλευρά του χρήστη. Στην περίπτωση που τα αρχεία βρίσκονται στον υπολογιστή και που πρέπει να φορτωθούν σε κάποια εξωτερική συσκευή player, αυτό γίνεται χωρίς ειδικό λογισμικό αφού οι μνήμες και όλο και πιο συχνά τα players φαίνονται από την πλευρά του υπολογιστή ως "removable storage device" όταν συνδέονται στο **USB**. Αντίθετα, από την πλευρά του player τα πράγματα δεν είναι σύνθετα, ένα **user interface** παρέχει πρόσβαση τόσο στην **δημιουργία όσο και διαχείριση ενός καταλόγου** με τα περιεχόμενα του player είτε αυτά βρίσκονται σε εξωτερική είτε σε εσωτερική μνήμη. Στο σημείο αυτό, οφείλουμε να αναφέρουμε ότι η διαδικασία την διαχείριση είναι απλή ή σύνθετη ανάλογα με τις δυνατότητες της συσκευής, ενώ από την στιγμή που ο χρήστης επιλέξει το τρακ και στην συνέχεια το play, τα δεδομένα του αντίστοιχου αρχείου mp3 οδηγούνται, στον **αποκωδικοποιητή, στον μετατροπέα d/a και στον ενισχυτή** ακουστικών για περισσότερη επεξεργασία. Εντούτοις, πολλά players διαθέτουν και **αναλογικές**

εισόδους έτσι ώστε ο ψηφιακός επεξεργαστής να διαθέτει μετατροπέα A/D και MP3 encoder. (Σχ.4.3.5)<sup>11</sup>



Σχ. 4.3.4 : Διάγραμμα Βαθμίδων ενός MP3

Επίσης, καλό θα ήταν να τυπωθεί ότι, το πιο γνωστό και διαδεδομένο πρόγραμμα αναπαραγωγής μουσικών αρχείων MP3 είναι το Winamp της εταιρίας Nullsoft, το οποίο και διανέμεται δωρεά και είναι εγκατεστημένο σε εκατομμύρια υπολογιστές του κόσμου. Εκτός από το Winamp, μπορεί να βρει κανείς στο διαδίκτυο και άλλα προγράμματα αναπαραγωγής MP3 όπως τα Media Juke Box, Music Match, Sonique, Ejay κ.α. Ο μεγαλύτερος όμως συναγωνιστής του Winamp είναι ο Media Player της Microsoft, ο οποίος διαχειρίζεται αποκωδικοποιητές MP3 του Ινστιτούτου Fraunhofer με καλύτερη ποιότητα ήχου, ενώ οι περισσότεροι mp3 player μπορούν να αναπαράγουν και άλλα πρότυπα ψηφιακής μουσικής όπως Wav, CD Audio, WMA κ.α , των οποίων ανάλυση γίνεται παρακάτω.

#### 4.4. ΆΛΛΑ ΠΡΟΤΥΠΑ

Παρά το γεγονός της ριζικής και ραγδαία εξέλιξη του MP3 το οποίο έχει γνωρίσει ευρεία αποδοχή μεταξύ χρηστών, εντούτοις, επικρατεί πληθώρα διαφορετικών τεχνολογιών που ολοένα και εξελίσσονται με πρωταρχικό στόχο την καλύτερη συμπίεση ψηφιακού ήχου. Τέτοιες προσπάθειες, έχουν υλοποιηθεί από την εταιρία Microsoft με το δικό της γνωστό και διαδεδομένο πρότυπο Windows Media Audio (WMA), το AAC (Advanced Audio Coding), το MP3 Pro, το OGG και το AC-3 Dolby Digital.

##### ❖ **WMA** (Windows Media Audio)

Η μεταφορά, η αποθήκευση και η χρήση ακουστικού υλικού με βάση την μορφή απωλεστικών συμπιεσμένων ηχητικών αρχείων μέσω υπολογιστή, ωθεί τον μεγαλύτερο κατασκευαστή λειτουργικών συστημάτων να μην μείνει έξω από το παιχνίδι. Το πρότυπο Windows Media Audio (WMA) της εταιρία Microsoft, προσφέρει όμοιες δυνατότητες με το MP3, με άριστη ποιότητα τόσο αναπαραγωγής όσο και μεγαλύτερη συμπίεση (64 kbps). Πιο συγκεκριμένα, το WMA αποτελεί ένα σύστημα κωδικοποίησης/αποκωδικοποίησης ήχου, επιτρέποντας την συμπίεση ψηφιακών δεδομένων ήχου στο 1/20 του αρχικού τους όγκου και την εγγραφή τους σε ένα μόνο δίσκο CD με επακόλουθο τα τραγούδια που είναι προστατευμένα να μην μπορούν να μεταδοθούν ελεύθερα. Συμπερασματικά, γι' αυτό ακριβώς το λόγο ο μεγαλύτερος αριθμός δισκογραφικών εταιριών χρησιμοποιεί στα πλαίσια υλοποίησης του έργου τους το πρότυπο αυτό.

##### ❖ **MP3 Pro**

Τον Ιανουάριο του 2001 στη CES, παρουσιάστηκε από την Coding Technologies η τεχνολογία Mp3 Pro, μια βελτιωμένη έκδοση του Mp3 με δυνατότητα να προσφέρει όμοια ποιότητα στο μισό μέγεθος των αρχείων, γεγονός που υλοποιείται με μεγαλύτερη συμπίεση δεδομένων. Συγκεκριμένα, η συμπίεση στα 64kbps και 96kbps, προσφέρει τη ίδια απόδοση ήχου με τα 128kbps και 192kbps του Mp3.

### ❖ **MP3 Surround**

Το 2004 το Ινστιτούτο Fraunhofer IIS παρουσίασε μία πολυκαναλική έκδοση του MP3, το MP3 Surround το οποίο βασίζεται στην τεχνολογία **Binaural Cue Coding της Agere. Η τεχνολογία αυτή**, επιτρέπει την μείξη σημάτων από πολλά κανάλια σε δύο, με στόχο την δημιουργία ενός σήματος συμβατού με τον απλό MP3 codec, ενώ κωδικοποιεί μία σειρά από παραμέτρους που περιγράφουν πλήρως το ηχητικό πεδίο surround. Τέτοιες παράμετροι είναι, οι χρονικές διαφορές μεταξύ των καναλιών, οι διαφορές στάθμης μεταξύ των καναλιών και η συσχέτιση μεταξύ των καναλιών.

### ❖ **AAC (Advance Audio Coding)**

Το πρότυπο ACC αναπτύχθηκε για πρώτη φορά το 1997 από το Ινστιτούτο Fraunhofer IIS και χρησιμοποιεί όπως και το MP3 το ψυχοακουστικό μοντέλο επικάλυψης, με σκοπό να καλύψει τα προβλήματα που υπήρχαν σχετικά με την ποιότητα των αρχείων MP3 στους μικρούς αριθμούς μετάδοσης των δεδομένων. Το πρότυπο ACC, ως ένα πραγματικό αριστούργημα κωδικοποίησης έχει την ικανότητα να αποδίδει εξαιρετικά υψηλή ποιότητα ήχου σε birate 64Kb/sec, επιτρέποντας τόσο την κωδικοποίηση 48 καναλιών ήχου και 16 καναλιών χαμηλής συχνότητας για εφέ όσο και την υποστήριξη πολλών γλωσσών ταυτόχρονα. Παράλληλα, το ACC διακατέχεται από τρεις διαφορετικές όψεις, την «κύρια», την «χαμηλής πολυπλοκότητα» και την «κλιμακούμενη συχνότητα δειγματοληψίας». Η «κύρια» όψη αναφέρεται σε εφαρμογές που η υπολογιστική ισχύει και εφαρμογές δεν είναι περιορισμένη, η «χαμηλής πολυπλοκότητα» σε εφαρμογές που η ισχύος και η μνήμη βρίσκονται σε μεγάλη ζήτηση, ενώ η τελευταία, είναι φτιαγμένη έτσι ώστε οι αποκωδικοποιητές να έχουν ελάχιστες απαιτήσεις σε μνήμη και ισχύ. Υποκειμενικά τεστ που έχουν πραγματοποιηθεί με καλά εκπαιδευμένους ακροατές, έδειξαν ότι η συγκεκριμένη κωδικοποίηση προσφέρει καλύτερη ποιότητα ήχου από οποιαδήποτε άλλη κωδικοποίηση ήχου με το μισό μόνο birate. Στο σημείο αυτό, οφείλουμε να αναφέρουμε ότι το πρότυπο ACC παρέχει καλύτερη απόδοση από το MP3, ενώ το 2003 η έκδοση του παρουσιάζεται συμβατή με τις προδιαγραφές του Mpeg ούτως ώστε το πρότυπο να αναφέρεται και ως Mpeg-4 ACC.

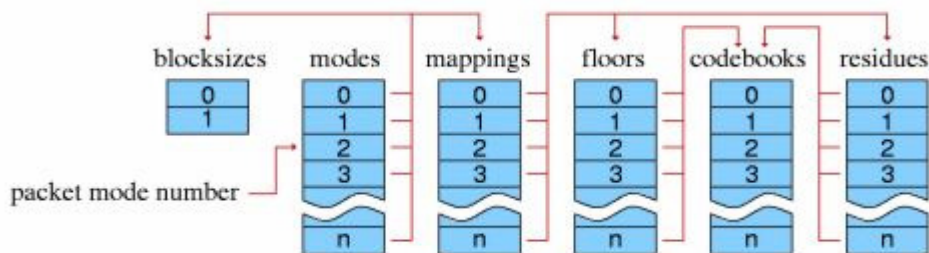
### ❖ **RA-Real Audio**



Το Real Audio ως κλειστό πρότυπο, δημιουργήθηκε και υποστηρίχθηκε από την εταιρία Real Network με σκοπό την αναπαραγωγή ήχων στο Διαδίκτυο χωρίς να προηγείται κατέβασμα των ήχων στο σκληρό δίσκο του υπολογιστή. Το πρότυπο Real Audio, είναι αρκετά δημοφιλές και αυτό εξαιτίας της ελεύθερης διάθεσης του λογισμικού ανάγνωσης των αρχείων ήχου τέτοιου τύπου υποστηρίζοντας, μεγάλη συμπίεση και κατακανόνα χαμηλή ποιότητα ήχου.

### ❖ OGG Vorbis

Ο codec Ogg Vorbis αναπτύχθηκε γύρω από το πρότυπο αρχείων Ogg και βασίζεται στη open source εφαρμογή απωλεστικής συμπίεσης με την ονομασία Vorbis. Ως προς τον τρόπο κωδικοποίησης, ο τρόπος μοιάζει με αυτό του Mp3 ενώ ταυτόχρονα ο Ogg Vorbis χρησιμοποιεί MDCT για τον μετασχηματισμό του σήματος από το πεδίο του χρόνου στο πεδίο της συχνότητας, καθώς και μία εναλλακτική μέθοδο επεξεργασίας του φάσματος, κατά την οποία κωδικοποιείται το φάσμα βάσης του οποίου η κατανομή είναι σχετικώς ομαλή και με περισσότερα ψηφία το απομένον φάσμα που η δομή και η χρονική εξέλιξη είναι πολύ πιο πολύπλοκη (σχ4.4.1)<sup>12</sup>. Συμπερασματικά, η τακτική αυτή σε συνδυασμό με την καλή ποιότητα ήχου που προσφέρει, ωθεί το πρότυπο Ogg σε ένα ανταγωνιστικό παιχνίδι ως προς το WMA και MP3.



Σχ. 4.4.1 : Διάγραμμα βαθμίδων του κωδικοποιητή Ogg Vorbis.

### ❖ AC3 Dolby Digital

Ένα από τα πιο διαδεδομένα πρότυπα για τον ψηφιακό πολυκάναλο ήχο είναι το AC3, που εκτός σημαντικού απροόπτου άρχισε να γίνεται το διεθνές πρότυπο για την συμπίεση ηχητικών δεδομένων. Στο ψηφιακό σύστημα ήχου AC3, ο ήχος κωδικοποιείται σε έξι συνολικά κανάλια στηριζόμενο στην μέθοδο 5.1. Συγκεκριμένα, υπάρχουν: (α) τρία κανάλια (αριστερό, κεντρικό, δεξί ) που αποσκοπούν στο να φέρουν την βασική ηχητική πληροφορία, (β) δύο συνοδευτικά κανάλια περιβάλλοντος ήχου και (γ) ένα κανάλι για τις υπόλοιπες συχνότητες (σύστημα 3/2/.1). Ως προς τον τρόπο λειτουργίας τους, τα πέντε πρώτα κανάλια διαχειρίζονται συχνότητες ήχου στο διάστημα 3-20000Hz, ενώ το έκτο συχνότητες 3-120Hz. Συμπερασματικά, ο ρυθμός δειγματοληψία είναι 48KHz μεγαλύτερος από το ρυθμό των 44KHz που χρησιμοποιείται στα CDs, ενώ η συμπίεση των ηχητικών δεδομένων ανέρχεται στην αναλογία 10:1.

#### 4.5. ΤΥΠΟΙ ΑΡΧΕΙΩΝ ΨΗΦΙΑΚΟΥ ΉΧΟΥ

Ανάλογα με την διαδικασία παραγωγής του, τα αρχεία ήχου διακρίνονται σε δύο κατηγορίες: τα αρχεία ήχων κυματομορφής και αρχεία MIDI. Τα αρχεία ήχων κυματομορφής, παράγονται με την διαδικασία της ψηφιοποίησης που σκοπό έχουν την παροχή ακριβής ψηφιακής εικόνας της κυματομορφής του ήχου. Πιο συγκεκριμένα, τα προγράμματα επεξεργασίας ήχων αυτής της κατηγορίας, αναπαριστάνουν τον ήχο ως κυματομορφή δίνοντας την ευκαιρία στο χρήστη να κόψει, να αντιγράψει, να εισάγει και να τροποποιήσει είτε ένα μέρος είτε ολόκληρη την κυματομορφή. Οι βασικές δυνατότητες αυτών των εργαλείων είναι:

- ♦ εισαγωγή ήχου
- ♦ αντιγραφή, αποκοπή, επικόλληση ηχητικού τμήματος
- ♦ ρύθμιση έντασης και συχνότητας
- ♦ απαλοιφή θορύβου
- ♦ εισαγωγή διάφορων εφέ
- ♦ τροποποίηση επιπέδων κβάντωσης
- ♦ τροποποίηση χρονικής διάρκειας
- ♦ συμπίεση

Τα πιο γνωστά προγράμματα επεξεργασίας ήχων κυματομορφής είναι το Sound Forge Sonic της Foundry, το Sound Edit Macromedia, το Coll Edit Syntrillium, Wave Studio της Creative κ.α.

Η δεύτερη κατηγορία αρχείων ήχου σχετικά με την διαδικασία παραγωγής τους είναι το αρχείο MIDI, το οποίο δεν καταγράφει το ηχητικό σήμα αλλά τα στάδια που πραγματοποιούνται για να παραχθεί αυτό. Συγκεκριμένα, αντί να αποθηκεύσουμε τους ήχους μετατρέποντας το αναλογικό σήμα σε ψηφιακό αποθηκεύουμε τέσσερις αριθμούς για κάθε νότα που παίζεται, δηλαδή ο πρώτος μας δίνει ποια νότα είναι, ο δεύτερος μας λέει πιο όργανο την παίζει, ο τρίτος μας πληροφορεί πόσο δυνατά παίζεται η νότα και ο

τέταρτος μας δίνει το χρόνο. Επειδή η διαδικασία αναπαράστασης μοιάζει με αυτή του κώδικα ASCII, πολλές φορές το MIDI αποκαλείται κα ως «κώδικα ASCII της μουσικής». Όσο αφορά τα προγράμματα επεξεργασίας αρχείων MIDI αυτά παρέχουν την δυνατότητα αναπαράστασης του ήχου ως παρτιτούρα καθώς και την δυνατότητα παροχής στο χρήστη, ώστε να τοποθετήσει νότες, παύσεις και άλλα σύμβολα της μουσικής σημειογραφίας πάνω στο πεντάγραμμο καθορίζοντας τον τρόπο εκτέλεσης της κάθε νότας.

Εν κατακλείδι, ανάλογα με την πλατφόρμα και το πρόγραμμα που χρησιμοποιείται για την παραγωγή των ψηφιακών ήχων, τα αρχεία αποθηκεύονται στον ηλεκτρονικό υπολογιστή με διαφορετική μορφή η οποία αναγνωρίζεται από τη έκταση του αρχείου (σχ. 4.5.1).

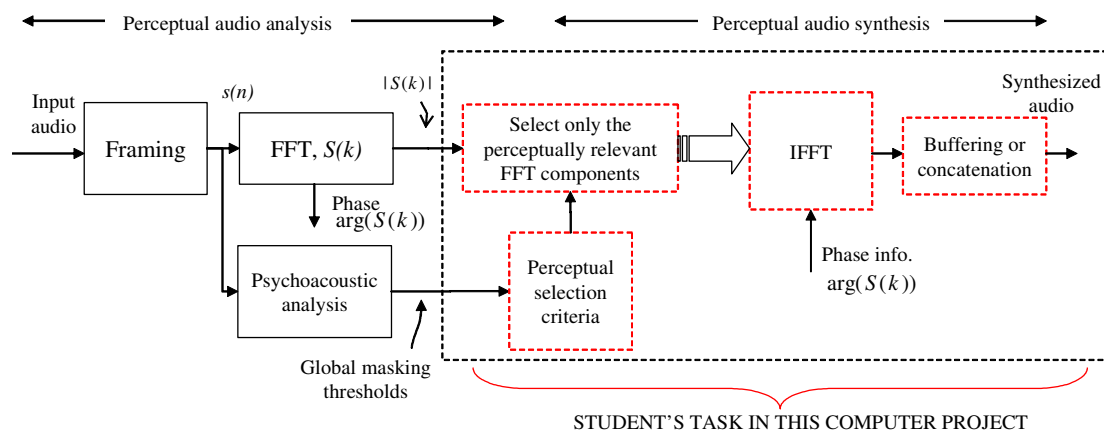
.rif	RIFF	Αναπτύχθηκε από την Microsoft και υποστηρίζει αρχεία ψηφιακού ήχου WAV, MIDI.
.wav	Wave	Αποτελεί το πρότυπο αποθήκευσης ψηφιακού ήχου και είναι υποσύνολο του προτύπου RIFF.
.mid	MIDI	Διεθνές πρότυπο για την αποθήκευση μουσικών αρχείων MIDI.
.aif	AIFF	Δημιουργήθηκε από την Apple αλλά υποστηρίζεται και από άλλες πλατφόρμες, υποστηρίζει δειγματοληψία 32 bit.
.rmi	RMI	Αναπτύχθηκε από την Microsoft για την υποστήριξη αρχείων MIDI.
.mp3	MPEG-layer3	Πρότυπο συμπίεσης αρχείων ήχου που χρησιμοποιείται κυρίως για την διακίνηση αρχείων μουσικής στο Διαδίκτυο
.wma	WMA	Windows Media Audio από την Microsoft. Στόχο έχει την υποστήριξη απωλεστικής συμπίεσης
.ra	Real Audio	Προορίζεται για την άμεση αναπαραγωγή ήχων μέσω του

		Διαδικύου
--	--	-----------

Σχ. 4.5.1: Τύποι Αρχείων Ήχου

## 5. ΚΕΦΑΛΑΙΟ 5 : ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

Στο πειραματικό μέρος της παρούσας εργασίας υπολοποιούμε την διαδικασία της "αντιληπτικής" (Perceptual) ανάλυσης και σύνθεσης ενός ακουστικού σήματος, επεξεργασμένου με την χρήση του Γρήγορου Μετασχηματισμού Fourier (FFT), χρησιμοποιώντας το Ψυχοακουστικό μοντέλο του προτύπου MP3. Η βήμα προς βήμα διαδικασία της προσομοίωσης φαίνεται στο παρακάτω σχήμα :



Σχήμα 5.1 : Αντιληπτική ανάλυση/ σύνθεση ακουστικού σήματος

Η προσομοίωση του μοντέλου υλοποιήθηκε με τη χρήση του λογισμικού Matlab.

Αρχικά με την εντολή "wavread" ένα ηχητικό αρχείο διαβάζεται και δειγματοληπτείται με ρυθμό 44.1 kHz. Το εισερχόμενο ακουστικό σήμα κατακερματίζεται σε πλαίσια (frames) των 512 δειγμάτων. Ακολούθως εφαρμόζεται ένας Γρήγορος Μετασχηματισμός Fourier 512 σημείων και υπολογίζεται το φάσμα πλάτους και η φάση του σήματος. Η ψυχοακουστική ανάλυση που εφαρμόζεται εν παραλλήλω έχει σκοπό να εκτιμήσει το κατώφλι ισχύος κάθε δείγματος. Η ψυχοακουστική ανάλυση σκοπό έχει γενικά την εύρεση παραμέτρων του σήματος και την αποκοπή όσων εξ αυτών παραβιάζουν κάποια κριτήρια και όρια της αντιληπτικής ικανότητας του ανθρώπινου βιολογικού συστήματος ακοής. Οι παράμετροι που μπορούν να υπολογιστούν ποικίλουν και η επιλογή τους εναπόκειται στην εκάστοτε έρευνα. Στο παρών μοντέλο που υλοποιούμε η παράμετρος που υπολογίζεται είναι η ισχύς των δειγμάτων που λαμβάνουμε. Αν φάσμα πλάτους του σήματος εισόδου σε dB πέφτει κάτω από το επίπεδο κατωφλιού επικάλυψης ισχύος (NMT) που έχουμε υπολογίσει, τα δείγματα αυτά αγνοούνται με την έννοια ότι θέτουμε σε αυτή την περίπτωση την ισχύ του δείγματος στην τιμή των -100 dB. Ακολούθως τα τροποποιημένα δείγματα FFT του σήματος συνδυάζονται κατάλληλα με την αντίστοιχη φάση. Τέλος εφαρμόζουμε αντίστροφο Μετασχηματισμό Fourier (IFFT) 512 σημείων για να ενσωματώσουμε όλα τα δείγματα και με τη χρήση ενός απομωνοτή (buffer) να ανασυστήσουμε το ηχητικό σήμα.

## 5.1 ΚΩΔΙΚΑΣ ΠΡΟΣΟΜΟΙΩΣΗΣ

Ο κώδικας προσομοίωσης αποτελείται από τρία αρχεία Matlab, δύο εκ των οποίων είναι αρχεία συναρτήσεων και το τρίτο αποτελεί το κύριο αρχείο του κώδικα.

1. **psymain.m** : Το αρχείο αυτό είναι το κύριο αρχείο μέσω του οποίου καθορίζεται η ροή του κώδικα και η εκτέλεση των συναρτήσεων.
2. **psychoacoustics.m** : Στην συνάρτηση αυτή εκτελείται η ψυχοακουστική ανάλυση που περιλαμβάνει την εύρεση του GMT και την συνθήκη ελέγχου της φασματικής πυκνότητας ισχύος κάθε πλαισίου.

**3. audio\_synthesis.solution.m** : Στο τμήμα αυτό του κώδικα εκτελείται η ανασύνθεση του ηχητικού σήματος καθώς και οι απαραίτητες επιπλέον λειτουργίες.

Αναλυτικότερα μπορούμε να θεωρήσουμε ότι όλη η διαδικασία χωρίζεται οργανικά σε δύο μεγάλα μέρη που εκτελούνται σε διαφορετικά μέρη του κώδικα.

## **ΜΕΡΟΣ Α: Audio analysis**

**Εισαγωγή αρχείου ήχου:** Ένα αρχείο ήχου τύπου .wav δειγματοληπτείται με συχνότητα δειγματοληψίας  $fs=44.1$  kHz, και ψηφιοποιείται με μια 16-bit Παλμοκωδική Διαμόρφωση (PCM) με την εντολή :

```
[s, fs, bits] = wavread('ch5_music.wav');
```

Η λειτουργία της ανάγνωσης του αρχείου εκτελείται στην `psymain.m`

- 1) Πλαισίωση (Framing)** : Χρησιμοποιήσαμε πλαίσια των 512 δειγμάτων ανα πλαίσιο το οποίο στο πεδίο του χρόνου μεταφράζεται σε:  
χρονική διάρκεια πλαισίου =  $(512 \text{ δείγματα}) / (44.1 \text{ δείγματα ανά δευτερόλεπτο}) = 11.61 \text{ms}$
- 2) Μετασχηματισμός Fourier (FFT)** : Εφαρμόζουμε Μετασχηματισμό FFT σε κάθε πλαίσιο με σκοπό τον υπολογισμό του φασματος πλάτους και της φάσης. Η λειτουργία αυτή εκτελείται στην `psymain.m`.

**Ψυχοακουστική Ανάλυση (Psychoacoustic Analysis)** : Στο τμήμα αυτό του κώδικα και την ομώνυμη συνάρτηση υπολογίζουμε το γενικό κατώφλι ισχύος GMT (Global masking threshold) (σε dB) καθώς και την φασματική πυκνότητα ισχύος ( power spectral density) (σε dB) .

## **ΜΕΡΟΣ Β : Audio Synthesis**

**1) Αντιληπτικά κριτήρια επιλογής (Perceptual selection criteria) :** Το κριτήριο αντιληπτικής ικανότητας που χρησιμοποιούμε αφορά την ισχύ του πλάτους των δειγμάτων. Αν η ισχύς του δείγματος πέσει κάτω από το GMT σε dB , τότε θέτουμε την τιμή ισχύος στην τιμή των -100 dB.

**2) Εφαρμογή FFT :** Στο σημείο αυτό υπολογίζεται η μετατροπή της φασματικής πυκνότητας ισχύος (PSD) του σήματος μέσω Γρήγορου Μετασχηματισμού Fourier.

**3) Αντίστροφος Μετασχηματισμός Fourier (IFFT) :** Αφορά την αντίστροφη διαδικασία του μετασχηματισμού Fourier εφόσον προηγουμένως έχουμε συνδυάσει τη φασματική πυκνότητα ισχύος με τη φάση του σήματος.

**Απομωνοτής (Buffering) :** Το τελικό στάδιο της επεξεργασίας γίνεται στο τμήμα του buffer (αλγοριθμικά υλοποιείται στην `psymain.m`) ,ο οποίος ενώνει όλα τα επεξεργασμένα τμήματα (frames) του σήματος ,από την συνάρτηση `audio_synthesis`, σε ένα απλό διάνυσμα.

Στα Παράρτημα Α παρουσιάζεται ο πλήρης κώδικας του κυρίου προγράμματος και των συναρτήσεων ενώ στο Παράρτημα Β παρουσιάζονται τα γραφήματα που προέκυψαν απο την επεξεργασία του ηχητικού σήματος. Τέλος σημειώνουμε οτι ο κώδικας δίνει τη δυνατότητα ακουστικής σύγκρισης του σήματος πριν και μετά την ψυχοακουστική ανάλυση.



## ΠΑΡΑΡΤΗΜΑ Α

### Psymain

```
clear all,  
close all,  
clc,  
[s,fs] = wavread('ch5_malespeech.wav');  
  
N = 512;  
fft_size = 512;  
No_frames = floor(length(s) / N);  
Output_audio = [ ];  
  
for i_c = 15 : 1 : No_frames-3  
    sprintf('Frame count = %d', i_c)  
  
        current_frame = s((i_c-1)*N+1 : i_c*N);
```

```

current_fr_FFT = fft(current_frame, fft_size);
P_spectrum = 10*log10( (abs(current_fr_FFT)).^2 );
FFT_phase = angle(current_fr_FFT);

view_fig = 0;
[P_SPL, JND] = psychoacoustics(current_frame, fft_size,view_fig);

Synth_audio = audio_synthesis_solution(P_SPL, JND,FFT_phase, fft_size);

Synth_audio = Synth_audio * fft_size;

Output_audio = [Output_audio, Synth_audio'];

figure(1),
plot(current_frame), hold on, plot(Synth_audio, 'k:'), hold off,

% pause,
end

fprintf('press any key to listen to the input audio \n'), pause,
fprintf('Playing input file... Wait until it is done.... \n')
sound(s, fs)
fprintf('After the input file playback is complete, press any key to
listen to the Output audio'), pause,
sound(Output_audio, fs)

```

## ΣΥΝΑΡΤΗΣΗ **psychoacoustics.m**

```

function [P, Thr_global]=psychoacoustics(current_frame, N, View_fig)

global fr_size  fft_size  fr_count
global fs  bit_res
global freq_hz  freq_bark  Abs_thr

fs = 44100;
fr_size = N;
fft_size = N;
freq_hz = [1 : fft_size/2+1] * (fs / fft_size);
freq_bark = 13 * atan(.00076*freq_hz) + 3.5 *
atan((freq_hz/7500).^2);

Abs_thr = 3.64.*(freq_hz./1000).^(-.8) - 6.5.*exp(-
0.6.*(freq_hz./1000-3.3).^2) + 0.001.*(freq_hz./1000).^4;

```

```

[P] = psy_step_1( current_frame);

[P_TM, P_NM] = psy_step_2( P );
if (View_fig == 1)
    figure(1), plot(freq_bark, P);
    hold on;
    hold on, plot(freq_bark, Abs_thr, 'r:');
    plot(freq_bark, P_TM, 'x');
    plot(freq_bark, P_NM, 'o');
    hold off,
    (gca, 'ylim', [-20, 100]);
    legend('PSD', 'Absolute Threshold', 'Tone maskers', 'Noise maskers');
    pause,
end

[P_TM_th, P_NM_th] = psy_step_3( P_TM, P_NM );
if (View_fig == 1)
    figure(2), plot(freq_bark, P);
    hold on, plot(freq_bark, Abs_thr, ':');
    plot(freq_bark, P_TM_th, 'rx');
    plot(freq_bark, P_NM_th, 'ko');
    set(gca, 'ylim', [-20, 100]);
    legend('PSD', 'Absolute Threshold', 'Selected tone maskers', 'Selected noise
maskers');
    hold off,
    pause,
    end

[Thr_TM, Thr_NM] = psy_step_4( P_TM_th, P_NM_th );

if (View_fig == 1)
    figure(3), plot(freq_bark, 10*log10(Thr_TM+eps), 'r:');
    hold on, plot(freq_bark, 10*log10(Thr_NM+eps), 'k:');
    legend('Tone masking threshold', 'Noise masking threshold');
    set(gca, 'ylim', [-20, 100]);
    hold off,
    pause,
end

[Thr_global] = psy_step_5( Thr_TM, Thr_NM );
Thr_global = 10*log10(Thr_global);

    if (View_fig == 1)
        figure(4), plot(freq_bark, 10*log10(Thr_global+eps), 'k');
        legend('JND curve');
        set(gca, 'ylim', [-20, 100]);
        hold off,
        pause
    end

function [ P ] = psy_step_1( s )

```

```

global fr_size fft_size fr_count bit_res
global freq_hz freq_bark Abs_thr

PN = 90.302;

x = s / fft_size;
w = ones(fr_size, 1);

P = PN + 10*log10( (abs(fft(w.*x, fft_size))).^2 );
P = P(1: fft_size/2+1);

function [P_TM, P_NM] = psy_step_2( P )

global fr_size fft_size fr_count fs bit_res
global freq_hz freq_bark

P_TM = zeros(1, length(P));
for k = 1 : length(P),
    if(tone_masker_check(P, k))
        P_TM(k) = 10*log10(10.^(0.1.*P(k1))+10.^(0.1.*P(k))+10.^(0.1.*P(k+1)));
    end
end

P_NM = zeros(1, length(P_TM));
lowbin = 1;
highbin = max(find(freq_bark < 1));
for band = 1:24,
    [noise_masker_at_loc, loc] = noise_masker_check(P, P_TM, lowbin, highbin);
    if (loc ~= -1)
        P_NM(floor(loc)) = noise_masker_at_loc;
    end
    lowbin = highbin;
    highbin = max(find(freq_bark<(band+1)));
end

function bool = tone_masker_check(P, k)

if ((k<=1) | (k>=250))
    bool = 0;

elseif ((P(k)<P(k-1)) | (P(k)<P(k+1))),
    bool = 0;

elseif ((k>2) & (k<63)),
    bool = ((P(k)>(P(k-2)+7)) & (P(k)>(P(k+2)+7)));
elseif ((k>=63) & (k<127)),
    bool = ((P(k)>(P(k-2)+7)) & (P(k)>(P(k+2)+7)) & (P(k)>(P(k3)+7)) &
(P(k)>(P(k+3)+7)));
elseif ((k>=127) & (k<=256)),
    bool = ((P(k)>(P(k-2)+7)) & (P(k)>(P(k+2)+7)) & (P(k)>(P(k-3)+7)) &
(P(k)>(P(k+3)+7)) ...

```

```

& (P(k)>(P(k-4)+7)) & (P(k)>(P(k+4)+7)) & (P(k)>(P(k-5)+7)) &
(P(k)>(P(k+5)+7)) ...
& (P(k)>(P(k-6)+7)) & (P(k)>(P(k+6)+7)));
else
    bool = 0;
end

function [noise_masker_at_loc, loc] = noise_masker_check(psd,
tone_masker, low, high)

    noise_members = ones(1,high-low+1);

    for k = low:high,

        if (tone_masker(k) > 0),

            if ((k>2) & (k<63))
                m = 2;
            elseif((k>=63) & (k<127))
                m = 3;
            elseif((k>=127) & (k<256))
                m = 6;
            else
                m = 0;
            end

            for n = (k-low+1)-m:(k-low+1)+m,
                if (n > 0)
                    noise_members(n) = 0;
                end
            end
        end
    end

    if (isempty(find(noise_members)))
        noise_masker_at_loc = 0;
        loc = -1;
    else
        temp = 0;
        for k = (low+find(noise_members)-1),
            temp = temp + 10.^(0.1.*psd(k));
        end
        noise_masker_at_loc = 10*log10(temp);
        loc = geomean(low+find(noise_members)-1);
    end

function [TM_above_thres, NM_above_thres] = psy_step_3( tone_masker,
noise_masker )

global fr_size fft_size fr_count fs bit_res
global freq_hz freq_bark Abs_thr

TM_above_thres = tone_masker.*(tone_masker>Abs_thr);
NM_above_thres = noise_masker.*(noise_masker>Abs_thr);

```

```

for j = 1 : length(Abs_thr),
    toneFound=0;
    noiseFound=0;

    if (TM_above_thres(j)>0)
        toneFound=1;
    end
    if (NM_above_thres(j)>0)
        noiseFound=1;
    end

    if (toneFound | noiseFound)
        masker_loc_barks = freq_bark(j);

        crit_bw_low = masker_loc_barks-0.5;
        crit_bw_high = masker_loc_barks+0.5;

        low_loc = max(find(freq_bark<crit_bw_low));
        if (isempty(low_loc))
            low_loc=1;
        else
            low_loc=low_loc+1;
        end
        high_loc = max(find(freq_bark<crit_bw_high));

        for k=low_loc:high_loc,
            if (toneFound)

                if ((TM_above_thres(j) < TM_above_thres(k)) & (k ~= j)),
                    TM_above_thres(j)=0;
                    break;
                elseif (k ~= j)
                    TM_above_thres(k)=0;
                end

                if (TM_above_thres(j) < NM_above_thres(k)),
                    TM_above_thres(j)=0;
                    break;
                else
                    NM_above_thres(k)=0;
                end

            elseif (noiseFound)

                if ((NM_above_thres(j) < NM_above_thres(k)) & (k ~= j)),
                    NM_above_thres(j)=0;
                    break;
                elseif (k ~= j)
                    NM_above_thres(k)=0;
                end

                if (NM_above_thres(j) < TM_above_thres(k)),
                    NM_above_thres(j)=0;
                    break;
                else
                    TM_above_thres(k)=0;
                end
            end
        end
    end
end

```

```

        end
    else
        disp('ERROR');
    end
end
end
end
end

```

```

function [Thr_TM, Thr_NM] = psy_step_4( P_TM_th, P_NM_th )

global freq_hz freq_bark Abs_thr

Thr_TM = zeros(1,length(P_TM_th));

for k = find(P_TM_th),

    [thres, start] = mask_threshold(1,k,P_TM_th(k),freq_bark);
    Thr_TM(start:start+length(thres)-1) =
Thr_TM(start:start+length(thres)-1)+10.^(0.1.*thres);

end

Thr_NM = zeros(1,length(P_NM_th));

for k = find(P_NM_th)

    [thres, start] = mask_threshold(0,k,P_NM_th(k),freq_bark);
    Thr_NM(start:start+length(thres)-1) =
Thr_NM(start:start+length(thres)-1)+10.^(0.1.*thres);
end

```

```

function [threshold, start] = mask_threshold(type, j, P,
bark)

maskerloc=bark(j);

low=maskerloc-3;
high=maskerloc+8;

lowbin=max(find(bark<low));

if (isempty(lowbin))
    lowbin = 1;
end
highbin=max(find(bark<high));

SF = spreading_function(j, P, lowbin, highbin, bark);

if (type==0)

    threshold=P-.175*bark(j)+SF-2.025;

```

```

else

    threshold=P-.275*bark(j)+SF-6.025;
end

start=lowbin;

function spread = spreading_function(masker_bin, power, low,
high, bark)

    masker_bark=bark(masker_bin);
    for i=low:high,
        maskee_bark=bark(i);
        deltaz=maskee_bark-masker_bark;
        if ((deltaz>=-3.5) & (deltaz<-1))
            spread(i-low+1)=17*deltaz-0.4*power+11;
        elseif ((deltaz>=-1) & (deltaz<0))
            spread(i-low+1)=(0.4*power+6)*deltaz;
        elseif ((deltaz>=0) & (deltaz<1))
            spread(i-low+1)=-17*deltaz;
        elseif ((deltaz>=1) & (deltaz<8.5))
            spread(i-low+1)=(0.15*power-17)*deltaz-0.15*power;
        end
    end
end

```

```

function [Thr_global] = psy_step_5( Thr_TM, Thr_NM )

    global freq_hz freq_bark Abs_thr

    temp = Thr_TM + Thr_NM;

    for k = 1 : length(Abs_thr),
        temp(k)=temp(k)+10.^(0.1.*Abs_thr(k));
    end
    Thr_global = temp;

```

## ΣΥΝΑΡΤΗΣΗ **audio\_synthesis\_solution.m**

```

function Synth_audio = peak_pick(P, Thr_global, cf_angle, fft_size)

```



```

count = 0;

for i = 1 : 1 : length(P)
    if ( P(i) < Thr_global(i) )
        count = count + 1;
        P(i) = -50;
    end
end
count
P = (10.^((P-90.302)/10)).^(0.5);

P = P .* cos(cf_angle(1:fft_size/2+1)) + j * P
.*sin(cf_angle(1:fft_size/2+1));

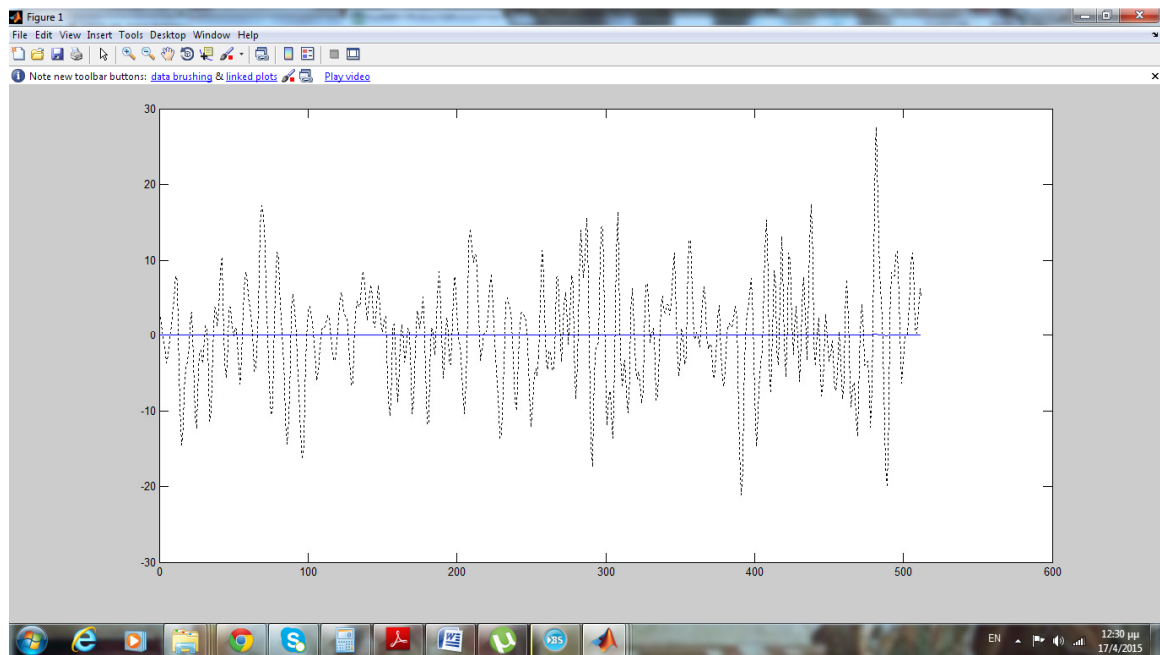
P = [P; conj(flipud(P(2:end-1)))];

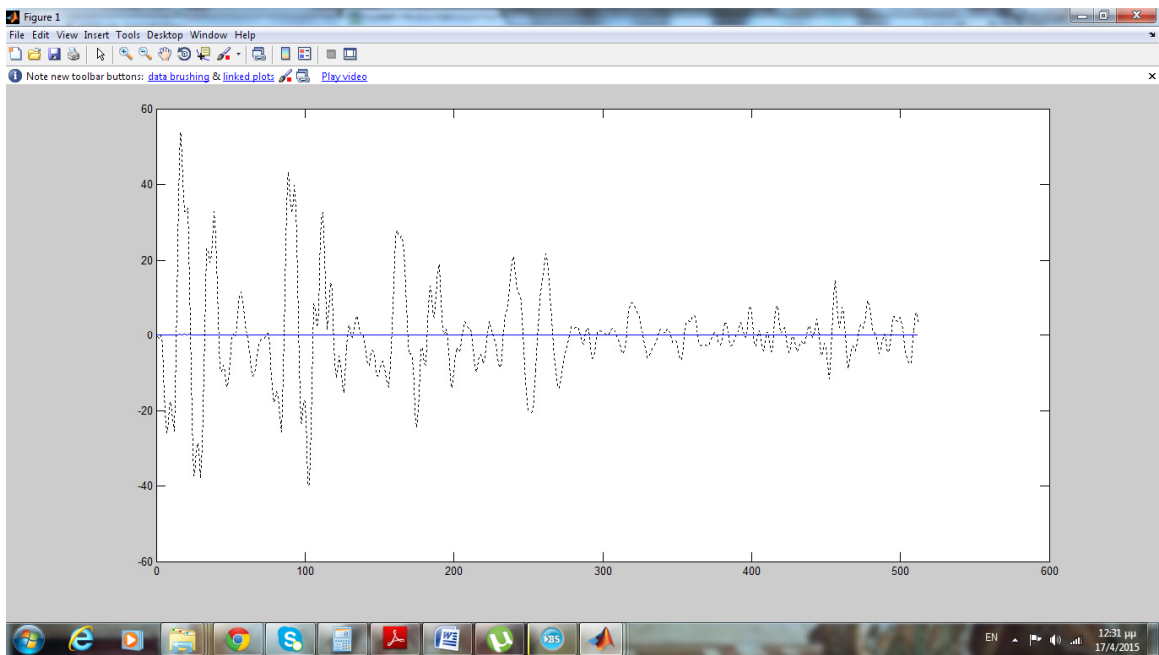
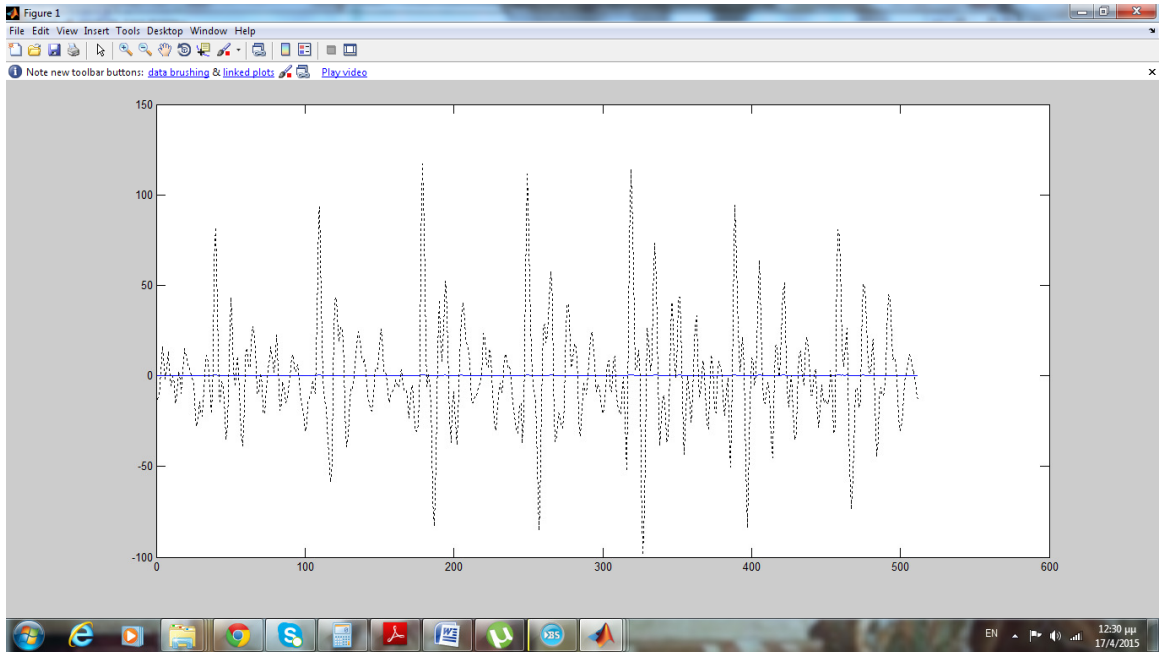
Synth_audio = real(ifft(P, fft_size)) * fft_size;

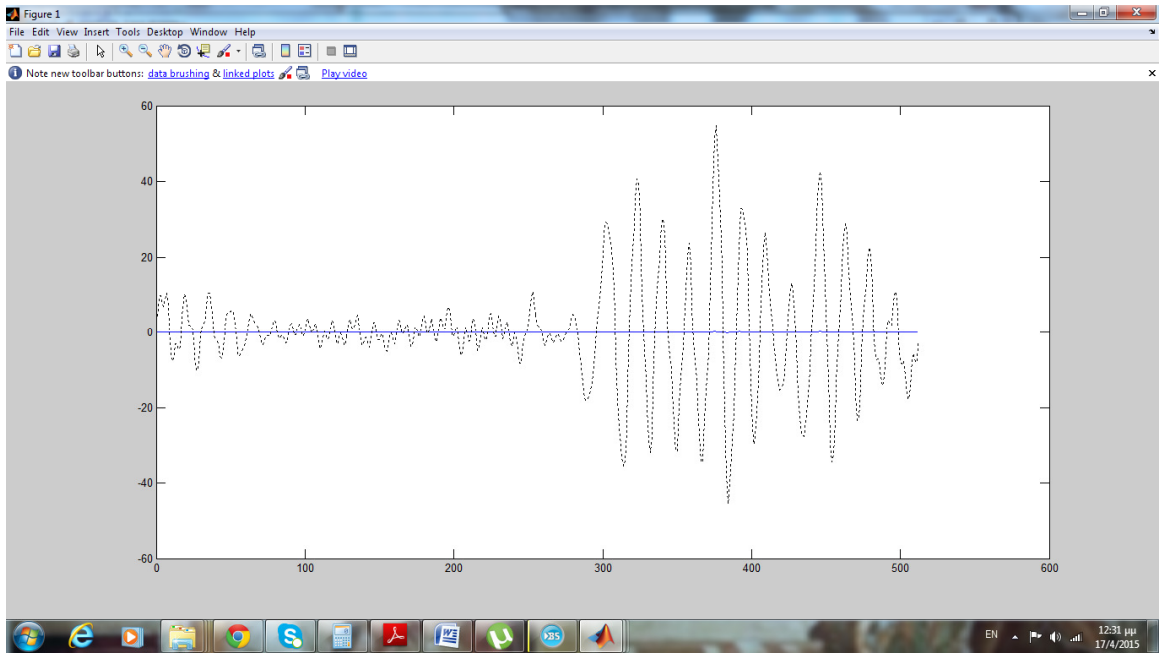
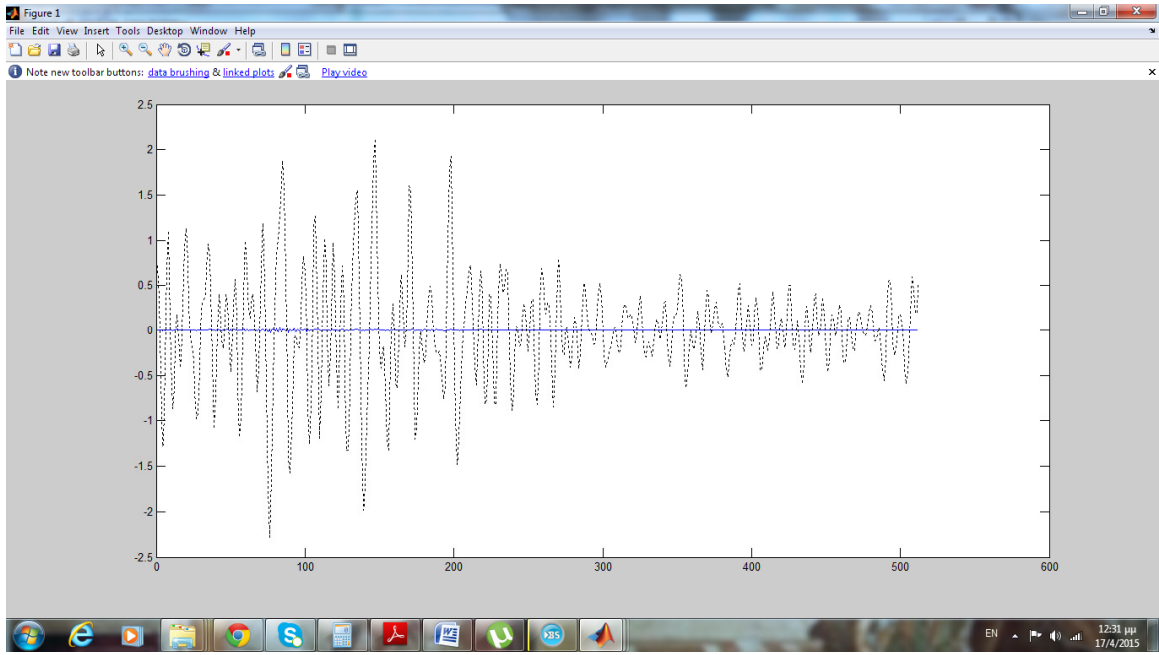
```

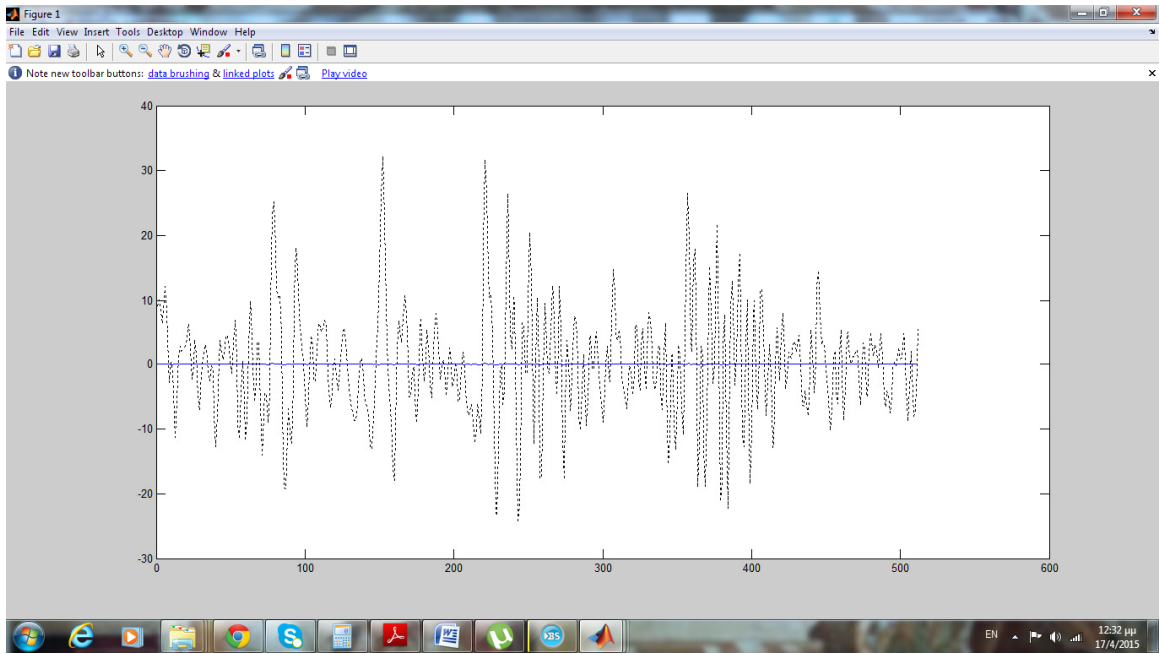
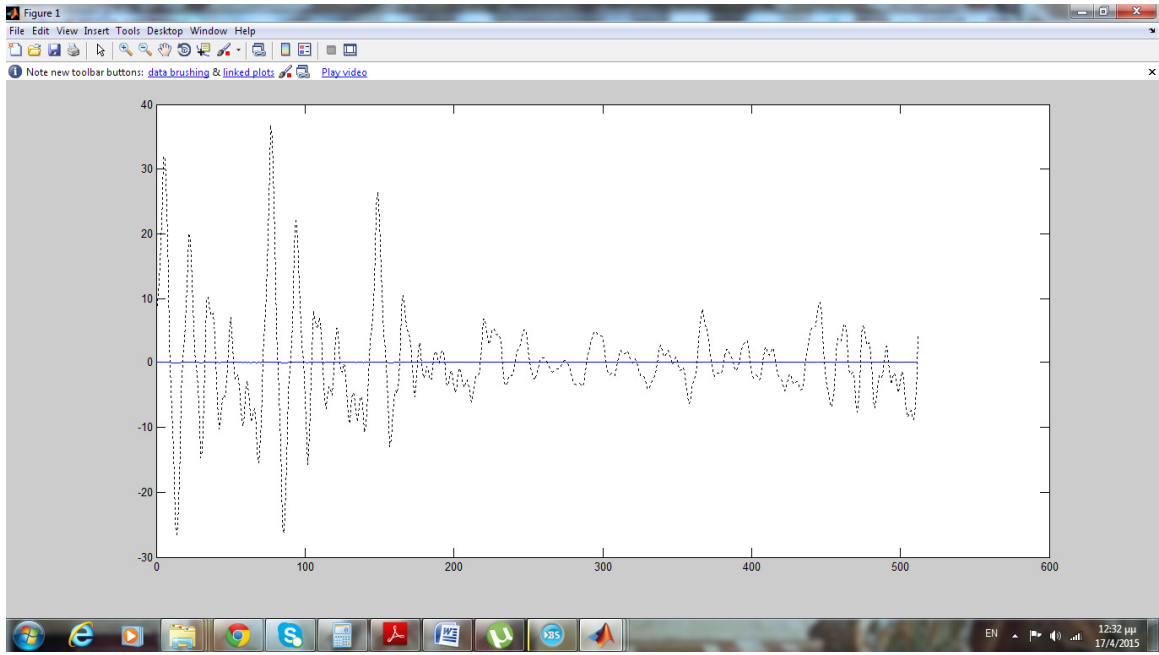
## ΠΑΡΑΡΤΗΜΑ Β

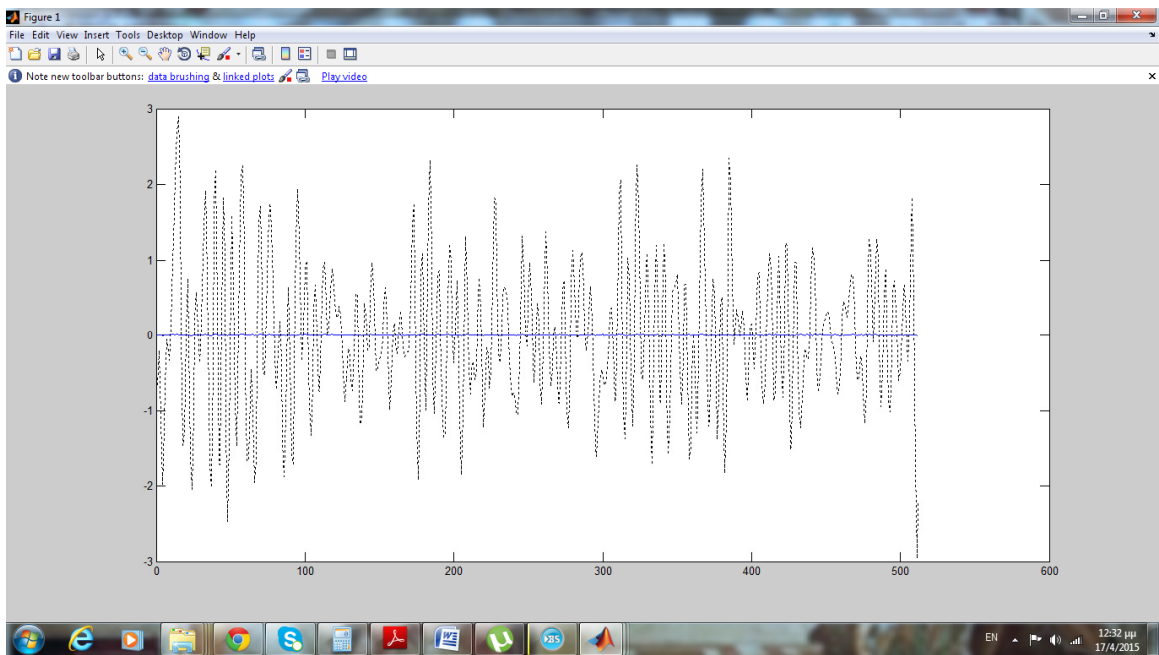
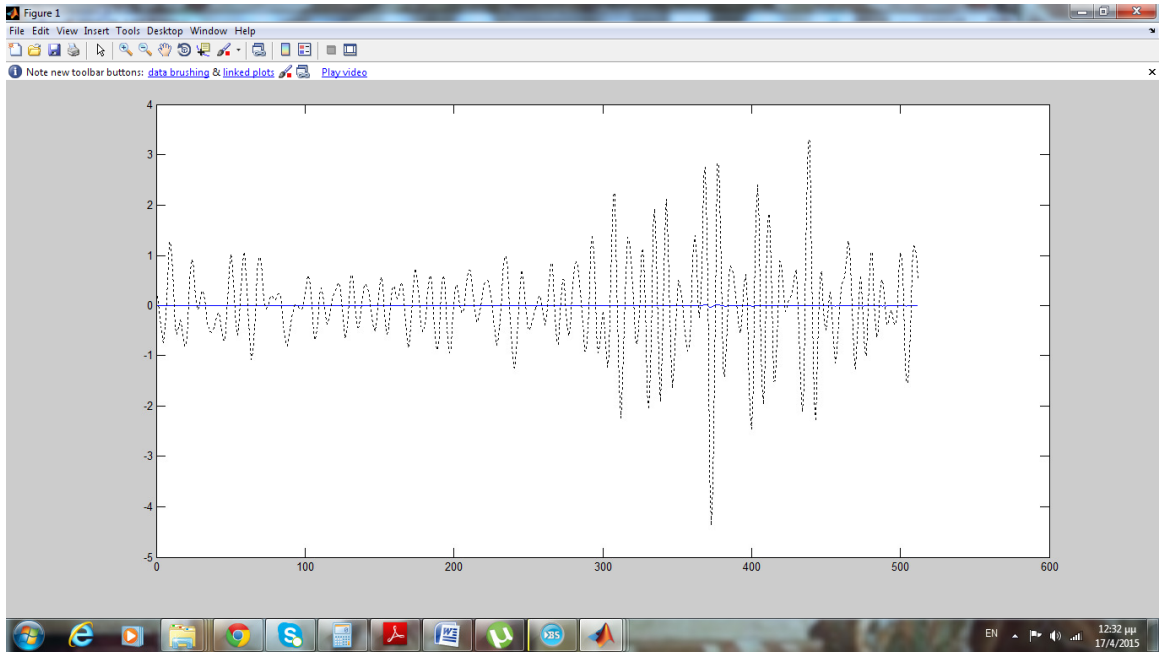
### ΔΙΑΓΡΑΜΜΑΤΑ

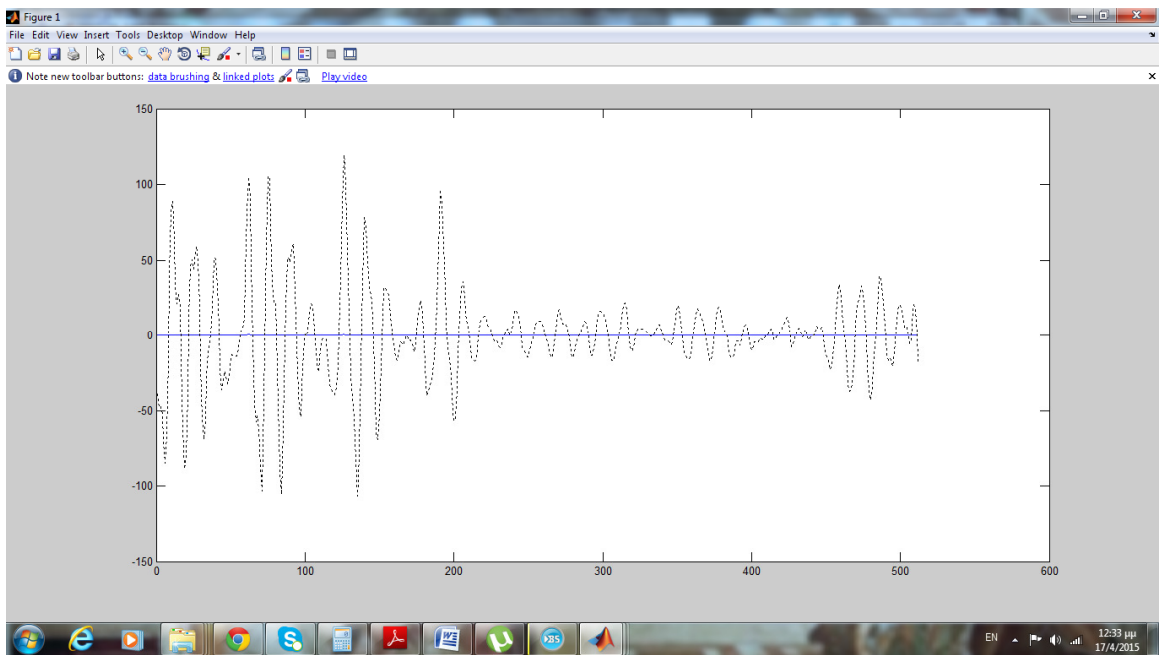
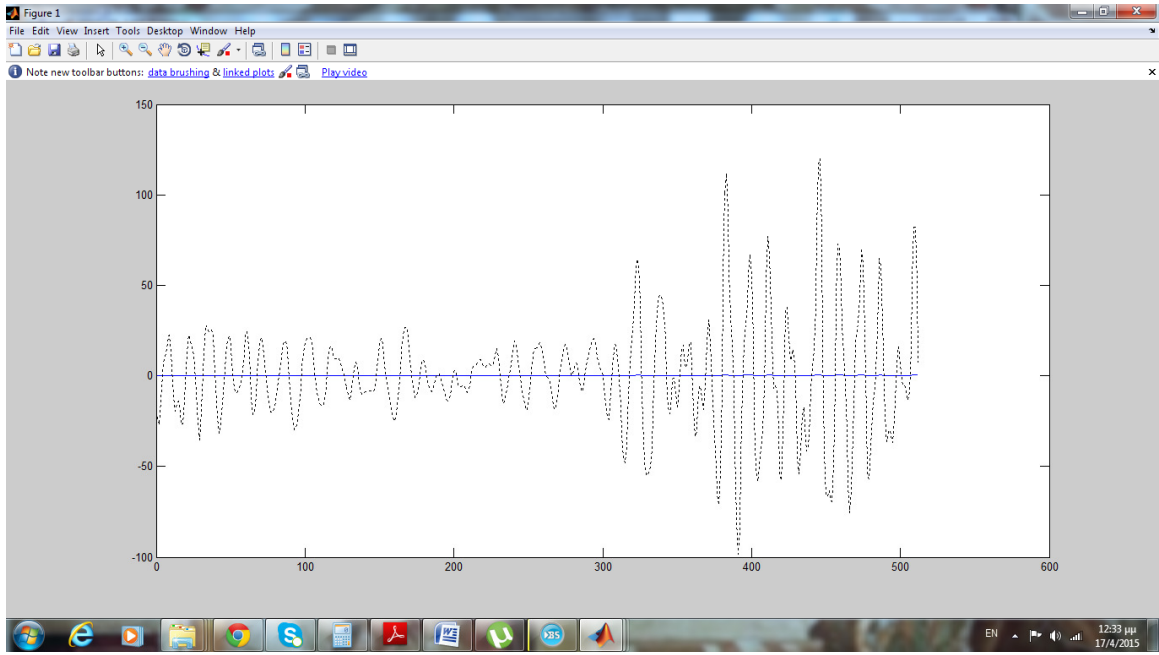


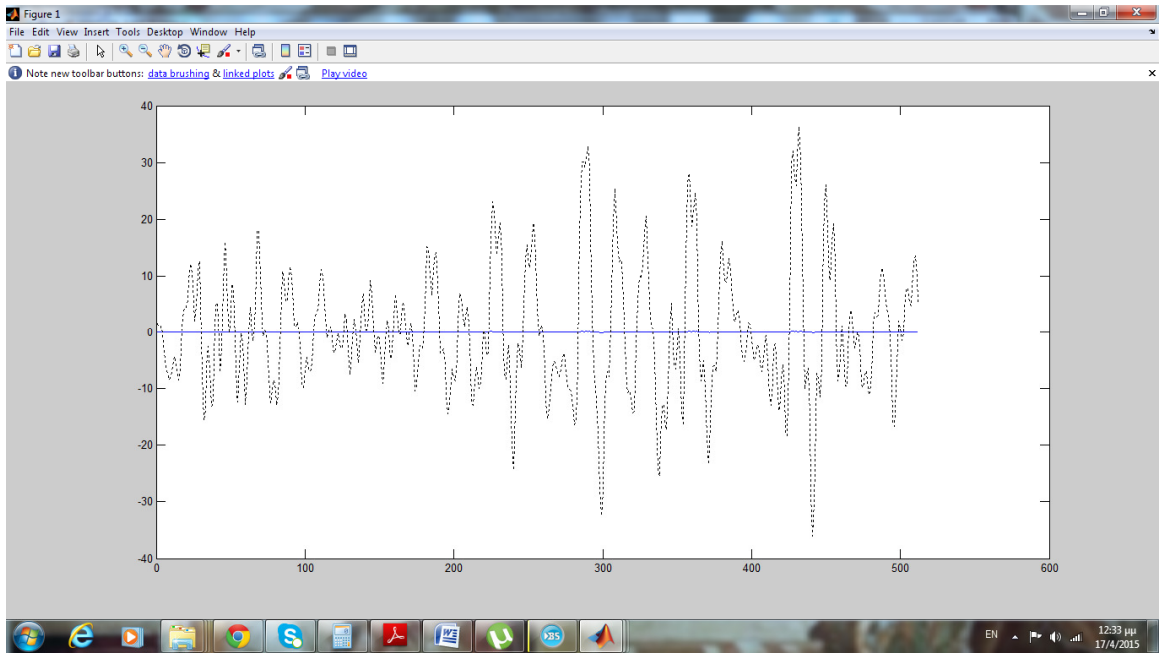
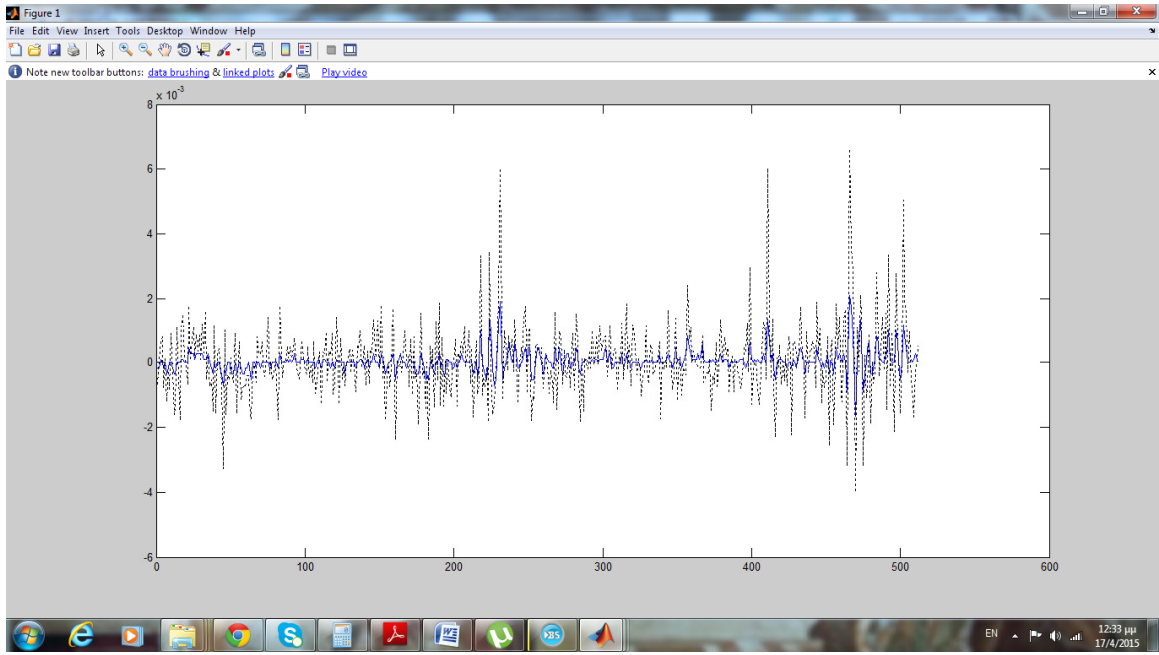


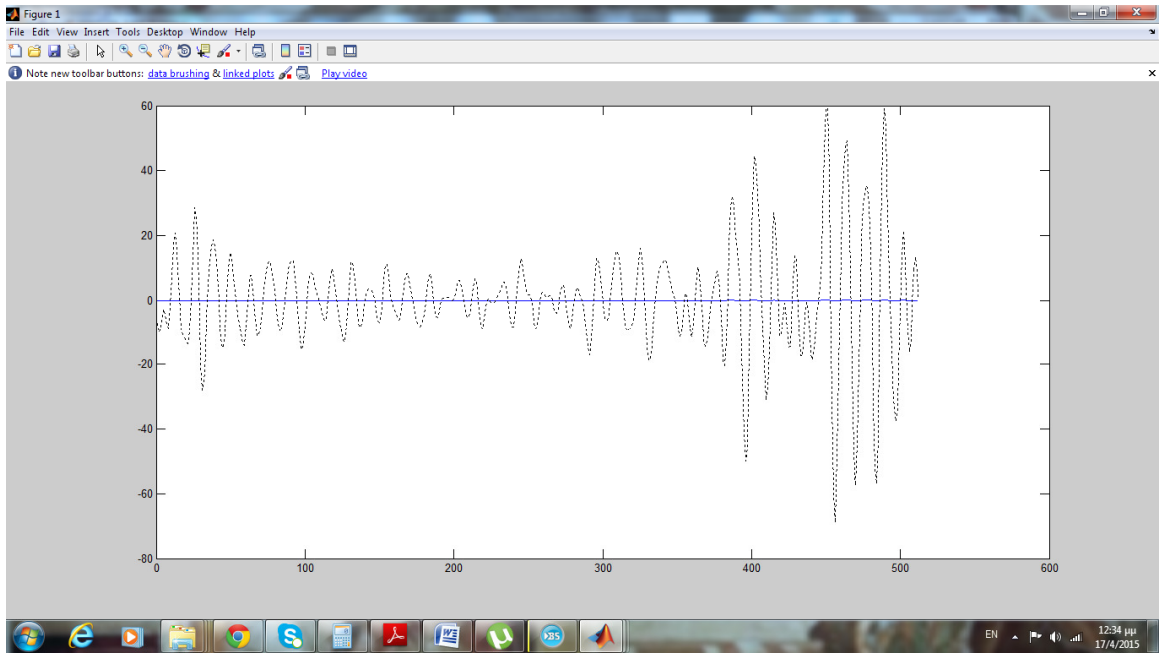
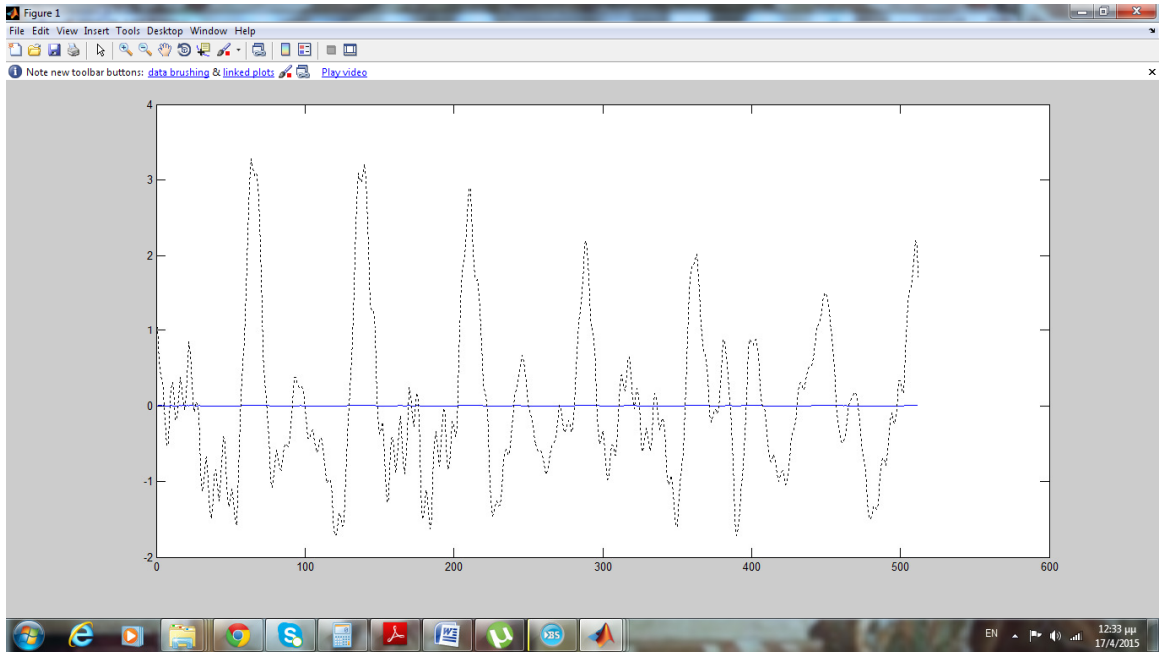




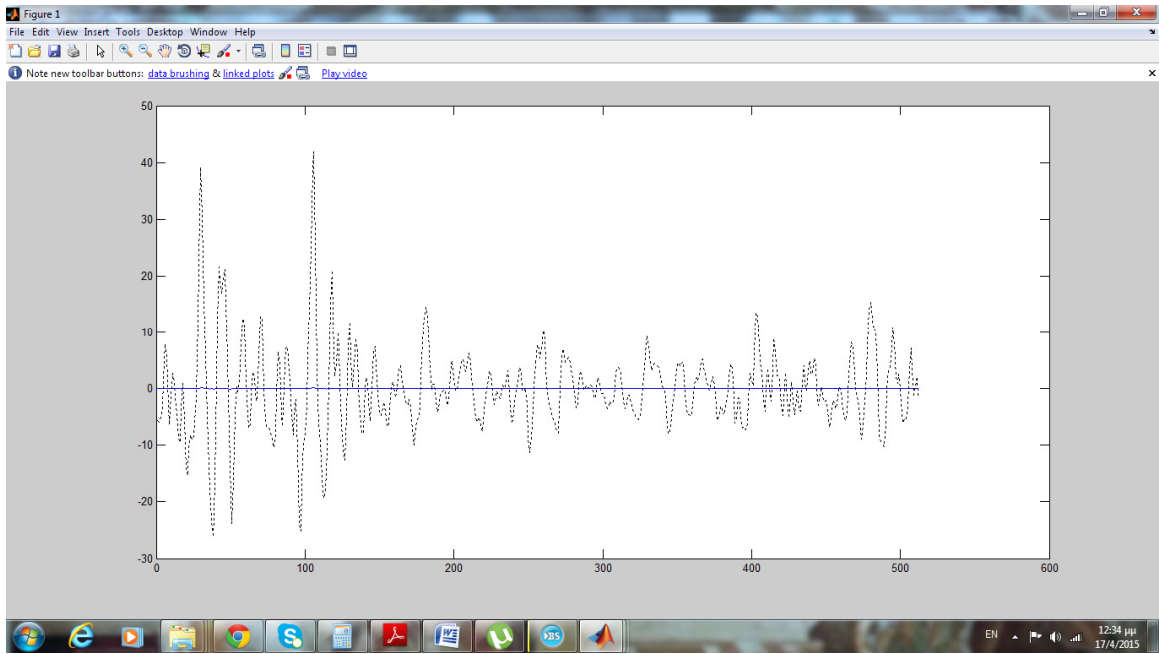
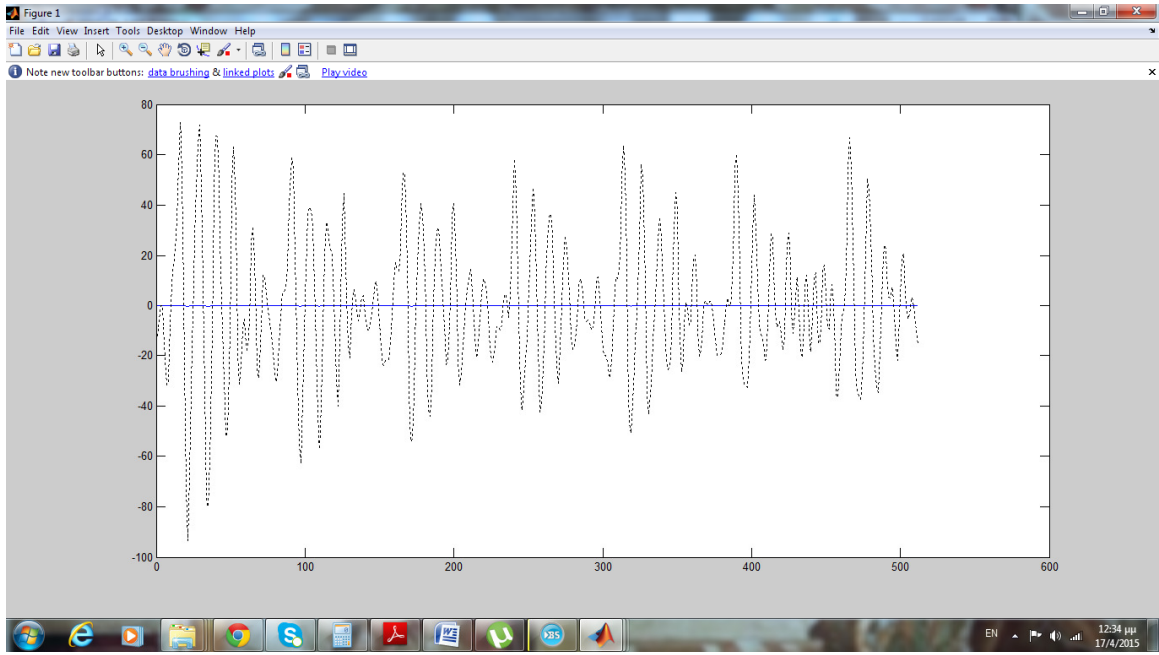


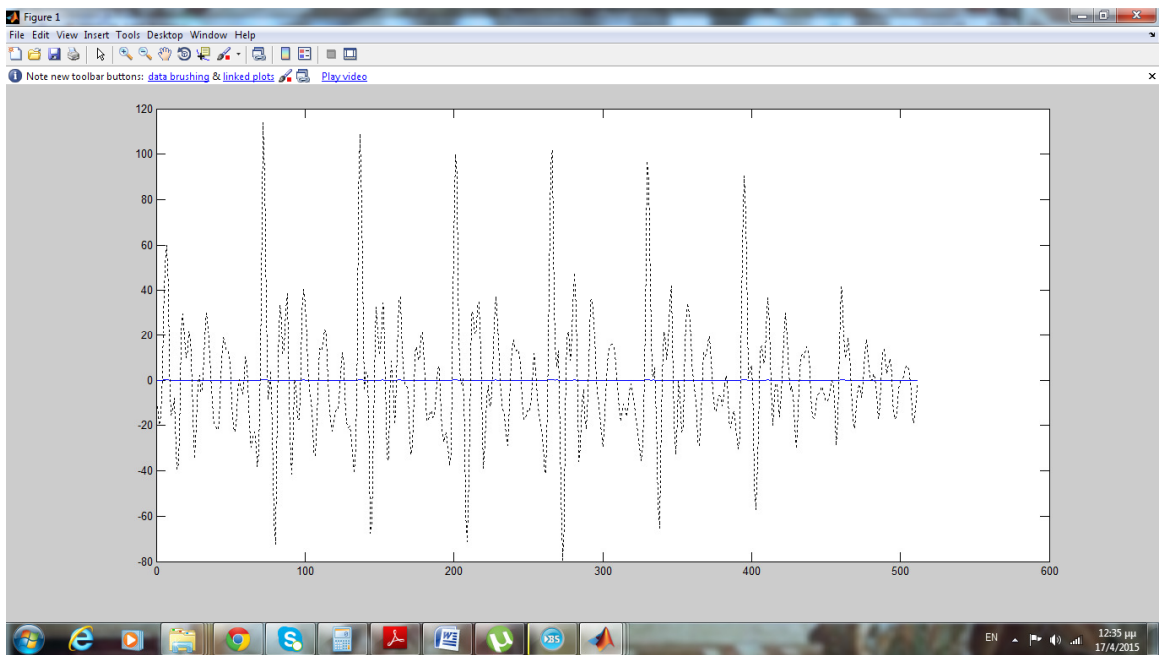
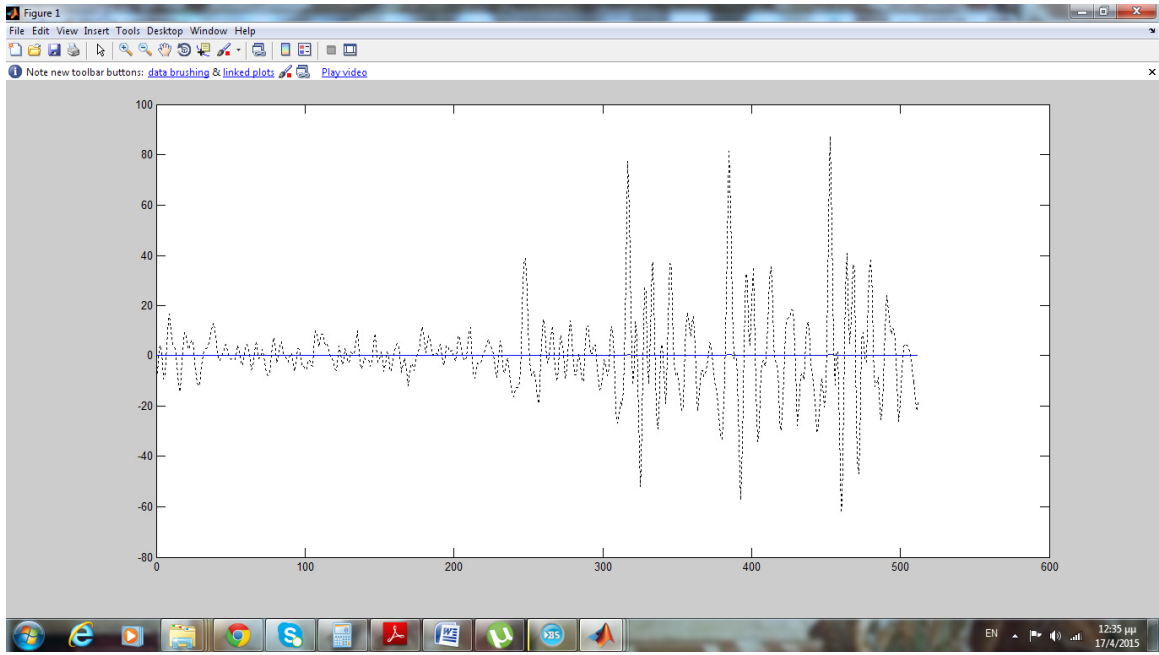


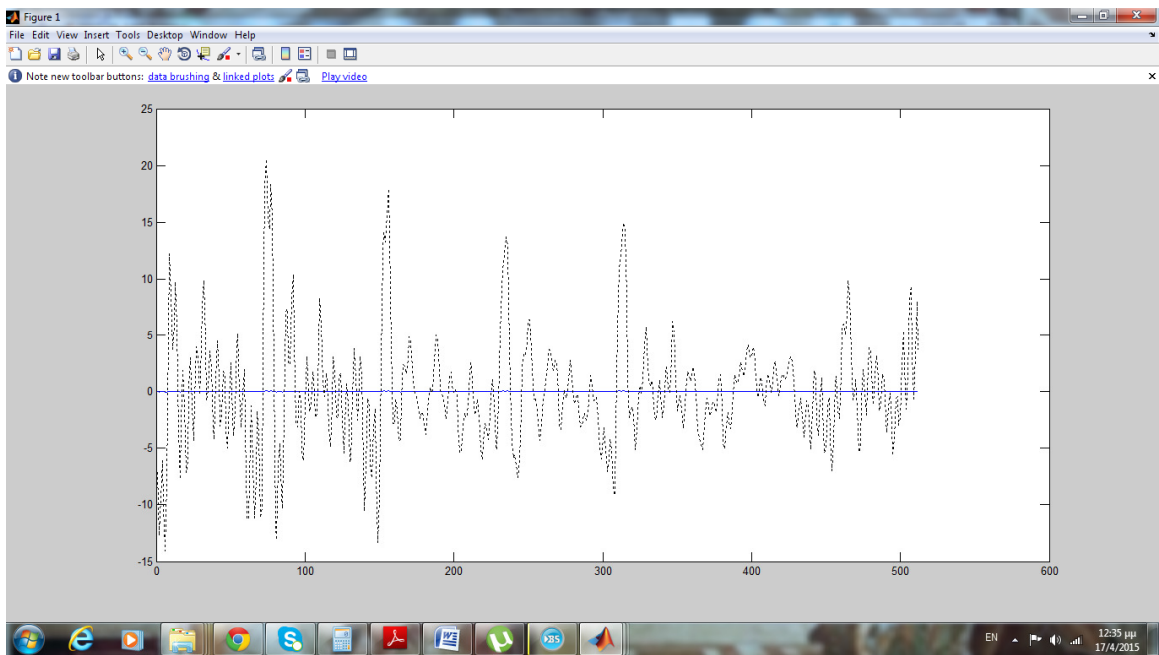
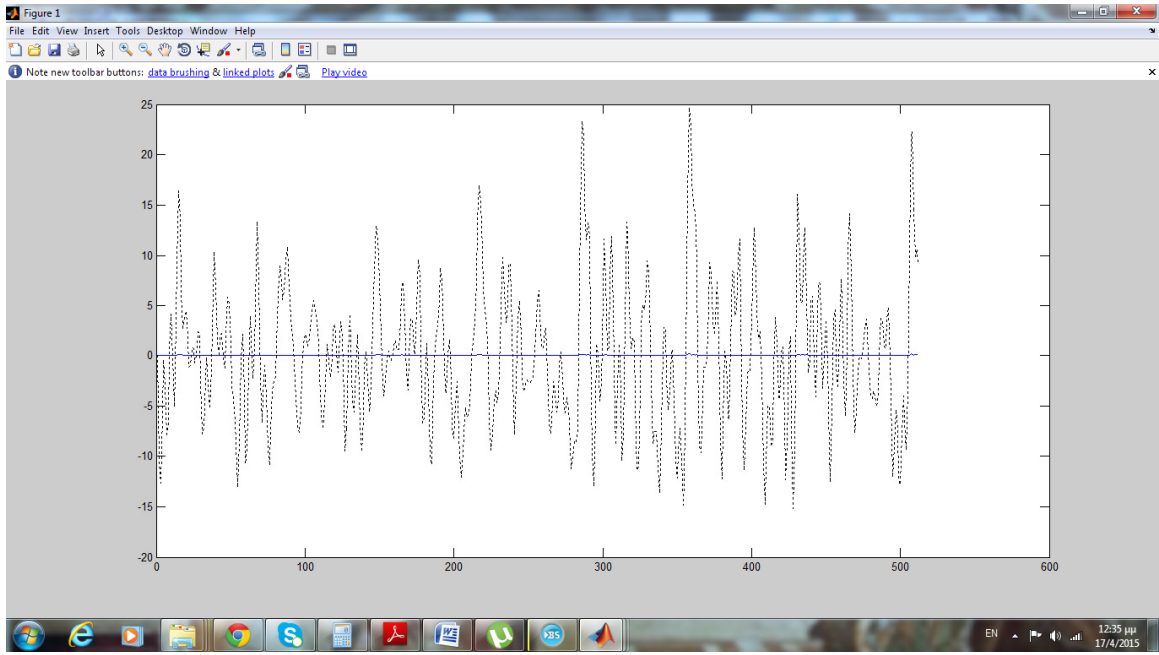


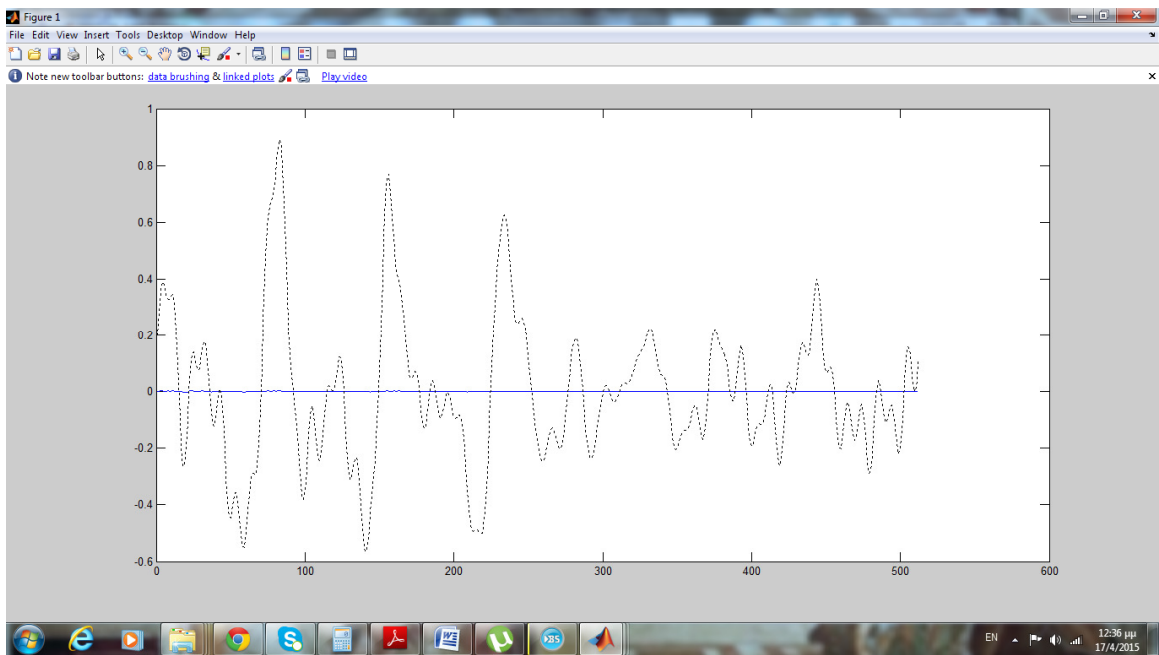
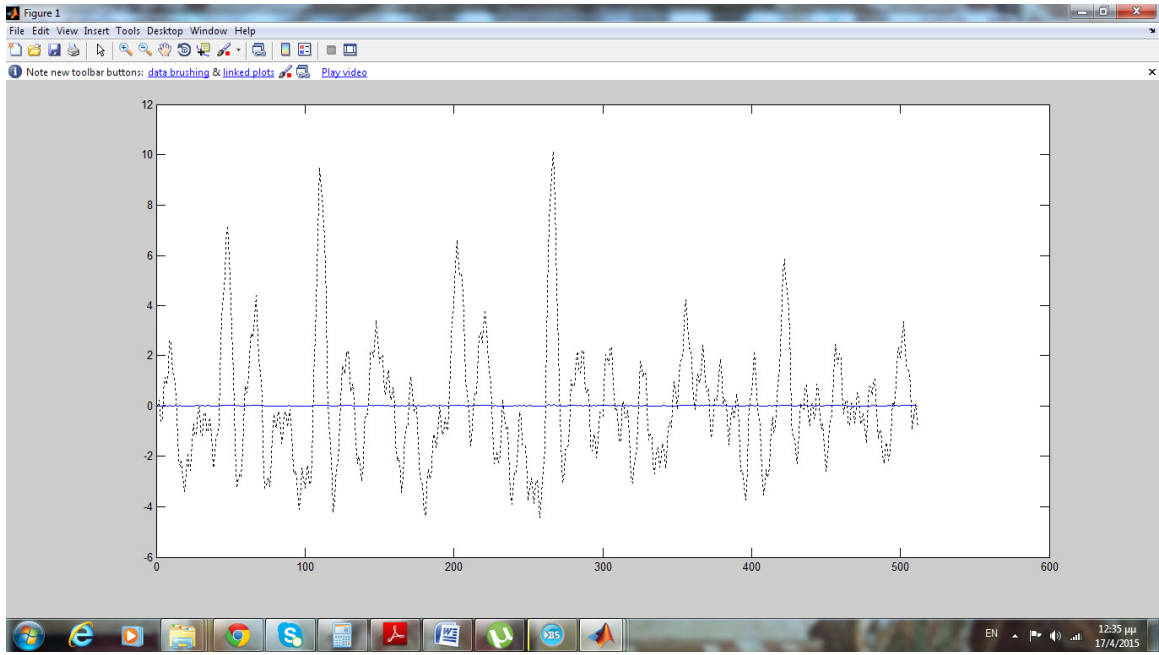


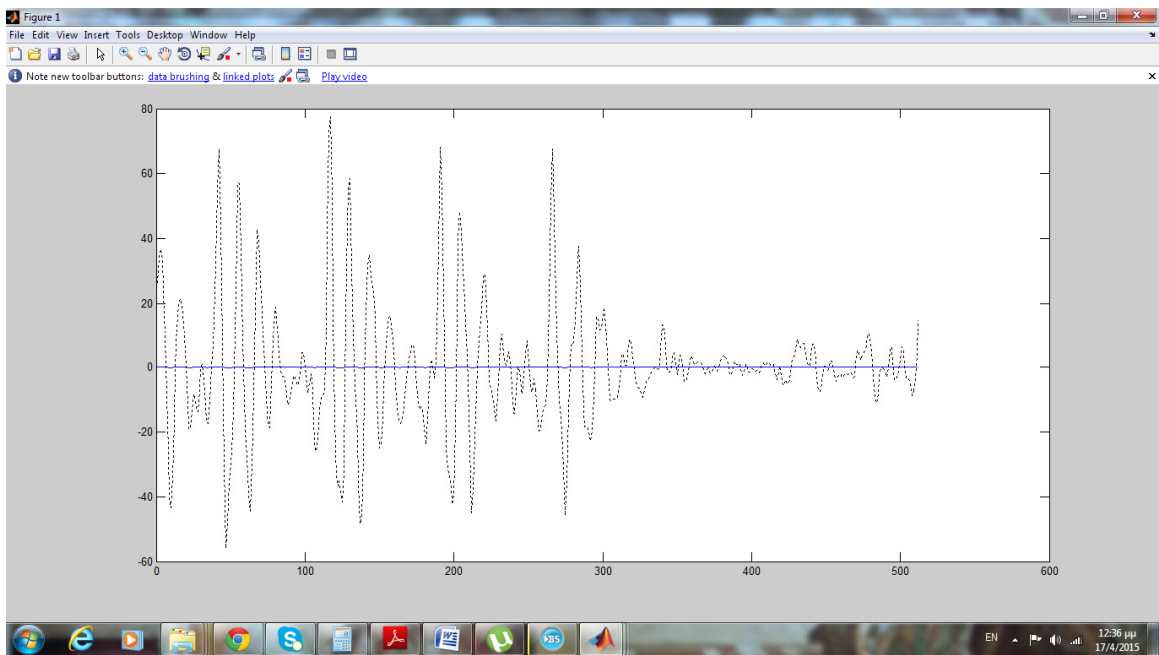
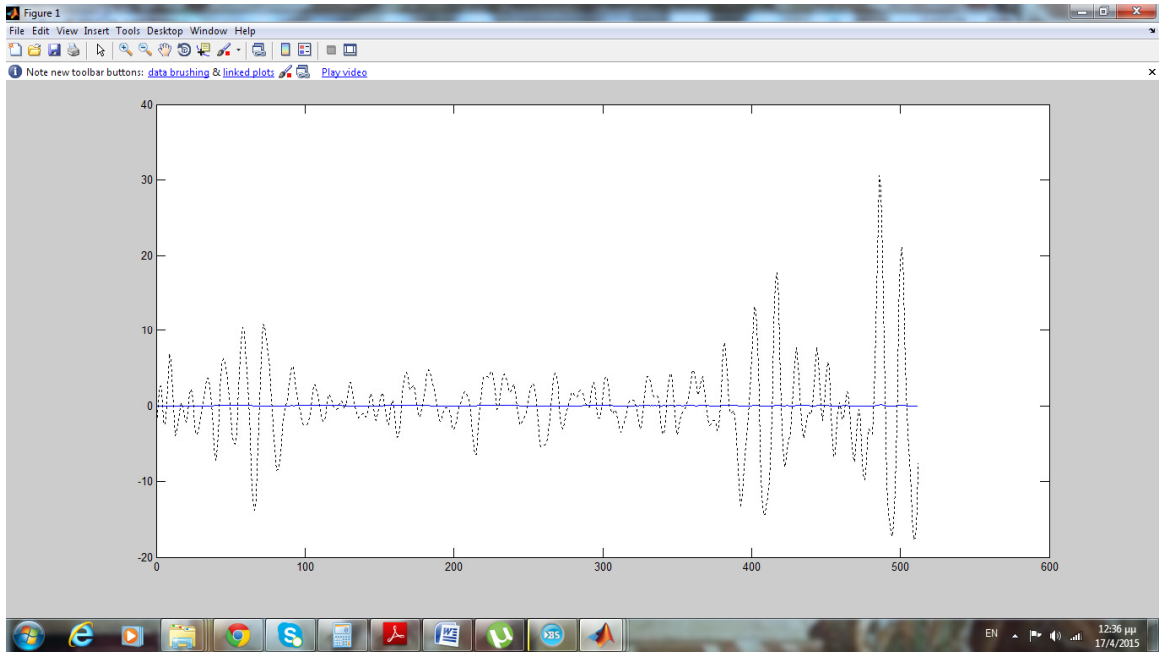


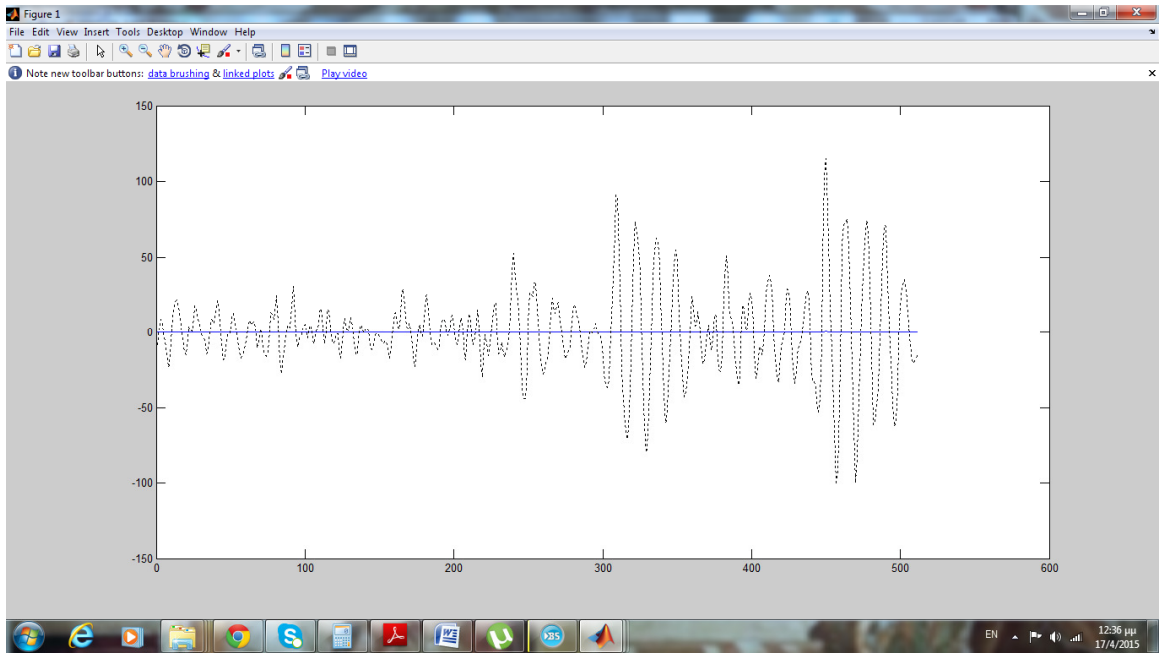
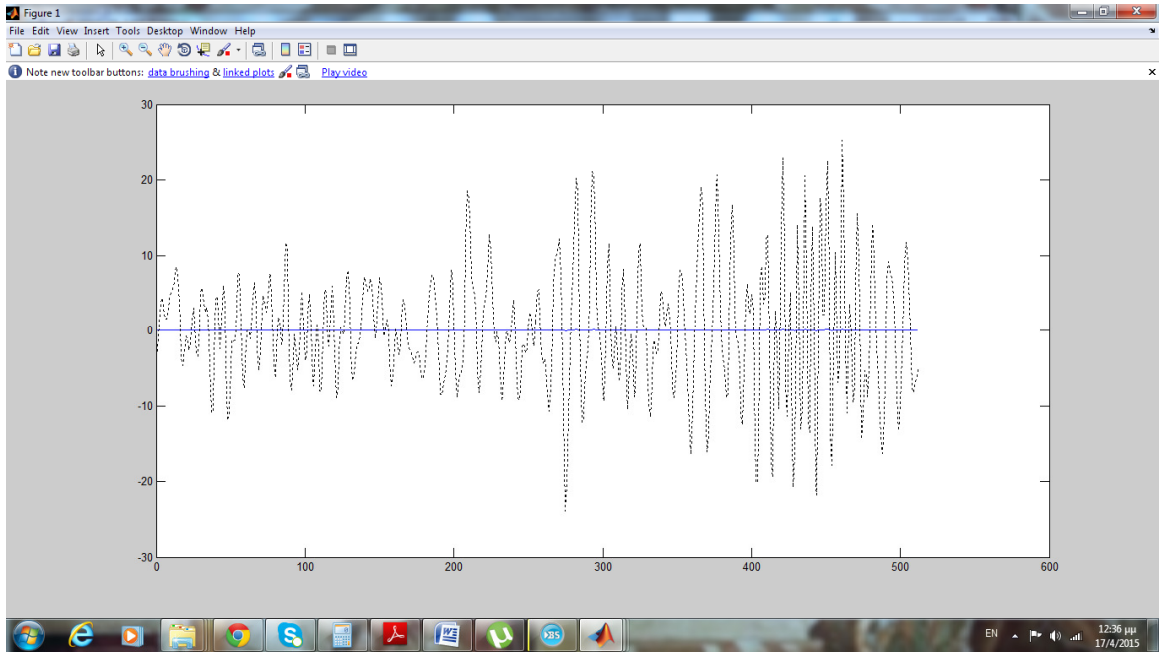


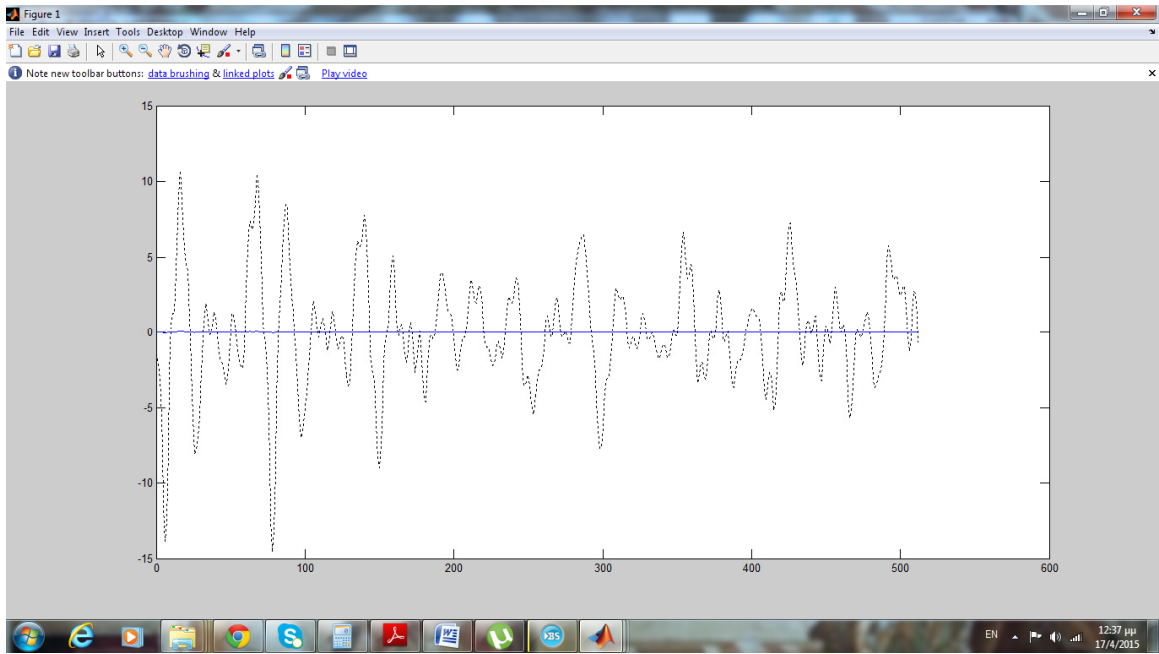
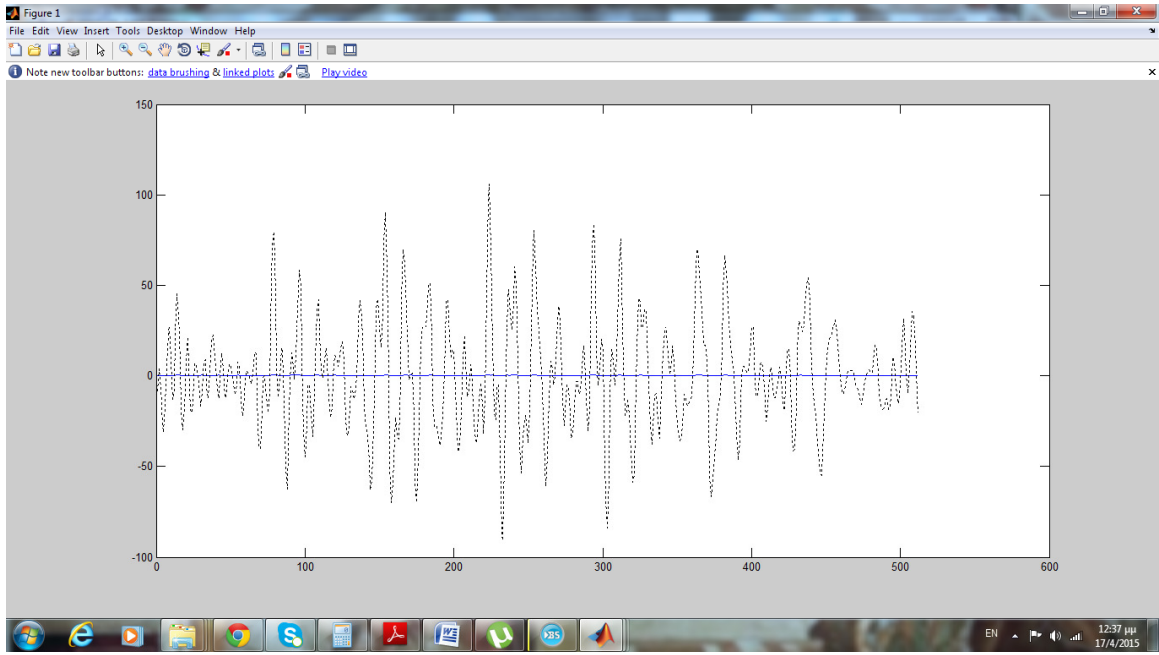


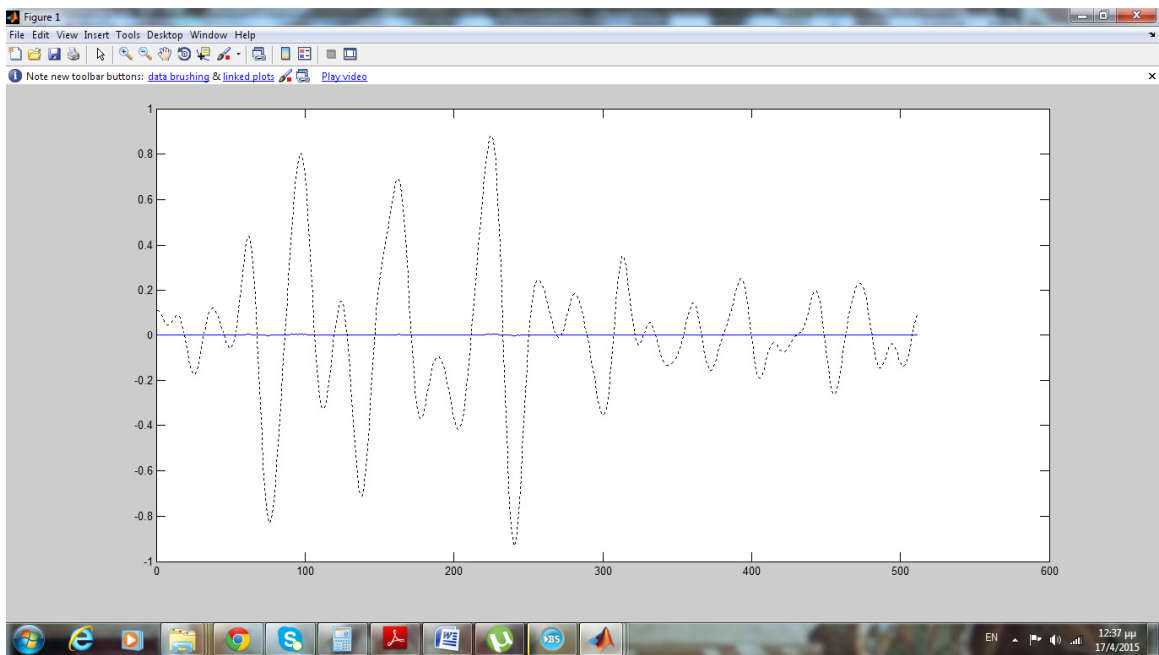
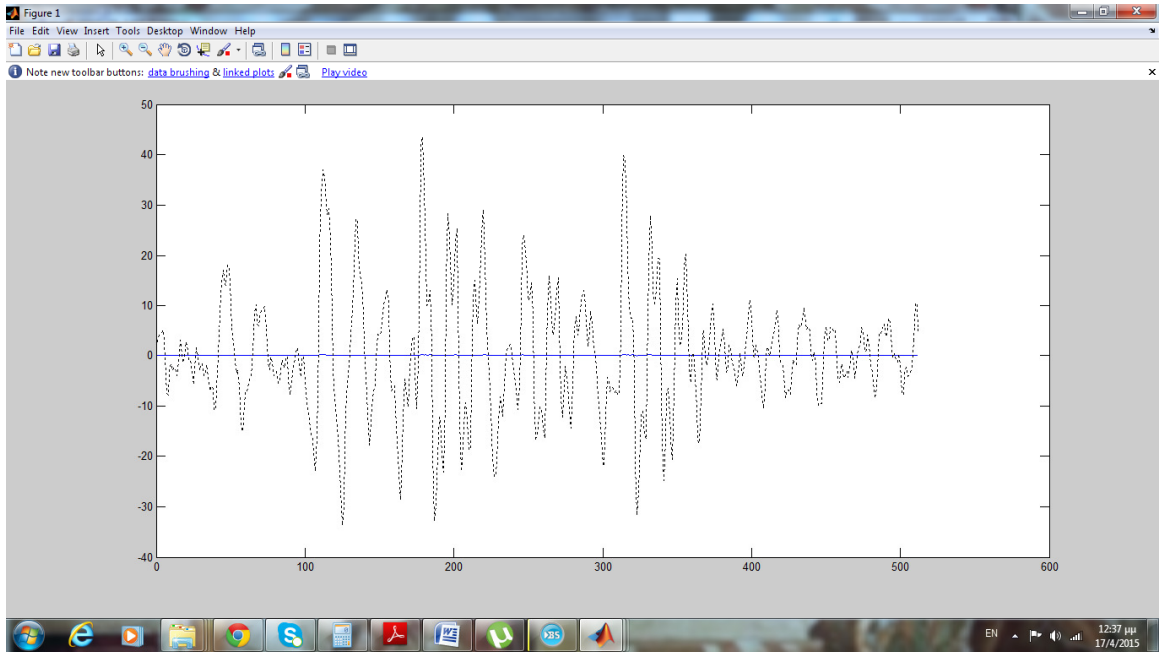




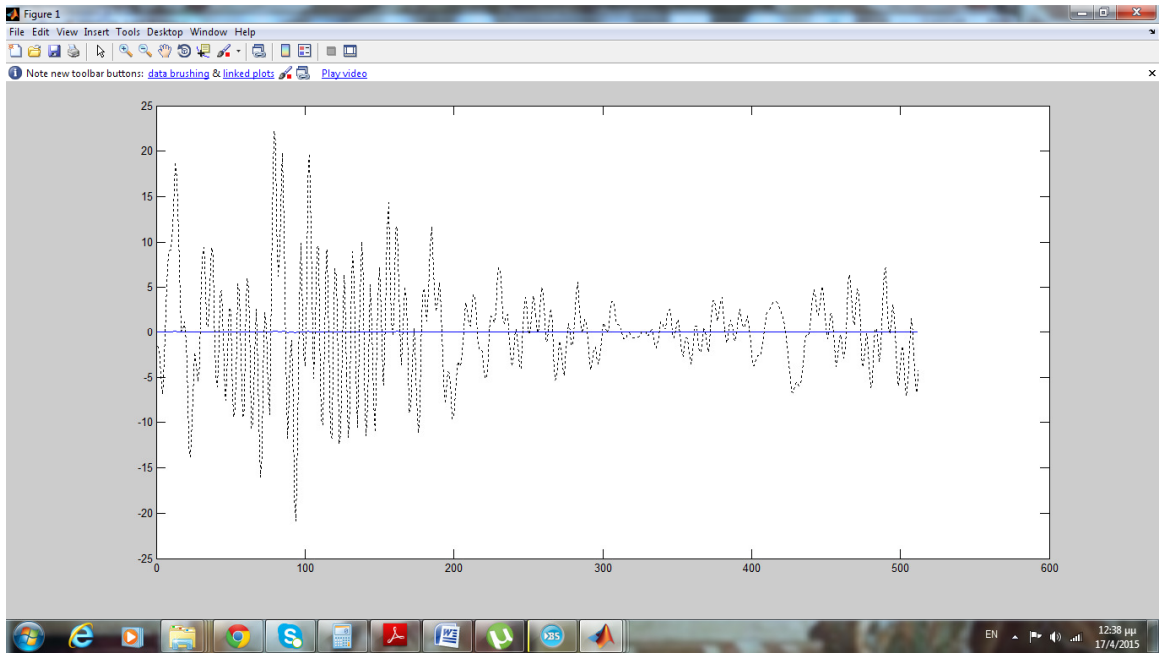
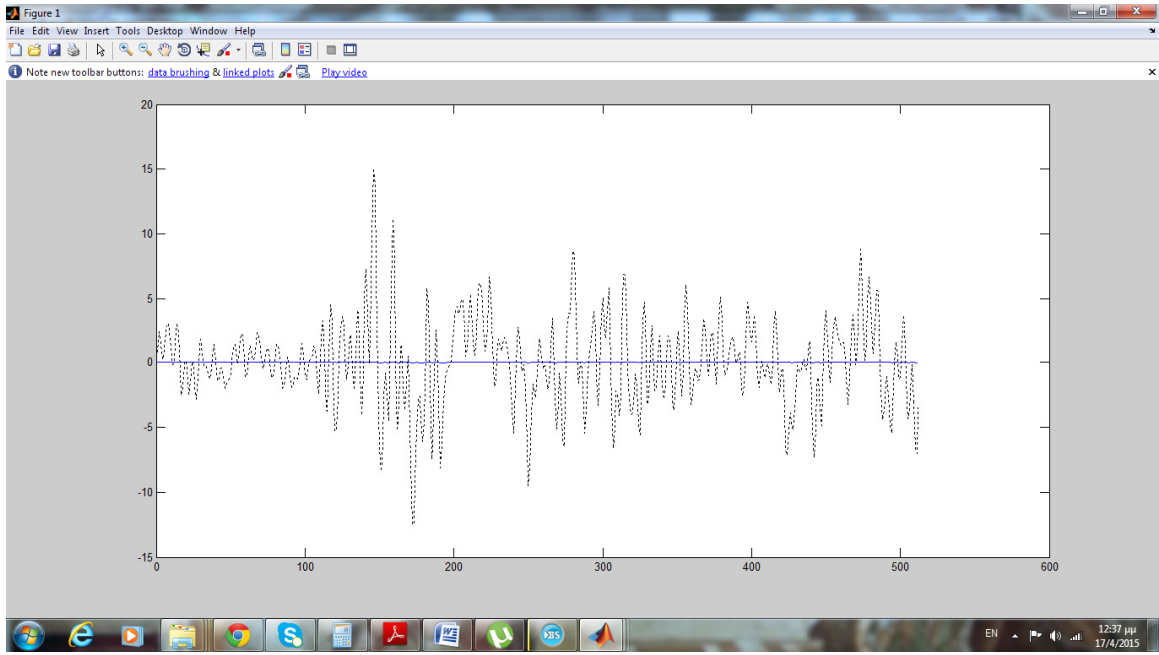


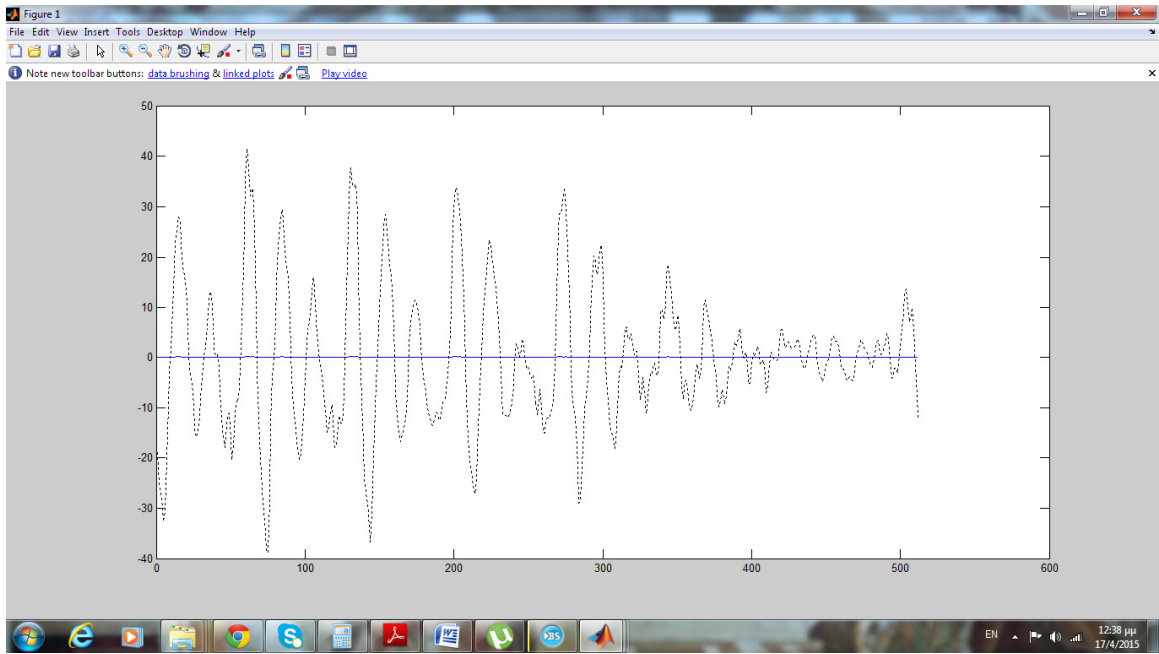
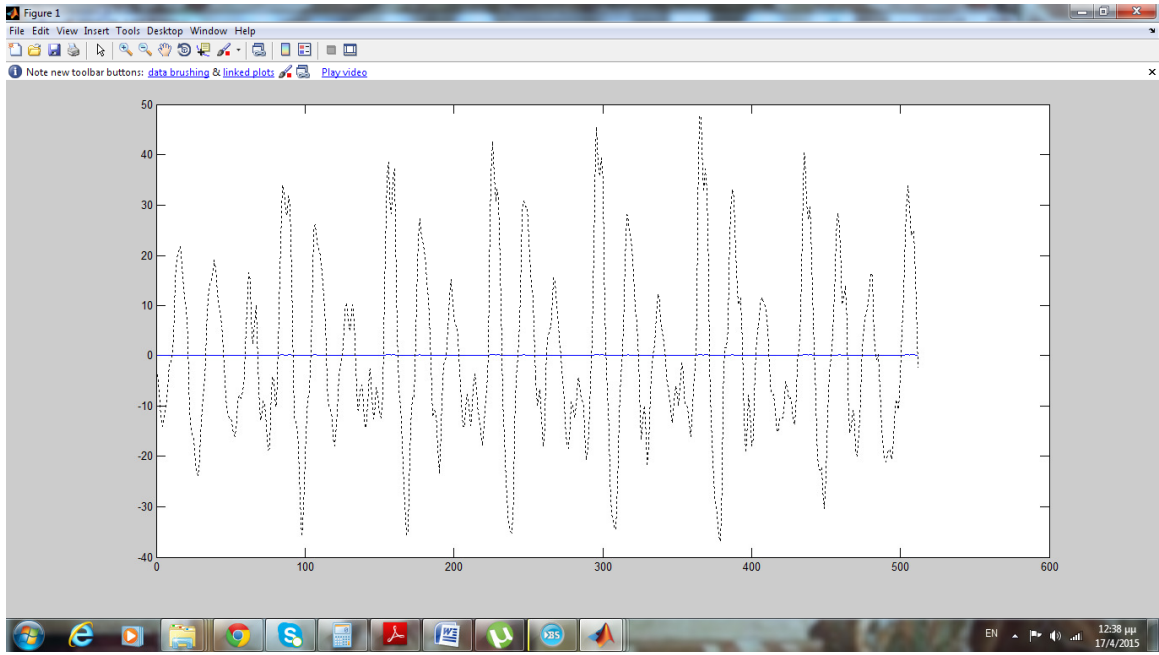


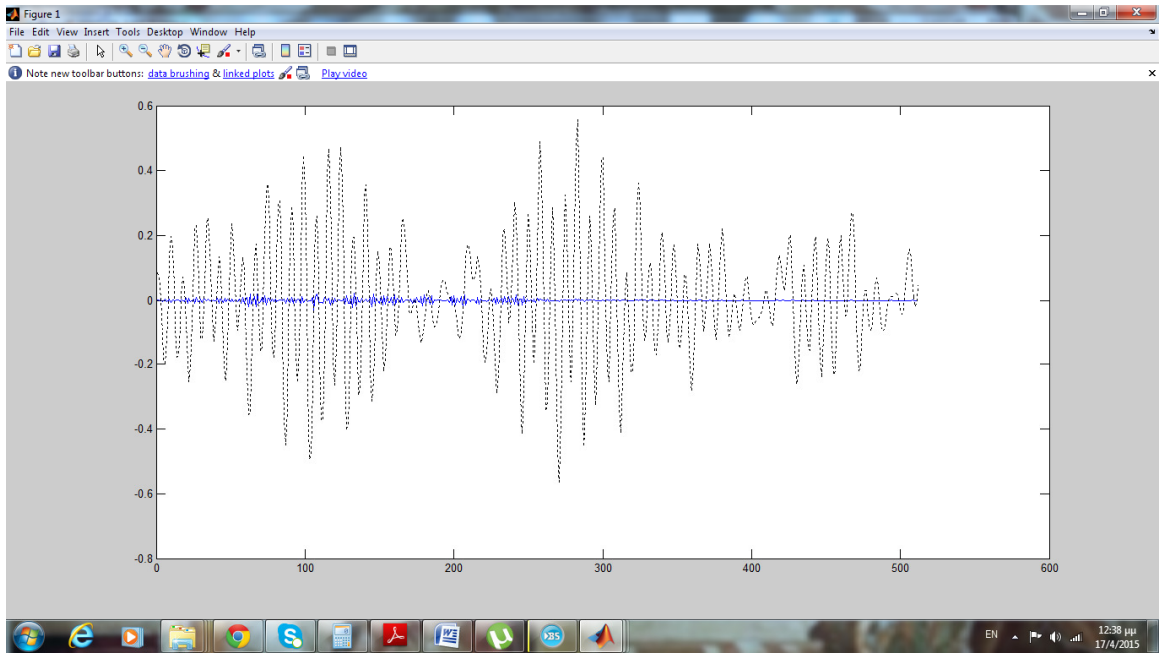
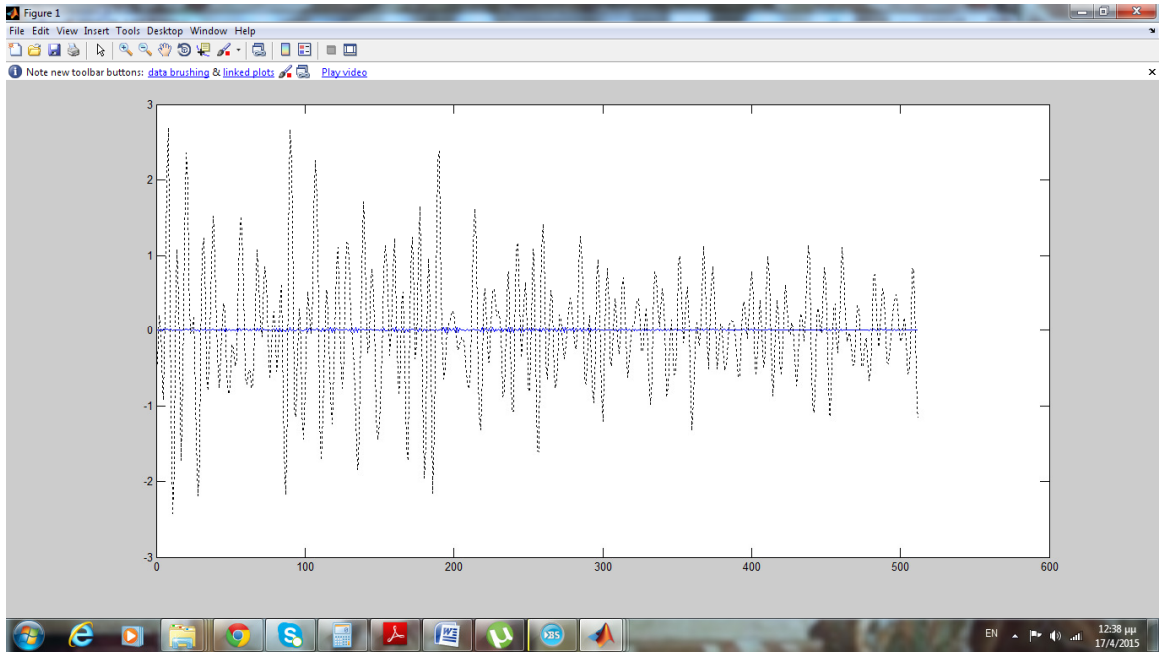


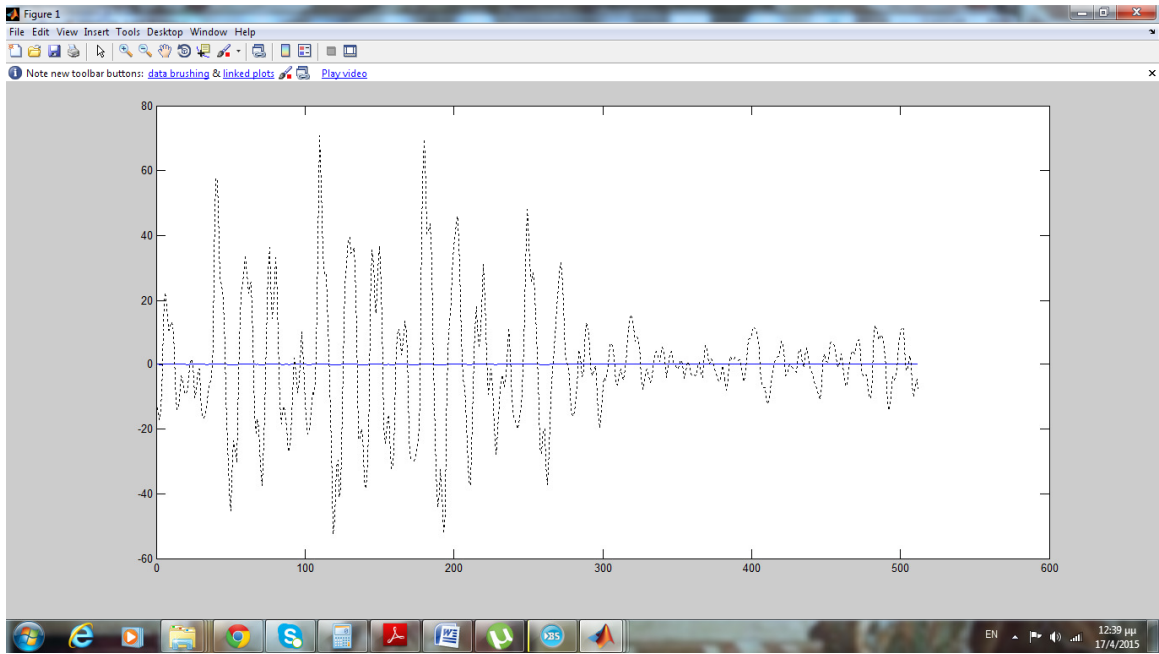
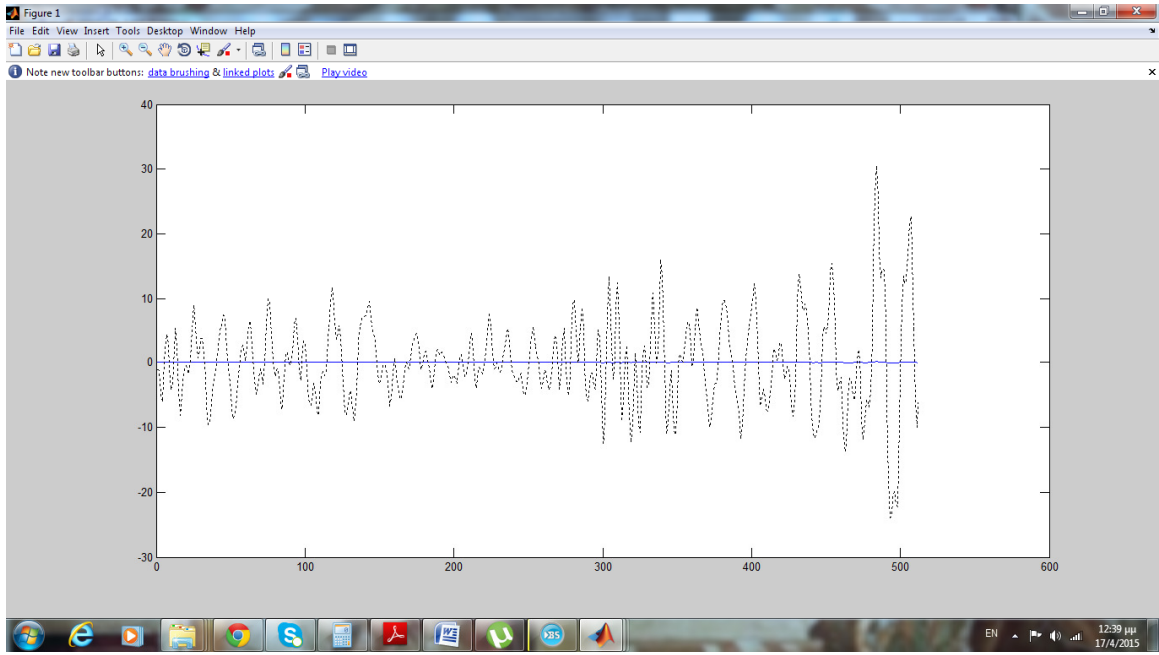


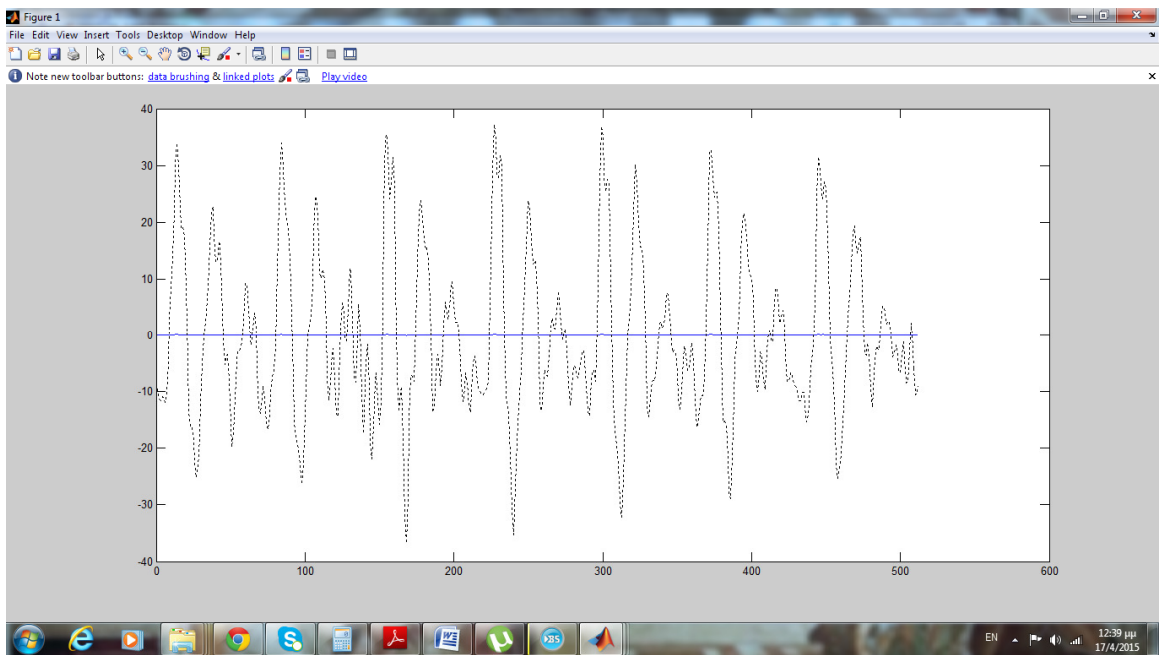
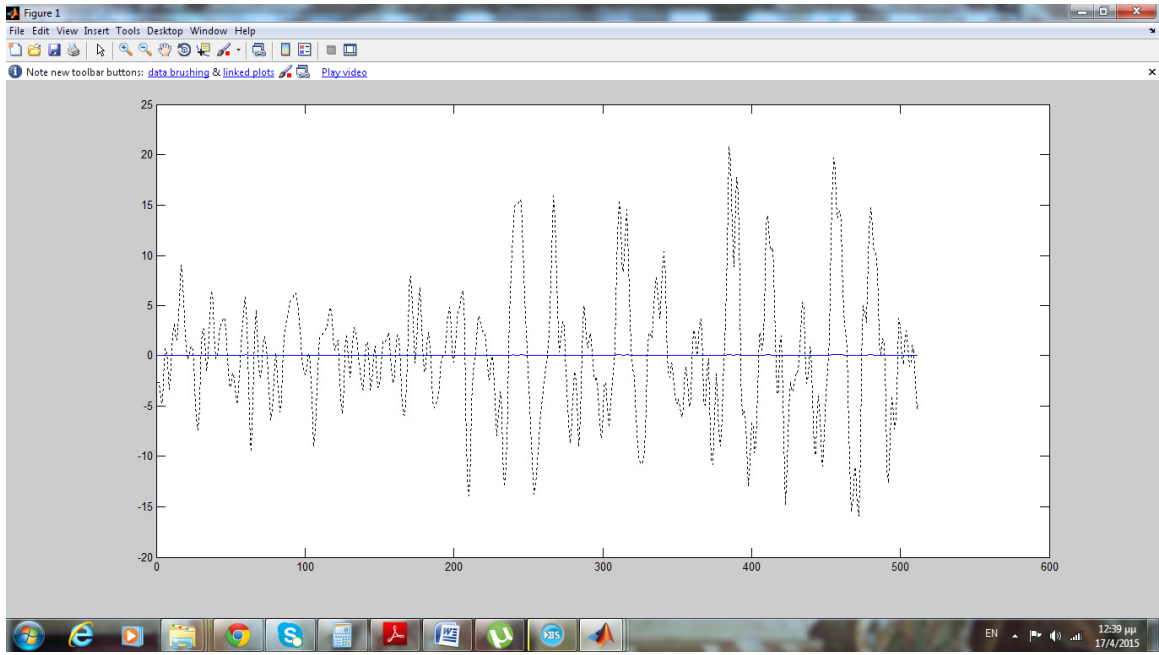


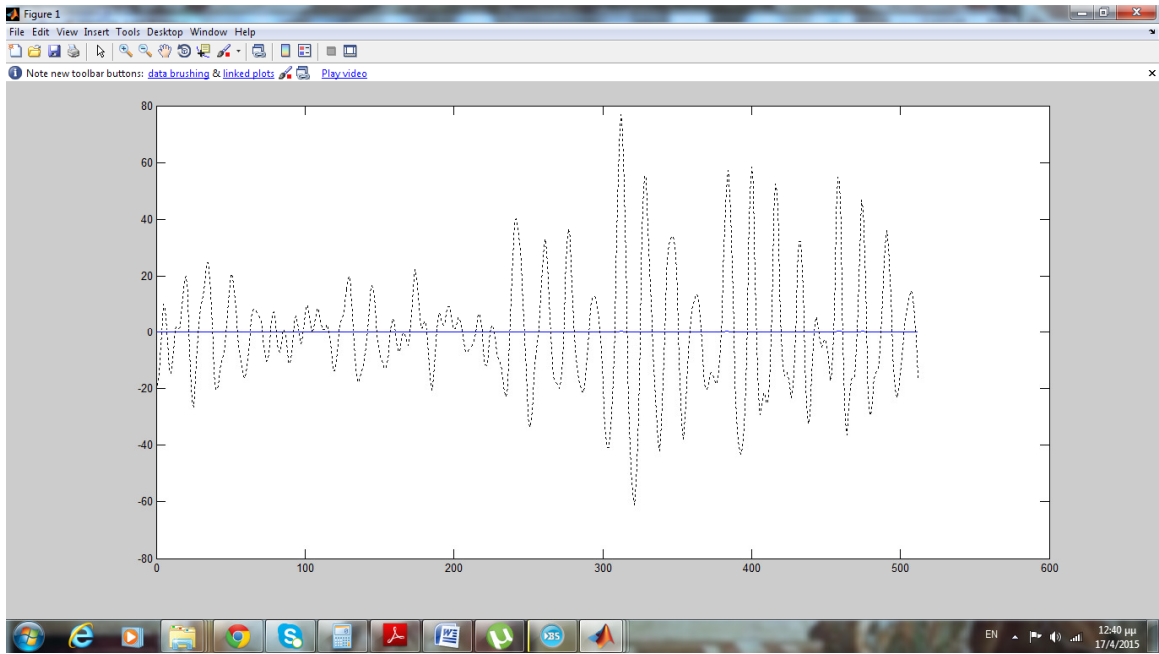
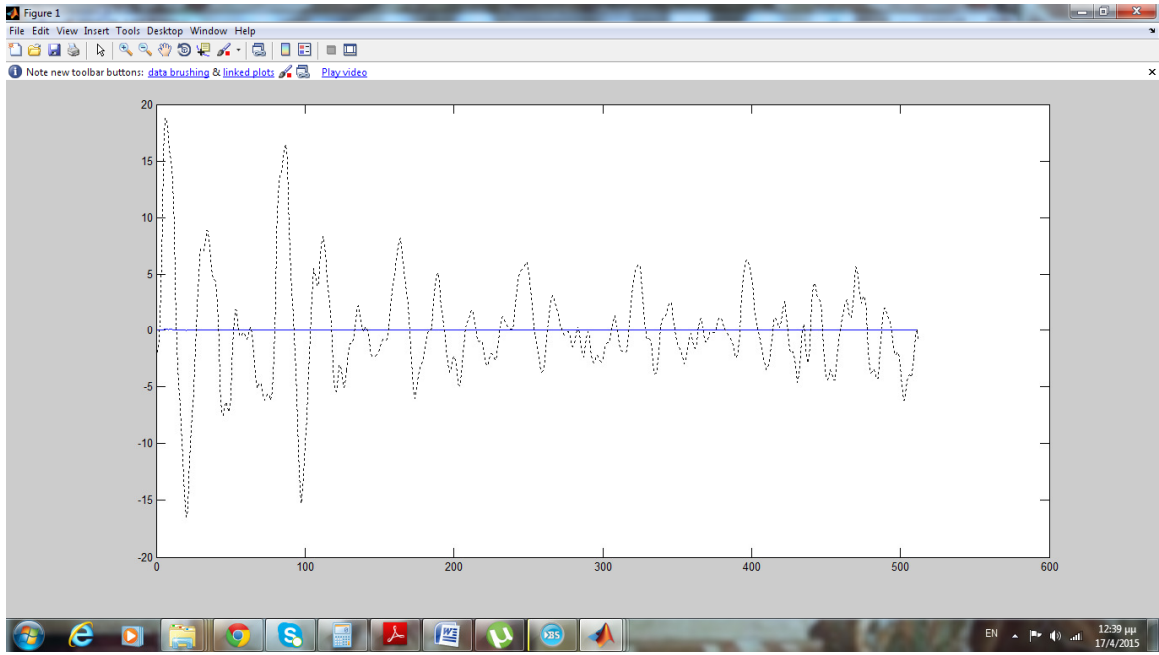


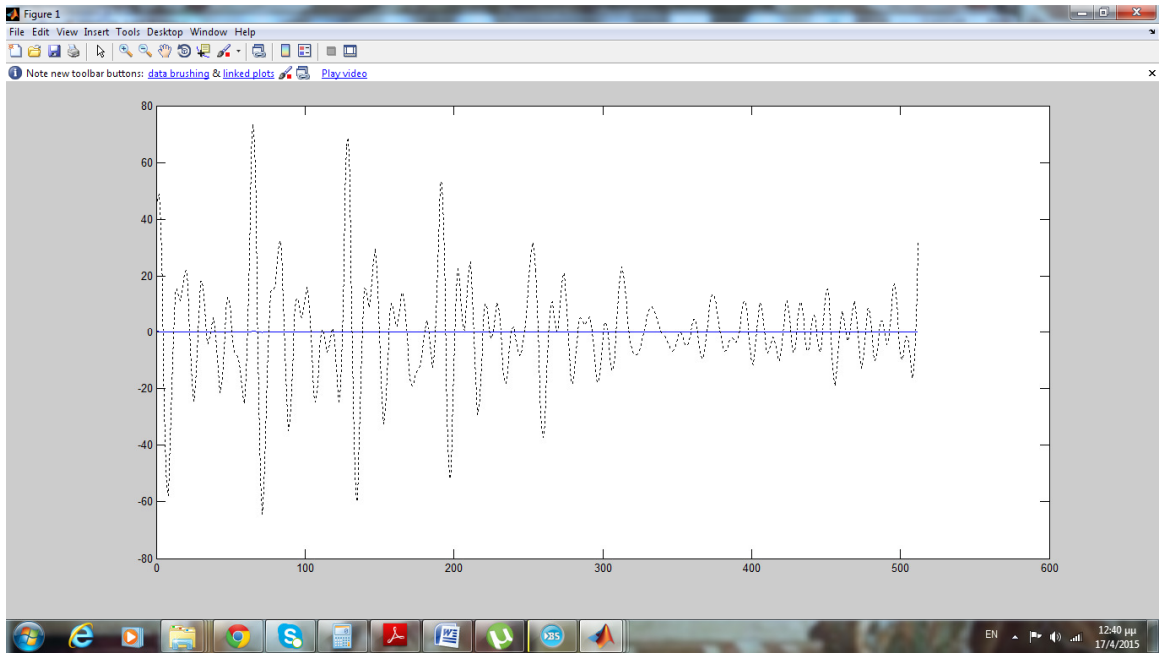
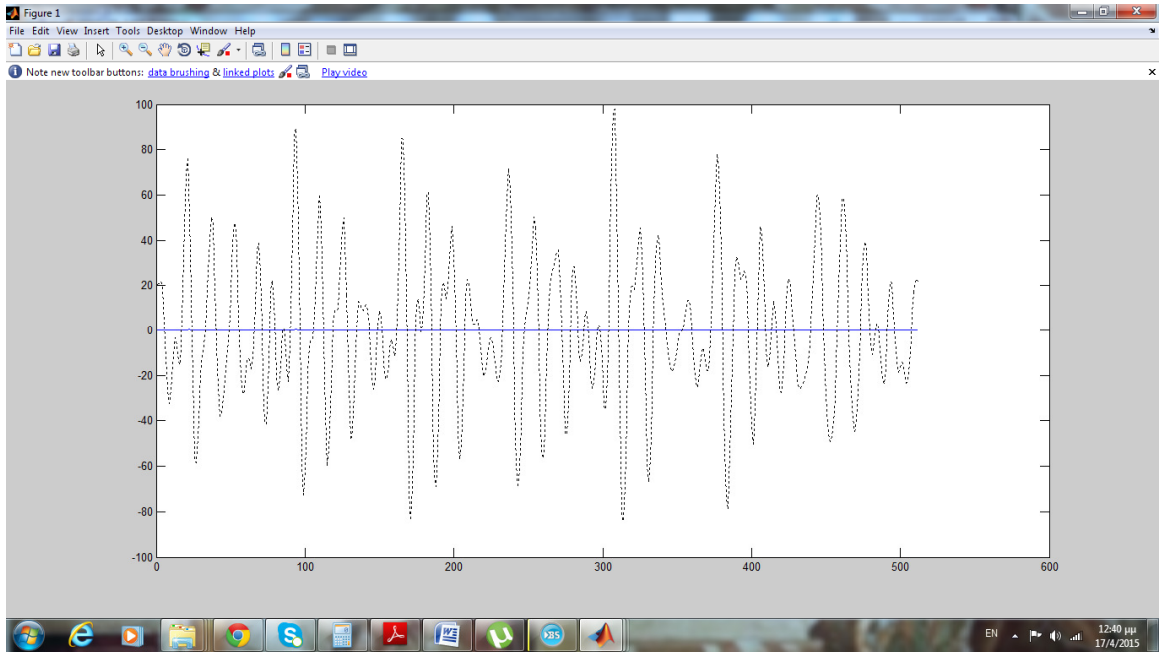


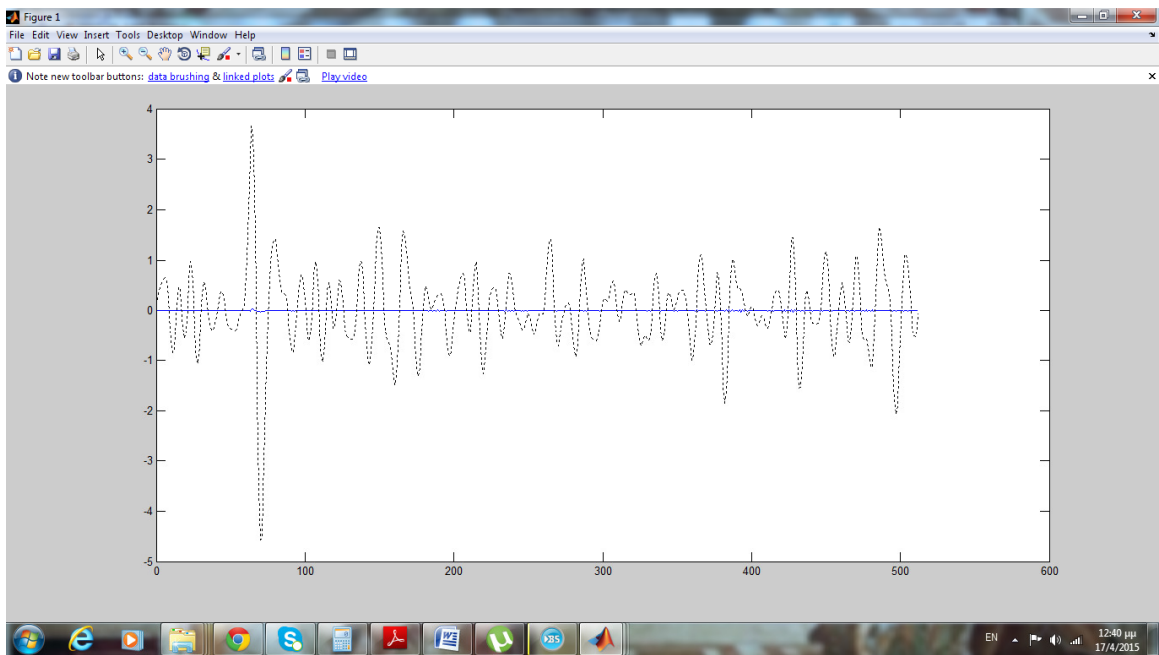
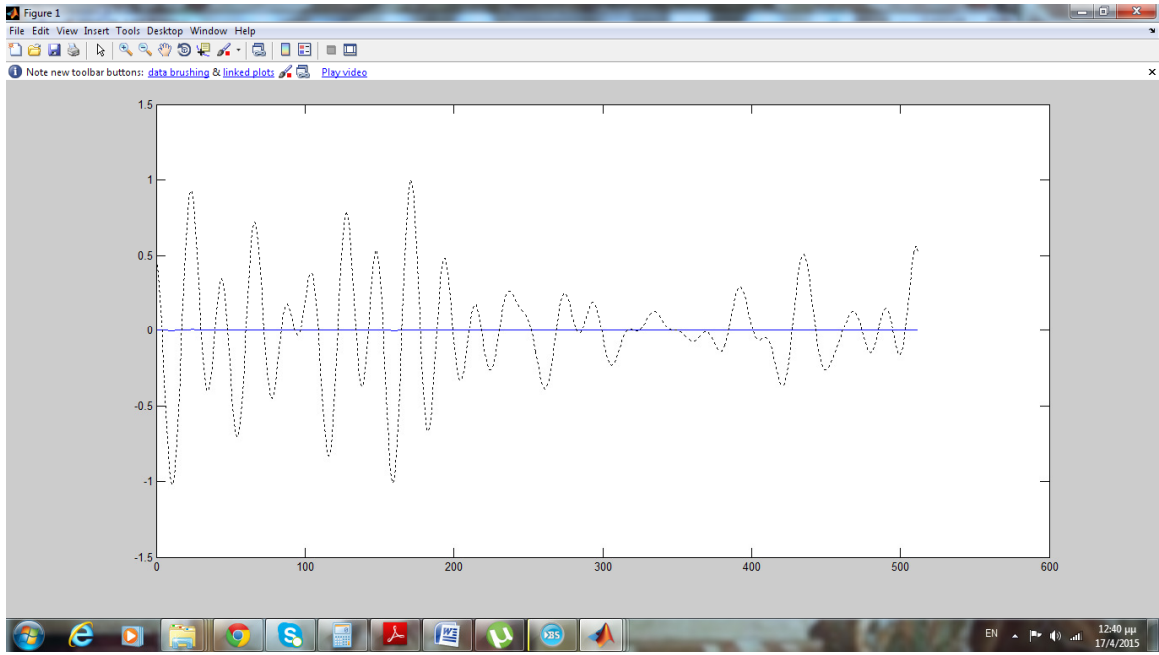














## **ΒΙΒΛΙΟΓΡΑΦΙΑ**

### **Ξενόγλωσση Βιβλιογραφία**

[1] Claude E. Shannon, A mathematical theory of communication, Bell System Technical Journal, 1948.

[2] Haykin S. Communication Systems , Wiley

[3] John Proakis. Digital Communications , Kindle Edition

### **Ελληνόγλωσση Βιβλιογραφία**

[4] Δημήτριος Αλέξανδρος Τουμπακάρης, Θεωρία Πληροφορίας, Σημειώσεις ομότιτλου μαθήματος, Πάτρα, 2010

[5] Ιστοσελίδα Wikipedia, <http://www.wikipedia.org>

[6] Ιστοσελίδα Matlab, <http://www.mathworks.org>

[7] Ιστοσελίδα [http://www.inf.teilam.gr/OLD/course\\_material](http://www.inf.teilam.gr/OLD/course_material)

[8] Ιστοσελίδα [www.ece.upatras.gr/gr/personnel/faculty.html?id=180](http://www.ece.upatras.gr/gr/personnel/faculty.html?id=180)

[9] Ιστοσελίδα [www.csd.uoc.gr/~hy430/HY430-2004\\_1.ppt](http://www.csd.uoc.gr/~hy430/HY430-2004_1.ppt)

[10] Σ. Λούβρος, Ι. Κούγιας. Το Δίκτυο Κινητής τηλεφωνίας GSM, Εκδόσεις Νέων Τεχνολογιών , 2010 . Ανακτήθηκε 02/06/2015 < <http://newtech-pub.com/wp-content/uploads/2013/10/%CE%9A%CE%B5%CF%86-GSM.pdf>>

[11] Κωνσταντίνου , Καψάλης, Κωττής , Εισαγωγή στις Τηλεπικοινωνίες, Εκδόσεις Παπασωτηρίου , 1995

[12] Παρασκευάς Μ , Μουρτζόπουλος Γ . Συμπύεση Ηχητικών Σημάτων από την ADPCM στο MPEG-4, Ομάδα Ήχου, Εργαστήριο Ενσύρματης Τηλεπικοινωνίας Τμήμα Ηλεκτρολόγων Μηχανικών & Τεχνολογίας Υπολογιστών Πανεπιστήμιο Πατρών