ΚΩΔΙΚΟΠΟΙΗΣΗ ΦΩΝΗΣ – ΚΩΔΙΚΟΠΟΙΗΤΕΣ ΚΥΜΑΤΟΜΟΡΦΗΣ

του

Γουμενίδη Θεόδωρου

Α.Τ.Ε.Ι. ΚΡΗΤΗΣ/ΠΑΡΑΡΤΗΜΑ ΧΑΝΙΩΝ/ΤΜΗΜΑ ΤΗΛΕΠΙΚΟΙΝΩΝΙΕΣ ΚΑΙ ΔΙΚΤΥΑ Η/Υ (Π.Σ.Ε.)

Φεβρουάριος 2004

Επιβλέπων Καθηγητής: Δρ. Γλεντής Γιώργος

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

ΕΙΣΑΓΩΓΗ

Στην παρούσα εργασία προτίθεμαι να παρουσιάσω μια μελέτη γύρω από την κωδικοποίηση φωνής και πιο συγκεκριμένα τους *κωδικοποιητές κυματομορφής*. Ποίες είναι όμως οι βασικές αρχές στις οποίες βασίζονται αυτοί οι κωδικοποιητές, πως λειτουργούν και σε ποίες κατηγορίες χωρίζονται; Όλα αυτά εξετάζονται εδώ.

Μπορούμε να θεωρήσουμε ότι η εργασία χωρίζεται σε τρία μέρη. Το πρώτο μέρος περιλαμβάνει τα τρία πρώτα κεφάλαια. Σε αυτά εξετάζουμε γενικά τους κωδικοποιητές φωνής, τον τρόπο με τον οποίο παράγεται η φωνή καθώς και τις ιδιότητες που αυτή παρουσιάζει. Επίσης, μελετάμε τα εργαλεία που έχουμε στην διάθεση μας για την αξιολόγηση των κωδικοποιητών και εξετάζουμε συνοπτικά και τις άλλες κατηγορίες εκτός από τους *κωδικοποιητές κυματομορφής*.

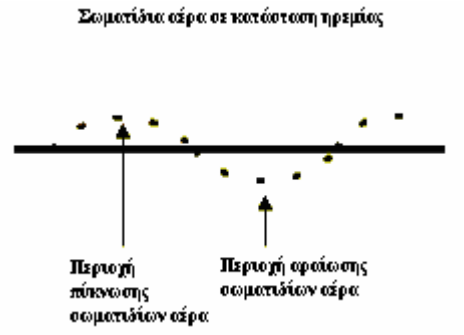
Το δεύτερο μέρος αποτελείται από τα Κεφάλαια 4 και 5. Επικεντρώνεται στους *κωδικοποιητές κυματομορφής* οι οποίοι ταξινομούνται ανάλογα με το πεδίο στο οποίο λειτουργούν. Η μελέτη τους ξεκινά από τον πιο απλό, όπως είναι το PCM και καταλήγει στους πλέον πολύπλοκους οι οποίοι μπορεί να περιλαμβάνουν προγνωστικές τεχνικές αλλά και τεχνικές μετασχηματισμών.

Στο τέλος ακολουθεί το τρίτο μέρος το οποίο περιλαμβάνει το $6^{ο}$ Κεφάλαιο. Σε αυτό γίνεται υλοποίηση επιλεγμένων κωδικοποιητών (PCM, m – law, DCPM) μέσα από το Matlab. Για τον λόγο αυτό έχει αναπτυχθεί και ένα αλληλεπιδραστικό γραφικό περιβάλλον το οποίο και μας επιτρέπει τον ευκολότερο χειρισμό των παραμέτρων αυτών των κωδικοποιητών καθώς επίσης και άμεση εξέταση των αποτελεσμάτων που μας δίνουν. Το περιβάλλον εμφανίζεται πληκτρολογώντας *guip1* στην prompt γραμμή του Matlab και βρίσκεται μαζί με τις συναρτήσεις των κωδικοποιητών με την μορφή m – file στην συνοδευτική δισκέτα στο οπισθόφυλλο της εργασίας.

ΚΕΦΑΛΑΙΟ 1

1.1 Η Ομιλία σαν Ηχητικό Σήμα

Ο ήχος προσδιορίζεται σαν η μεταβολή της πίεσης ή της ταχύτητας των σωματιδίων ενός ελαστικού μέσου, η οποία διαδίδεται κυματικά εντός του μέσου αυτού. Η παραγωγή και η διάδοση από ένα σώμα ή μια δονούμενη μάζα των ακουστικών κυμάτων εδώ θεωρείται ότι γίνεται μέσα στον αέρα και σε αυτήν την περίπτωση διαδίδονται με διαμήκη κύματα[1]. Το δονούμενο σώμα προκαλεί μια αναταραχή των σωματιδίων του αέρα με αποτέλεσμα την δημιουργία μεταβολών στην πυκνότητα και την πίεση του αέρα (Σχήμα 1.1).



Σχήμα 1.1: Διάταξη σωματιδίων ελαστικού μέσου σε διαμήκες κύμα.

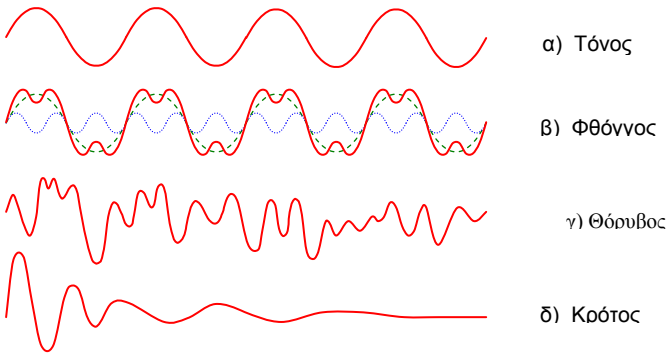
Οι ήχοι τώρα διακρίνονται στις εξής κατηγορίες:

Απλοί ήχοι ή τόνοι: Με τον όρο αυτό καθορίζεται ένα ηχητικό σήμα μιας ορισμένης συχνότητας.

Σύνθετοι ήχοι ή φθόγγοι: Με τον όρο αυτό καθορίζεται ένα ηχητικό σήμα όχι μιας συχνότητας αλλά ενός πλήθους συχνοτήτων. Σε αυτή την κατηγορία ήχων ανήκει και η ανθρώπινη ομιλία. Οι σύνθετοι ήχοι περιέχουν άπειρο αριθμό απλών τόνων, των οποίων οι συχνότητες είναι ακέραια πολλαπλάσια (αρμονικές) μιας θεμελιώδους συχνότητας.

Θόρυβοι: Είναι ηχητικά σήματα που αντιστοιχούν σε ακανόνιστα κύματα και δεν παρουσιάζουν καμία περιοδικότητα.

Κρότοι: Με τον όρο αυτό καθορίζονται τα ηχητικά σήματα των οποίων το πλάτος από μια πολύ μεγάλη τιμή ελαττώνεται απότομα σε ελάχιστο χρονικό διάστημα. (Σχήμα 1.2).



Σχήμα 1.2: Απλοί και σύνθετοι ήχοι.

[1] *Διαμήκη κύματα* είναι εκείνα στα οποία η διαδιδόμενη αναταραχή προκαλεί κίνηση των σωματιδίων του ελαστικού μέσου κατά μήκος της διεύθυνσης μετάδοσης του κύματος.

Κάθε ήχος τώρα παρουσιάζει μια σειρά από χαρακτηριστικά τα οποία και διακρίνονται σε υποκειμενικά και αντικειμενικά. Αντικειμενικά χαρακτηριστικά είναι εκείνα τα οποία βασίζονται σε επιστημονικές μετρήσεις και υποκειμενικά είναι εκείνα τα οποία βασίζονται στον τρόπο αντίληψης του ήχου με το αισθητήριο της ακοής. Τα αντικειμενικά χαρακτηριστικά είναι τα εξής:

Η ένταση: Εξαρτάται από το πλάτος των ταλαντώσεων του ήχου. Αν το πλάτος των ταλαντώσεων αυξάνει τότε λέμε ότι ο ήχος ακούγεται δυνατότερα.

Η συχνότητα: Είναι το πλήθος των ταλαντώσεων που εκτελούν τα μόρια του υλικού μέσα σε ένα δευτερόλεπτο.

Το φασματικό περιεχόμενο: Εκφράζει το πλήθος και τη σχετική ένταση των απλών ήχων που απαρτίζουν ένα σύνθετο ήχο.

Τα υποκειμενικά χαρακτηριστικά είναι τα εξής:

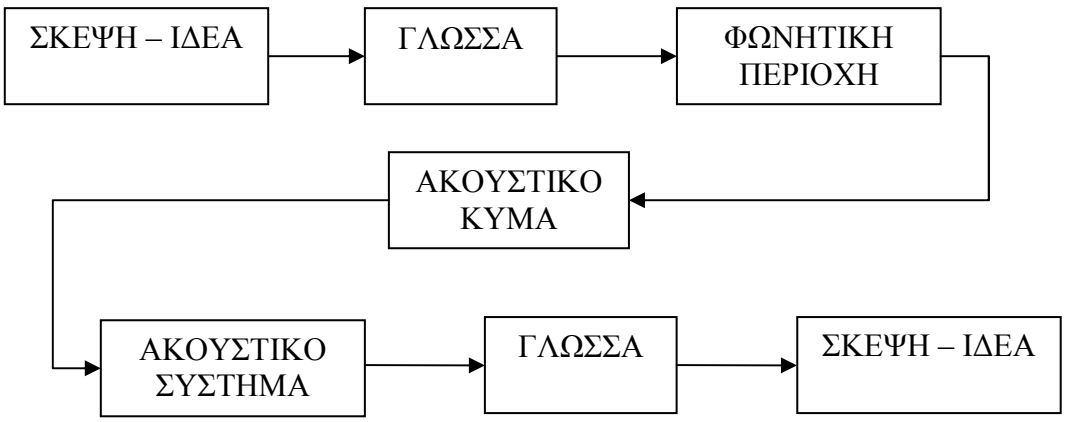
Η ακουστότητα: Είναι το γνώρισμα εκείνο το οποίο μας επιτρέπει να χαρακτηρίσουμε έναν ήχο ισχυρό ή αδύναμο. Εξαρτάται κυρίως από την ένταση του ήχου αλλά και από την συχνότητα. Έτσι ήχοι με σταθερή ένταση που έχουν χαμηλή ή υψηλή συχνότητα (π.χ. 100Hz ή 10.000Hz) ακούγονται με μικρότερη ακουστότητα απ' ότι ήχοι με ενδιάμεση συχνότητα (π.χ. 1.000Hz).

Το ύψος: Το γνώρισμα αυτό μας επιτρέπει να χαρακτηρίσουμε έναν ήχο οξύ ή βαρύ. Εξαρτάται κυρίως από την συχνότητα αλλά δευτερευόντως και από την ένταση. Έτσι ήχοι ίδιας συχνότητας ακούγονται λιγότερο οξείς όσο μεγαλώνει η ένταση τους.

Η χροιά: Το γνώρισμα αυτό μας επιτρέπει να ξεχωρίσουμε δύο ήχους με το ίδιο ύψος και ακουστότητα, οι οποίοι παράγονται από δύο διαφορετικές πηγές. Εξαρτάται κυρίως από το φασματικό περιεχόμενο αλλά και από την ένταση του.

1.2 Η Ομιλία σαν Μέσο Επικοινωνίας

Σε γενικές γραμμές, η επικοινωνία με ομιλία είναι η διαδικασία εκείνη όπου προφορικά μεταδίδεται μια ιδέα - μήνυμα από ένα ¨πομπό¨ σε έναν ¨δέκτη¨. Η διαδικασία υλοποίησης αυτής της επικοινωνίας ξεκινά με την μετατροπή της σκέψης – ιδέας σε λογική πρόταση (proposition)[2] η οποία στην συνέχεια μετατρέπεται σε ¨μυϊκές εκφράσεις¨ της φωνητικής περιοχής, του λάρυγγα και των πνευμόνων. Η φωνητική περιοχή είναι αυτή στην οποία η λογική πρόταση γίνεται ακουστικό κύμα με την πίεση του αέρα. Παράγονται έτσι *φωνήματα*[3] και *φθόγγοι*. Στη συνέχεια αυτά μεταδίδεται στο ακουστικό σύστημα του ακροατή όπου γίνεται αντιληπτή η ομιλία. Χρησιμοποιώντας της γνωστικές πηγές και γνωρίζοντας την γλώσσα ο ακροατής επεξεργάζεται, κατανοεί την λογική πρόταση και εξάγει το συμπέρασμα.



Σχήμα 1.3: Σχηματικό διάγραμμα μηχανισμού ομιλίας σαν μέσο επικοινωνίας.

[2] Λογικές προτάσεις σύμφωνα με τους γλωσσολόγους ονομάζονται οι προτάσεις που περιέχουν μια περιγραφή πραγμάτων ή γεγονότων.

[3] Σαν *φώνημα* καθορίζεται η ελάχιστη μονάδα που διαφοροποιεί το νοηματικό περιεχόμενο της ομιλίας στο ηχητικό σύστημα μιας γλώσσας.

1.3 Ανάπτυξη της Επεξεργασίας της Ομιλίας

Η τεράστια ανάπτυξη της τεχνολογίας στις μέρες μας δημιούργησε την ανάγκη της αποτελεσματικής επεξεργασίας των σημάτων που διακινούταν μέσα από τα διάφορα κανάλια μετάδοσης. Έτσι λοιπόν πέρα από την αρχική αναλογική μετάδοση η οποία παρουσίαζε πολλά μειονεκτήματα αναπτύχθηκε η ψηφιακή μετάδοση η οποία προσέφερε νέες δυνατότητες στον τομέα αυτό.

Η ανάπτυξη όμως αυτής της ψηφιακής μετάδοσης βασίστηκε κατά ένα μεγάλο ποσοστό στην ψηφιακή επεξεργασία σήματος (DSP). Ένας από τους πιο ολοκληρωμένα μελετημένους και ώριμους κλάδους της είναι η κωδικοποίηση φωνής. Αυτό έγινε για τρεις κυρίους λόγους. Ο πρώτος λόγος είναι ότι το σήμα της φωνής είναι ένα σήμα με πολύ μικρό εύρος ζώνης. Έτσι τα τηλεφωνικής ποιότητας σήματα έχουν εύρος ζώνης μόνο 3.2kHz ενώ τα υψηλής ποιότητας μπορεί να φτάσουν μέχρι 5 – 6kHz οπότε με μικρής συχνότητας δειγματοληψία μπορούμε να τα επεξεργαστούμε σε πραγματικό χρόνο πολύ αποτελεσματικά.

Ένας δεύτερος παράγοντας είναι η ραγδαία εξέλιξη της τεχνολογίας των VLSI (very large – scale integrated) κυκλωμάτων η οποία μας έδωσε την δυνατότητα να υλοποιήσουμε ισχυρότατους επεξεργαστές σε πολύ μικρό μέγεθος και με πολύ μικρό κόστος. Αυτό είχε σαν αποτέλεσμα την ευρεία εξάπλωση και χρησιμοποίηση τους στον τομέα της κωδικοποίησης της ομιλίας.

Και τέλος ο τρίτος παράγοντας είναι η αποτελεσματικότητα που παρουσίασαν οι διάφοροι αλγόριθμοι DSP στην αντιμετώπιση των βασικών προβλημάτων που εμφανίζουν τα συστήματα κωδικοποίησης φωνής. Η τεχνικές λοιπόν ψηφιακής επεξεργασίας σήματος αποδείχθηκαν πολύ αποτελεσματικές στην μοντελοποίηση του τρόπου παραγωγής και αντίληψης της ομιλίας.

1.4 Στόχος των Κωδικοποιητών Φωνής

Ο στόχος όλων των κωδικοποιητών φωνής είναι να αναπαραστήσουν την ομιλία σε ψηφιακή μορφή με την μεγαλύτερη δυνατή ποιότητα και με τον μικρότερο δυνατό αριθμό bits . Οι περισσότεροι από τους κωδικοποιητές φωνής βασίζονται σε αλγόριθμους συμπίεσης με απώλειες (lossy algorithms) όπου το σημασιολογικό περιεχόμενο ουσιαστικά δεν μεταβάλλεται αλλά υπεισέρχεται η έννοια της μείωσης της ποιότητας. Αυτοί οι αλγόριθμοι βέβαια είναι αποδεκτοί στην κωδικοποίηση ομιλίας επειδή η μείωση της ποιότητας δεν γίνεται αντιληπτή από το ακουστικό σύστημα του ανθρώπου Εδώ πρέπει να σημειωθεί ότι υπάρχει ένα κατώτερο όριο στον ρυθμό των bits με το οποίο μπορεί να κωδικοποιηθεί η ομιλία και το οποίο καθορίζεται τόσο από τον ρυθμό της *φωνημικής πληροφορίας*, ο οποίος έχει υπολογιστεί περίπου στα 50 bits ανά δευτερόλεπτο όσο και από τον συνολικό ρυθμό *γνωστικής πληροφορίας* της ομιλίας και ο οποίος έχει υπολογιστεί περίπου στα 400 bits ανά δευτερόλεπτο. Ο ρυθμός της *φωνημικής πληροφορίας* υπολογίζεται τόσο με βάση τους φυσικούς περιορισμούς των αρθρωτών του ανθρώπου όσο και με βάση τον αριθμό των *φωνημάτων* κάθε γλώσσας. Ο δε ρυθμός της *γνωστικής πληροφορίας* υπολογίζεται με βάση άλλους παράγοντες όπως είναι η ταυτότητα, η συναισθηματική κατάσταση του ομιλητή, η ταχύτητα με την οποία μιλά, το πόσο δυνατά μιλά κ.α.

Βέβαια ακόμα και αν είχε επιτευχθεί η υλοποίηση αυτού του ιδανικού κωδικοποιητή στα 400 bits ανά δευτερόλεπτο με ιδανική ποιότητα θα ήταν δύσκολο να επιλεγεί στην πράξη σε σχέση με άλλους όχι και τόσο καλούς κωδικοποιητές. Αυτό θα γινόταν επειδή τέτοιου είδους κωδικοποιητές είναι ιδιαίτερα ευαίσθητοι σε μη ακουστικά σήματα, σε ακουστικά σήματα με θόρυβο και στην ύπαρξη πολλαπλών ομιλητών. Επιπρόσθετα είναι μη – ανθεκτικοί και σε τυχόν λάθη του καναλιού μετάδοσης και εισάγεται μεγάλη καθυστέρηση κατά την επεξεργασία της φωνής από αυτούς.

Επίσης πρέπει να επισημάνουμε ότι υπάρχουν διαφορετικές απαιτήσεις από τους κωδικοποιητές όταν αυτοί πρόκειται να χρησιμοποιηθούν για μετάδοση ή αποθήκευση. Έτσι στην μετάδοση πρέπει η καθυστέρηση που εισάγουν να είναι ιδιαίτερα μικρή ιδίως όταν σε αυτήν υπάρχουν διάφορες άλλες επιπρόσθετες καθυστερήσεις. Επίσης στις μεταδόσεις συνήθως υπάρχουν

και μηχανισμοί που ανιχνεύουν και διορθώνουν τα λανθασμένα bits οι οποίοι και καταλαμβάνουν ένα μέρος του bit rate με αποτέλεσμα να αφήνονται ακόμα λιγότερα bits για χρησιμοποιηθούν από τους κωδικοποιητές.

Στις εφαρμογές αποθήκευσης τώρα η παράμετρος της καθυστέρησης είναι μικρής σημασίας και συνήθως αυτές είναι σε λιγότερο θορυβώδη περιβάλλοντα. Έτσι υπάρχει η δυνατότητα να μην έχουμε πρόγνωση και διόρθωση σφαλμάτων κάτι που σαφώς ευνοεί τον ρυθμό μετάδοσης.

Όλα τα παραπάνω βέβαια είναι μερικοί μόνο από τους παράγοντες με βάση τους οποίους αποτιμούμε του κωδικοποιητές φωνής και οι οποίοι θα παρουσιαστούν στο σύνολο τους σε επόμενη παράγραφο.

1.5 Μοντέλα στην Κωδικοποίηση Φωνής

Η κωδικοποίηση φωνής βασίζεται σε ορισμένους βασικούς μηχανισμούς (μοντέλα) παραγωγής της ομιλίας οι οποίοι κατατάσσονται ως εξής:

Γνωστικό Μοντέλο: Εδώ βασιζόμαστε στις αρχικές διεργασίες που λαμβάνουν χώρα τον εγκέφαλο για την παραγωγή της ομιλίας. Η χρήση του είναι περιορισμένη και η ερεύνα είναι ακόμα σε αρχικό στάδιο.

Γλωσσικό Μοντέλο: Εδώ βασιζόμαστε στη θέση ότι ο προφορικός λόγος παρουσιάζει μια ακριβή και ολοκληρωμένη συντακτική δόμηση η οποία σχετίζεται με ένα μοτίβο συλλαβικής έντασης. Σε αυτόν, υπάρχουν πληροφορίες όχι μόνο για το νοηματικό περιεχόμενο αλλά και για την στάση του ομιλητή απέναντι σε αυτό. Οι κωδικοποιητές που βασίζονται στο μοντέλο αυτό παράγουν ομιλία χαμηλής ποιότητας, ωστόσο μας δίνουν την δυνατότητα της απομόνωσης των άϋλων στοιχείων της γλώσσας (π.χ. νόημα της πρότασης, στάση του ομιλητή κτλ) έτσι ώστε να μπορούμε να εστιάζουμε αποκλειστικά στα χαρακτηριστικά της κυματομορφής του λόγου που μας ενδιαφέρουν και που πρέπει να διατηρηθούν.

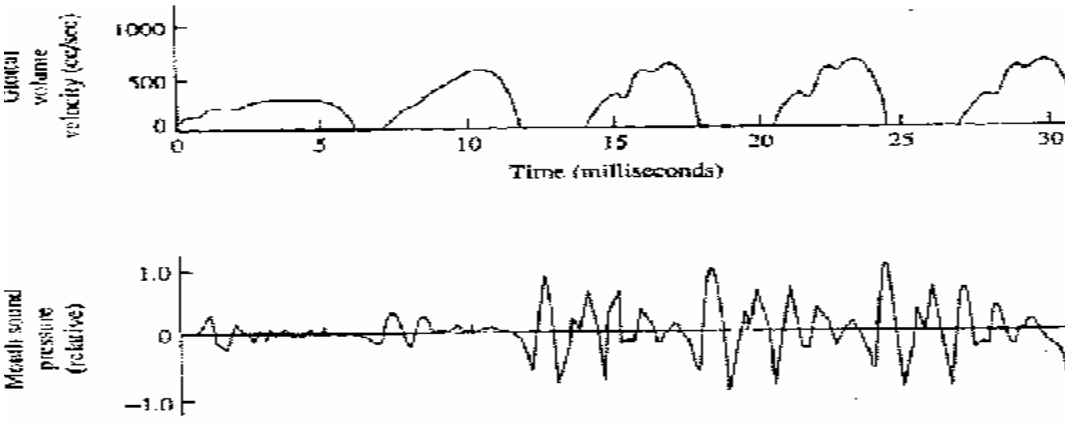
Μοντέλο Φωνητικού Σωλήνα: Εδώ βασιζόμαστε στον τρόπο με τον οποίο λειτουργεί η φωνητική περιοχή του ανθρώπου και στους βασικούς ήχους που αυτή παράγει, *έμφωνα* – άφωνα. Οι κωδικοποιητές που λειτουργούν με βάση αυτό το μοντέλο προσπαθούν να μοντελοποιήσουν τη διεργασία παραγωγής της φωνής με ένα δυναμικό σύστημα και επιπλέον προσπαθούν να ποσοτικοποιήσουν συγκεκριμένους περιορισμούς γι' αυτό το σύστημα. Βασικές λειτουργίες τους είναι οι εξής: Πραγματοποιούν ανάλυση του σήματος φωνής στον πομπό, μεταδίδουν τις παραμέτρους που προκύπτουν από την ανάλυση και έπειτα χρησιμοποιώντας τις παραμέτρους αυτές συνθέτουν τη φωνή στο δέκτη. Κατά κανόνα αποδίδουν μικρό ρυθμό bits με όχι όμως απαραίτητα καλή ποιότητα.

Λίγο πιο αναλυτικά μπορούμε να πούμε ότι σε αυτό το μοντέλο ¨αναπαράγονται¨ *η φωνητική περιοχή* του ανθρώπινου συστήματος, *η ακτινοβολία του ήχου στα χείλη* καθώς και *η πηγή διέγερσης*. Πιο συγκεκριμένα η φωνητική περιοχή εξομοιώνεται με ένα διακριτού χρόνου μοντέλο απωλεστικού σωλήνα του οποίου η συνάρτηση μεταφοράς δίνεται από την εξίσωση

$V(z)=\dfrac{G}{1-\sum_{k=1}^{N} a_k z^{-k}}$ όπου G είναι το κέρδος και z^{-k} η καθυστέρηση. Σε αυτή την περίπτωση οι συντονισμοί του σήματος της φωνής αντιστοιχούν στους πόλους της συνάρτησης $V(z)$ ωστόσο για μια πληρέστερη αναπαράσταση απαιτείται η ύπαρξη πόλων και μηδενικών. Βασιζόμενοι όμως σε ένα θεώρημα του Atal[4] μπορούμε να πούμε ότι όμοια με την επίδραση των μηδενικών στην συνάρτηση μεταφοράς μπορεί να επιτευχθεί και με την εισαγωγή επιπλέον πόλων. Το μοντέλο που προκύπτει με αυτό τον τρόπο είναι γνωστό ως *autoregressive (AR)* και για να είναι σταθερό πρέπει όλοι οι πόλοι να βρίσκονται μέσα στον μοναδιαίο κύκλο. Η υλοποίηση του τώρα μπορεί να γίνει είτε παράλληλα, είτε διαδοχικά, είτε με διάφορες άλλες υλοποιήσεις οι οποίες ισχύουν στα

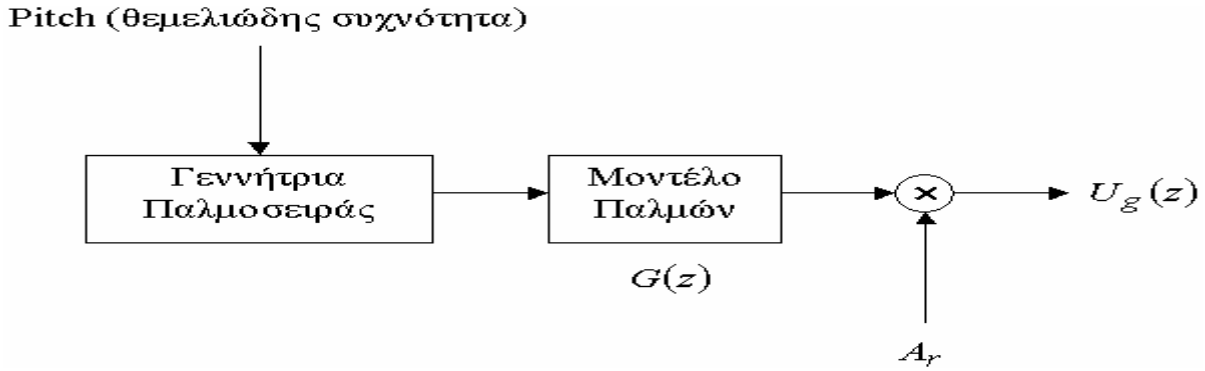
[4] B. S. Atal, and S. L. Hanauer, ¨Speech Analysis And Synthesis By Linear Prediction Of The Speech Wave¨, *J. Acoust. Soc. Am.,* Vol 50, No 2, pp 637 – 655.

ψηφιακά φίλτρα. Μια συνήθη μορφή είναι η $V(z)=\dfrac{1}{\prod_{k=1}^{p}(1-c_k z^{-1})}$. Πέρα όμως από την εξομοίωση της φωνητικής περιοχής έχουμε την εξομοίωση τόσο της ακτινοβολίας του ήχου στα χείλη όσο και της πηγής διέγερσης. Για την πρώτη από αυτές χρησιμοποιείται ένα υψιπερατό φίλτρο πρώτης τάξης της μορφής $R(z)=R_0(1-z^{-1})$. Για την περίπτωση της πηγής διέγερσης τώρα επειδή το σήμα της ομιλίας όπως θα δούμε και παρακάτω ταξινομείται σε *έμφωνα* και *άφωνα* τμήματα πρέπει η διέγερση μας να μπορεί να παράγει τόσο ημι – περιοδικούς παλμούς οι οποίοι αντιστοιχούν στα *έμφωνα*, όσο και τυχαίο θόρυβο ο οποίος και αντιστοιχεί στα άφωνα.



Σχήμα 1.4: Διέγερση σε έμφωνους ήχους

Πιο συγκεκριμένα στην περίπτωση των *έμφωνων* ήχων, η διέγερση πρέπει να μοιάζει με τη πρώτη κυματομορφή του Σχήματος 1.4. Για να ικανοποιήσουμε την απαίτηση αυτή χρησιμοποιούμε το συνδυασμό του Σχήματος 1.5. Σύμφωνα με αυτό η γεννήτρια παλμοσειράς παράγει μοναδιαίους παλμούς με βάση την θεμελιώδη συχνότητα (pitch) της ομιλίας. Στη συνέχεια το σήμα αυτό διεγείρει ένα γραμμικό σύστημα $G(z)$ του οποίου η απόκριση έχει την επιθυμητή μορφή (ίδια μορφή με εκείνη του σήματος που παράγει ο ανθρώπινος μηχανισμός), ενώ η ένταση της διέγερσης καθορίζεται από το κέρδος A_γ.



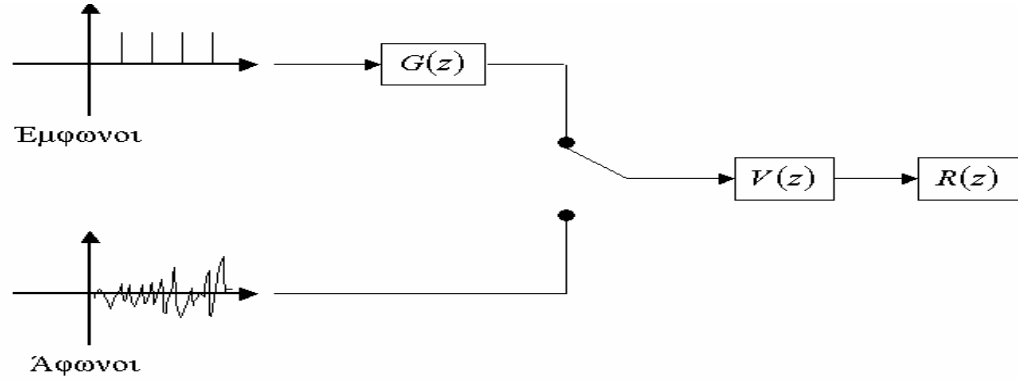
Σχήμα 1.5: Παραγωγή του σήματος διέγερσης για έμφωνους ήχους

Για το $g(n)$ όμως έχει βρεθεί ότι η επιλογή της μορφής του δεν είναι τόσο κρίσιμη όσο οι ιδιότητες του μετασχηματισμού Fourier του. Έτσι αυτό μπορεί να αντικαταστήσει με επιτυχία την ανθρώπινη *πηγή παλμών* όταν ισχύει:

$$g(n) = \frac{1}{2}[1 - \cos(\pi n / N_1)] \quad 0 \le n \le N_1$$
$$= \cos(\pi(n - N_1) / 2N_2) \quad N_1 \le n \le N_1 + N_2$$
$$= 0 \text{ αλλιώς}$$

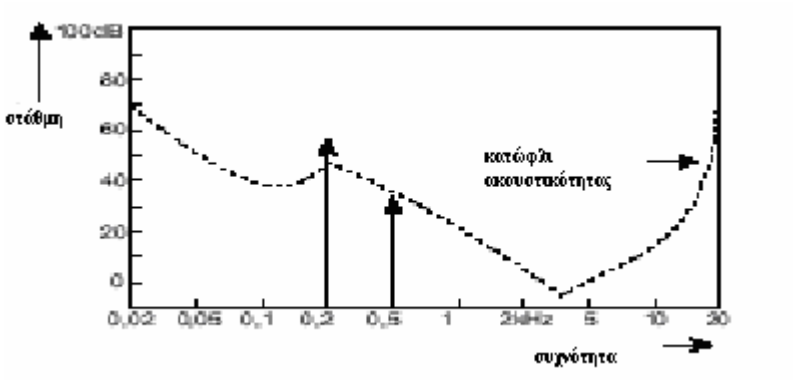
Εκτός όμως του μοντέλου διέγερσης για τα *έμφωνα* τμήματα υπάρχει και το μοντέλο διέγερσης για τα άφωνα τμήματα το οποίο είναι πολύ πιο απλό και αποτελείται από μια πηγή τυχαίου θορύβου και μια παράμετρο κέρδους, ενώ στην περίπτωση του διακριτού χρόνου αποτελείται από μια γεννήτρια τυχαίων αριθμών.

Η χρήση όλων των παραπάνω βέβαια μας δίνει το τελικό αποτέλεσμα που είναι το *συνδυασμένο μοντέλο φωνητικού σωλήνα* το οποίο περιγράφεται μέσα από την εξίσωση $H(z) = G(z)V(z)R(z)$ και φαίνεται στο Σχήμα 1.6.



Σχήμα 1.6: Συνδυασμένο μοντέλο φωνητικού σωλήνα.

Ακουστικό Μοντέλο: Η ανθρώπινη ακοή καθορίζεται από δύο φαινόμενα. Το φαινόμενο του *Ελάχιστου Επιπέδου Ακοής (Κατώφλι Ακοής)* και το φαινόμενο της *Απόκρυψης Πλάτους (Amplitude Masking)*.Η καμπύλη που περιγράφει το *Κατώφλι Ακοής* σε σχέση με τη συχνότητα καθορίζει τον ελάχιστο σε ένταση ήχο που γίνεται αντιληπτός από το ανθρώπινο αυτί σε μια δεδομένη συχνότητα. Το *Κατώφλι Ακοής* ορίζεται σαν 0dB στα 1KHz. Το αυτί είναι περισσότερο ευαίσθητο ανάμεσα στο 1 και 5 KHz. Σε γενικές γραμμές, δύο διαφορετικής συχνότητας τόνοι, ίδιας ισχύος δεν θα γίνουν αντιληπτοί σαν έχοντες ίδια ένταση. Όμοια, η παραμόρφωση και ο θόρυβος δεν γίνονται αντιληπτοί το ίδιο σε όλο το φάσμα συχνοτήτων. Η ευαισθησία μας μειώνεται στις μεγάλες και τις μικρές συχνότητες. Για παράδειγμα, ένας ήχος 20Hz θα πρέπει να είναι 70dB δυνατότερος από ένα ήχο του 1KHz για να γίνει απλά αντιληπτός. Αξιοποιώντας το φαινόμενο αυτό, ένας κωδικοποιητής αναλύει ένα σήμα σε σχέση με το *Κατώφλι Ακοής* σε κάθε συχνότητα και αφαιρεί όλους τους ήχους που θεωρεί πως δεν θα γινόταν αντιληπτοί ούτως ή άλλως. Η *Απόκρυψη Πλάτους* συμβαίνει όταν ένας τόνος μετατοπίζει το *Κατώφλι Ακοής* προς τα πάνω σε ένα φάσμα συχνοτήτων που τους περιβάλλει. Όταν τόνοι ακούγονται συγχρόνως, οι δυνατοί τόνοι αποκρύπτουν πλήρως τους πιο αδύνατους. Για παράδειγμα, ένας τόνος των 200Hz μπορεί να αποκρύψει πλήρως ένα τόνο των 500Hz χαμηλότερης έντασης (Σχήμα 1.7). Με απλά λόγια, η απλή παρουσία ενός ήχου δεν σημαίνει απαραίτητα πως αυτός θα γίνει και ακουστός. Η Απόκρυψη Πλάτους αξιοποιείται για την απαλοιφή του θορύβου κβάντισης.



Σχήμα 1.7: Απόκρυψη Πλάτους: Το υψηλής στάθμης σήμα επικαλύπτει το χαμηλής στάθμης σήμα

Η *Χρονική Απόκρυψη (Temporal Masking)* είναι ένα φαινόμενο που εμφανίζεται όταν ήχοι δημιουργούνται πολύ κοντά χρονικά αλλά όχι συγχρόνως. Ένας δυνατός τόνος που εμφανίζεται είτε μόλις πριν (Pre-Masking ή Backward Masking) είτε αμέσως μετά (Post-Masking ή Forward Masking) από ένα πιο μαλακό τόνο, αποκρύπτει πλήρως τον μαλακό τόνο. Όπως στην περίπτωση της *Απόκρυψης Πλάτους*, το φαινόμενο εντείνεται όσο οι διαφορές σε συχνότητα μικραίνουν, έτσι και στη *Χρονική Απόκρυψη*, το φαινόμενο εντείνεται όσο ο\η χρονική διαφορά στην εμφάνιση των τόνων μικραίνει. Το φαινόμενο της *Χρονικής Απόκρυψης* οδηγεί στην υπόθεση πως ο ανθρώπινος εγκέφαλος ολοκληρώνει τους ήχους που λαμβάνει σε ένα διάστημα χρόνου πριν τους αναλύσει στην ουσία σε "πακέτα". Εναλλακτικά, είναι πιθανόν απλά ο εγκέφαλος να επεξεργάζεται ταχύτερα τους δυνατούς ήχους παρά τους αδύνατους.

1.6 Προσδιορισμός Ενός Κωδικοποιητή Φωνής

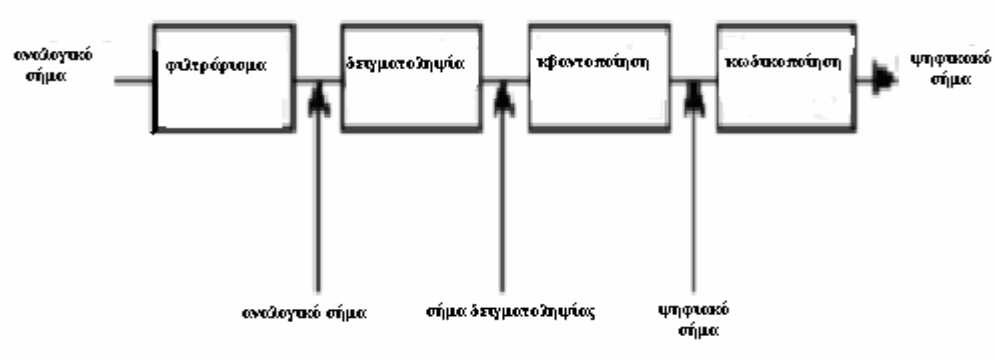
Ένας κωδικοποιητής φωνής συνήθως συντίθεται από μέρη που εκτελούν τρεις βασικές διαδικασίες: την ανάλυση της φωνής, την παραμετρική κβάντιση και την παραμετρική κωδικοποίηση. Κατά την πρώτη διαδικασία σε ορισμένους κωδικοποιητές η φωνή μπορεί να αναλυθεί και να εξαχθούν διάφορες παράμετροι, όπως στο LPC, που την μοντελοποιούν, ενώ σε άλλους, όπως το PCM που θα δούμε παρακάτω, να μην υποστεί καμία επεξεργασία.

Στην συνέχεια μετά την ανάλυση το αποτέλεσμα που προκύπτει θα κβαντιστεί έτσι ώστε το σήμα να πάρει περιορισμένο αριθμό τιμών στάθμης και έτσι να μειωθεί ο αριθμός των bit που απαιτούνται για την αναπαράσταση της φωνής. Η έξοδος μπορεί να θεωρηθεί σαν μια θορυβημένη αναπαράσταση της εισόδου και είναι μια μη αντιστρεπτή διαδικασία. Το σήμα τώρα μετά και από αυτή των διαδικασία θα κωδικοποιηθεί, δηλαδή κάθε στάθμη θα αντιστοιχηθεί με ένα μοναδικό δυαδικό αριθμό. Συνήθως αυτοί οι αριθμοί συνδυάζονται σε πακέτα για πιο αποτελεσματική μετάδοση ή αποθήκευση.

Ο αποκωδικοποιητής φωνής τώρα αντιστρέφει τις λειτουργίες του κωδικοποιητή. Αφού το κωδικοποιημένο σήμα αποκωδικοποιηθεί θα εξαχθούν από αυτό οι στάθμες των παραμέτρων, μέσω του αντιστρόφου του κβαντιστή. Οι στάθμες αυτές, απουσία bit error, θα συντεθούν και θα μας δώσουν το αρχικό σήμα.

1.7 Παλμοκωδική Διαμόρφωση (Pulse Code Modulation)

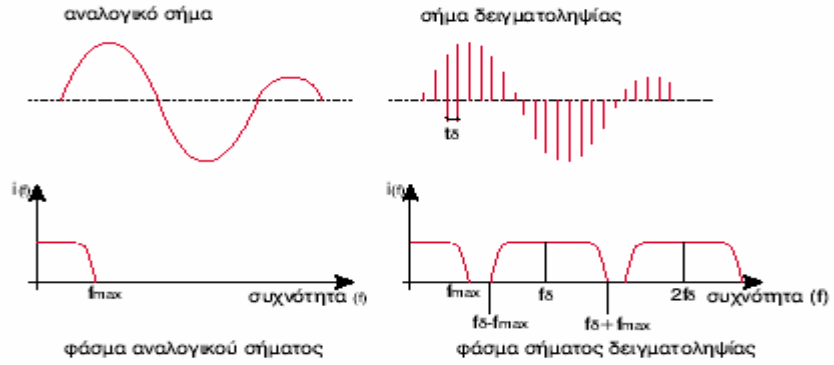
Οι βασικές επεξεργασίες στην παλμοκωδική διαμόρφωση είναι η δειγματοληψία, η κβαντοποίηση και η κωδικοποίηση (Σχήμα 1.8).



Σχήμα 1.8: Διάταξη ενός PCM συστήματος.

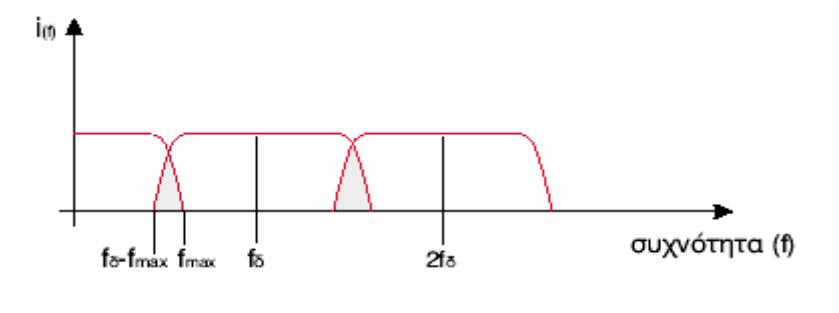
Ας δούμε την κάθε μία από τις διαδικασίες ξεχωριστά:

Δειγματοληψία: Η δειγματοληψία ενός σήματος γίνεται για την μετατροπή ενός συνεχούς σήματος σε διακριτό. Ως γνωστό πρέπει να ισχύει το θεώρημα του Nyquist το οποίο και λέει ότι η συχνότητα δειγματοληψίας πρέπει να είναι τουλάχιστον διπλάσια της μέγιστης συχνότητας της πληροφορίας . Κατά αυτό τον τρόπο η πράξη της δειγματοληψίας αφήνει άθικτο το φάσμα του μηνύματος και απλώς το επαναλαμβάνει περιοδικά στο πεδίο των συχνοτήτων δίνοντας μας έτσι την δυνατότητα να μπορούμε να έχουμε ανασύσταση του σήματος με φιλτράρισμα (Σχήμα 1.9). Το θεώρημα αυτό ισχύει και για σήματα περιορισμένου εύρους ζώνης και για αυτό μπορούμε να το εφαρμόσουμε στο σήμα της φωνής.



Σχήμα 1.9: Φάσμα αναλογικού σήματος και δειγματοληψίας.

Στην πράξη τώρα συναντάμε διάφορα προβλήματα κατά την δειγματοληψία. Έτσι η επίδραση της μη ιδανικότητας των φίλτρων αλλά και το γεγονός ότι σήματα πεπερασμένου χρονικού διαστήματος δεν έχουν απαραίτητα και περιορισμένο εύρος ζώνης έχει ως αποτέλεσμα στο σήμα που έχουμε ανασυστήσει να έχουμε επιπρόσθετα και κάποιες παρασιτικές συνιστώσες με συχνότητα μεγαλύτερη από την μέγιστη συχνότητα του σήματος μας οι οποίες βρίσκονται έξω από τη ζώνη του μηνύματος (Σχήμα 1.10).



Σχήμα 1.10: Φαινόμενο αναδίπλωσης συχνοτήτων (Foldover/Aliasing).

Αυτές πρέπει να είναι σημαντικά υποβιβασμένες έτσι ώστε η παρουσία τους να μην είναι ενοχλητική. Η επίδραση αυτή που ονομάζεται και αλλοίωση – aliasing μπορεί να ελαχιστοποιηθεί περιορίζοντας το εύρος ζώνης του σήματος με ένα φίλτρο πριν τη δειγματοληψία και δειγματοληπτώντας στη συνέχεια με ρυθμό ελαφρά μεγαλύτερο από τον ονομαστικό ρυθμό Nyquist.

Όσο αφορά το σήμα της φωνή τώρα η πληροφορία που αυτό μεταφέρει βρίσκεται στην χαμηλής συχνότητας ζώνη μεταξύ 300 – 3400Hz. Παρόλα αυτά τα συστήματα δειγματοληψίας φωνής συνήθως χρησιμοποιούν φίλτρα με υποβιβασμό 3dB γύρω στα 3.4Hz. Αυτό αναιρεί την ανάγκη για αυστηρά καθορισμένα φίλτρα αντί – aliasing.

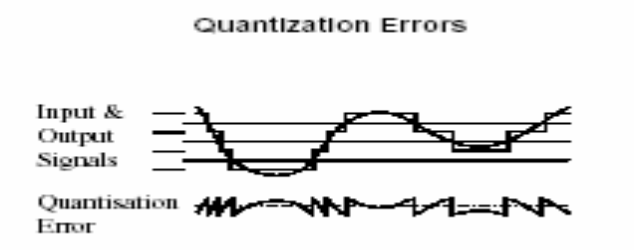
Ομοιόμορφη Κβάντιση: Ως γνωστό με την κβάντιση αναπαριστούμε τις δειγματοληπτημένες τιμές με ένα πεπερασμένο σύνολο σταθμών και με αυτό τον τρόπο μετατρέπεται ένα δείγμα συνεχούς πλάτους σε δείγμα διακριτού πλάτους. Κατά την ομοιόμορφη κβάντιση, η περιοχή των δειγμάτων χωρίζεται σε Q διαστήματα (στάθμες κβάντισης) ίσου πλάτους Δ. Για να πάρουμε συγκεκριμένο αριθμό σταθμών κβάντισης Q χρειαζόμαστε $B = \log_2 Q$ bits και το bit rate για μια συχνότητα δειγματοληψίας f_s είναι $f_s B$. Εάν τώρα η εκάστοτε τιμή πλάτους βρίσκεται στο i *– στο* διάστημα, τότε παίρνουμε σαν κβαντισμένη τιμή το μέσο του διαστήματος. Δηλαδή αν τα α και β είναι τα άκρα του διαστήματος τότε το μήκος του *βήματος κβάντισης (step size)* είναι

$$\Delta = \frac{(\beta - \alpha)}{Q}$$

και η κβαντισμένη έξοδος παράγεται ως εξής σύμφωνα με τα παραπάνω

$$X_q = m_i \text{ αν } x_{i-1} < X \le x_i$$

όπου $x_i = a + i\Delta$ και $m_i = \frac{x_{i-1} + x_i}{2}, i = 1,2,....Q$. Το κβαντισμένο σήμα που προκύπτει είναι μια προσέγγιση του αρχικού και η ακρίβεια της προσέγγισης μπορεί να μεγαλώσει ελαττώνοντας το διάστημα κβάντισης, δηλαδή αυξάνοντας των αριθμό των σταθμών κβάντισης, με αυτόν τον τρόπο όμως αυξάνεται το bit rate του κβαντιστή.



Σχήμα 1.11: Σφάλματα κβάντισης.

Η διαφορά μεταξύ του σήματος εισόδου και του κβαντισμένου σήματος ονομάζεται *σφάλμα ή θόρυβος κβάντισης* (Σχήμα 1.11) και η ισχύς του δίνεται από την σχέση $N = \frac{\Delta^2}{12}$ ενώ η ισχύς του σήματος δίνεται από τον τύπο $S = \frac{A^2}{2}$, όπου A το πλάτος του σήματος εισόδου. Με βάση αυτούς του τύπους ορίζεται ο λόγος SNR με τον οποίο έχουμε μια αποτίμηση του συστήματος κβάντισης και για τον οποίο προκύπτει η σχέση

$$SNR_{(dB)} = 6{,}02B + 1{,}76$$

Επειδή στον ομοιόμορφο κβαντιστή η τιμή του θορύβου κβάντισης είναι σταθερή και ανεξάρτητη από την τιμή του δείγματος που κβαντίζεται, στην περίπτωση όπου το σήμα προς κβάντιση είναι μικρό το πηλίκο SNR θα είναι πολύ μικρό. Για τον λόγο αυτό και επειδή τμήματα μικρού πλάτους είναι πιθανότερο να εμφανιστούν στην ομιλία, για την κβάντιση τέτοιων σημάτων είναι προτιμότερο να χρησιμοποιούμε ανομοιόμορφη κβάντιση.

Ανομοιόμορφη κβάντιση: Ο ανομοιόμορφος κβαντιστής αποδεικνύεται καταλληλότερος για την περίπτωση του σήματος της ομιλίας. Αυτό συμβαίνει γιατί η χρησιμοποίηση μεταβαλλόμενου *βήματος κβάντισης* (Σχήμα 1.12) μας οδηγεί σε καλύτερο λόγο SNR.

Σχήμα 1.12: Ανομοιόμορφη κβάντιση.

Στην πράξη, η ανομοιόμορφη κβάντιση πραγματοποιείται με μια συμπίεση των δειγμάτων μετά την οποία μπαίνει ένας ομοιόμορφος κβαντιστής. Η επιλογή του συμπιεστή (compressor) γίνεται έτσι ώστε η έξοδος του να μας δίνει ένα σήμα με σχετικά ομοιόμορφη κατανομή. Το συμπιεσμένο αυτό σήμα στη συνέχεια κβαντίζεται ομοιόμορφα και μεταδίδεται. Στο δέκτη μετά την αποκωδικοποίηση το σήμα αποσυμπιέζεται (expanded) χρησιμοποιώντας την αντίστροφη διαδικασία.

Οι συμπιεστές/αποσυμπιεστές που χρησιμοποιούνται πιο συχνά κάνουν λογαριθμική συμπίεση. Συνήθως χρησιμοποιούνται δύο νόμοι λογαριθμικής συμπίεσης που λέγονται νόμος συμπίεσης μ και νόμος συμπίεσης A και οδηγούν σε μια μέση ισχύ θορύβου κβάντισης που είναι εντελώς ανεξάρτητη από τη στατιστική του σήματος (στην ουσία είναι ανάλογη προς την στιγμιαία τιμή του δείγματος με αποτέλεσμα να σκεπάζεται η επίδραση του θορύβου κβάντισης).

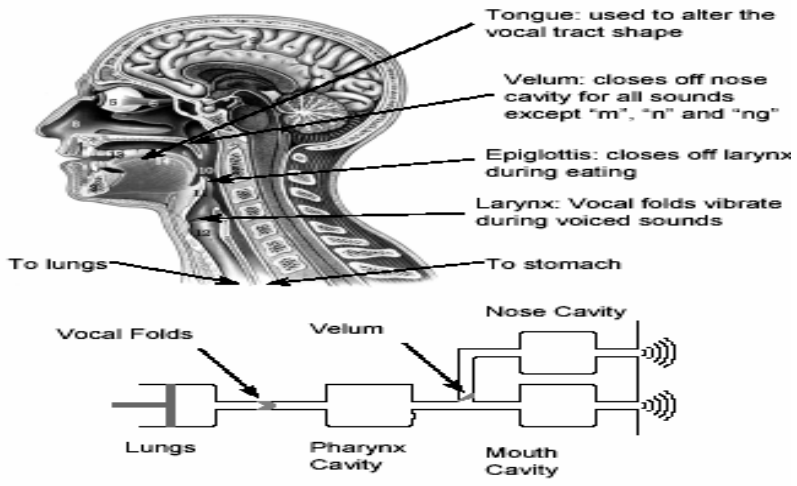
Κωδικοποίηση: Η κωδικοποίηση είναι το τελικό στάδιο της μετατροπής ενός αναλογικού σήματος (όπως είναι το σήμα της ομιλίας) σε ψηφιακό. Σε αυτό το στάδιο αντιστοιχούμε σε κάθε τιμή του δείγματος ένα δυαδικό αριθμό. Εδώ επιγραμματικά να πούμε ότι υπάρχουν δύο είδη δυαδικών κωδικών: οι μονοπολικοί και οι διπολικοί. Οι μονοπολικοί προσδιορίζουν μόνο το μέτρο στάθμης του δείγματος, ενώ οι διπολικοί και το πρόσημο της και για αυτό το λόγο χρησιμοποιούν ένα επιπλέον ψηφίο, το ψηφίο πρόσημου.

ΚΕΦΑΛΑΙΟ 2

2.1 Ανθρώπινος Μηχανισμός Παραγωγής της Ομιλίας

Ανεξάρτητα την ομιλούμενη γλώσσα όλοι οι άνθρωποι χρησιμοποιούν σχετικά την ίδια ανατομία για να παράγουν ήχους. Η ομιλία σε γενικές γραμμές παράγεται από τον αέρα που ωθείται από τους πνεύμονες – οι οποίοι μπορούν να θεωρηθούν ως πηγή – και ο οποίος περνά μέσα από την φωνητική περιοχή και τον στόμα. Η φωνητική περιοχή εκτείνεται από τις φωνητικές χορδές έως τον στόμα και για ένα μέσο άνθρωπο έχει μήκος περίπου 17 εκ. Εισάγει βραχυπρόθεσμους συσχετισμούς (της τάξης των 1 ms) στο σήμα της ομιλίας και μπορεί να θεωρηθεί σαν ένα φίλτρο που παράγει διάφορους τύπους ήχων οι οποίοι και αποτελούν την ομιλία. Μέσα στο σήμα της ομιλίας υπάρχουν συχνότητες στις οποίες παρουσιάζεται συγκεντρωμένη ενέργεια και που ονομάζονται *formants*. Οι συχνότητες αυτές ελέγχονται από την μεταβολή του σχήματος της περιοχής για παράδειγμα αλλάζοντας την θέση της γλώσσας.

Επίσης υπάρχει η ρινική περιοχή η οποία είναι και αυτή ένας μη – ομοιόμορφος ακουστικός αγωγός πεπερασμένου μεγέθους η οποία τερματίζεται από μπροστά από τα ρουθούνια και από πίσω από ένα μετακινούμενο πτερύγιο από δέρμα που ονομάζεται μαλακή υπερώα, η οποία ελέγχει την ακουστική σύζευξη μεταξύ της στοματικής και της ρινικής περιοχής.



Σχήμα 2.1:Φωνητικός μηχανισμός του ανθρώπου.

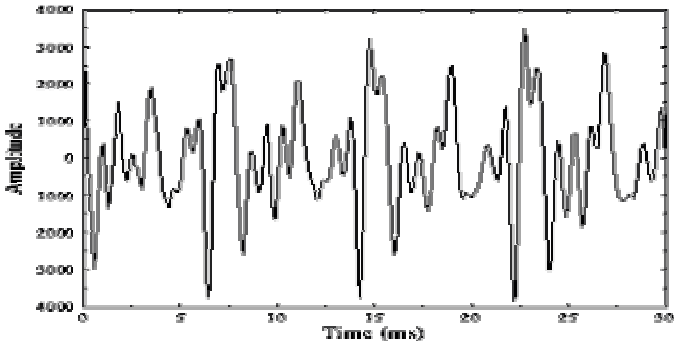
Για να κατανοήσουμε τώρα καλύτερα πώς η φωνητική περιοχή μετατρέπει τον αέρα από τους πνεύμονες σε ήχο είναι σημαντικό να γίνουν ορισμένες διευκρινήσεις. Για τον λόγο αυτό κατατάσσουμε τους φθόγγους[5] ανάλογα με των τρόπο με τον οποίο αρθρώνονται σε τρεις ευρύς κατηγορίες:

Έμφωνα (Voiced): Είναι συνήθως φωνήεντα (π.χ. [a], [e], [j]) και συχνά έχουν υψηλό μέσο ενεργειακό επίπεδο και ευδιάκριτες συχνότητες *formants*. Παράγονται από τον αέρα των πνευμόνων που περνά μέσα από τις φωνητικές χορδές. Στην περίπτωση αυτή οι φωνητικές χορδές δονούνται με κάποια περιοδικότητα γεγονός που παράγει μια σειρά από παλμούς αέρα. Η συχνότητα, η οποία καθορίζεται από την πίεση του αέρα στην τραχεία, και με την οποία οι φωνητικές χορδές πάλλονται είναι αυτό που καθορίζει τον τόνο του ήχου που παράγεται. Αυτός ο τόνος μπορεί να ρυθμιστεί μεταβάλλοντας το σχήμα και το τέντωμα των φωνητικών χορδών. Οι παλμοί αέρα που δημιουργούνται από τις δονήσεις τελικά περνούν κατά μήκος της υπόλοιπης φωνητικής περιοχής όπου κάποιες συχνότητες ενισχύονται. Είναι γενικά γνωστό ότι οι γυναίκες και

[5] Φθόγγοι (ή φωνές = phones) ονομάζονται τα ελάχιστα φθογγικά στοιχεία που συνθέτουν την έκφραση, τη χαρακτηριστική προφορά της γλώσσας και που λειτουργούν σε ένα φωνητικό σύστημα.

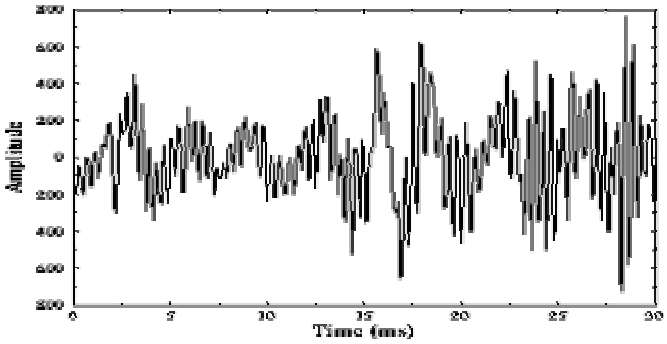
τα παιδία έχουν υψηλότερους τόνους φωνής (50 Hz έως 500 Hz) από τους άντρες (50 Hz έως 250 Hz) σαν αποτέλεσμα του πιο γρήγορου ρυθμού δόνησης κατά την διάρκεια της παραγωγής του ήχου.

Οι *έμφωνοι* ήχοι παρουσιάζουν μεγάλο βαθμό περιοδικότητας μεταξύ των τόνων της φωνής η οποία κυμαίνεται τυπικά μεταξύ 2 και 20 ms. Αυτή η μακράς διάρκειας περιοδικότητα φαίνεται στο Σχήμα 2.2 το οποίο έχουμε τμήματα *έμφωνης* ομιλίας που έχει δειγματοληπτηθεί στα 8 kHz και που η περίοδος είναι περίπου 8 ms ή 64 δείγματα.



Σχήμα 2.2: Τυπικά *έμφωνα* τμήματα ομιλίας**.**

*Άφωνα:*Είναι συνήθως σύμφωνα (π.χ. [l], [r]), γενικά έχουν λιγότερη ενέργεια και μεγαλύτερες συχνότητες από τα *έμφωνα* και οι κυματομορφές τους είναι χαοτικής και τυχαίας μορφής (Σχήμα 2.3). Η παραγωγή των unvoiced γίνεται από το πέρασμα του αέρα μέσα από τις φωνητικές χορδές οι οποίες εδώ δεν πάλλονται αντιθέτως μένουν ανοιχτές . Ο ήχος παράγεται μέσω ενός σφιξίματος στην φωνητική περιοχή. Ο τόνος της φωνής είναι ένας ασήμαντος παράγοντας αφού δεν υπάρχει δόνηση των φωνητικών χορδών. Τέλος και εδώ εμφανίζονται μικρής διάρκειας συσχετισμοί λόγω της φωνητικής περιοχής.



Σχήμα 2.3: Τυπικά άφωνα τμήματα ομιλίας.

Plosive sounds: Είναι αποτέλεσμα της δημιουργίας πίεσης στην φωνητική περιοχή όταν το στόμα είναι κλειστό και στην συνέχεια της απότομης απελευθέρωσης του αέρα η οποία προκαλεί προσωρινή διέγερση στην φωνητική περιοχή.

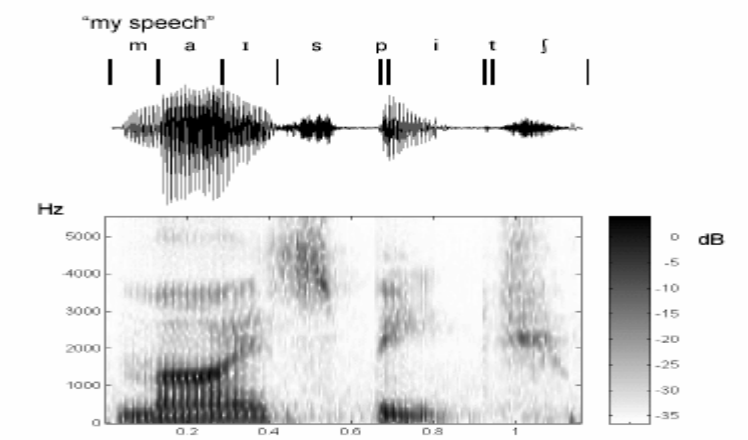
Πέρα από τις παραπάνω γενικές κατηγορίες υπάρχουν και άλλες πιο εξειδικευμένες ως προς τον τρόπο παραγωγής του ήχου καθώς επίσης υπάρχουν και ήχοι οι οποίοι δεν εμπίπτουν σε καμία από αυτές. Σαν παράδειγμα μπορούμε να αναφέρουμε τα διαρκή σύμφωνα (π.χ. [f], [θ], [t])τα οποία είναι αποτέλεσμα τόσο της δόνησης των φωνητικών χορδών όσο και του σφιξίματος στη φωνητική περιοχή.

Σε γενικές γραμμές παρόλο την μεγάλη ποικιλία ήχων που μπορούν να παραχθούν, το σχήμα της φωνητικής περιοχής και τρόπος της διέγερσης αλλάζουν σχετικά αργά και έτσι η ομιλία μπορεί να θεωρηθεί σαν να είναι ημι – στατική σε μικρές χρονικές περιόδους (της τάξης των 20 ms).

Επίσης η ομιλία παρουσιάζει μεγάλο βαθμό προβλεψιμότητας που οφείλετε στις ημι – περιοδικές δονήσεις των φωνητικών χορδών.

2.2 Επεξεργασία της Ομιλίας με Σπεκτρόγραμμα

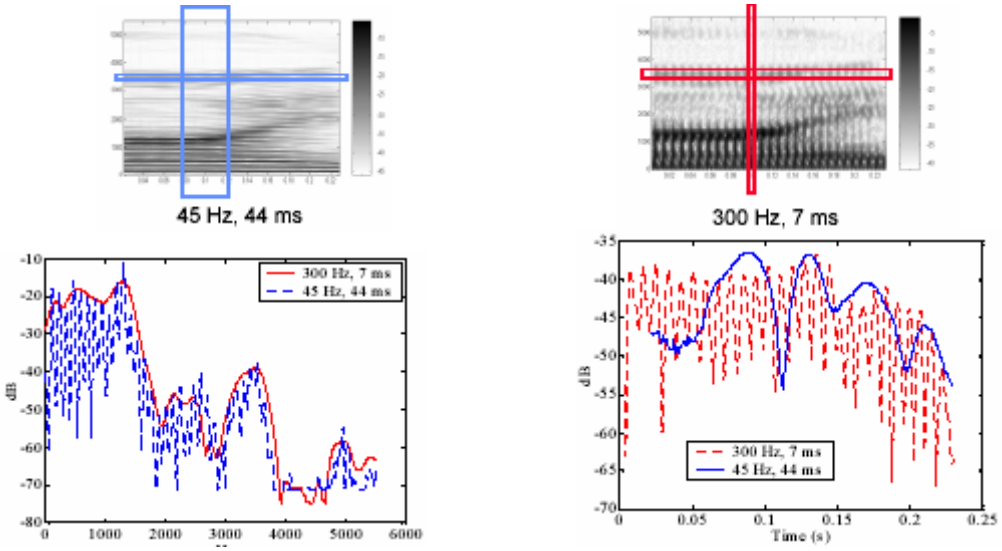
Το σπεκτρόγραμμα αποτελεί το πλέον βασικό εργαλείο στην επεξεργασία της ομιλίας και αυτό γιατί μας δίνει την δυνατότητα να έχουμε μια τρισδιάστατη γραφική αναπαράσταση της ομιλίας σε σχέση με την συχνότητα, τον χρόνο και την ενέργεια. Έτσι ο κάθετος άξονας σε ένα σπεκτρόγραμμα αναπαριστά την συχνότητα για το φάσμα της ομιλίας (από 0 – 4 Khz), ενώ ο οριζόντιος άξονας αναπαριστά τον χρόνο και τέλος η τρίτη διάσταση αναπαριστά την ενέργεια για διαφορετικές ζώνες συχνοτήτων για κάθε χρονική στιγμή (Σχήμα 2.4)



Σχήμα 2.4: Τα σκούρα τμήματα στο σπεκτρόγραμμα δείχνουν μεγάλη ενέργεια.

Ορισμένες βασικές έννοιες σε ένα σπεκτρόγραμμα είναι το *κατώφλι* το οποίο ορίζει την ενέργεια σε dB κάτω από την οποία σε ένα ασπρόμαυρο σπεκτρόγραμμα η απεικόνιση θα είναι άσπρη και η *δυναμική περιοχή* η οποία καθορίζει το εύρος πέρα από το οποίο έχουμε άπλωμα της γκρι κλίμακας μας. Για την δημιουργία του σπεκτρογράμματος σε ένα σήμα ομιλίας βασιζόμαστε στο Short Time Fourier Transform (STFT) όπου το σήμα της φωνής χωρίζεται σε αλληλο – επικαλυπτόμενα τμήματα (frame blocks) και σε κάθε ένα από αυτά εφαρμόζεται κάποιο παράθυρο το οποίο έχει σαν στόχο την εξάλειψη της ασυνέχειας από το ένα frame στο άλλο. Το παράθυρο που συνήθως χρησιμοποιείται είναι το Hamming και για την δημιουργία του σπεκτρογράμματος σχεδιάζεται το τετράγωνο της απόλυτης τιμής του STFT σε dB.

Εδώ πρέπει να αναφέρουμε ότι δεν είναι δυνατόν να έχουμε ταυτόχρονα καλή ανάλυση στο άξονα του χρόνου και στον άξονα των συχνοτήτων στο ίδιο σπεκτρόγραμμα και για αυτό τον λόγο διακρίνουμε τα σπεκτρογράμματα σε δύο κατηγορίες. Στα ¨ευρύς ζώνης¨ σπεκτρογράμματα τα οποία προκύπτουν από STFT με μικρό αριθμό δειγμάτων και στα οποία πετυχαίνουμε υψηλή ανάλυση στο πεδίο του χρόνου. Εδώ είναι εμφανείς οι οριζόντιες μαύρες ζώνες οι οποίες αναπαριστούν τα formants της φωνητικής περιοχής καθώς και οι κάθετες γραμμές οι οποίες είναι γνωστές σαν ¨ραβδώσεις¨ και οι οποίες αναπαριστούν τις ξεχωριστές αντηχήσεις στην φωνητική περιοχή. Μετρώντας την απόσταση μεταξύ των ¨ραβδώσεων¨ είναι δυνατόν υπολογίσουμε την θεμελιώδη συχνότητα. Όταν τώρα χρησιμοποιούμε πολύ περισσότερα δείγματα παίρνουμε ¨στενής ζώνης¨ σπεκτρογράμματα στα οποία έχουμε υψηλή ανάλυση στο πεδίο των συχνοτήτων. Εδώ τα formants είναι λιγότερο ευδιάκριτα και οι ¨ραβδώσεις¨ δεν είναι ορατές αλλά η αναπαράσταση των αρμονικών του σήματος είναι ορατή με μεγαλύτερη λεπτομέρεια (Σχήμα 2.5).



Σχήμα 2.5: Στο σχήμα αυτό εμφανίζονται δύο σπεκτρογράμματα ένα ¨στενής ζώνης – 45 Hz¨ και ένα ¨ευρύς ζώνης – 300 Hz¨ για την φράση /ai/. Όπως βλέπουμε με το ¨στενής ζώνης¨ πετυχαίνουμε καλύτερη ανάλυση στο πεδίο των συχνοτήτων ενώ με το ¨ευρύς ζώνης¨ πετυχαίνουμε καλύτερη ανάλυση στο πεδίο του χρόνου.

Για να αντιληφθούμε κάθε τμήμα του σπεκτρογράμματος σε τι φθόγγο αντιστοιχεί βασιζόμαστε σε ορισμένα βασικά βήματα ανάγνωσης των οποίων η σειρά εφαρμογής δεν είναι απόλυτη. Τα βήματα λοιπόν έχουν ως εξής:

(1) Αναγνωρίζουμε τις μεταβολές μεταξύ γενικών φθόγγων. Οι φθόγγοι οι οποίοι είναι εύκολο να αναγνωριστούν είναι τα φωνήεντα/ημίφωνα, έρρινα σύμφωνα, συνεχόμενα διαρκή και κλειστά διαρκή.

(2) Κάθε περιοχή ευρείας τάξης μπορεί να περιέχει ένα ή περισσότερα τμήματα. Έτσι αναζητούμε ενδείξεις που θα μας βοηθήσουν να αποφασίσουμε για το μήκος της κάθε περιοχής καθώς και για το αν υπάρχουν μεταβολές μεταξύ των formants.

(3) Ελέγχουμε εάν το φθογγικό στοιχείο είναι έμφωνο ή άφωνο.

(4) Μετράμε την συχνότητα των formants και κάνουμε 2 – 3 υποθέσεις για την ταυτότητα του κάθε τμήματος.

(5) Λαμβάνουμε υπόψη και την επίδραση του ¨περιβάλλοντος¨ στο οποίο βρίσκεται ο κάθε φθόγγος και κάνουμε επιπρόσθετες υποθέσεις εάν το τμήμα είναι ασαφή.

(6) Προσπαθούμε να βρούμε λέξεις οι οποίες ταιριάζουν στα εν λόγω τμήματα.

Οι τάξεις στις οποίες μπορούμε να κατατάξουμε τους φθόγγους ανάλογα με *τρόπο* με τον οποίο αρθρώνονται είναι σε *αντηχητικούς* και *μη – αντηχητικούς*[6] (με εμπόδιο), όταν έχουμε formant (χαμηλής συχνότητας) και όταν δεν έχουμε καθόλου formant αντίστοιχα. Μια άλλη γενική διάκριση είναι μεταξύ *έμφωνων* και *άφωνων*, η οποία βασίζεται στο μήκος των ¨ραβδώσεων¨ όπου στα *έμφωνα* είναι συνήθως μεγαλύτερο από ότι στα άφωνα. Επίσης η διάκριση των *φωνηέντων/ημίφωνων* (που είναι μια υποδιαίρεση της παραπάνω κατηγορίας) γίνεται από το γεγονός ότι αυτά ανήκουν στην κατηγορία των αντηχητικών, ότι η δομή των formants[7] τους είναι καθαρή και ότι οι μεταβολές μεταξύ τους παρουσιάζουν μια συνέχεια. Τα *έρρινα* μπορούμε να τα αντιληφθούμε από το ότι ανήκουν στην κατηγορία των αντηχητικών, ότι έχουν εμφανές formants

[6] Οι *αντηχητικοί φθόγγοι* παράγονται στη στοματική κοιλότητα όταν είναι διαμορφωμένη έτσι ώστε η πίεση του αέρα μέσα στην κοιλότητα να είναι σχεδόν όμοια με την πίεση του αέρα έξω από την κοιλότητα. Αντίθετα οι *μη – αντηχητικοί* φθόγγοι παράγονται με περισσότερη σύσφιξη των αρθρωτών μέσα στην στοματική κοιλότητα έτσι ώστε η πίεση του αέρα μέσα στο στόμα να είναι πολύ μεγαλύτερη από εκείνη του εξωτερικού περιβάλλοντος.

[7] Πρέπει να επισημάνουμε ότι τα *ημίφωνα* δεν είναι ούτε άφωνα, αφού το πέρασμα του αέρα κατά την παραγωγή τους είναι ελεύθερο αλλά ούτε και έμφωνα αφού η φασματική ανάλυση δείχνει ότι στερούνται της ακουστικής ιδιότητας των formants που χαρακτηρίζει τα φωνήεντα.

χαμηλής συχνότητας (το πρώτο formant) και μη εμφανή υψηλότερης συχνότητας, καθώς επίσης και από το γεγονός οι μεταβολές μεταξύ τους είναι ασυνεχής. Τους *plosive ήχους* μπορούμε να τους αντιληφθούμε σαν μια περίοδο απόλυτης σιγής η οποία ακολουθείτε από ένα ¨καταιγισμό¨ θορύβου ευρείας ζώνης ή υψηλών συχνοτήτων λόγο της απότομης απελευθέρωσης του αέρα ο οποίος και δημιουργεί μια προσωρινή διέγερση των φωνητικών χορδών για 5 – 100 ms. Τέλος τα *διαρκή* χαρακτηρίζονται από την απουσία formants και από ενέργεια υψηλής συχνότητας.

Εδώ πρέπει να πούμε σε μια αρκετά μεγάλη περιοχή του σπεκτρογράμματος είναι δυνατόν να περιέχονται δύο ή και περισσότερα συνεχόμενα φθογγικά στοιχεία. Για να αναγνωρίσουμε την ύπαρξη των ξεχωριστών αυτών φθογγικών στοιχείων βασιζόμαστε στο γεγονός ότι στα συνήθη φωνήεντα και ημίφωνα αντιστοιχούν ορισμένα formants. Στηριζόμενοι λοιπόν στα formants αυτά και στον αριθμό των μεταβολών τους μπορούμε να έχουμε μια εκτίμηση του αριθμού των φθογγικών στοιχείων.

Ο τελικός προσδιορισμός που πρέπει να γίνει σε ένα φθόγγο εκτός από τον *τρόπο* είναι και το *σημείο* στο οποίο αρθρώνεται. Σημεία άρθρωσης είναι τα χείλια, τα δόντια, τα φατνία (η περιοχή των ούλων πίσω από τα δόντια), ο ουρανίσκος, η υπερώα, η γλωσσίδα και ο φάρυγγας. Έτσι λοιπόν διακρίνουμε τα *εμπρόσθια σύμφωνα* τα οποία αποτελούνται από τα αυτά στα οποία το σημείο άρθρωσης είναι το *πρώτο μισό του στόματος* δηλαδή τα χείλια, τα δόντια ή τα φατνία π.χ. [p], [δ] και οι συχνότητες των formants τους είναι $F_1 = 180Hz$, $F_2 = 1000Hz$ και $F_3 = 2000Hz$. Επίσης στα *εμπρόσθια σύμφωνα* έχουμε και εκείνα τα οποία σχηματίζονται στον *ουρανίσκο* π.χ. [j] στο ελληνικό [jatrós] ή στην *υπερώα* όπως [k], [g], [γ] κ.α. και για τα οποία οι συχνότητες των formants είναι $F_1 = 200Hz$ και F_2, F_3 αρκετά κοντά μεταξύ τους και μέσα στην περιοχή των 1500 – 2200 Hz. Μία άλλη κατηγορία είναι τα *κορονικά* τα οποία σχηματίζονται με το πάνω μέρος της γλώσσας όπως είναι τα [t], [d], [n] και για τα οποία ισχύουν οι συχνότητες $F_1 = 180Hz$, $F_2 = \begin{cases} 1700Hz\ για\ αντρες \\ 2000Hz\ για\ γυναικες \end{cases}$, ενώ η συχνότητα F_3 δεν συγκλίνει με την F_2 αλλά βρίσκεται μέσα στην περιοχή των 2500 -3000Hz. Στα *τριβόμενα* διακρίνουμε τα *συριστικά* ([z], [s]) τα οποία συνήθως μοιάζουν στο σπεκτρόγραμμα σαν λευκός θόρυβος φιλτραρισμένος από ένα υψηλών συχνοτήτων φίλτρο και το σημείο άρθρωσης τους καθορίζεται από συχνότητα αποκοπής ενώ τα *μη – συριστικά* ([θ],[δ], [v]) δεν είναι δυνατόν να αναγνωριστούν εύκολα γιατί έχουν την τάση να αφομοιώνουν τα χαρακτηριστικά των διπλανών ήχων. Τέλος να πούμε ότι και για τα φωνήεντα ισχύει το χαρακτηριστικό *εμπρόσθιο* που χρησιμοποιήσαμε στα σύμφωνα. Και εδώ τα διακρίνουμε εκείνα που σχηματίζονται με την ανύψωση του μπροστινού μέρους της γλώσσας προς το μπροστινό τμήμα της στοματικής κοιλότητας π.χ. [i], [e] και στα υπόλοιπα [a], [o], [u].

2.3 Πλεονασμοί στο Σήμα της Ομιλίας

Όπως αναφέρθηκε στην Παράγραφο 1.7 ένα τυπικό PCM είναι ιδιαίτερα ικανό να κωδικοποιήσει ένα σήμα τυχαίας κυματομορφής αφού κωδικοποιεί κάθε δείγμα της κυματομορφής εισόδου ξεχωριστά από όλα τα άλλα δείγματα. Στη περίπτωση όμως της κυματομορφής του σήματος της ομιλίας εμφανίζονται πλεονασμοί τόσο μεταξύ των γειτονικών όσο και μεταξύ των απομακρυσμένων δειγμάτων. Οι πλεονασμοί γειτονικών δειγμάτων εμφανίζονται μεταξύ δειγμάτων που είναι κοντά το ένα στο άλλο και ο βαθμός συσχέτισης τους για 8 Khz είναι περίπου 0.85 ή και μεγαλύτερος. Επίσης υπάρχουν τόσο οι πλεονασμοί μεταξύ απομακρυσμένων δειγμάτων (που οφείλονται στη έμφυτη περιοδικότητα λόγω των *έμφωνων* τμημάτων της ομιλίας) αλλά και οι πλεονασμοί (εκτός από το πεδίο του χρόνου όπως είναι οι παραπάνω) και στο πεδίο των συχνοτήτων. Οι πλεονασμοί αυτοί δεν είναι ανεξάρτητοι από τους πλεονασμούς στο πεδίο του χρόνου και προσφέρουν έναν εναλλακτικό τρόπο ανάλυσης και επεξεργασίας των πλεονασμών. Έτσι λοιπόν κατατάσσονται ως εξής:

(1) Πεδίο του Χρόνου

a. Μη ομοιόμορφη πιθανοτική κατανομή του πλάτους (non – uniform amplitude distridution).
b. Συσχετισμός γειτονικών δειγμάτων (sample to sample correlation)
c. Συσχετισμός μεταξύ περιόδων (cycle to cycle correlation)
d. Συσχέτιση μεταξύ παύσεων ανάμεσα στις *θεμελιώδης συχνότητες (pitch)* του σήματος της ομιλίας (pitch interval to pitch interval correlations)

(2) Πεδίο των Συχνοτήτων

e. Μη επίπεδη φύση του φάσματος (nonuniform long – time spectral densities)
f. Sound – specific short – time spectral densities
g. Το σήμα της ομιλίας είναι ζωνοπερατό

2.3.1 Μη Ομοιόμορφη Πιθανοτική Κατανομή Πλάτους

Τα δείγματα με χαμηλό πλάτος (τα οποία είναι συνήθως το αποτέλεσμα των παύσεων σε μια συνομιλία) έχουν πιο μεγάλη πιθανότητα να συμβούν από τα δείγματα με υψηλό πλάτος. Αυτό το χαρακτηριστικό χρησιμοποιείτε στο λογαριθμικό PCM. Ωστόσο η πιο κατάλληλη μέθοδος τις επεξεργασίας των πλατών του σήματος , ώστε να μειωθεί ο ρυθμός των κωδικοποιημένων bits, περικλείει κάποια μορφή προσαρμοστικού ελέγχου.

2.3.2 Συσχετισμός Γειτονικών Δειγμάτων

Όπως αναφέρθηκε πιο πάνω τα δείγματα έχουν ένα υψηλό βαθμό συσχέτισης της τάξης του 0.85 (για ρυθμό δειγματοληψίας 8Khz). Για να μειώσουμε λοιπόν τον ρυθμό των bits μπορούμε να εκμεταλλευτούμε αυτή την συσχέτιση μεταξύ γειτονικών δειγμάτων. Στην πράξη για τα 8Khz υπάρχει υψηλή συσχέτιση για δύο και τρία δείγματα μακριά (Πίνακας 2.1) και συνήθως αυξάνει όσο αυξάνει και η συχνότητα δειγματοληψίας.

Ο πιο απλός τρόπος για να εκμεταλλευτούμε την συσχέτιση μεταξύ των δειγμάτων του σήματος της ομιλίας είναι να κωδικοποιούμε μόνο την διαφορά μεταξύ γειτονικών δειγμάτων. Οι διαφορές αυτές συγκεντρώνονται στον αποκωδικοποιητή και χρησιμοποιούνται για την ανασύσταση του σήματος.

Απόσταση μεταξύ δειγμάτων	**Μέσος βαθμός συσχέτισης για μεγάλο χρονικό τμήμα**
1 δείγμα	0.825
2 δείγματα	0.562
3 δείγματα	0.308
4 δείγματα	0.004
5 δείγματα	-0.243

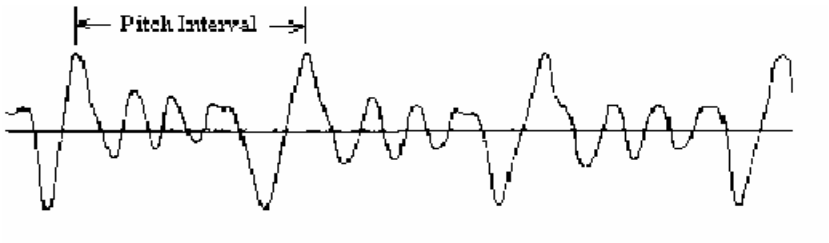
Πίνακας 2.1: Τιμές της συσχέτισης μεταξύ πλατών των δειγμάτων για 8Khz δειγματοληπτημένη ομιλία.

2.3.3 Συσχέτιση Μεταξύ Περιόδων

Η συσχέτιση αυτή βασίζεται στην βασική περίοδο των *έμφωνων* τμημάτων της ομιλίας. Ως γνωστό αυτά είναι περιοδικά γεγονός που οφείλετε στον τρόπο παραγωγής τους. Με βάση λοιπόν αυτή την περιοδική φύση τους, (Σχήμα 2.6) εφαρμόζεται μια πρόγνωση των επόμενων δειγμάτων η οποία προσπαθεί να εκμεταλλευτεί την περιοδικότητα της διέγερσης και είναι γνωστή σαν *Long Term Prediction (LTP)*. Η *LT* πρόγνωση μεταβάλλεται με το χρόνο έτσι ώστε να ταιριάζει με την μορφή του φάσματος του σήματος της ομιλίας . Οι κωδικοποιητές τώρα που βασίζονται σε αυτή, όπως θα δούμε, είναι πολύ περισσότερο πολύπλοκοι από αυτούς που εκμεταλλεύονται τον πλεονασμό μεταξύ γειτονικών δειγμάτων.

2.3.4 Συσχέτιση Μεταξύ Παύσεων

Ως γνωστό σε ένα σήμα ομιλίας εκτός από τα *έμφωνα* και άφωνα τμήματα υπάρχουν και οι παύσεις μεταξύ των *θεμελιωδών συχνοτήτων* (Σχήμα 2.6). Αυτές οφείλονται στις αντίστοιχες παύσεις του αέρα που διεγείρει την φωνητική περιοχή.



Σχήμα 2.6: Κυματομορφή στο πεδίο του χρόνου για έμφωνο τμήμα ομιλίας.

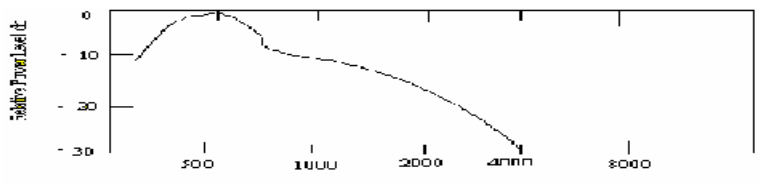
Παρουσιάζεται λοιπόν στα τμήματα αυτά εκτός από την συσχέτιση μεταξύ των περιόδων και ένα επαναλαμβανόμενο μοτίβο στις παύσεις . Για τον λόγο αυτό ο πιο αποτελεσματικός τρόπος κωδικοποίησης των *έμφωνων* τμημάτων είναι να κωδικοποιήσουμε μια τέτοια παύση και να χρησιμοποιήσουμε αυτή την κωδικοποίηση σαν φόρμα για κάθε διαδοχική παύση. Η δε διάρκεια αυτών των παύσεων είναι τυπικά 5 με 20 msec για τους άντρες και 2.5 με 10 msec για τις γυναίκες και αφού η διάρκεια ενός έμφωνου τμήματος είναι περίπου 100 msec σημαίνει ότι μέσα σε αυτό περιέχονται περί τις 20 με 40 παύσεις. Παρόλο όμως που η κωδικοποίηση των παύσεων αυτών μας προσφέρει σημαντικές μειώσεις στο bit rate, επειδή υπάρχει μια δυσκολία στην αναγνώριση της θεμελιώδους συχνότητας. Αυτό έχει ως αποτέλεσμα αν αυτή κωδικοποιηθεί λάθος να δημιουργείται ένα περίεργο ηχητικό αποτέλεσμα.

Μια ενδιαφέρουσα τώρα παράμετρος της κωδικοποίησης αυτών των παύσεων είναι ότι μπορούμε να επιταχύνουμε την ομιλία διατηρώντας την αναγνωρισιμότητα. Αυτό γίνεται εφικτό αφαιρώντας κάποιο ποσοστό των παύσεων από τα φωνήματα και έτσι ο ρυθμός με τον οποίο παράγεται ο ήχος αυξάνεται με τρόπο που είναι αντίστοιχος με την πιο γρήγορη δημιουργία λέξεων. Η θεμελιώδης συχνότητα ωστόσο παραμένει αμετάβλητη. Σε αντίθεση αν ο ρυθμός της αναδόμησης αυξηθεί μερικώς όλες οι συχνότητες περιλαμβανόμενης και της θεμελιώδους αυξάνονται ανάλογα και με μεγαλύτερες αυξήσεις το σήμα γίνεται ακατανόητο.

2.3.5 Μη Επίπεδη Φύση του Φάσματος

Ένα τυχαίο σήμα έχει ένα φάσμα συχνοτήτων το οποίο είναι επίπεδο στο εύρος ζώνης που μας ενδιαφέρει. Για αυτό τα σήματα που έχουν ασυσχέτιστα δείγματα στο πεδίο του χρόνου απλώνονται ομοιόμορφα σε όλο το εύρος ζώνης τους. Από την άλλη μεριά η μη ομοιόμορφη φασματική πυκνότητα δηλώνει μια αναποτελεσματική χρήση του εύρους ζώνης και είναι ενδεικτικό της ύπαρξης πλεονασμών στην κυματομορφή του σήματος.

Όπως βλέπουμε τώρα στο Σχήμα 2.7 κυρίως τα τμήματα εκείνα που έχουν συχνότητα πάνω από 3kHz έχουν χαμηλά επίπεδα ισχύος τα οποία είναι άμεση συνέπεια της συσχέτισης μεταξύ δειγμάτων στο πεδίο του χρόνου. Τα μεγάλου πλάτους σήματα δεν μπορούν να μεταβληθούν απότομα γιατί αποτελούνται κατά μέσο όρο από συνιστώσες χαμηλών συχνοτήτων.

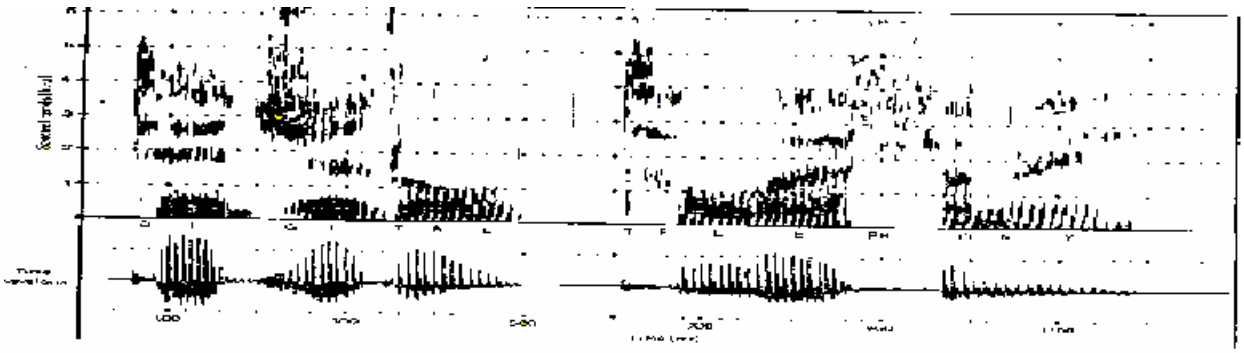


Σχήμα 2.7: Φασματική πυκνότητα ισχύος για μακράς διάρκειας τμήμα του σήματος της ομιλίας**.**

Από την ανομοιόμορφη αυτή κατανομή και τα ιδιαίτερα χαμηλά επίπεδα ισχύος για υψηλές συχνότητες 2 έως 3.4kHz οδηγούμαστε στο συμπέρασμα ότι στο σήμα της ομιλίας κατανέμεται σε μεγαλύτερο εύρος ζώνης από αυτό που πραγματικά χρειάζεται και ότι θα μπορούσαμε ίσως να αγνοήσουμε αυτές τις συνιστώσες. Αυτό βέβαια είναι λάθος για τον λόγο ότι οι συνιστώσες αυτές αντιστοιχούν κατά κύριο λόγο σε σύμφωνα τα οποία και περιέχουν το μεγαλύτερο πληροφοριακό περιεχόμενο σε ένα σήμα ομιλίας. Αυτό συμβαίνει γιατί αν αφαιρεθούν κάποια φωνήεντα τα οποία και συνήθως απαιτούν περισσότερη ενέργεια και βρίσκονται στα χαμηλότερα τμήματα της ζώνης των συχνοτήτων το νόημα της φράσης μπορεί έστω και με κάποια δυσκολία να γίνει αντιληπτό κάτι το οποίο όμως δεν μπορεί να γίνει αν αφαιρεθούν τα σύμφωνα.

2.3.6 Sound – Specific Short – Time Spectral Densities

Στην προηγούμενη παράγραφο αναφερθήκαμε σε κατά μέσο όρο μακράς διάρκειας τμήματα του σήματος της ομιλίας. Αν εξετάσουμε τώρα την φασματική πυκνότητα για ένα μικρότερης διάρκειας τμήμα της ομιλίας θα δούμε ότι σε αυτό η φασματική πυκνότητα μεταβάλλεται αξιοσημείωτα και ότι παρουσιάζονται ενεργειακές ¨ενισχύσεις¨ σε συγκεκριμένες συχνότητες αλλά και ενεργειακές ¨υποβαθμίσεις¨ σε άλλες συχνότητες. Οι συχνότητες στις οποίες πραγματοποιούνται αυτές οι ¨ενισχύσεις¨ ονομάζονται *συχνότητες formant* ή πιο απλά *formants* (όπως έχουμε δει και σε άλλη παράγραφο) και υπάρχουν συνήθως δύο ή τρία στα *έμφωνα* τμήματα της ομιλίας. Όλα αυτά τα χαρακτηριστικά φαίνονται σε ένα σπεκτρόγραμμα (π.χ. Σχήμα 2.8).



Σχήμα 2.8: Σπεκτρόγραμμα της φράσης “digital telephony”.

2.3.7 Το Ζωνοπερατό του Σήματος της Ομιλίας

Ίσως η πιο σημαντική ιδιότητα του σήματος της φωνής είναι ότι, το φάσμα συχνοτήτων του είναι *ζωνοπερατό* στη βασική ζώνη. Αυτή την ιδιότητα εκμεταλλεύονται όλοι οι κωδικοποιητές φωνής γιατί ένα τέτοιο σήμα μπορεί με ευκολία να δειγματοληπτηθεί με συγκεκριμένο ρυθμό και να ανακτηθεί πλήρως από τα δείγματα του. Απαραίτητη προϋπόθεση βέβαια η συχνότητα δειγματοληψίας να είναι μεγαλύτερη από το διπλάσιο της μέγιστης συχνότητας του.

2.4 Ιδιότητες του Ακουστικού Συστήματος του Ανθρώπου

Όλες οι παραπάνω ιδιότητες του σήματος της ομιλίας είναι ιδιαίτερα σημαντικές γιατί μας παρέχουν την δυνατότητα να κατασκευάσουμε κωδικοποιητές με ¨ικανοποιητική¨ ποιότητα ομιλίας συνδυασμένη με χαμηλό bit rate. Βέβαια κατά την υλοποίηση αυτή πρέπει να λάβουμε υπόψη την ψυχοακουστική η οποία εξετάζει την ανθρώπινη αίσθηση της ακοής. Η κωδικοποίηση που βασίζεται σε αυτήν ονομάζεται perceptual κωδικοποίηση. Ένα βασικό αξίωμα της ψυχοακουστικής είναι η *φασματική επικάλυψη (spectral masking)*, κατά την οποία η παρουσία ενός ηχητικού σήματος επικαλύπτει την αίσθηση κάποιου άλλου, τα σήματα που επικαλύπτονται είναι κυρίως μικρής ισχύος σε γειτονικές συχνότητες.

Χρησιμοποιώντας την επεξεργασία ψηφιακού σήματος στο επίπεδο της συχνότητας, οι perceptual κωδικοποιητές εξαφανίζουν ¨άχρηστα¨ κομμάτια από το ηχητικό σήμα που

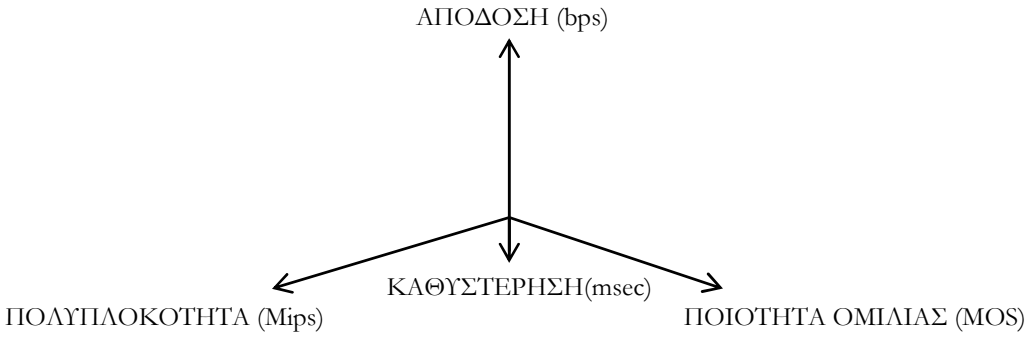
επικαλύπτονται από άλλα, πιο δυνατά, έτσι μειώνονται οι απαιτήσεις για μεγάλο bit rate. Σαν παράδειγμα μπορούμε να αναφέρουμε την περίπτωση των formants που αντιστοιχούν σε φωνήεντα και τα οποία καλύπτουν τον θόρυβο σε μια περιοχή γύρω από αυτά, όσο βέβαια ο θόρυβος αυτός είναι 15 dB μικρότερος από το σήμα. Με αυτό τον τρόπο μπορούν γίνουν ανεκτά μεγαλύτερα λάθη και κατά συνέπεια μπορούμε να μειώσουμε το coding rate άρα κατ' επέκταση και το bit rate. Ο κωδικοποιητής βέβαια μπορεί να μειώσει ακόμη περισσότερο το μέγεθος του σήματος μειώνοντας επιλεκτικά ακόμη και τα μη επικαλυπτόμενα σήματα. Φυσικά όσο περισσότερο μειώνουμε την ανάλυση του ψηφιακού ηχητικού σήματος τόσο αυξάνονται ο θόρυβος και οι παραμορφώσεις. Αλλά όσο οι κωδικοποιητές κρατούν αυτές τις δυσμορφίες κάτω από το κατώφλι επικαλύψεων παραμένουν ανεπαίσθητες.

Μια άλλη βασική ιδιότητα είναι η *aural frequency resolution* σύμφωνα με την οποία η ευαισθησία του ανθρώπινου αυτιού μεταβάλλεται ανάλογα με την περιοχή των συχνοτήτων, δηλαδή ένας ήχος με δεδομένη στάθμη και συχνότητα μπορεί να γίνεται αντιληπτός, ενώ κάποιος άλλος με μεγαλύτερη στάθμη αλλά με διαφορετική συχνότητα να μην γίνεται. Μπορεί, επομένως στο συνολικό φάσμα συχνοτήτων του ήχου να καθοριστεί το απόλυτο *κατώφλι (κάτω όριο) ακουστικότητας*. Αναλύοντας το σήμα της ομιλίας μπορούμε να απορρίψουμε τα μέρη, που βρίσκονται κάτω από το κατώφλι ακουστικότητας, και έτσι να μειώσουμε το bit rate. Επίσης επειδή είναι γνωστό ότι ένας θόρυβος χαμηλής συχνότητας επηρεάζει περισσότερο την αντιληπτή ομιλία από ένα θόρυβο υψηλής συχνότητας πρέπει ο αριθμός των bits κωδικοποίησης για τις χαμηλές συχνότητες να είναι μεγαλύτερος από ότι για τις υψηλές συχνότητες.

ΚΕΦΑΛΑΙΟ 3

3.1 Οι Διαστάσεις της Απόδοσης των Κωδικοποιητών Φωνής

Οι αλγόριθμοι κωδικοποίησης φωνής αποτιμούνται με βάση το πόσο καλά συνδυάζουν ορισμένα βασικά χαρακτηριστικά που έχει κάθε αλγόριθμος . Αυτά είναι ο ρυθμός των bits, η ποιότητα του επανασυντεθειμένης ομιλίας, η πολυπλοκότητα υλοποίησης (κατανάλωση ισχύος) του κωδικοποιητή, η καθυστέρηση που εισάγει και η ¨αντοχή¨ του στα λάθη του καναλιού και στα ακουστικά παράσιτα. Σε γενικές γραμμές για να επιτευχθεί υψηλή ποιότητα ομιλίας σε χαμηλούς ρυθμούς bits απαιτούνται πιο πολύπλοκοι αλγόριθμοι πράγμα που σημαίνει αύξηση του κόστους υλοποίησης τους αλλά και της καθυστέρησης.



Σχήμα 3.1: Σχηματική αναπαράσταση της αλληλεπίδρασης των παραγόντων απόδοσης ενός κωδικοποιητή.

Στη συνέχεια ακολουθεί μια σύντομη παρουσίαση των παραμέτρων της απόδοσης.

Ρυθμός των Bits: Είναι η ικανότητα κωδικοποίησης και εκφράζεται σε bits per second (bps).

Καθυστέρηση: Συχνά οι κωδικοποιητές επεξεργάζονται την ομιλία σε blocks και η αυτή διαδικασία εισάγει καθυστέρηση. Εάν καθυστέρηση ενός συστήματος επικοινωνίας ξεπερνά τα 50 msec τότε η ηχώ μπορεί να δημιουργήσει σημαντικά προβλήματα εάν δεν αντιμετωπιστεί με άλλα μέσα. Ανάλογα με την εφαρμογή η συνολική καθυστέρηση κυμαίνεται από το 1 msec (τηλεφωνικό δίκτυο) έως τα 500 msec (βίντεο - τηλεφωνία). Η ελαχιστοποίηση της καθυστέρησης είναι ένας σημαντικός παράγοντας κατά την σύγκριση των κωδικοποιητών φωνής.

Πολυπλοκότητα: Είναι η επεξεργαστική προσπάθεια που απαιτείται για την υλοποίηση του αλγόριθμου. Η υλοποίηση αυτή συνήθως γίνεται σε επεξεργαστές ψηφιακού σήματος (DSP's – Digital Signal Processors) και μετρείται από τον αριθμό των πολλαπλασιασμών και των προσθέσεων που απαιτούνται για την κωδικοποίηση της φωνής. Εκφράζεται σε αριθμό υπολογισμών ανά δευτερόλεπτο (millions of instruction per second - MIP's).

Ποιότητα Ομιλίας: Η ποιότητα της ομιλίας είναι δύσκολο να προσδιοριστεί πόσο μάλλον να μετρηθεί. Ο σκοπός της μέτρησης είναι να περιγραφεί πλήρως η ποιότητα ενός κωδικοποιητή ομιλίας σε ένα απλό νούμερο. Η μέτρηση αυτή γίνεται τόσο με υποκειμενικά όσο με αντικειμενικά κριτήρια. Στην συνέχεια ακολουθεί μια πιο εκτενής ανάλυση.

3.2 Αξιολόγηση της Απόδοσης Κωδικοποιητών Φωνής

Για να αξιολογήσουμε την απόδοση ενός κωδικοποιητή φωνής είναι απαραίτητο να έχουμε κάποιο δείκτη της *κατανοησιμότητας* και της *ποιότητας* της ομιλίας που παράγεται. Ο όρος *κατανοησιμότητα* συνήθως αναφέρεται στο εάν η ομιλία εξόδου είναι εύκολα κατανοητή ενώ ο όρος *ποιότητα* είναι ένας δείκτης που δείχνει το πόσο φυσικά ακούγεται η ομιλία . Είναι πιθανό για ένα κωδικοποιητή να παράγει υψηλά κατανοητή ομιλία που είναι όμως χαμηλής ποιότητας με αποτέλεσμα ο ομιλητής να μην είναι αναγνωρίσιμος. Από την άλλη πλευρά είναι απίθανο μια μη

κατανοητή ομιλία να είναι υψηλής ποιότητας αλλά υπάρχουν καταστάσεις στις οποίες ο ευχάριστα αντιληπτός λόγος δεν έχει υψηλή κατανοησιμότητα..

Οι τεχνικές αξιολόγησης της απόδοσης ενός κωδικοποιητή φωνής όσον αφορά την κατανοησιμότητα και της ποιότητα της ομιλίας που αυτός παράγει χωρίζονται σε δύο κατηγορίες τις υποκειμενικές και τις αντικειμενικές.

3.2.1 Υποκειμενικές Τεχνικές Αξιολόγησης Κωδικοποιητών Φωνής

Στις τεχνικές αυτές ο απόλυτος κριτής είναι ο άνθρωπος. Πραγματοποιούνται παίζοντας ένα δείγμα φωνής σε έναν αριθμό από άτομα, ζητώντας τους να κρίνουν την ποιότητα της φωνής. Επειδή οι κωδικοποιητές φωνής εξαρτώνται σημαντικά από τον ομιλητή η ποιότητα φωνής που παράγουν ποικίλει ανάλογα με την ηλικία, το φύλο και την ταχύτητα ομιλίας του ομιλητή. Τα υποκειμενικά τεστ δίνουν αποτελέσματα σε σχέση με την συνολική ποιότητα, την προσπάθεια ακρόασης, το βαθμό καταληπτότητας και την φυσικότητα της φωνής. Υπάρχουν διάφορα τέτοια τεστ και τα πιο διαδομένα από αυτά είναι τα εξής:

Diagnostic Rhyme Test (DTR). Το *διαγνωστικό τεστ ρίμας* χρησιμοποιείται κυρίως σε εκείνους τους κωδικοποιητές οι οποίοι παράγουν ομιλία χαμηλής ποιότητας. Τα τεστ ρίμας ονομάστηκαν έτσι γιατί ο ακροατής πρέπει να καθορίσει πιο σύμφωνο χρησιμοποιείτε όταν του παρουσιάζεται ένα ζεύγος λέξεων που κάνουν ρίμα. Δηλαδή ζητείται από τον ακροατή να ξεχωρίσει μεταξύ ζευγάρια λέξεων όπως είναι

γ-onos, m-onos, p-onos, t-onos, f-onos,
p-aros, f-aros, p-ali, z-ali, και
p –eras, t-eras.

Πιο συγκεκριμένα παρουσιάζεται στον ακροατή μια ομιλούμενη λέξη από ένα ζεύγος και του ζητείτε να αποφασίσει πια λέξη ειπώθηκε. Το τελικό αποτέλεσμα του *διαγνωστικού τεστ ρίμας* είναι οι επί τοις εκατό και υπολογίζεται σύμφωνα με τον τύπο

$$P = \frac{R - W}{T} \times 100$$

όπου R είναι ο αριθμός αυτών που επιλέχθηκαν σωστά , W είναι ο αριθμός των λάθος επιλογών και T είναι ο συνολικός αριθμός των ζευγαριών λέξεων που ελέχθησαν. Συνήθως ισχύει $75 \le DRT \le 95$ με το 90 να ανταποκρίνεται σε ένα καλό σύστημα.

Diagnostic Acceptability Measure (DAM). Το *μέτρο διαγνωστικής αποδεκτικότητας* είναι μια προσπάθεια να γίνει η μέτρηση της ποιότητας ομιλίας περισσότερο συστηματική. Σε αυτό σημαντικό ρόλο παίζει οι ακροατές να έχουν ¨ακουστική¨ εκπαίδευση. Σε αυτούς δίνονται φωνητικά ισορροπημένες προτάσεις όπως είναι *"Cats and dogs each hate the other"* και *"The pipe began to rust while new"*. Οι προτάσεις αυτές είναι παρμένες από μια λίστα προτάσεων του Harvard και έχουν επεξεργαστεί κάθε φορά από τον κωδικοποιητή του ενδιαφέροντος μας. Από τον ακροατή ζητείται να προσδώσει έναν αριθμό μεταξύ του 0 και του 100 σε τρεις κατηγορίες: ποιότητα σήματος, ποιότητα υπόβαθρου και ολική επίδραση. Οι εκτιμήσεις κάθε κατηγορίας σταθμίζονται και χρησιμοποιούνται κατάλληλα. Στο τέλος γίνονται κάποιες διορθώσεις για να αντισταθμιστεί η απόδοση του ακροατή. Ένα τυπικό *DAM* αποτέλεσμα είναι 45 – 55% με το 50% να ανταποκρίνεται σε ένα καλό σύστημα.

Mean Opinion Score (MOS). Η κλίμακα *MOS* είναι το πιο δημοφιλές σύστημα αξιολόγησης. Για να οριστεί η τιμή του *MOS* για ένα κωδικοποιητή , ζητείται από τους ακροατές να ταξινομήσουν την ποιότητα της κωδικοποιημένης ομιλίας σε μια από πέντε κατηγορίες: *έξοχη (5), καλή (4), μέτρια (3), φτωχή (2)* και *κακή (1)*. Εναλλακτικά, μπορεί να ζητηθεί από τους ακροατές να ταξινομήσουν την κωδικοποιημένη ομιλία σύμφωνα με την αναλαμβανόμενη παραποίηση σε μια από τις υποδεέστερες κατηγορίες: *ανεπαίσθητη (5), με δυσκολία αντιληπτή αλλά όχι ενοχλητική (4), αντιληπτή και ενοχλητική (3)*, *ενοχλητική αλλά όχι απαράδεκτη (2) ή πολύ ενοχλητική και δυσάρεστή (1)*. Οι αριθμοί στις παρενθέσεις χρησιμοποιούνται για να προσδώσουν μια αριθμητική αξία στις υποκειμενικές εκτιμήσεις. Στο τέλος υπολογίζεται ο μέσος όρος από την βαθμολόγια

όλων των ακροατών για να δημιουργηθεί το *MOS* για τον κωδικοποιητή. Ένα *MOS* μεταξύ 4.0 και 4.5 συνήθως υποδηλώνει υψηλή ποιότητα. Είναι πολύ βασικός παράγοντας στον υπολογισμό των διαφορών των τιμών του *MOS* είναι η *διακύμανση* της οποίας μεγάλες τιμές έχουν ως αποτέλεσμα την αναξιοπιστία του τεστ. Μεγάλες *διακυμάνσεις* μπορούν να συμβούν γιατί οι ακροατές δεν γνωρίζουν την σημασία των κατηγοριών π.χ. καλή ή κακή , και είναι χρήσιμο ορισμένες φορές να παρουσιάζονται σε αυτούς παραδείγματα καλής ή κακής ομιλίας έτσι ώστε να βαθμονομείται με επιτυχία η κλίμακα των 5 σημείων.

Τιμή του MOS	**Ποιότητα**	**Παραποίηση**
5	Έξοχη	Ανεπαίσθητη
4	Καλή	Με δυσκολία αντιληπτή (όχι ενοχλητική)
3	Μέτρια	Αντιληπτή (ενοχλητική)
2	Φτωχή	Ενοχλητική (όχι απαράδεκτη)
1	Κακή	Πολύ ενοχλητική (δυσάρεστη)

Πίνακας 3.1: Ερμηνεία των τιμών της κλίμακας του MOS.

3.2.2 Αντικειμενικές Τεχνικές Αξιολόγησης Κωδικοποιητών Φωνής

Σε αυτές συγκρίνεται το αρχικό σήμα με αυτό στην έξοδο και γίνονται μετρήσεις βασισμένες στο λόγο σήματος προς θόρυβο (SNR) που δίνεται από τον τύπο:

$$SNR = 10\log_{10}\left\{\frac{\sum_{n=0}^{M} s^2(n)}{\sum_{n=0}^{M}(s(n)-\hat{s}(n))^2}\right\}$$

όπου $s(n)$ είναι το αρχικό σήμα ομιλίας και $\hat{s}(n)$ το κωδικοποιημένο σήμα. Το *SNR* είναι μια *μακροπρόθεσμη* μέτρηση της απόδοσης του συστήματος για αυτό μια καλύτερη αποτίμηση μπορεί να γίνει χρησιμοποιώντας έναν *βραχυπρόθεσμο* λόγο σήματος προς θόρυβο υπολογίζοντας το *SNR* για N – τμήματα της ομιλίας. Έτσι λοιπόν ένα άλλο μέτρο της απόδοσης είναι το *τμηματικό SNR (SEGSNR)* το οποίο δίνεται από τον τύπο:

$$SEGSNR = \frac{10}{L}\sum_{i=0}^{L-1}\log_{10}\left\{\frac{\sum_{n=0}^{N-1} s^2(iN+n)}{\sum_{n=0}^{N-1}(s(iN+n)-\hat{s}(iN+n))^2}\right\}$$

Τέλος σαν κατακλείδα και για τις δύο τεχνικές μπορούμε να πούμε ότι παρόλο που οι αντικειμενικές μετρήσεις είναι ευαίσθητες στις μεταβολές του κέρδους και της καθυστέρησης δεν μπορούν να ληφθούν υπ' όψιν σε σχέση με την αντιληπτική ικανότητα του αυτιού. Επειδή λοιπόν η έννοια της ¨καλής ποιότητας¨ ομιλίας είναι υψηλά ατομική και υποκειμενική για αυτό παρόλο που και τα δύο είδη τεχνικών είναι χρήσιμα, σαν <u>πιο κατάλληλες μέθοδοι μέτρησης κρίνονται οι υποκειμενικές</u>.

3.3 Κατηγορίες Ποιότητας της Ομιλίας

Στις ψηφιακές επικοινωνίες η ποιότητα της ομιλίας ταξινομείται σε τέσσερις (4) γενικές κατηγορίες, αυτές είναι:

Broadcast: Η *broadcast* – ευρείας ζώνης ομιλία αναφέρεται σε υψηλής ποιότητας ομιλία που μπορεί να επιτευχθεί με ρυθμούς bit πάνω από 64 kbits/s.

Network ή toll: Αναφέρεται σε ποιότητα η οποία είναι συγκρίσιμη με την κλασική αναλογική ομιλία (200 – 300 Hz) και μπορεί να επιτευχθεί με ρυθμούς bit πάνω από 16 kbits/s.

Communication: Εδώ εισάγεται σε κάποιο βαθμό μια αλλοίωση της ποιότητας παραμένει ωστόσο φυσική και υψηλά κατανοητή. Μπορεί να επιτευχθεί με ρυθμούς bit πάνω από 4.8 kbits/s.

Synthetic: Είναι σε γενικές γραμμές κατανοητή αλλά μπορεί να αποκτήσει ¨αφύσικη¨ χροιά. Συνδέεται επίσης με την μη αναγνωρισιμότητα του ομιλητή.

3.4 Κατηγορίες Κωδικοποιητών Φωνής

Οι κωδικοποιητές φωνής για να λειτουργήσουν και να επιτύχουν το επιθυμητό αποτέλεσμα χρησιμοποιούν ένα μεγάλο πλήθος τόσο των ιδιοτήτων του σήματος ομιλίας (Παράγραφος 2.3) όσο και των χαρακτηριστικών του ακουστικού συστήματος (Παράγραφος 2.4) αλλά και της φωνητικής περιοχής του ανθρώπου (Παράγραφος 1.5). Επίσης μεγάλη διαφορά μεταξύ τους παρουσιάζει και τρόπος με τον οποίο γίνεται η κβάντιση έτσι έχουμε την *απευθείας κβάντιση* στην οποία έχουμε απευθείας δυαδική αναπαράσταση των δειγμάτων και την *παραμετρική κβάντιση* στην οποία έχουμε δυαδική αναπαράσταση του μοντέλου φωνητικής περιοχής και/ή των φασματικών παραμέτρων. Άλλες διακρίσεις που μπορούν να γίνουν στην κβάντιση είναι *με μνήμη* ή *χωρίς μνήμης*, *προσαρμοστική* ή *μη προσαρμοστική*, *διανυσματική* ή *γραμμική* και *ομοιόμορφη* ή *μη ομοιόμορφη*. Όλες αυτές τις κατηγορίες θα τις διακρίνουμε εκτενέστερα μέσα στις επόμενες παραγράφους.

Με βάση λοιπόν όλες τις παραπάνω παραμέτρους οι κωδικοποιητές φωνής μπορούν να ταξινομηθούν ανάλογα με τον τρόπο λειτουργίας τους σε τρεις γενικές κατηγορίες οι οποίες και είναι:

Κωδικοποιητές Κυματομορφής: Είναι οι κωδικοποιητές οι οποίοι κατά κανόνα βασίζονται στην ύπαρξη πλεονασμού μέσα στο σήμα της ομιλίας και επικεντρώνονται στην ¨έγκυρη¨ αναπαράσταση της κυματομορφής του σήματος της ομιλίας.

Vocoders: Είναι οι κωδικοποιητές οι οποίοι βασίζονται στη *παραμετρική περιγραφή* της ομιλίας και επικεντρώνονται στο να παράγουν αντιληπτή ομιλία χωρίς απαραίτητα να διατηρείτε η κυματομορφή της. Κατά κανόνα έχει αφύσικη ή συνθετική χροιά και το bit rate που επιτυγχάνετε είναι χαμηλότερο από αυτό των κωδικοποιητών κυματομορφής.

Υβριδικοί Κωδικοποιητές: Οι κωδικοποιητές αυτοί συνδυάζουν χαρακτηριστικά και από τις δύο πιο πάνω κατηγορίες. Έτσι έχουν την ικανότητα κωδικοποίησης των *vocoders* και την ποιότητα των *κωδικοποιητών κυματομορφής*. Αυτό το πετυχαίνουν μοντελοποιώντας τις φασματικές ιδιότητες του σήματος της ομιλίας και εκμεταλεύωντας τις ακουστικές ιδιότητες του ανθρώπινου ακουστικού συστήματος μαζί με την ταυτόχρονη διατήρηση της κυματομορφής του σήματος της ομιλίας. Έτσι τελικά επιτυγχάνουν ¨ικανοποιητική¨ ποιότητα ομιλίας σε πολύ χαμηλά bit rate. Στην επόμενη παράγραφο θα δούμε λίγο πιο αναλυτικά τις διάφορες κατηγορίες κωδικοποιητών φωνής και τις διάφορες υποκατηγορίες τους για να αποκτήσουμε μια πληρέστερη εικόνα.

3.5 Κωδικοποιητές Κυματομορφής

Οι κωδικοποιητές κυματομορφής χωρίζονται σε δύο μεγάλες κατηγορίες στους *κωδικοποιητές κυματομορφής χρόνου* και στους *κωδικοποιητές κυματομορφής συχνότητας*. Οι κωδικοποιητές και οι αποκωδικοποιητές της πρώτης κατηγορίας χρησιμοποιούν συνήθως έναν αλγόριθμο πρόγνωσης ο οποίος βασίζεται σε στατιστικές ιδιότητες του σήματος, και οι πιο εξελιγμένες μορφές αυτών κβαντίζουν μόνο την πρόγνωση λάθους, επίσης η λειτουργία τους βασίζεται και σε όλες εκείνες τις παραμέτρους που αναφέρουμε στην προηγούμενη παράγραφο. Ξεκινώντας λοιπόν από το απλό γραμμικό PCM βλέπουμε ότι σε αυτό δεν γίνονται υποθέσεις για το προς κωδικοποίηση σήμα και για αυτό και απαιτεί το μεγαλύτερο bit rate για να μας δώσει *toll* ποιότητα ομιλίας. Προχωρώντας τώρα προς το λογαριθμικό PCM βλέπουμε ότι για την υλοποίηση του γίνεται και χρήση των ιδιοτήτων της *φασματικής επικάλυψης (spectral masking)* στην ανθρώπινη ακοή *(aural noise masking)* με αποτέλεσμα να επιτυγχάνουμε λίγο χαμηλότερο bit rate για *toll* ποιότητα ομιλίας από ότι στο γραμμικό PCM. Στους *Διαφορικούς* (differential) *Κωδικοποιητές* εκμεταλλευόμαστε την συσχέτιση μεταξύ των δειγμάτων (Παράγραφος 2.3.2) και

το μοντέλο φωνητικού σωλήνα (Παράγραφος 1.5). Τους διακρίνουμε σε *Προβλέψεις Μικρής Διάρκειας (Sort Term Prediction)* όπως είναι η *Διαφορική Παλμοκωδική Διαμόρφωση (ADPCM), η Διαμόρφωση Δέλτα (DM)*, η *Προσαρμοστική Διαφορική Παλμοκωδική Διαμόρφωση* (ADPCM) κ.α. και σε *Προβλέψεις Μεγάλης και Μικρής Διάρκειας* όπως είναι ο *Προσαρμοστική Προβλεπτική Κωδικοποίηση (APC).*

Στους *κωδικοποιητές κυματομορφής συχνότητας* τώρα κωδικοποιείτε μια μετασχηματισμένη έκδοση του σήματος ομιλίας σε αντίθεση με του κωδικοποιητές χρόνου όπου κωδικοποιείτε το ίδιο το σήμα. Εδώ χρησιμοποιούνται τα χαρακτηριστικά των σημάτων ομιλίας στο πεδίο της συχνότητας για να εξαλείψουν τον πλεονασμό. Σαν παράδειγμα μπορούμε να αναφέρουμε τον *Κωδικοποιητή Υπό – ζωνών (SBC)* όπου διαιρείτε το φάσμα του σήματος εισόδου σε ξεχωριστές ζώνες χρησιμοποιώντας φίλτρα διέλευσης ζώνης. και το σήμα το οποίο περνά από κάθε μια από αυτές τις ζώνες κωδικοποιείτε ξεχωριστά. Μια πιο πολύπλοκη τεχνική κωδικοποίησης συχνότητας είναι η *Προσαρμοστική Κωδικοποίηση Μετασχηματισμού (ATC)*. Αυτή βασίζεται στον μετασχηματισμό σε μπλοκ, των τμημάτων εισόδου της κυματομορφής της ομιλίας. Κάθε τμήμα αναπαριστάτε από ένα σύνολο από συντελεστές μετασχηματισμού, σε σύγκριση με λιγότερο σημαντικούς συντελεστές. Στον δέκτη ένας αντίστροφος μετασχηματισμός χρησιμοποιείτε για την ανασύνθεση του σήματος ομιλίας. Τα περισσότερο πρακτικά συστήματα κωδικοποίησης μετασχηματισμού για ομιλία είναι προσαρμοστικά στο γεγονός ότι η κατανομή των bits σε κάθε συντελεστή αλλάζει από πλαίσιο σε πλαίσιο. Αυτή η δυναμική κατανομή των bits ελέγχεται από τις χρονικά μεταβαλλόμενες στατιστικές τους σήματος ομιλίας , οι οποίες έχουν μεταδοθεί σαν πλευρική πληροφορία.

3.6 Vocoders

Μια γενική αναφορά στα μοντέλα στα οποία στηρίζονται οι vocoders έχει γίνει στη Παράγραφο 1.5 *(μοντέλο φωνητικού σωλήνα)*. Όπως είναι γνωστό ένας vocoder κωδικοποιεί μόνο της αντιληπτές και βασικές παραμέτρους της φωνής. Αυτή η παραμετρική περιγραφή της ομιλίας μπορεί να πάρει μια ποικιλία μορφών για παράδειγμα, είτε σε τιμές πλάτους που αποτιμούνται σε συγκεκριμένες συχνότητες, για ένα μικρό διάστημα του φάσματος της φωνής (όπως γίνεται στον *channel vocoder*), είτε σε τιμές συχνότητας των βασικών αντηχήσεων (*formant vocoder*), είτε σε γραμμικούς προβλεπτικούς συντελεστές οι οποίοι προβλέπουν το δείγμα της φωνής (*LP vocoder*), είτε τέλος σε αριθμό αρμονικών που προκύπτουν από την επεξεργασία του σήματος (*homomorphic vocoder)*.

Έτσι στους *channel vocoders* μεταδίδονται τα επίπεδα της ισχύος του σήματος μαζί με πληροφορίες για το τμήμα της φωνής (δηλαδή αν αυτό είναι έμφωνο ή άφωνο) για στενές υπό – ζώνες του φάσματος ενός μικρού χρονικού τμήματος της ομιλίας. Στο συνθέτη τώρα έχουμε ένα χρονικά μεταβαλλόμενο φίλτρο (το οποίο αποτελείται από φίλτρα στενής ζώνης) το οποίο λειτουργεί με τέτοιο τρόπο, ώστε η μορφή του φάσματος του συντεθειμένου σήματος να ¨ανταποκρίνεται¨ στη μορφή του φάσματος του αρχικού σήματος. Η είσοδος αυτού του φίλτρου ενισχύεται με μια ακολουθία από ψευδό – περιοδικούς παλμούς για τα *έμφωνα* τμήματα και με λευκό θόρυβο για τα άφωνα τμήματα. Το αποτέλεσμα είναι υψηλής ποιότητας κατανοητή συνθετική[8] ομιλία για bit rate της τάξης των 1 – 2 kb/s.

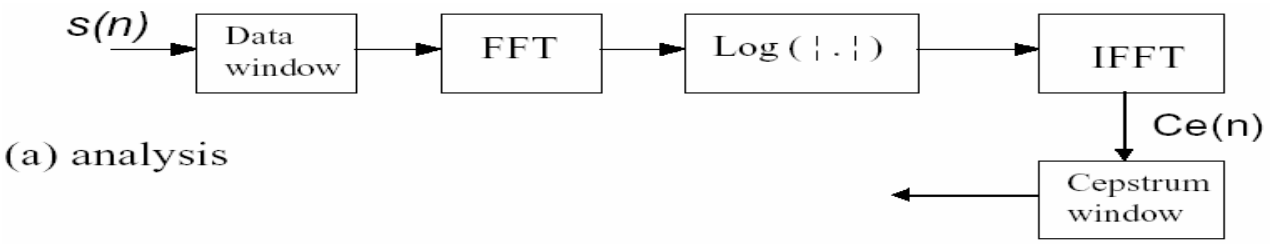
Οι λειτουργία των *formant vocoders* τώρα βασίζεται στο γεγονός ότι οι η φασματική πυκνότητα ενός τμήματος ομιλίας βραχύς διάρκειας τείνει να είναι συγκεντρωμένη σε τρία ή τέσσερα formants. Έτσι η ομιλία σε αυτού του είδους τους κωδικοποιητές παριστάνεται σαν μια χρονικά μεταβαλλόμενη πρόγνωση των formants, λαμβάνονται δηλαδή μόνο οι συχνότητες των formants και τα εύροι τους αντί για ολόκληρη την μορφή του φάσματος. Για να το επιτύχουν αυτό οι *formant vocoders* πρέπει να εντοπίσουν ακριβώς τις αλλαγές στις formant συχνότητες. Η

[8] Επειδή σε μια συνηθισμένη ομιλία δεν περιλαμβάνονται μόνο έυφωνα ή άφωνα τμήματα

διαδικασία αυτή είναι πιο πολύπλοκη από την διαδικασία των *channel vocoders* και μας δίνει υψηλή ποιότητα κωδικοποιημένης φωνής με bit rate της τάξης των 500 – 1500 kb/s.

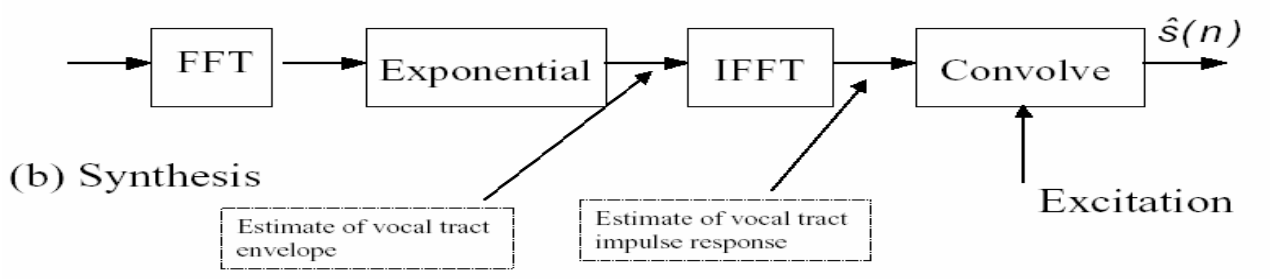
Οι LP vocoders βασίζονται σε ένα χρονικά μεταβαλλόμενο *μοντέλο φωνητικού σωλήνα* του οποίου οι παράμετροι και οι συντελεστές ανανεώνονται περιοδικά. Η κωδικοποίηση του σήματος της ομιλίας εδώ γίνεται με τον διαχωρισμό του σε τμήματα και με την εξέταση του κάθε τμήματος. Έτσι προκύπτει ένα πλήθος παραμέτρων οι οποίοι προσδιορίζουν αν το τμήμα είναι έμφωνο ή άφωνο, τη θεμελιώδη συχνότητα του τμήματος,, το κέρδος του κ.α.. Όλοι αυτοί οι παράμετροι μεταδίδονται περιγραφόμενοι από κάποιο αριθμό συντελεστών στον δέκτη (¨συνθέτη¨), ο οποίος προσπαθεί να ξαναδημιουργεί την ομιλία περνώντας τις παραμέτρους μέσα από ένα μαθηματικό μοντέλο (*synthesis filter*) της φωνητικής περιοχής. Η αδυναμία και σε αυτόν το κωδικοποιητή εντοπίζεται στο γεγονός ότι η ομιλία δεν περιέχει μόνο *έμφωνα* ή άφωνα τμήματα, ωστόσο μας δίνει ιδιαίτερα χαμηλά bit rate.

Τέλος οι *homomorphic vocoders* μεταδίδουν τον αντίστροφο μετασχηματισμό Fourier του λογαρίθμου του φάσματος της ομιλίας (ονομάζεται *cepstrum)* (Σχήμα 3.2) για διάφορα τμήματα ο οποίος περιέχει πληροφορίες για την φωνητική περιοχή και τους γλωττικούς παλμούς. Τα δείγματα από το *cepstrum* ονομάζονται *quefrencies*. Οι χαμηλές *quefrencies* συσχετίζονται με την κρουστική απόκριση του συστήματος ενώ οι υψηλότερες συσχετίζονται με την διέγερση και μας δίνουν πληροφορίες για την βασική συχνότητα.



Σχήμα 3.2: Μπλοκ διάγραμμα ενός κωδικοποιητή *homomorphic.*

Στον αποκωδικοποιητή ακολουθείτε η αντίστροφη διαδικασία (Σχήμα 3.3) όπου μετά τον μετασχηματισμό Fourier εφαρμόζουμε έναν εκθετικό παράγοντα με τον οποίο εξάγουμε την μορφή του φάσματος της φωνητικής περιοχής στην συνέχεια ακολουθεί ένα αντίστροφος μετασχηματισμός Fourier από τον οποίο παίρνουμε την κρουστική απόκριση της φωνητικής περιοχής. Το τελικό αυτό σήμα με συγκερασμό με την διέγερση μας δίνει την τελική συντιθέμενη ομιλία.

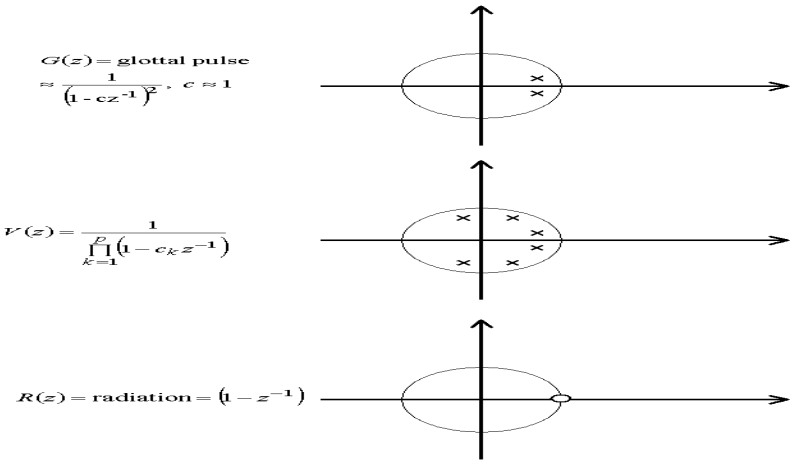


Σχήμα 3.3: Μπλοκ διάγραμμα ενός αποκωδικοποιητή *homomorphic.*

3.7 Γραμμική Πρόγνωση (Linear Prediction)

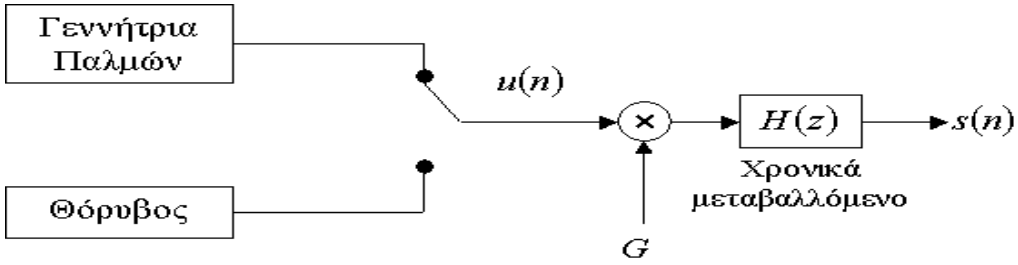
Η γραμμική πρόγνωση βασίζεται στο *μοντέλο φωνητικού σωλήνα* (Παράγραφος 1.5) το οποίο όπως έχουμε δει μας δίνεται μέσα από την εξίσωση $H(z) = G(z)V(z)R(z)$ με

$G(z)=\dfrac{1}{(1-cz^{-1})^2}, c\approx 1$ *απόκριση γλωττικού παλμού (πηγή διέγερσης),* $R(z)=(1-z^{-1})$ *συνάρτηση ακτινοβολίας* και $V(z)=\dfrac{1}{\prod_{k=1}^{p}(1-c_k z^{-1})}$ *συνάρτηση μεταφοράς φωνητικού σωλήνα.*



Σχήμα 3.4: Γραφική απεικόνιση πόλων και μηδενικών των τριών συναρτήσεων του μοντέλου φωνητικού σωλήνα.

Το ψηφιακό αυτό μοντέλο παριστάνεται γραφικά στο Σχήμα 3.3 και όπως ξέρουμε αποτελείται από μια γεννήτρια παλμών, μια γεννήτρια θορύβου (τα οποία και τα δύο μαζί απαρτίζουν την πηγή διέγερσης), τον διακόπτη επιλογής της διέγερσης, το κέρδος και το χρονικά μεταβαλλόμενο φίλτρο - *synthesis filter* (που παίζει τον ρόλο του προγνώστη).



Σχήμα 3.5: Γραφική απεικόνιση ψηφιακού μοντέλου παραγωγής φωνής.

Η πιο γενική μορφή τώρα του προγνώστη αυτού, είναι ένα μοντέλο γνωστό ως autoregressive moving average (ARMA) στο οποίο η ομιλία προκύπτει τόσο από *p* προηγούμενα δείγματα πρόγνωσης $\hat{s}[n-1],...,\hat{s}[n-p]$ όσο και το σήμα της διέγερσης σύμφωνα με την σχέση $\hat{s}(n)=\sum_{k=1}^{p}a_k\hat{s}(n-k)+G\sum_{l=0}^{q}b_l u[n-l], b_o=1$ (Σχήμα 3.5). Εδώ τα a_i, b_l είναι οι συντελεστές πρόγνωσης, p η τάξη του προγνώστη και G η παράμετρος του κέρδους. Στο πεδίο της συχνότητας τώρα το *synthesis filter* δίνεται μέσα από την σχέση $H(z)=\dfrac{1+\sum_{l=1}^{p}b_l z^{-l}}{1-\sum_{k=1}^{p}a_k z^{-k}}$ σύμφωνα με την οποία προκύπτει ένα μοντέλο με πόλους και μηδενικά. Στην περίπτωση τώρα όπου ισχύει $a_k=0$ για

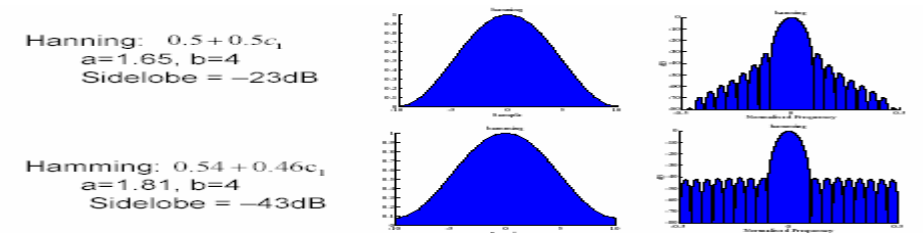
$1 \le k \le p$ το H(z) γίνεται ένα μοντέλο μηδενικών ή *moving average (MA)* μιας και η έξοδος είναι η ¨σταθμισμένη¨ μέσος όρος των q προηγούμενων εισόδων. Αντίστοιχα στην περίπτωση όπου $b_l = 0$ για $1 \le l \le q$ το H(z) γίνεται ένα μοντέλο πόλων ή *autoregressive (AR)* στο οποίο η πρόγνωση γίνεται με βάση την εξίσωση $\hat{s}(n) = \sum_{k=1}^{p} a_k s(n-k)$ η οποία μεταφράζεται στο πεδίο της συχνότητας $\mathrm{H(z)} = \frac{1}{1 - \mathrm{A(z)}}$ όπου $A(z) = \sum_{k=1}^{p} a_k z^{-k}$. Το σφάλμα τώρα το οποίο προκύπτει λόγω του προγνώστη στο *AR* μοντέλο (ονομάζεται και *σφάλμα πρόγνωσης*) δίνεται μέσα από την εξίσωση $e(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k)$ και μπορούμε να το πάρουμε με την χρήση του αντίστροφου φίλτρου $\frac{1}{H(z)} = 1 - A(z)$.

Με βάση την $H(z)$ μπορούμε να έχουμε μια καλή μοντελοποίηση του φάσματος. Τα μειονεκτήματα βέβαια εδώ είναι ότι δεν υπάρχει ικανοποιητική αναπαράσταση όταν έχουμε ¨φασματικά μηδενικά¨ (spectral zeros) τα οποία και δεν συνεισφέρουν στο πλάτος, και ότι δεν υπάρχει ικανοποιητική πρόγνωση για τα άφωνα τμήματα. Ωστόσο με ένα αριθμό συντελεστών p περίπου στο 10 μπορούμε να πούμε ότι έχουμε αποδεκτά αποτελέσματα.

Για την επιτυχή μετάδοση τώρα όλων των απαραίτητων παραμέτρων στον δέκτη χρησιμοποιούμε *ανάλυση βραχέως χρόνου* (short – time analysis) θεωρούμε δηλαδή ότι το σήμα της ομιλίας είναι στατικό για μικρά τμήματα (των 20 ms) και κατά αυτό τον τρόπο μπορούμε να πάρουμε ένα παράθυρο N δειγμάτων αρκεί το N να είναι αρκετά μικρό. Πετυχαίνουμε έτσι να μοντελοποιήσουμε το σήμα της φωνής με διαδοχικά φίλτρα $H(z)$ των οποίων οι συντελεστές παραμένουν σταθεροί μέσα σε ένα παράθυρο. Ο υπολογισμός τώρα των συντελεστών γίνεται με βάση το κριτήριο του τετραγωνικού σφάλματος πρόγνωσης, επιλέγονται δηλαδή έτσι ώστε να ελαχιστοποιηθεί το *συνολικό τετραγωνικό σφάλμα πρόγνωσης* $E = \sum_{n} e^2(n)$. Με βάση το κριτήριο αυτό υπάρχουν διάφορες προσεγγίσεις από τις οποίες οι ποιο διαδεδομένες είναι η *μέθοδος αυτοσυσχέτισης* και η *μέθοδος συμμεταβλητότητας*.

3.7.1 Μέθοδος Αυτοσυσχέτισης

Στην μέθοδο της αυτοσυσχέτισης το επιλεγμένο παράθυρο το οποίο είναι συνήθως παράθυρο *Hamming* (ή υβριδικό *Hamming – cosine*) ολισθαίνει πάνω στο σήμα της φωνής, έτσι ώστε οι συντελεστές να υπολογιστούν σε τακτά χρονικά διαστήματα (της τάξης των 10 ms). Το παράθυρο *Hamming* είναι της μορφής $w(n) = \begin{cases} 0{,}54 - 0{,}46\cos(\frac{2\pi n}{N-1}), 0 \le n \le N-1 \\ 0, \text{αλλού} \end{cases}$ και έχει επιλεγεί στην συγκεκριμένη περίπτωση γιατί ελαχιστοποιεί τις παρεμβολές (ο κύριος λοβός απέχει κατά πολύ από τους πλευρικούς - Σχήμα 3.6).



Σχήμα 3.6: Διάφορα παράθυρα Hamming.

Το γινόμενο αυτό του σήματος επί το παράθυρο ορίζει ένα νέο σήμα το οποίο έχει άπειρη έκταση, αλλά είναι μηδέν έξω από το παράθυρο αυτό. Αυτό μας δίνει την δυνατότητα να μπορούμε να υπολογίσουμε την πραγματική συνάρτηση αυτοσυσχέτισης για όλο το σήμα. Αν τώρα υποθέσουμε ότι έχουμε ένα παράθυρο μήκος *N* το οποίο και βρίσκεται στην θέση μηδέν τότε το γινόμενο είναι $s_w(n) = w(n)s(n)$ και το συνολικό τετραγωνικό σφάλμα μπορεί να υπολογιστεί με βάση την εξίσωση:

$$E = \sum_{n=-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} \left(s_w(n) - \sum_{k=1}^{p} a_k s_w(n-k) \right)^2$$

Όμως επειδή το συνολικό σφάλμα είναι τετραγωνική συνάρτηση των συντελεστών a_k υπάρχει ένα ολικό ακρότατο το οποίο μπορεί να επιτευχθεί παραγωγίζοντας το σφάλμα ως προς του συντελεστές a_k, $\frac{\partial \mathrm{E}}{\partial a_k} = 0$ για $k = 1,2,3,..., p$. Αυτό μας οδηγεί σε *p* γραμμικές εξισώσεις με *p* άγνωστα a_k οι οποίες είναι γνωστές και σαν εξισώσεις *Yule – Walker*:

$$\sum_{k=1}^{p} a_k \sum_{n=-\infty}^{\infty} s_w(n-i)s_w(n-k) = \sum_{n=-\infty}^{\infty} s_w(n-i)s_w(n),\ 0 \le i \le p$$

Από τον ορισμό τώρα της συνάρτησης αυτοσυσχέτισης για το $s_w(n)$ ως εξής $R(i) = \sum_{n=i}^{N-1} s_w(n)s_w(n-i),\ 0 \le i \le p$ και λόγω της αρτιότητας της, $R(i) = R(-i)$ οι εξισώσεις μπορούν να γραφούν $\sum_{k=1}^{p} R(|i-k|)a_k = R(i),\ 0 \le i \le p$ ή σε μορφή πινάκων

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \mathrm{K} & R(p-2) \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ R(p-1) & R(p-2) & \mathrm{K} & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \mathrm{M} \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \mathrm{M} \\ R(p) \end{bmatrix}$$

Από τις εξισώσεις αυτές βλέπουμε ότι έχουμε ένα πίνακα με ειδική δομή, είναι συμμετρικός και Toeplitz (τα στοιχεία στη διαγώνιο και όλες τις αντιδιαγωνίους έχουν σταθερή τιμή). Αυτό μας δίνει την δυνατότητα να εφαρμόσουμε για την επίλυση πολύ γρήγορους αλγόριθμους όπως είναι ο *Levinson – Durbin* και ο *Schur* των οποίων η πολυπλοκότητα είναι $O(p^2)$ αντί για $O(p^3)$ που είναι υπό κανονικές συνθήκες. Επίσης λόγω του πίνακα Toeplitz το $A(z)$ έχει τους μηδενισμούς του μέσα στον μοναδιαίο κύκλο, με αποτέλεσμα το *synthesis filter* $\mathrm{H(z)} = \frac{1}{1 - \mathrm{A(z)}}$ να είναι σταθερό. Αυτό μας δίνει ένα πρόσθετο κίνητρο για εφαρμογή της μεθόδου αυτοσυσχέτισης στην γραμμική πρόγνωση.

3.7.2 Μέθοδος Συμμεταβλητότητας

Στην μέθοδο της συμμεταβλητότητας σε αντίθεση με την μέθοδο της αυτοσυσχέτισης δεν πολλαπλασιάζεται το σήμα με κάποιο παράθυρο, αλλά η ακολουθία του *σφάλματος πρόγνωσης* $e(n)$, οπότε έχουμε την εξίσωση $E = \sum_{n=-\infty}^{\infty} {e_w}^2(n) = \sum_{n=-\infty}^{\infty} e^2(n)w(n)$. Για την εύρεση τώρα του ελαχίστου εφαρμόζουμε πάλι την μέθοδο της μερικής παραγώγου $\frac{\partial \mathrm{E}}{\partial a_k} = 0$ για $k = 1,2,3,..., p$ η οποία σε συνδυασμό με την εξίσωση της συνάρτησης συμμεταβλητότητας η οποία ορίζεται ως εξής

$\phi(i,k) = \sum_{n=-\infty}^{\infty} w(n)s(n-i)s(n-k)$ μας δίνει p γραμμικές εξισώσεις της μορφής $\sum_{k=1}^{p} \phi(i,k)a_k = \phi(i,0)$ με $1 \le i \le p$. Οι εξισώσεις βέβαια αυτές μπορούν να γραφούν με την μορφή πίνακα

$$\begin{bmatrix} \phi(1,1) & \phi(1,2) & \text{K} & \phi(1,p) \\ \phi(2,1) & \phi(2,2) & \text{K} & \phi(2,p) \\ \text{M} & \text{M} & \text{O} & \text{M} \\ \phi(p,1) & \phi(p,2) & \text{K} & \phi(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \text{M} \\ a_p \end{bmatrix} = \begin{bmatrix} \psi(1) \\ \psi(2) \\ \text{M} \\ \psi(p) \end{bmatrix}$$

με $\psi(i) = \psi(i,0)$ για $i = 1,2,...,p$ ο οποίος είναι συμμετρικός όχι όμως και Toeplitz, πράγμα το οποίο κάνει την συγκεκριμένη μέθοδο λιγότερο αποτελεσματική από την μέθοδο της αυτοσυσχέτισης, μιας και η πολυπλοκότητα εδώ είναι ανάλογη του p^3 και όχι του p^2. Βέβαια, η μέθοδος συμμεταβλητότητας δίνει καλύτερα αποτελέσματα, χωρίς όμως να εγγυάται σταθερότητα. Για τον λόγο αυτό και έχουν αναπτυχθεί επιπρόσθετες τεχνικές όπως είναι η μέθοδος *closed – phase covariance.*

3.7.3 Τάξη του Προγνώστη

Εκτός από τις παραπάνω μεθόδους που εφαρμόζονται για τον υπολογισμό των συντελεστών υπάρχουν και ένα πλήθος επιμέρους λεπτομερειών στην *γραμμική πρόγνωση* οι οποίες πρέπει να ληφθούν υπόψη. Μια από αυτές είναι η *τάξη του προγνώστη* η οποία καθορίζεται με βάση την ανάγκη να είναι αρκετά μεγάλη έτσι ώστε το να περιγράφεται κάθε *formant* της ομιλίας με δύο πόλους. Η ανάγκη αυτή προκύπτει από την σύγκριση την οποία έκαναν οι *Atal and Schroeder* για την αναπαράσταση των ένρινων ήχων τόσο με μηδενισμούς όσο και με πόλους.

Βέβαια εδώ υπάρχουν και πρακτικοί περιορισμοί, όπως το γεγονός ότι η ενέργεια του *σφάλματος πρόγνωσης* μειώνεται όσο ο αριθμός *P* των πόλων του φίλτρου $H(z)$ αυξάνει, αλλά και ότι πρέπει διατηρούμε την πολυπλοκότητα σε χαμηλά επίπεδα. Έτσι ένας τρόπος επιλογής του *P*, είναι να ορίσουμε ένα κατώφλι μετά το οποίο το σφάλμα δεν αυξάνεται σημαντικά. Αν λοιπόν το κατώφλι αυτό είναι t_e και αν $1 - \frac{E_{p+1}}{E_p} < t_e$ τότε επιλέγουμε $P = p$. Στην πράξη για σήμα ομιλίας των 8 kHz χρησιμοποιούνται προγνώστες της τάξης από 10 έως 16.

3.7.4 Προέμφαση

Η χρήση της *προέμφασης* έχει σαν σκοπό, την ενίσχυση των υψηλών συχνοτήτων του σήματος της φωνής. Αυτή η ανάγκη προκύπτει λόγο της ιδιότητας που παρουσιάζει το φάσμα της ανθρώπινης ομιλίας να έχουμε μείωση κατά 6 dB/οκτάβα πράγμα το οποίο πρακτικά σημαίνει ότι το πλάτος ελαττώνεται κατά ½ για κάθε διπλασιασμό της συχνότητας. Το γεγονός αυτό οδηγεί σε ανακριβείς προσεγγίσεις των υψηλότερων *formants* και καμία φορά σε προβληματικές γραμμικές εξισώσεις. Για την ελαχιστοποίηση λοιπόν αυτών των προβλημάτων χρησιμοποιείται ένα σταθερό *φίλτρο προέμφασης* πριν τον υπολογισμό των συντελεστών του *synthesis filter* $H(z)$. Η συνάρτηση μεταφοράς τώρα αυτού του *φίλτρου προέμφασης* είναι $V_{pre}(z) = 1 - az^{-1}$ με $0,9 \le a \le 1,0$, ενώ η βέλτιστη τιμή του a βρίσκεται με στατιστικό τρόπο και είναι διαφορετική για διαφορετικούς ομιλητές. Ωστόσο μια τυπική τιμή του είναι α=0,95.

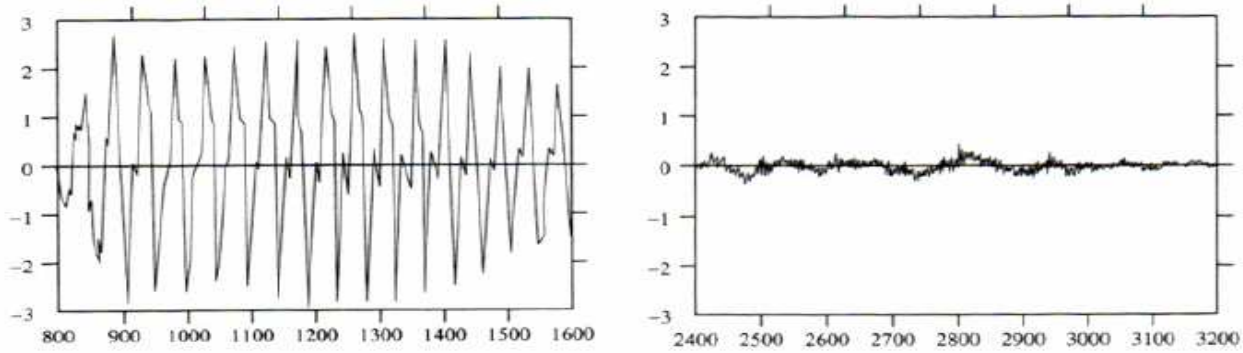
3.7.5 Υπολογισμός Κέρδους

Εκτός όμως από τους παραπάνω ένας τρίτος σημαντικός παράγοντας που πρέπει να αναφέρουμε είναι το *κέρδος*. Ο υπολογισμός του γίνεται τέτοιο τρόπο ώστε η ενέργεια του αρχικού σήματος εισόδου *s(n)* να είναι ίση με την ενέργεια της κρουστικής απόκρισης *h(n)* η οποία δίνεται

από την εξίσωση $h(n)=\sum_{k=1}^{p}a_k h(n-k)+G\delta(n)$. Πρέπει δηλαδή να ισχύει $\sum_n s^2(n)=\sum_n h^2(n)$ με βάση την οποία καταλήγουμε στην ακόλουθη έκφραση για το κέρδος $G^2=R(0)-\sum_{k=1}^{p}a_k R(k)$.

3.7.6 Καθορισμός Έμφωνου/Άφωνου Τμήματος

Ο προσδιορισμός του τμήματος της φωνής που εξετάζεται κάθε φορά σε έμφωνο ή άφωνο είναι ο επόμενος σημαντικός παράγοντας. Ο προσδιορισμός αυτός μπορεί να γίνει με πολλούς τρόπους μιας και όπως έχουμε αναφέρει αυτά διαφέρουν δραματικά. Έτσι μια από τις συνήθεις μέθοδοι που ακολουθούνται είναι η εξέταση της ενέργειας για σήματα βραχέως χρόνου. Η ενέργεια σε αυτή την περίπτωση δίνεται από τον τύπο $E_n=\sum_{n=1}^{N}s^2{}_w(n)$ με N το μήκος του τμήματος (παραθύρου). Τα τμήματα εκείνα τώρα που αντιστοιχούν σε μεγάλη ενέργεια είναι *έμφωνα* ενώ *άφωνα* είναι τα τμήματα με μικρή ενέργεια. Μια δεύτερη μέθοδος λαμβάνει υπόψη της τον αριθμό που τέμνει το συγκεκριμένο τμήμα της φωνής των άξονα των x. Σε γενικές γραμμές περιμένουμε η κυματομορφή ενός άφωνου τμήματος να περάσει τον άξονα των x πιο πολλές φορές από την κυματομορφή ενός *έμφωνου* τμήματος (Σχήμα 3.7). Για αυτό και η διάκριση της ομιλίας μπορεί να γίνει μετρώντας το πόσες φορές περνά η κυματομορφή από των άξονα των x και συγκρίνουμε αυτή την τιμή με τις συνήθεις τιμές των *έμφωνων* και άφωνων τμημάτων.



Σχήμα 3.7: Σύγκριση έμφωνων και άφωνων τμημάτων στα οποία διακρίνεται ότι ο αριθμός τομής του άξονα *x* από το έμφωνο τμήμα είναι κατά πολύ μεγαλύτερος από αυτόν για το άφωνο. Το έμφωνο τμήμα αντιστοιχεί στο γράμμα ¨e¨ της λέξης ¨test¨ και το άφωνο στο γράμμα ¨s¨ της ίδια λέξης.

Επιπρόσθετα μια τρίτη μέθοδος η οποία μπορεί να χρησιμοποιηθεί σαν συμπληρωματική είναι η εξέταση της θέσης του αναγνωρισμένου τμήματος. Σύμφωνα με αυτή ελέγχουμε π.χ. εάν ένα τμήμα που χαρακτηρίζεται ως άφωνο βρίσκεται στην μέση *έμφωνων* τμημάτων, γεγονός που υποδηλώνει εσφαλμένη εκτίμηση.

3.7.7 Pitch Detection

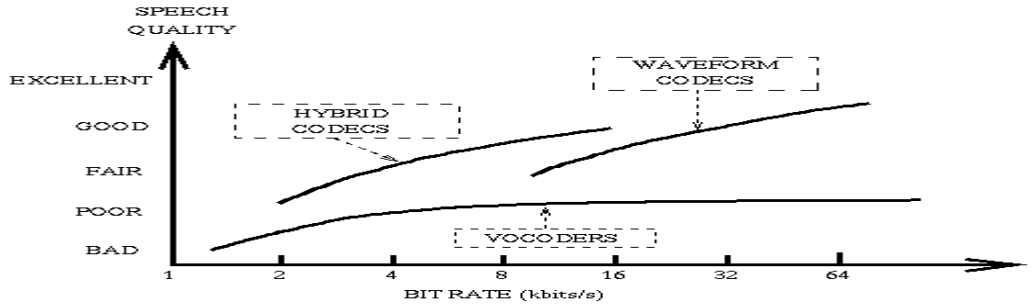
Μια ακόμα σημαντική παράμετρος είναι και η *θεμελιώδης συχνότητα* έμφωνων τμημάτων της ομιλίας η οποία μεταδίδεται μαζί με τους υπολογιζόμενους συντελεστές του $H(z)$ και το κέρδος G. Για την εύρεση της υπάρχουν διάφορες προσεγγίσεις οι οποίες μπορούν να χωριστούν τόσο στο πεδίο του χρόνου όσο και στο πεδίο της συχνότητας. Συνήθως στο πεδίο του χρόνου το σήμα της ομιλίας υφίσταται μια επεξεργασία έτσι ώστε να βρεθεί η *θεμελιώδης συχνότητα* ενώ στο πεδίο της συχνότητας αυτή βρίσκεται κυρίως με βάση την φασματική πληροφορία. Ενδεικτικά μπορούμε να αναφέρουμε μερικές από τις μεθόδους που χρησιμοποιούνται για αυτό τον σκοπό όπως είναι η Average Magnitude Difference Function (AMDF) Pitch Extractor, η peak-picking the autocorrelation sequence of center-clipped speech., η peak-picking the cepstrum και άλλες.

3.7.8 Κβάντιση των Παραμέτρων της Γραμμικής Πρόγνωσης

Τέλος πρέπει να πούμε ότι στην γραμμική πρόγνωση ιδιαίτερη σημασία αποδίδεται και στη κβάντιση των μεταδιδόμενων παραμέτρων (συντελεστές του *synthesis filter* $H(z)$, περίοδος των έμφωνων τμημάτων, κέρδος, παράγοντας ο οποίος καθορίζει εάν το τμήμα της φωνής είναι έμφωνο ή άφωνο). Έτσι αν και ο κβαντισμός της περιόδου, του κέρδους και του παράγοντα άφωνου/έμφωνου τμήματος γίνεται εύκολα με βαθμωτούς κβαντιστές, η κβάντιση των συντελεστών πρόγνωσης ακολουθεί διαφορετική προσέγγιση. Αυτό οφείλεται στο γεγονός ότι οι συντελεστές αυτοί απαιτούν μεγάλο αριθμό bit για να περιγραφούν επειδή είναι εξαιρετικά ευαίσθητοι στα σφάλματα κβάντισης. Για τον λόγο αυτό και έχουν προταθεί διάφορες ισοδύναμες μορφές οι οποίες είναι λιγότερο ευαίσθητες στα σφάλματα αυτά.

3.8 Υβριδικοί Κωδικοποιητές

Όπως έχουμε στις δύο παραπάνω παραγράφους οι κωδικοποιητές κυματομορφής παρουσιάζουν καλή ποιότητα ομιλίας για υψηλά όμως bit rate ενώ οι vocoders παρουσιάζουν χαμηλά bit rate με μέτρια (toll) ποιότητα ομιλίας (Σχήμα 3.8). Το χάσμα μεταξύ αυτών των δύο τύπων κωδικοποιητών έρχονται να μειώσουν οι υβριδικοί κωδικοποιητές οι οποίοι συνδυάζουν δύο ή και περισσότερες τεχνικές κωδικοποίησης σε μία.



Σχήμα 3.8: Ταξινόμηση των κωδικοποιητών σε σχέση με το bit rate και την ποιότητα ομιλίας.

Και εδώ λοιπόν, όπως και στους vocoders βασιζόμαστε στο *μοντέλο φωνητικού σωλήνα* (Παράγραφος 1.5). Με βάση το μοντέλο αυτό ως γνωστό μεταδίδονται διάφορες παράμετροι της ομιλίας οι οποίες εδώ όμως ιδίως για την *πηγή διέγερσης* δεν περιορίζονται στην ¨απλουστευμένη¨ περιγραφή των *έμφωνων* τμημάτων με περιοδικά σήματα και των *άφωνων* με θόρυβο, άλλα το σήμα διέγερσης βελτιστοποιείτε και κωδικοποιείτε αποτελεσματικά με κωδικοποίηση κυματομορφής.

Οι υβριδικοί κωδικοποιητές ανήκουν στην κατηγορία των *analysis – by – synthesis (AbS)* κωδικοποιητών (Σχήμα 3.9) οι οποίοι αρχικά χωρίζουν την ομιλία σε πλαίσια, συνήθως των 20ms. Στην συνέχεια για κάθε πλαίσιο καθορίζονται οι παράμετροι για το *synthesis filter* καθώς και το σήμα διέγερσης, το οποίο είναι ένα πλήθος διαφορετικών προσεγγίσεων του αρχικού σήματος, έτσι ώστε να έχουμε το μικρότερο δυνατό λάθος μεταξύ του δύο σημάτων (συντιθέμενου και αρχικού). Το *synthesis filter* είναι της όπως είδαμε και στην *γραμμική πρόγνωση* μορφής

$$H(z)=\frac{1}{A(z)}$$

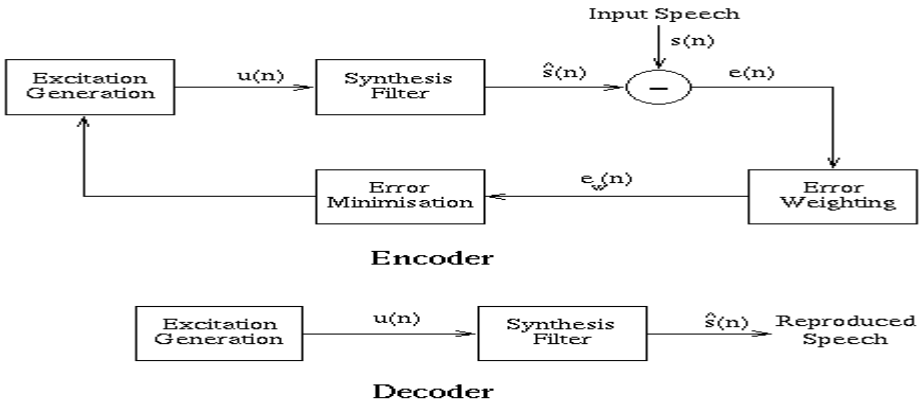
όπου

$$A(z)=1-\sum_{i=1}^{p}a_i z^{-i}$$

και το $\sum_{i=1}^{p}a_i z^{-i}$ είναι ο *βραχύς διάρκειας προγνώστης (short term predictor).* Η μεταβλητή z^{-1} αναπαριστά την καθυστέρηση κατά ένα δείγμα, οι συντελεστές a_i είναι οι συντελεστές που

προκύπτουν με βάση την *γραμμική πρόγνωση* (Παράγραφος 3.7) ενώ η *τάξη του προγνώστη p* και έχει άμεση επίδραση στην ποιότητα της συντιθέμενης ομιλίας.

Εκτός όμως από τα παραπάνω, στο φίλτρο είναι δυνατόν να περιέχεται και ένας *μακράς διάρκειας προγνώστης (long term predictor – LTP)* ο οποίος έχει ως αποτέλεσμα το παραμένων σήμα να μοιάζει σαν θόρυβος. Έτσι μπορούμε να αναζητήσουμε σε ένα ¨λεξικό¨ από σήματα θορύβου για εκείνο το σήμα το οποίο ταιριάζει καλύτερα με το παραμένων.



Σχήμα 3.9: Μπλοκ διάγραμμα ενός AbS κωδικοποιητή/αποκωδικοποιητή.

Πέρα όμως από όλα τα παραπάνω η μεγαλύτερη διαφορά που παρατηρείτε στους *AbS* κωδικοποιητές είναι στο πώς καθορίζεται και πως κωδικοποιείτε το *σήμα διέγερσης*. Οι κύριες προσεγγίσεις είναι οι:

Multi – Pulse Excitation (MPE). Εδώ το σήμα διέγερσης δίνεται μέσα από σταθερό αριθμό μη μηδενικών παλμών για κάθε πλαίσιο και η θέση καθώς και το πλάτος τους καθορίζεται δυναμικά. Θεωρητικά είναι δυνατόν να βρεθούν οι καλύτερες τιμές για την θέση και το πλάτος αυτών των παλμών όμως κάτι τέτοιο εισάγει μεγάλη πολυπλοκότητα. Για να πετύχουμε στο *MPE* την καλύτερη ποιότητα ομιλίας με το χαμηλότερο bit rate θέλουμε μικρό αριθμό παλμών ανά πλαίσιο. Οι τυπικές λοιπόν τιμές είναι 4 – 5 παλμοί ανά πλαίσιο των 5 msec κάτι το οποίο μας δίνει καλή ποιότητα συντεθειμένης ομιλίας με bit rate γύρω στα 10 kbits/s. Περαιτέρω μείωση σε 2 παλμούς ανά πλαίσιο πάλι των 5 msec προϋποθέτει την ύπαρξη ενός *LTP* προγνώστη για να διατηρηθούμε στην ίδια ποιότητα ομιλίας με πάνω χωρίς όμως *LTP*.

Regular Pulse Excitation (RPE). Και σε αυτή την προσέγγιση χρησιμοποιούνται μη μηδενικοί παλμοί για μας δώσουν το σήμα διέγερσης, όμως η τοποθέτηση τους εδώ καθορίζονται με βάση την θέση του αρχικού παλμού. Έτσι αφού υπολογιστεί το σημείο του αρχικού παλμού οι υπόλοιποι τοποθετούνται μετά από σταθερή απόσταση. Με αυτό τον τρόπο χρειάζεται να μεταδοθούν μόνο η αρχική θέση και τα πλάτη των παλμών με αποτέλεσμα να μειώνεται η απαιτούμενη πληροφορία. Ο αριθμός βέβαια αυτών των παλμών πρέπει να είναι μεγαλύτερος από ότι στο *MPE* γιατί η τοποθέτηση τους είναι προκαθορισμένη. Έτσι για παράδειγμα για bit rate στα 10 kbits/s απαιτούνται γύρω στους 10 παλμούς ανά πλαίσιο των 5 msec σε αντίθεση με το *MPE* όπου απαιτούνται γύρω στους 4 – 5 παλμούς. Αυτό προσδίδει στην *RPE* μια ελαφρώς καλύτερη ποιότητα όμως εισάγει και μεγαλύτερη πολυπλοκότητα.

Codebook Excitation Linear Predictive Coding (CELP). Ο αλγόριθμος αυτός δημιουργήθηκε από την ανάγκη ύπαρξης bit rate κάτω από τα 10 kbits/s κάτι για το οποίο δεν ήταν κατάλληλοι οι *MPE* και *RPE* γιατί είναι αναγκασμένοι να μεταδίδουν σημαντική ποσότητα πληροφορίας. Έτσι για αυτούς μια περαιτέρω μείωση του bit rate σημαίνει απότομη μείωση της ποιότητας της συντεθειμένης ομιλίας. Για να μπορέσουμε να μειώσουμε λοιπόν την ποσότητα της μεταδιδόμενης πληροφορίας προτάθηκε σαν μέθοδος η διανυσματική κβάντιση (Παράγραφος 4.2.6) για το σήμα διέγερσης. Σύμφωνα με την μέθοδο αυτή επειδή το υπόλοιπο του σήματος

ομιλίας μετά από έναν *βραχύς διάρκειας* και έναν *μακράς διάρκειας προγνώστη* έχει την μορφή τυχαίου σήματος μπορεί να μοντελοποιηθεί με μια Gaussian μηδενικού μέσου όρου ακολουθία. Για τον λόγο αυτό χρησιμοποιείτε ένα ¨λεξικό¨ από Gaussian ακολουθίες του οποίου συνήθως (για πλαίσιο 40 δειγμάτων) ο δείκτης είναι της τάξης των 10 bits (με αποτέλεσμα το μέγεθος του λεξικού να είναι 1024 θέσεων). Κατά αυτόν τον τρόπο απαιτείται να μεταδίδεται μόνο ο δείκτης και το κέρδος (της τάξης των 5 bits) για την κάθε ακολουθία κάτι το οποίο σημαίνει μέγεθος μεταδιδόμενης πληροφορίας στα 15 bits σαφώς δηλαδή μικρότερη από την μεταδιδόμενη πληροφορία των παραπάνω προσεγγίσεων.

Παρόλα αυτά για την εύρεση της βέλτιστης ακολουθίας απαιτείται μεγάλη υπολογιστική ικανότητα και για αυτό τον λόγο έχουν προταθεί διάφορες τεχνικές. Έτσι οι πιο διαδεδομένες, στην περίπτωση του CELP, είναι η *εξαντλητική αναζήτηση* (όπου δοκιμάζονται όλες οι τιμές μέχρι να βρεθεί η βέλτιστη), τεχνικές που βασίζονται σε *μεθόδους αυτοσυσχέτισης* οι οποίες μειώνουν κατά πολύ το πλήθος των αναζητήσεων και η χρήση *δομημένων ¨λεξικών¨* τα οποία επιτρέπουν διαδικασίες γρήγορης αναζήτησης.

Εκτός όμως από τον ¨παραδοσιακό¨ CELP κωδικοποιητή που περιγράψαμε παραπάνω, υπάρχουν και διάφορες παραλλαγές του. Οι βασικότερες από αυτές είναι ο *VCELP (Vector Sum Excited Linear Prediction)* και ο *LD – CELP (Low Delay – CELP)*, οι οποίες βρίσκουν εφαρμογή σε διάφορα πεδία.

Στην δομή βέβαια ενός CELP κωδικοποιητή μπορούν να υπεισέλθουν πλήθος βελτιώσεων, βασιζόμενοι λοιπόν σε αυτές διακρίνουμε τις ακόλουθες κατηγορίες κωδικοποιητών:

MBE (MultiBand Excitation) κωδικοποιητές στους οποίους χωρίζουμε το φάσμα του σήματος της φωνής σε υπό – ζώνες με κεντρικές συχνότητες τις αρμονικές που παράγονται κατά την διέγερση των φωνητικών χορδών. Στην συνέχεια καθορίζεται αν το τμήμα της ομιλίας που περιέχεται σε κάθε υπό – ζώνη είναι έμφωνο ή άφωνο. Για αυτές που περιέχουν έμφωνα τμήματα γίνεται κωδικοποίηση της φάσης και της ενέργειας τους έτσι ώστε να μπορούν να μεταδοθούν ενώ για τις υπό – ζώνες εκείνες οι οποίες περιέχουν άφωνα τμήματα παριστάνονται με κατάλληλα φιλτραρισμένα τμήματα λευκού θορύβου. Η επισταμένη επιλογή των συχνοτήτων και των φάσεων σε κάθε τμήμα είναι ο καθοριστικός παράγοντας για την ελαχιστοποίηση των λαθών και επομένως για την επιτυχή υλοποίηση ενός *MBE* κωδικοποιητή.

Mixed – Excitation Linear Prediction (MELP) κωδικοποιητές οι οποίοι βασίζονται στο μοντέλο του LP vocoder όμως εδώ η πηγή διέγερσης έχει χαρακτηριστικά τα οποία είναι πιο κοντά στην ανθρώπινη ομιλία. Πιο συγκεκριμένα το σήμα διέγερσης εδώ δεν είναι μόνο παλμοί ή θόρυβος αλλά μπορεί να είναι συνδυασμός και των δύο ενώ η μοντελοποίηση του φάσματος του γίνεται με την χρήση συντελεστών Fourier. Η παραγωγή τώρα του σήματος διέγερσης πετυχαίνετε με αντίστροφο μετασχηματισμό Fourier.

Εκτός όμως από τις παραπάνω παραλλαγές των υβριδικών κωδικοποιητών υπάρχει και ένα μεγάλο πλήθος άλλων διαφορετικών αρχιτεκτονικών και παραλλαγών. Για τον λόγο αυτό η παράγραφος αυτή επικεντρώθηκε στους κυριότερους από αυτούς.

ΚΕΦΑΛΑΙΟ 4

4.1 Εισαγωγή

Ως γνωστό η λειτουργία των κωδικοποιητών κυματομορφής βασίζεται στην όσο γίνεται πιο ¨πιστή¨ αναπαράσταση της κυματομορφής του σήματος εισόδου. Σε γενικές γραμμές η ποιότητα ομιλίας που παράγουν είναι καλή κάτι όμως το οποίο επιτυγχάνεται με υψηλό bit rate. Οι κατηγορίες που μπορούν να χωριστούν όπως έχουμε πει, ανάλογα με τον τρόπο λειτουργίας τους αν είναι στο *πεδίο του χρόνου* ή στο *πεδίο της συχνότητας*. Στο κεφάλαιο λοιπόν αυτό θα αναφερθούμε στους κωδικοποιητές οι οποίοι λειτουργούν στο *πεδίο του χρόνου* και στο επόμενο κεφάλαιο θα ακολουθήσουν οι κωδικοποιητές στο *πεδίο των συχνοτήτων.*

4.2 Κβάντιση στους Κωδικοποιητές Κυματομορφής

Η κβάντιση των κωδικοποιητών μπορεί να διακριθεί σε *ομοιόμορφη* ή *σε ανομοιόμορφη* (*uniform* ή *nonuniform)* και σε *σταθερή* ή *προσαρμοστική (fixed* ή *adaptive)* καθώς και σε *βαθμωτή* ή *διανυσματική (scalar* ή *vector)*. Η πιο απλή μορφή της είναι η *ομοιόμορφη σταθερή* την οποία και συναντούμε στην βασική μορφή του PCM το οποίο και έχει περιγραφεί στην Παράγραφος 1.7. Επειδή όμως παίζει σημαντικό ρόλο στους κωδικοποιητές κυματομορφής θα δούμε αναλυτικά ξανά μερικούς από τους τύπους της

4.2.1 Ομοιόμορφη Κβάντιση

Κατά την *ομοιόμορφη κβάντιση* το σήμα εισόδου $s(n)$ κβαντίζεται σε Q στάθμες με *σταθερό βήμα κβάντισης* Δ . Για την αναπαράσταση αυτών των διαστημάτων απαιτούνται $B = [\log_2 Q]$ bits ενώ για την κβαντισμένη έξοδο S_q έχουμε $S_q = m_i$ αν για την τιμή του σήματος εισόδου ισχύει $x_{i-1} < s(n) \le x_i$. Ο δείκτης *i* αναφέρεται στο *i – στο* διάστημα κβάντισης ενώ ισχύει $x_i = i\Delta - S_{\max}$ και $m_i = \frac{x_{i-1} + x_i}{2}$ με $i = 1,2,...,Q$. Στον ομοιόμορφο κβαντιστή μπορούμε να διακρίνουμε δύο τύπου κβαντίσεις την *mid – tread* στην οποία μπορούμε να έχουμε μηδενική έξοδο αλλά ο αριθμός των θετικών και αρνητικών επιπέδων κβάντισης διαφέρει και την *mid – riser* στην οποία δεν μπορούμε να έχουμε μηδενική έξοδο όμως είναι συμμετρική ως προς το μηδέν. Για τις δύο αυτές κβαντίσεις (*mid – tread* και *mid – zero)* η τιμή του βήματος κβάντισης δίνεται αντίστοιχα από τους τύπους $\Delta = \frac{2S_{\max}}{2^B - 1}$ και $\Delta = \frac{2S_{\max}}{2^B}$ με $S_{\max}$ το μέγιστο πλάτος του σήματος εισόδου.

Το σφάλμα $q(n)$ τώρα της κβάντισης ορίζεται σύμφωνα με τον τύπο $q(n) = s(n) - \hat{s}(n)$ όπου $\hat{s}(n)$ το κβαντισμένο σήμα. Επίσης το σφάλμα είναι ομοιόμορφα κατανεμημένο μεταξύ του $\Delta/2$ και του $-\Delta/2$. Ο λόγος SNR υπολογίζεται από τον τύπο $SNR = \frac{\sigma^2{}_s}{\sigma^2{}_q} = \frac{E[s^2(n)]}{E[q^2(n)]} = \frac{\sum_n s^2(n)}{\sum_n q^2(n)}$.

Λόγο όμως της ομοιόμορφης κατανομής του *σφάλματος,* ισχύει $\sigma_q{}^2 = \frac{\Delta^2}{12} = \frac{S^2{}_{\max}}{(3)2^{2B}}$ οπότε και έχουμε $SNR = \frac{(3)2^{2B}}{\left(\frac{S_{\max}}{\sigma_q}\right)^2}$ από τον οποίο προκύπτει ο τύπος

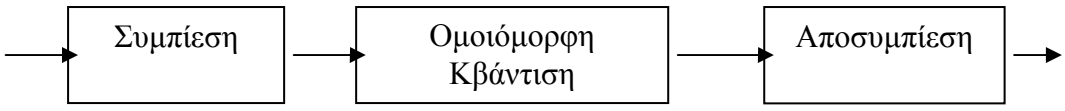
$SNR(dB) = 10\log_{10}\left(\frac{\sigma_s^{\,2}}{\sigma_q^{\,2}}\right) = 6B + 4.77 - 20\log_{10}\left(\frac{S_{\max}}{\sigma_s}\right)$. Αν υποθέσουμε τώρα $S_{\max} = 4\sigma_x$ καταλήγουμε στην σχέση $SNR(dB) = 6B - 7.2$ σύμφωνα με την οποία βλέπουμε ότι αυξάνοντας των αριθμό των bits κατά ένα αυξάνεται ο λόγος SNR κατά 6 dB, δηλαδή εξαρτάται από τον αριθμό των bits. Για την ακρίβεια όσο μεγαλύτερος είναι ο αριθμός των σταθμών κβάντισης τόσο το λάθος μεταξύ του κβαντισμένου και του αρχικού σήματος ελαττώνεται. Εδώ το $S_{\max}$ δηλώνει την μέγιστη τιμή του σήματος εισόδου, ενώ το σ_x την τυπική απόκλιση του, η οποία βέβαια είναι και η *rms* τιμή.

Στην ομοιόμορφη κβάντιση όμως επειδή η αντίστοιχη *rms* τιμή του θορύβου είναι σταθερή και ίση με $\Delta/\sqrt{12}$, όταν το σήμα εισόδου είναι μικρό για μακρές περιόδους, το πηλίκο *SNR* είναι και αυτό μικρό, πράγμα το οποίο αποτελεί ένα ουσιαστικό πρόβλημα το οποίο και θα αναλύσουμε διεξοδικότερα στα παρακάτω.

4.2.2 Ανομοιόμορφη – Λογαριθμική Κβάντιση

Ως γνωστό ο ομοιόμορφος κβαντιστής δίνει άριστο πηλίκο σήματος προς θόρυβο όταν το σήμα έχει ομοιόμορφη κατανομή κάτι όμως που δεν ισχύει στο σήμα της ομιλίας στο οποίο η ισχύς είναι ανομοιόμορφα κατανεμημένη. Την απόδοση όμως της κβάντισης επηρεάζει και το γεγονός ότι τιμή του θορύβου είναι σταθερή ενώ το σήμα ομιλίας εμφανίζει πολύ μεγάλες διακυμάνσεις μέσα σε κλάσματα δευτερολέπτου. Αυτό έχει ως αποτέλεσμα η απόδοση της να είναι καλή σε τμήματα της φωνής τα οποία έχουν έντονο ήχο και κακή στα τμήματα εκείνα τα οποία αντιστοιχούν σε παύσεις. Έτσι οδηγούμαστε σε μεταβαλλόμενο SNR. Επίσης λόγω του της ιδιότητας του ακουστικού συστήματος να παρουσιάζετε *φασματική επικάλυψη (spectral masking)* ο θόρυβος στα τμήματα με παύσεις δεν θα επικαλύπτεται με αποτέλεσμα να γίνεται περισσότερο αντιληπτός από ότι στα τμήματα όπου έχουμε ομιλία.

Για την αντιμετώπιση λοιπόν όλων αυτών των προβλημάτων με τέτοιο τρόπο ώστε να έχουμε περίπου σταθερό SNR χρησιμοποιούμε την ανομοιόμορφη κβάντιση στην οποία η τιμή του θορύβου είναι ανάλογη του πλάτους του σήματος ομιλίας για κάθε τμήμα (μικρός θόρυβος για μικρά πλάτη και μεγάλος θόρυβος για μεγάλα πλάτη). Σε μια τέτοια κβάντιση ο σηματοθορυβικός λόγος δίνεται από την σχέση $SNR = \frac{1}{\sigma_q^{\,2}}$ από την οποία και βλέπουμε ότι είναι ανεξάρτητος της διακύμανσης του σήματος εισόδου.



Σχήμα 4.1: Γενική μορφή ενός ανομοιόμορφου κβαντιστή.

Για να υλοποιήσουμε τώρα ένα ανομοιόμορφο κβαντιστή χρησιμοποιούμε συμπιεστές/αποσυμπιεστές (compressor/expander) με ενδιάμεση παρεμβολή ενός ομοιόμορφου κβαντιστή (Σχήμα 4.1) . Οι συμπιεστές/αποσυμπιεστές αυτοί είναι γνωστοί ως *compander* και χρησιμοποιούν για τον σκοπό αυτό δύο τροποποιημένους λογαριθμικούς τύπους. Οι τύποι αυτοί ονομάζονται *νόμος συμπίεσης μ* και *νόμος συμπίεσης Α* και το κύριο χαρακτηριστικό τους είναι ότι έχουν πεπερασμένο εύρος (κάτι το οποίο δεν ισχύει για τον απλό λογάριθμο). Οι νόμοι αυτοί ορίζονται αντίστοιχα από τις

$$t(n) = S_{\max} \frac{\log(1+\mu\left|\frac{s(n)}{s_{\max}}\right|)}{\log(1+\mu)} \sin g(s(n))$$

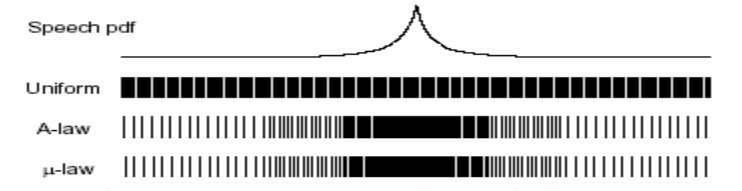
και

$$t(n) = \begin{cases} \dfrac{As(n)}{1+\log A} & 0 \le s(n) \le \dfrac{S_{\max}}{A} \\ S_{\max} \dfrac{1+\log\left(A|s(n)|/S_{\max}\right)}{1+\log A} \sin g(s(n)) & \dfrac{S_{\max}}{A} \le s(n) \le S_{\max} \end{cases}$$

όπου μ και A οι συντελεστές συμπίεσης, *sing(s(n))* το πρόσημο του σήματος εισόδου *s(n)* το οποίο έχει τιμή ± 1 και *log()* ο φυσικός λογάριθμος. Για η κβάντιση με *νόμο συμπίεσης μ* μπορούμε να πούμε ότι είναι περίπου γραμμική για μικρές τιμές του σήματος εισόδου γιατί $\log(1+\mu s(n)) \approx \mu s(n)$ ενώ ο *νόμος συμπίεσης A* είναι γραμμικός για τιμές πλάτους μικρότερες από $S_{\max}/A$ και λογαριθμικός πάνω από αυτές. Στην περίπτωση μάλιστα του *νόμου συμπίεσης μ* έχουμε τον τύπο

$$SNR(dB) = 6B + 4.77 - 20\log_{10}[\ln(1+\mu)] - 10\log_{10}\left[1+\left(\frac{S_{\max}}{\mu\sigma_x}\right)^2 + \sqrt{2}\left(\frac{S_{\max}}{\mu\sigma_x}\right)\right]$$ από τον οποίο μας δίνεται το SNR όπου και μπορούμε να συμπεράνουμε ότι εδώ είναι πολύ λιγότερο ευαίσθητο στις μεταβολές τις ποσότητας $\frac{S_{\max}}{\sigma_x}$.

Μάλιστα χρησιμοποιώντας μεγάλες τιμές συντελεστών συμπίεσης *μ (A)* οδηγούμαστε σε ένα σταθερό SNR (στην ουσία ο σηματοθορυβικός λόγος είναι ανεξάρτητος από την στατιστική του σήματος). Ωστόσο η συνολική βελτίωση που πετυχαίνεται είναι μικρή σε σχέση με το SNR του κβαντισμένου σήματος είναι μικρή.



Σχήμα 4.2: Σύγκριση ομοιόμορφου κβαντιστή με companders νόμου *A* και *μ*.

4.2.3 Βέλτιστη Κβάντιση

Η χρήση βέλτιστου κβαντιστή μπορεί να θεωρηθεί σαν μια δεύτερη προσέγγιση της ανομοιόμορφης κβάντισης. Ο σκοπός του είναι να μας δώσει για δοσμένο αριθμό επιπέδων κβάντισης σήμα εξόδου με το μεγαλύτερο δυνατό SNR. Οι δύο κβαντιστές που αναφέρουμε παραπάνω *νόμου συμπίεσης μ* και *νόμου συμπίεσης A* δεν μπορούν να θεωρηθούν βέλτιστοι στην περίπτωση όπου η κατανομή του σήματος είναι γνωστή και για αυτό το λόγο αναζητούμε μια νέα συνάρτηση στην οποία τα διαστήματα και οι στάθμες κβάντισης θα εκλεγούν έτσι ώστε ο σηματοθορυβικός λόγος να μεγιστοποιηθεί. Για να συμβεί αυτό ο θόρυβος κβάντισης ο οποίος μας δίνεται από την σχέση

$$\sigma_q^{\,2} = \sum_{i=1}^{Q} \int_{x_{i-1}}^{x_i} (x - m_i)^2 f_x(x) dx$$

πρέπει να ελαχιστοποιηθεί. Όπου $f_x(x)$ η συνάρτηση πυκνότητας πιθανότητας του σήματος εισόδου $s(n) = X$. Η ελαχιστοποίηση του θορύβου επιτυγχάνεται παραγωγίζοντας το $\sigma_q^{\,2}$ ως προς

τα m_i και x_i και μηδενίζοντας τις παραγώγους αυτές. Με αυτόν τον τρόπο προκύπτει ότι η βέλτιστη θέση για κβάντιση είναι η m_Q στατιστική μέση τιμή (κέντρο βάρους) του αντίστοιχου Q – οστού διαστήματος. Για την εύρεση αυτής της λύσης δεν υπάρχει αναλυτική διαδικασία αλλά βασίζεται σε μια επαναληπτική μέθοδο σύμφωνα με την οποία για μια δοσμένη $f_x(x)$ ορίζουμε το m_1 και αν αυτό έχει εκλεγεί σωστά τότε το υπολογιζόμενο m_Q θα είναι η μέση τιμή. Σε αντίθετη περίπτωση όπου το m_Q δεν είναι η μέση τιμή του Q – οστού διαστήματος εκλέγουμε άλλο m_1 και επαναλαμβάνουμε την διαδικασία.

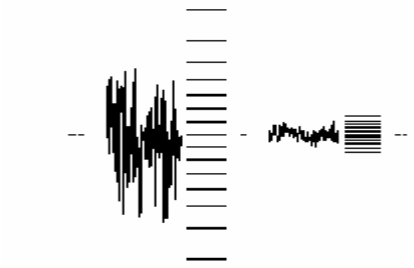
Τέλος πρέπει να αναφέρουμε ότι η επειδή η ομιλία έχει μια συνάρτηση πυκνότητας η οποία είναι σχετικά κοντά στην Laplace κατανομή πυκνότητας ο βέλτιστος κβαντιστής για τέτοιου είδους σήματα είναι ο *νόμος συμπίεσης m* ο οποίος και δίνεται από την σχέση

$$t_{Laplace}(n) = S_{\max} \frac{1 - e^{-m|s(n)|/S_{\max}}}{1 - e^{-m}} sign(s(n))$$

και στον οποίο είναι $m_{βελτιστο} = \frac{\sqrt{2}S_{\max}}{3\sigma_s}$, με σ_s την τυπική απόκλιση του σήματος εισόδου.

4.2.4 Προσαρμοστική Κβάντιση

Στις παραπάνω περιπτώσεις των ανομοιόμορφων κβαντιστών το σήμα της ομιλίας αντιμετωπίζεται σαν σήμα σταθερό, μεγάλου εύρους. Όμως στην πραγματικότητα το σήμα της ομιλίας είναι χρονικά μεταβαλλόμενο για τον λόγο αυτό η καλύτερη μέθοδος ¨διαχείρισης¨ του είναι η προσαρμοστική κβάντιση. Η βασική αρχή της προσαρμοστική κβάντισης είναι ότι τα χαρακτηριστικά του κβαντιστή (βήμα κβάντισης) μεταβάλλονται ανάλογα με τις μεταβολές της ομιλίας (Σχήμα 4.3). Οι στρατηγικές προσαρμοστικής κβάντισης μπορούν να βασίζονται είτε στο σήμα εισόδου είτε στο σήμα εξόδου και με βάση αυτό το κριτήριο διακρίνονται σε *adaptive quantization with forward (AQF) and backward (AQB) estimation* αντίστοιχα.



Σχήμα 4.3:Παράδειγμα προσαρμοστικής κβάντισης όπου το βήμα κβάντισης προσαρμόζεται ανάλογα με τις μεταβολές του σήματος.

Στην περίπτωση του *AQF* με βάση το σήμα εισόδου εξάγεται μια παράμετρος η οποία και ελέγχει τον κβαντιστή. Η παράμετρος όμως αυτή μεταδίδεται και στον αποκωδικοποιητή σαν *πλευρική πληροφορία* γεγονός το οποίο έχει ως αποτέλεσμα την εισαγωγή ενός πλεονασμού και μιας επιπρόσθετης καθυστέρησης. Σε αντίθεση με τα παραπάνω στην περίπτωση του *AQB* μεταδίδεται μόνο το κωδικοποιημένο σήμα και χρησιμοποιείται τόσο στον κβαντιστή όσο και στον αποκωδικοποιητή ανάδραση έτσι ώστε να πετυχαίνεται προσαρμοστική κβάντιση και αποκωδικοποίηση αντίστοιχα.

Επιπρόσθετα μια άλλη διάκριση η οποία γίνεται στην προσαρμοστική κβάντιση είναι σε *στιγμιαία* και σε *συλλαβική (instantaneous – syllabic)*. Στην *στιγμιαία κβάντιση* οι παράμετροι του κβαντιστή ανανεώνονται σε κάθε δείγμα ή ανά μερικά δείγματα ενώ αντίθετα στην *συλλαβική κβάντιση* η ανανέωση των παραμέτρων του κβαντιστή γίνεται πιο αργά συνήθως κάθε 10 – 20 ms γιατί το σήμα της ομιλίας μπορεί να θεωρηθεί σταθερό σε αυτό το χρονικό διάστημα. Συνήθως στα

συστήματα με *εμπρόσθια ανάδραση (feed – forward)* χρησιμοποιείται *συλλαβική κβάντιση* γιατί οι παράμετροι της κβάντισης μεταδίδονται μετά από κάθε ανανέωση ενώ στα συστήματα με *οπίσθια ανάδραση (feedback)* όπου δεν έχουμε μετάδοση πλευρικής πληροφορίας χρησιμοποιείται *στιγμιαία κβάντιση*.

4.2.5 Διαφορική Κβάντιση

Μέχρι τώρα σε όλες τις τεχνικές κβάντισης που είδαμε κάθε δείγμα τις ακολουθίας κβαντιζόταν ανεξάρτητα από την τιμή του προηγούμενου δείγματος. Σε αντίθεση λοιπόν με τα προηγούμενα στη διαφορική κβάντιση, βασιζόμενοι στο γεγονός ότι στο σήμα ομιλίας υπάρχει υψηλός βαθμός συσχέτισης μεταξύ των γειτονικών δειγμάτων (Παράγραφος 2.3.2), το κάθε δείγμα μπορεί να υπολογιστεί από τα προηγούμενα δείγματα. Για τον λόγο αυτό είναι δυνατόν να μην μεταδίδεται κάθε φορά το δείγμα αυτούσιο αλλά η διαφορά ανάμεσα σε γειτονικά δείγματα. Επίσης υπάρχει η δυνατότητα να μεταδίδεται η διαφορά ανάμεσα στην προβλεπόμενη τιμή, η οποία είναι ένας γραμμικός συνδυασμός ενός ή περισσότερων προηγούμενων δειγμάτων, και στην πραγματική τιμή του δείγματος. Η διαφορά αυτή ονομάζεται *διαφορά πρόγνωσης ή παράγοντας λάθους*.

Από πλευράς απόδοσης τώρα στη διαφορική πρόγνωση το σήμα το οποίο προκύπτει είναι σε γενικές γραμμές μικρότερου πλάτους από το αρχικό σήμα και με μικρότερο εύρος. Το γεγονός αυτό οδηγεί στο να μπορεί το σήμα εισόδου να κωδικοποιηθεί ακριβέστερα και με χαμηλότερη τιμή ισχύος *θορύβου κβάντισης* συγκριτικά με τα συστήματα απευθείας κβάντισης. Πιο συγκεκριμένα η διαφορική κβάντιση οδηγεί σε μέσο τετραγωνικό σφάλμα τόσο μικρότερο συγκριτικά με την απευθείας κβάντιση όσο τα δείγματα είναι εντονότερα συσχετισμένα. Η ελάττωση αυτή σφάλματος είναι πάντα δυνατή αρκεί η συσχέτιση από δείγμα σε δείγμα να μην είναι μηδέν. Βέβαια η μεγαλύτερη ελάττωση σφάλματος παρουσιάζεται όταν η διαφορική κβάντιση δρα στη διαφορά μεταξύ της πραγματικής τιμής του δείγματος και της προβλεπόμενης τιμής. Εδώ όταν ο προγνώστης είναι καλός, πράγμα που συμβαίνει όταν τα δείγματα έχουν έντονη συσχέτιση, το μέσο τετραγωνικό σφάλμα του διαφορικού κβαντιστή είναι πολύ μικρό.

Τέλος να πούμε ότι στην διαφορική κβάντιση μπορούμε να διακρίνουμε διάφορες υλοποιήσεις. Η πιο απλή είναι αυτή στην οποία χρησιμοποιείται σταθερός προγνώστης και σταθερός ομοιόμορφος κβαντιστής ενώ στα ποιο πολύπλοκα συστήματα χρησιμοποιούνται προσαρμοστικοί κβαντιστές (π.χ. ADPCM) ή προσαρμοστικοί προγνώστες (π.χ. APC) ή και οι δύο. Τα συστήματα αυτά θα αναλυθούν περαιτέρω παρακάτω.

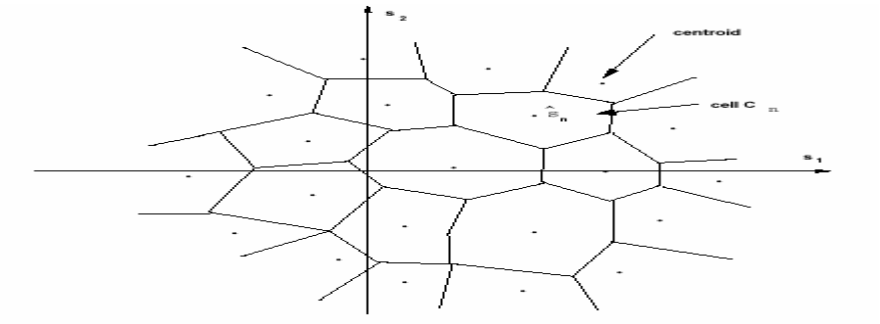
4.2.6 Διανυσματική Κβάντιση

Όλες οι παραπάνω μέθοδοι κβάντισης που αναφέρθηκαν ανήκουν στην ευρύτερη κατηγορία της *βαθμωτής κβάντισης*. Σε αντίθεση λοιπόν με αυτή μπορούμε να διακρίνουμε την *διανυσματική κβάντιση* στην οποία σχηματίζονται *διανύσματα* μεγέθους $N \times 1$ της μορφής $s_n = [s_n(0), s_n(1), ..., s_n(N-1)]^T$ και τα οποία αποτελούνται από συνεχόμενα δείγματα. Επίσης τόσο στον πομπό όσο και στον δέκτη υπάρχει ένα ¨λεξικό¨ μήκους L στο οποίο είναι αποθηκευμένα ένα πλήθος από *πρότυπα* διανύσματα $\{\overset{)}{s}_n = [\overset{)}{s}_n(0), \overset{)}{s}_n(1), ..., \overset{)}{s}_n(N-1)]^T, n = 1,2,...L\}$. Η ύπαρξη αυτή του ¨λεξικού¨ βοηθά έτσι ώστε για τα σχηματιζόμενα διανύσματα s_n, να βρίσκεται με βάση κάποιο κριτήριο, το πλησιέστερο *πρότυπο* διάνυσμα $\overset{)}{s}_n$ του ¨λεξικού¨. Το συνηθέστερο κριτήριο το οποίο χρησιμοποιείτε σε αυτήν την περίπτωση είναι η τιμή της *διακύμανσης* η οποία προκύπτει με βάση τον παρακάτω τύπο:

$$\varepsilon(s_n, \overset{)}{s}_n) = \sum_{k=0}^{N-1} \left(s(k) - \overset{)}{s}(k)\right)^2$$

Αφού λοιπόν βρεθεί αυτό το ¨πλησιέστερο¨ διάνυσμα η διεύθυνση του στο ¨λεξικό¨ καθορίζει και το σύμβολο $\{u_n, n = 1,2,..., L\}$ το οποίο θα σταλεί μέσα από το κανάλι μετάδοσης.. Η διαδικασία η οποία εξηγείτε και γραφικά στο Σχήμα 4.4 έχει ως εξής: καθορίζονται L περιοχές

(κύτταρα) απόφασης για τα L διανύσματα κατά τέτοιο τρόπο ώστε να μην επικαλύπτονται μεταξύ τους. Το κάθε κύτταρο τώρα αντιστοιχίζεται σε ένα *πρότυπο* διάνυσμα $\hat{s}_n$. Αν το διάνυσμα s_n ανήκει στο κύτταρο C_n θα επιλεγεί το πρότυπο διάνυσμα $\hat{s}_n$ το οποίο και είναι το κέντρο του κυττάρου, ενώ το σύμβολο u_n το οποίο θα σταλεί είναι συνήθως η δυαδική μορφή της διεύθυνσης ή του δείκτη του $\hat{s}_n$.



Σχήμα 4.4: Δύο διαστάσεων διανυσματική κβάντιση.

Σαν την πιο απλή μορφή διανυσματικού κβαντιστή κατά αναλογία με την βαθμωτή κβάντιση μπορούμε να αναφέρουμε το διανυσματικό PCM (VPCM). Στο *VPCM* για κάθε εισερχόμενο διάνυσμα έχουμε εξαντλητική αναζήτηση (full search VQ ή F – VQ) και ο αριθμός των bits ανά δείγμα δίνεται από την σχέση $B = \frac{\log_2 L}{N}$ ενώ ο σηματοθορυβικός λόγος από την $SNR_N = 6B + K_N$. Για N=1 πέφτουμε στην περίπτωση του απλού *βαθμωτού PCM* ωστόσο η *VPCM* παρουσιάζει βελτιωμένο SNR καθώς εκμεταλλεύεται την συσχέτιση μεταξύ των διανυσμάτων.

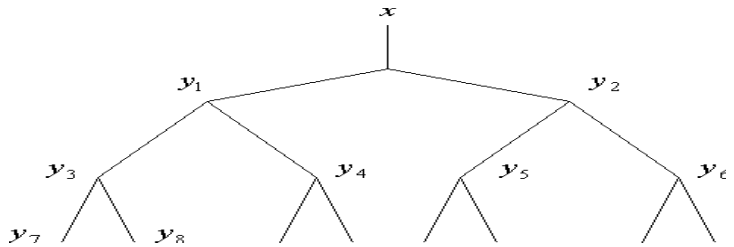
Παρόλο όμως που η διανυσματική κβάντιση προσφέρει σημαντικό κέρδος κατά την κωδικοποίηση η μνήμη που απαιτείται και η πολυπλοκότητα της αυξάνονται εκθετικά σε σχέση με το N (*πολλαπλασιασμοί/προσθέσεις για τον υπολογισμό d(x,y)).* Πιο συγκεκριμένα ο αριθμός των υπολογισμών είναι της τάξης 2^{RN} ενώ ο αριθμός των θέσεων μνήμης αφού *1 θέση μνήμης/διάσταση* δίνεται με βάση τον τύπο $N \cdot 2^{RN}$ με *R=bits/διάσταση*. Για την μείωση λοιπόν αυτών των απαιτήσεων έχουν αναπτυχθεί διάφορες τεχνικές οι οποίες κατά κανόνα σχετίζονται με την επιλογή του ¨λεξικού¨[9] - διαδικασία εκπαίδευσης. Η πιο συνηθισμένη από αυτές είναι ο αλγόριθμος *K – means (Generalized – Llyod)* σύμφωνα με τον οποίο επιλέγονται τα αρχικά πρότυπα ενώ στην συνέχεια τα δείγματα που έχουν επιλεγεί ως *δείγματα εκπαίδευσης*[10] ταξινομούνται στα κύτταρα με βάση τον κανόνα του πλησιέστερου γείτονα (nearest – neighbor). Μετά από αυτή την ταξινόμηση ακολουθεί ο υπολογισμός νέων προτύπων σε κάθε κύτταρο και αν η *μέση διακύμανση* δεν έχει μειωθεί ικανοποιητικά ο αλγόριθμος συνεχίζει να επαναλαμβάνετε. Κατά αυτό τον τρόπο συγκλίνουμε σε ένα τοπικό ελάχιστο της *ολικής διακύμανσης* ενώ για την εύρεση του ολικού ελαχίστου απαιτείται να τρέξει ο αλγόριθμος από πολλά διαφορετικά αρχικά σημεία.

Εκτός όμως από τον αλγόριθμο *K – means* για την μείωση της πολυπλοκότητας σαν άλλες τεχνικές μπορούμε να αναφέρουμε τα *δομημένα* ¨λεξικά¨ στα οποία μας δίνεται η δυνατότητα για αποτελεσματικότερη αναζήτηση με ταυτόχρονη όμως απώλεια της απόδοσης και σε μερικές περιπτώσεις αύξηση του αριθμού των θέσεων μνήμης. Σαν τέτοιους κβαντιστές μπορούμε να αναφέρουμε τους *tree – structured* και τους *multi - stage* διανυσματικούς κβαντιστές. Στην

[9] Το *¨λεξικό¨* μπορεί να είναι σταθερό ή μεταβαλλόμενο.

[10] Πρέπει να είναι όσο το δυνατόν αντιπροσωπευτικά των δειγμάτων που αναμένεται να έχουμε στον κβαντιστή και ο αριθμός αυτών, εξαρτάται από το μέγεθος του *¨λεξικού¨*. Συνήθως ο ελάχιστος αριθμός τους είναι δέκα ενώ προτιμότερο είναι να είναι πενήντα και πάνω.

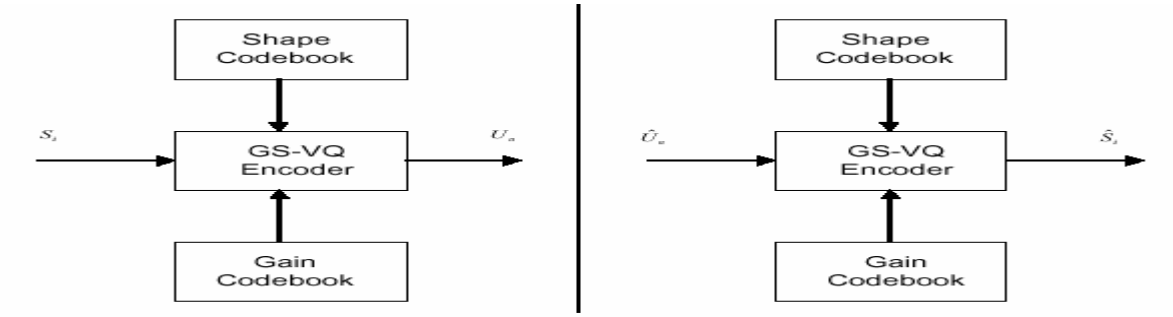
περίπτωση του *tree – structure* (Σχήμα 4.5) η πολυπλοκότητα του είναι ανάλογη του $\log_2 L$ αντί του L ενώ η διαδικασία για η επιλογή του ¨λεξικού¨ γίνεται με χωρισμό του *N – διάστατου* χώρου σε δύο περιοχές με βάση τον *K – means*, με *K=2* .Στην συνέχεια τα δείγματα εκπαίδευσης ταξινομούνται σε μια από τις δύο περιοχές. Ακολουθεί η επανάληψη του *K – means*, με *K=2* και για τις περιοχές αυτές, έως ότου δημιουργηθούν τέσσερα κύτταρα. Το τελευταίο βήμα επαναλαμβάνεται μέχρι την δημιουργία $L = 2^B$ κυττάρων. Τα κέντρα σε όλα τα επίπεδα και ο κβαντισμός ενός διανύσματος γίνεται ιεραρχικά, διασχίζοντας το δέντρο. Εδώ όμως για την ταξινόμηση σε κάθε επίπεδο χρειαζόμαστε *2N προσθέσεις/πολλαπλασιασμούς ο*πότε συνολικά $2N \log_2 L = 2N \log_2 2^B = 2NB$ σε αντίθεση με την πλήρη αναζήτηση (F – VQ) όπου θέλουμε όπως είδαμε παραπάνω $N \cdot 2^B$. Όλα αυτά έχουν σαν αποτέλεσμα η *tree – structured quantization* συγκρινόμενη με την *F – VQ (Full Vector Quantization)* να προσφέρει βελτίωση περίπου κατά 1 dB του λόγου SNR στην περίπτωση του 1 bit/sample.



Σχήμα 4.5: Η μορφή του ¨λεξικού¨ ενός *tree – structured* διανυσματικού κβαντιστή.

Όπως είπαμε όμως και πιο πάνω στην κατηγορία των *δομημένων* ¨λεξικών¨ ανήκουν και οι *multi - stage* διανυσματικοί κβαντιστές οι οποίοι υλοποιούνται από δύο ή περισσότερους διαδοχικούς κβαντιστές όπου ο κάθε ένας κωδικοποιεί το σφάλμα ή το υπόλοιπο του προηγούμενου. Εδώ το συνολικό ¨λεξικό¨ είναι το καρτεσιανό γινόμενο των διαφορετικών ¨λεξικών¨ τα οποία ανήκουν στους διαδοχικούς κβαντιστές. Επίσης η διάσταση του συνολικού ¨λεξικού¨ είναι το άθροισμα των διαστάσεων όλων των συστατικών ¨λεξικών¨ ενώ το μέγεθος του είναι το γινόμενο των διαστάσεων των επιμέρους αυτών ¨λεξικών¨. Τέλος να αναφέρουμε ότι και οι *MSVQ (Multi – Stage VQ)* όπως και οι *TSVQ (Tree – Structured VQ)* παρουσιάζουν σε σχέση με τους *F – VQ (Full VQ)* βελτίωση περίπου κατά 1 dB του λόγου SNR στην περίπτωση του 1 bit/sample.

Πέρα όμως από την μέθοδο των *δομημένων* ¨λεξικών¨ μια άλλη αποτελεσματική μέθοδος είναι η *Gain/Shape VQ (GS – VQ).* Σύμφωνα με την μέθοδο αυτή (Σχήμα 4.6) υπάρχουν δύο ¨λεξικά¨ όπου το ένα περιέχει πρότυπα για το σχήμα της κυματομορφής ενώ το άλλο περιέχει πρότυπα για το κέρδος. Κατά αυτό τον τρόπο είναι δυνατόν να μεταδίδεται ξεχωριστά το σχήμα και το κέρδος των διανυσμάτων κάτι το οποίο μας οδηγεί σε μια βελτίωση σε σχέση με την *F – VQ* κατά 0.7 dB πάντα για 1 bit/sample.



Σχήμα 4.6:Μπλοκ διάγραμμα ενός *Gain/Shape VQ*.

Άλλες παραλλαγές στην διανυσματική κβάντιση και στην διαδικασία της εκπαίδευσης υπάρχουν αρκετές από τις οποίες οι κυριότερες ονομαστικά είναι η *vector sum quantization*, η *adaptive VQ* την οποία και θα δούμε αναλυτικότερα σε άλλη παράγραφο και η *finite – state VQ* η οποία είναι μια ειδική περίπτωση της *A – VQ*.

4.3 Προσαρμοστική Παλμοκωδική Διαμόρφωση (Adaptive PCM)

Ως γνωστό Adaptive PCM σύστημα ονομάζεται ένα σύστημα PCM με προσαρμοστικό βήμα κβάντισης. Όπως είδαμε και στην παράγραφο 4.2.4 η βασική διάκριση η οποία μπορεί να γίνει σε ένα τέτοιο σύστημα είναι σε *adaptive quantization with forward (AQF) and backward (AQB) estimation*. Ας δούμε τώρα λίγο πιο αναλυτικά τις δύο αυτές περιπτώσεις.

4.3.1 Adaptive Quantization with Forward Estimation (AQF)

Στην περίπτωση της *AQF* η *διακύμανση* του σήματος της ομιλίας παίζει βασικό ρόλο στον καθορισμό της στρατηγικής που θα ακολουθηθεί για την προσαρμογή του βήματος κβάντισης. Η *διακύμανση* αυτή μπορεί να βρεθεί υπολογίζοντας την ενέργεια βραχέως χρόνου του σήματος εισόδου. Κατά τον υπολογισμό αυτό μπορούν να χρησιμοποιηθούν διάφορα παράθυρα τα οποία έχουν σαν σκοπό να μειώσουν τις ασυνέχειες, το πλήθος των δεδομένων και την διάρκεια του σήματος της ομιλίας.

Το παράθυρο που χρησιμοποιείται συνήθως σε αυτή την περίπτωση είναι το *ορθογώνιο παράθυρο* με βάση το οποίο ο τύπος της διακύμανσης παίρνει την μορφή

$$\sigma^2[n] = \frac{1}{M} \sum_{m=n}^{m=n+M-1} s^2[m]$$

όπου *M* είναι το μέγεθος του παραθύρου και η επιλογή του έχει άμεση σχέση με τον υπολογισμό της διακύμανσης. Έτσι μεγάλες τιμές του *M* επιδρούν στον υπολογισμό της στην περίπτωση της *συλλαβικής κβάντισης* ενώ μικρές τιμές του επιδρούν στην περίπτωση της *στιγμιαίας κβάντισης*. Βέβαια όπως αναφέρεται και στην παράγραφο 4.2.4 στην περίπτωση των συστημάτων με *εμπρόσθια ανάδραση (feed – forward)* χρησιμοποιείται *συλλαβική κβάντιση* γιατί το ποσό της πλευρικής πληροφορίας που μεταδίδεται είναι πολύ μικρό (περίπου 1% του συνολικού bit rate).

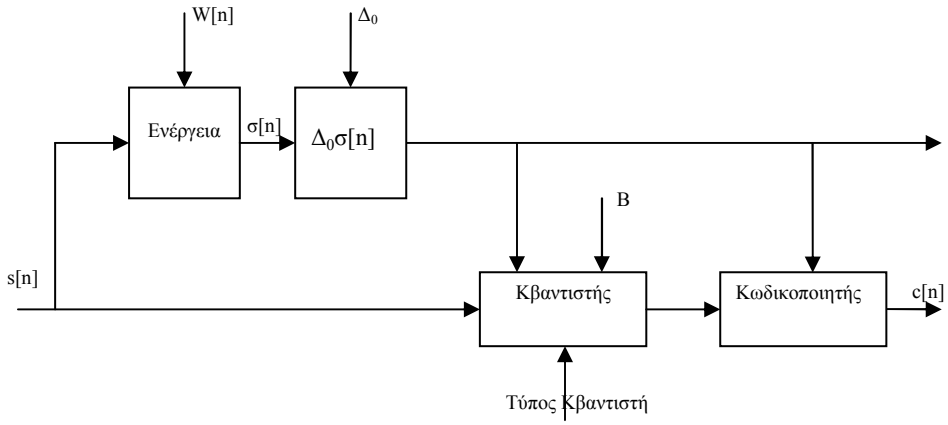
Με βάση λοιπόν αυτή την υπολογιζόμενη *διακύμανση* υπάρχουν δύο βασικές προσεγγίσεις ως προς την χρησιμοποίηση της. Η πρώτη από αυτές είναι η *προσαρμογή του βήματος κβάντισης* (Σχήμα 4.7) κατά την οποία το βήμα κβάντισης είναι ανάλογο με την *διακύμανση* και δίνεται μέσα από τον τύπο

$$\Delta[n] = \frac{\Delta_0 \sigma[n]}{2^{n-1}}$$

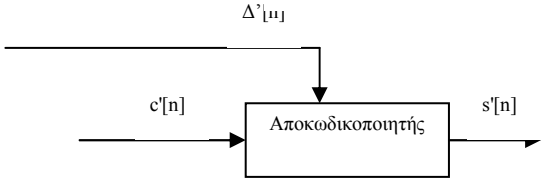
όπου Δ_0 είναι μια σταθερή παράμετρος κλιμάκωσης. Με τον τρόπο αυτό για μεγάλες διακυμάνσεις έχουμε μεγάλα βήματα ενώ για μικρές διακυμάνσεις έχουμε μικρά βήματα οπότε πετυχαίνουμε αποτελεσματικότερη χρήση των διαθέσιμων bits. Επειδή όμως είναι δυνατόν το σήμα εισόδου να έχει αξιοσημείωτες μεταβολές για την αποφυγή εξαιρετικά μεγάλων ή μικρών βημάτων προκαθορίζονται για τον κβαντιστή συγκεκριμένα όρια ως εξής

$$\Delta_{\min} \le \Delta[n] \le \Delta_{\max}$$

Η επιλογή του ελαχίστου Δ_{min} και του μεγίστου Δ_{max} γίνεται κατά τέτοιο τρόπο ώστε να ελαχιστοποιείτε ο θόρυβος καναλιού και η αποκοπή του κβαντιστή αντίστοιχα. Τέλος ο λόγος $\Delta_{\max} / \Delta_{\min}$ καθορίζει το δυναμικό εύρος του κβαντιστή, ενώ το Δ_0 ορίζει το βήμα κβάντισης για μοναδιαία μεταβολή του σήματος εισόδου.



α) Μεταδότης



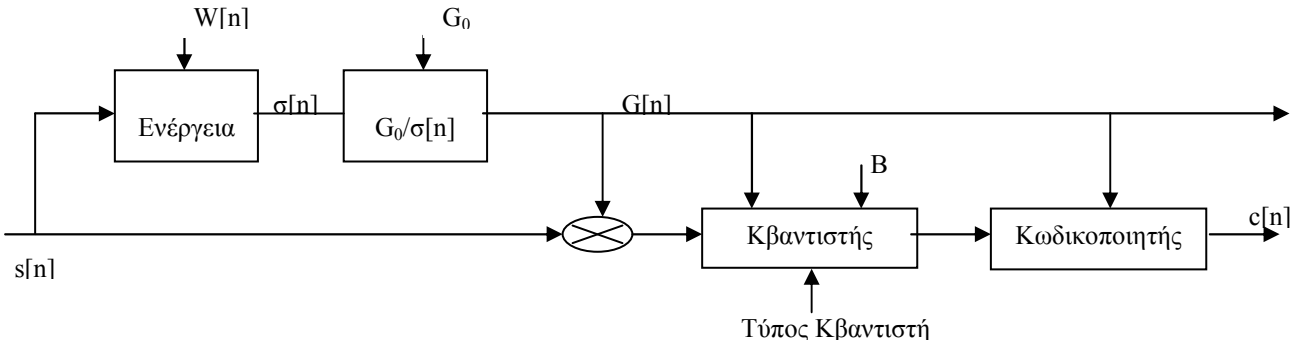
β) Δέκτης

Σχήμα 4.7:Μπλοκ διάγραμμα πομπού (α) και δέκτη (β) ενός forward estimation APCM με προσαρμογή του βήματος κβάντισης.

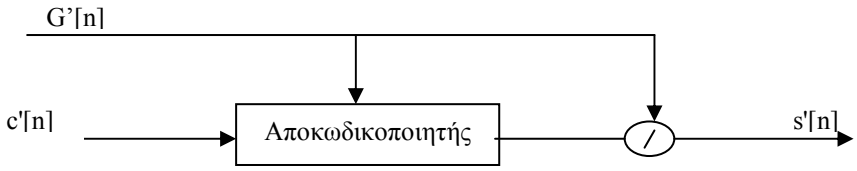
Η δεύτερη προσέγγιση τώρα ως προς την χρήση της διακύμανσης ονομάζεται *προσαρμογή κέρδους* (Σχήμα 4.8). Εδώ το σήμα της ομιλίας διαμορφώνεται από έναν χρονικά μεταβαλλόμενο παράγοντα κέρδους $G[n]$ ο οποίος και επιλέγετε κατά τέτοιο τρόπο ώστε να είναι αντιστρόφως ανάλογος με την *διακύμανση*. Ο παράγοντας αυτός δίνεται από τον τύπο

$$G[n] = \frac{G_0}{\sigma[n]}$$

όπου G_0 η τιμή του για μοναδιαία διακύμανση. Με τον τρόπο αυτό πετυχαίνουμε μεγάλο κέρδος για σήματα χαμηλής ενέργειας και μικρό κέρδος για σήματα μικρής ενέργειας με αποτέλεσμα ένα σήμα με σχετικά ομοιόμορφο εύρος το οποίο είναι πιο κατάλληλο για τον *ομοιόμορφο* κβαντιστή που ακολουθεί. Και εδώ όπως και στην πιο πάνω περίπτωση υπάρχουν ορισμένες μέγιστες και ελάχιστες τιμές ενώ ο λόγος $G_{\max}/G_{\min}$ καθορίζει το δυναμικό εύρος του ¨διαμορφωμένου¨ σήματος. Τέλος με κατάλληλη επιλογή του G_0 μπορούμε να μορφοποιήσουμε το σήμα σε οποιοδήποτε επιθυμητό επίπεδο.



α) Μεταδότης



β) Δέκτης

Σχήμα 4.8:Μπλοκ διάγραμμα πομπού (α) και δέκτη (β) ενός forward estimation APCM με προσαρμογή κέρδους.

4.3.2 Adaptive Quantization with Backward Estimation (AQB)

Στην περίπτωση της *AQB* σε αντίθεση με την *AQF* οι παράμετροι προσαρμογής της κβάντισης (βήμα κβάντισης, κέρδος) δεν μεταδίδονται σαν πλευρική πληροφορία αλλά προκύπτουν μέσα από το μεταδιδόμενο κωδικοποιημένο σήμα ομιλίας. Επίσης ο υπολογισμός της *διακύμανσης* του σήματος της ομιλίας εδώ γίνεται με βάση τις προηγούμενες τιμές του σήματος και για τον υπολογισμό αυτό εφαρμόζεται ένα *εκθετικό παράθυρο* το οποίο και οδηγεί στην παρακάτω διαφορική εξίσωση

$$\sigma^2[n] = a\sigma^2[n-1] + (1-a)s^2[n]$$

(από την οποία και μπορούμε να παρατηρήσουμε ότι η διακύμανση βασίζεται σε προηγούμενες τιμές του κωδικοποιημένου σήματος).

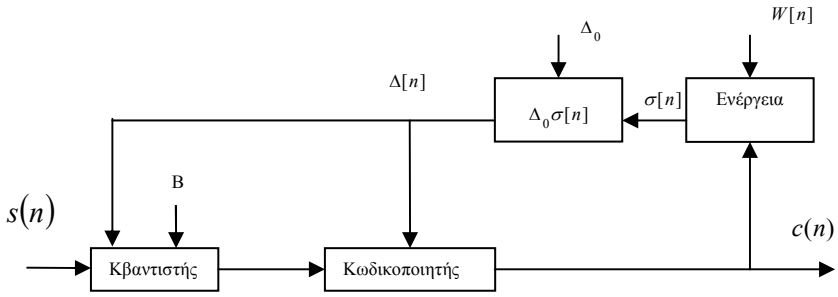
Ως προς αυτή την υπολογιζόμενη διακύμανση υπάρχουν και εδώ δύο βασικές προσεγγίσεις για την υλοποίηση κβαντιστών οι οποίες είναι η *προσαρμογή του βήματος κβάντισης* (Σχήμα 4.9) κατά την οποία το βήμα κβάντισης δίνεται πάλι μέσα από τον τύπο

$$\Delta[n] = \frac{\Delta_0 \sigma[n]}{2^{n-1}}$$

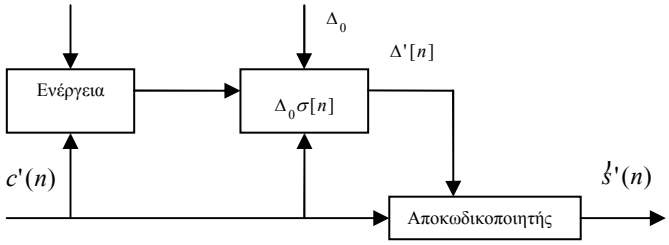
και η *προσαρμογή κέρδους* (Σχήμα 4.10) όπου το κέρδος δίνεται μέσα από τον τύπο

$$G[n] = \frac{G_0}{\sigma[n]}$$

Τέλος στην περίπτωση των backward estimation συστημάτων χρησιμοποιείται *στιγμιαία κβάντιση* γιατί εδώ δεν μας απασχολεί το μέγεθος της μεταδιδόμενης πλευρικής πληροφορίας.

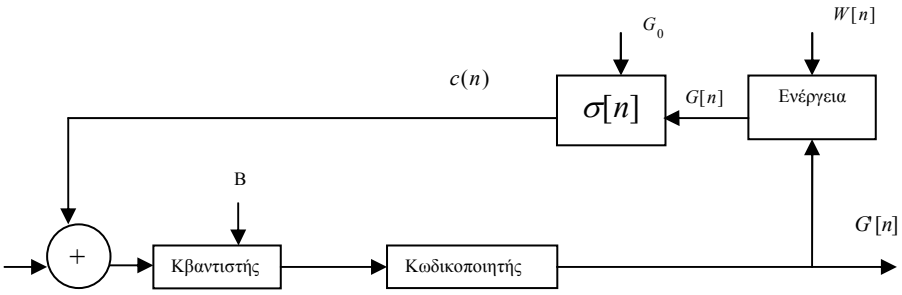


α) Πομπός

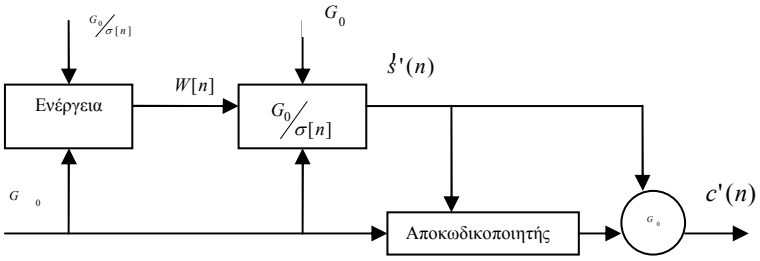


β) Δέκτης

Σχήμα 4.9:Μπλοκ διάγραμμα πομπού (α) και δέκτη (β) ενός backward estimation APCM με προσαρμογή του βήματος κβάντισης.



α) Πομπός



β) Δέκτης

Σχήμα 4.10:Μπλοκ διάγραμμα πομπού (α) και δέκτη (β) ενός backward estimation APCM με προσαρμογή κέρδους.

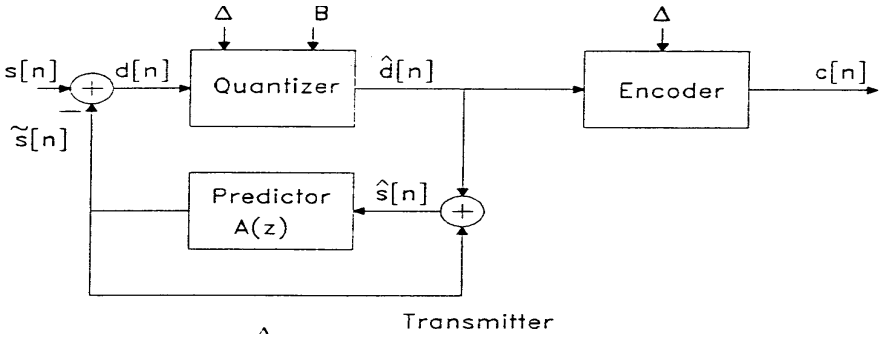
4.4 Προσαρμοστική Διανυσματική Διαμόρφωση (Adaptive VQ)

Η προσαρμοστικότητα στους διανυσματικούς κωδικοποιητές βασίζεται στο γεγονός ότι το ¨λεξικό¨ τους δεν είναι σταθερό αλλά προσαρμόζεται ανάλογα με τις μεταβολές της ομιλίας. Και εδώ μπορούμε να διακρίνουμε δύο τύπους κωδικοποιητών με *εμπρόσθια* και με *οπίσθια*

προσαρμογή. Στην περίπτωση της *εμπρόσθιας προσαρμοστικής διανυσματικής κβάντισης* η ανανέωση του ¨λεξικού¨ βασίζεται στα τρέχοντα (και μερικές φορές στα μελλοντικά) δεδομένα και πάντα κάποια επιπρόσθετη πληροφορία πρέπει να μεταδοθεί. Ενώ στην *οπίσθια προσαρμοστική διανυσματική κβάντιση* η αναπροσαρμογή του ¨λεξικού¨ γίνεται με βάση προηγούμενα δεδομένα τα οποία είναι διαθέσιμα και στον αποκωδικοποιητή. Σε γενικές γραμμές υπάρχει μια μεγάλη αναλογία μεταξύ των δυο αυτών κβαντίσεων με τις αντίστοιχες *AQF* (Παράγραφος 4.3.1) και *AQB* (Παράγραφος 4.3.2) της βαθμωτής κωδικοποίησης.

4.5 Διαφορική Παλμοκωδική Διαμόρφωση (Differential PCM)

Στην διαφορική παλμοκωδική διαμόρφωση θεωρούμε ότι το σήμα της ομιλίας είναι χρονικά αμετάβλητο και κατά αυτή την έννοια οδηγούμαστε σε μη προσαρμοστικές μεθόδους κωδικοποίησης. Η όλη λειτουργία της βασίζεται στο γεγονός ότι το σήμα της φωνής είναι βαθυπερατό και ότι τα γειτονικά δείγματα έχουν μεγάλη συσχέτιση. Η διαφορική παλμοκωδική διαμόρφωση αφαιρεί αυτόν τον συσχετισμό κατά το μέγιστο δυνατό και για τον σκοπό αυτό χρησιμοποιεί μια προσέγγιση $\tilde{s}(n)$ του υπάρχοντος σήματος $s(n)$ η οποία προσέγγιση προκύπτει με βάση έναν *γραμμικό προγνώστη*. Ο γραμμικός αυτός προγνώστης βασίζεται στον γραμμικό συνδυασμό ενός ή περισσότερων προηγούμενων δειγμάτων και καθορίζει την πολυπλοκότητα του *DPCM* κωδικοποιητή. Έτσι η πιο απλή μορφή του είναι ο *πρώτης τάξεως προγνώστης* όπου η διαφορά μεταξύ των διαδοχικών δειγμάτων είναι και το σήμα προσέγγισης ενώ σε μια πιο πολύπλοκη μορφή του χρησιμοποιούνται περισσότερα από ένα δείγματα τα οποία και συνδυάζονται για τον υπολογισμό ενός καλύτερου σήματος προσέγγισης. Η διαφορά τώρα μεταξύ αυτό του σήματος προσέγγισης $\tilde{s}(n)$ και του υπάρχοντος σήματος $s(n)$ ονομάζεται *σφάλμα πρόγνωσης ή διαφορικό σήμα* $d(n)$ και είναι αυτό το οποίο κβαντίζεται και μεταδίδεται (Σχήμα 4.11). Το σήμα προσέγγισης μας δίνεται μέσα από την εξίσωση $\tilde{s}(n) = \sum_{i=1}^{P} a_i \hat{s}(n-i)$ όπου $a_i, i = 1,...,P$ οι συντελεστές του προγνώστη και $\hat{s}(n)$ οι προηγούμενες κβαντισμένες τιμές, ενώ ο προγνώστης μας δίνεται μέσα από την εξίσωση η οποία είναι όμοια με αυτή ενός *FIR* φίλτρου: $A(z) = \sum_{i=1}^{P} a_i z^{-i}$ με P την τάξη μεγέθους του. Οι συντελεστές του προγνώστη υπολογίζονται κατά τέτοιο τρόπο ώστε να έχουμε το ελάχιστο δυνατό *σφάλμα πρόγνωσης*. Το *σφάλμα πρόγνωσης* τώρα μας δίνεται μέσα από τον τύπο $d(n) = s(n) - \tilde{s}(n)$ ενώ το σφάλμα του κβαντιστή $q_d(n)$ από τον τύπο $\hat{d}(n) = Q[d(n)] = d(n) + q_d(n) \Rightarrow q_d(n) = \hat{d}(n) - d(n)$ (Σχήμα 4.11).

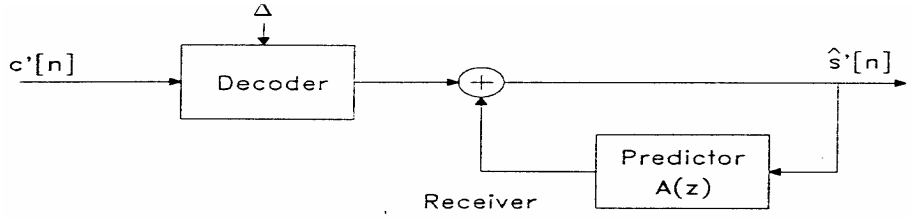


Σχήμα 4.11:Μπλοκ διάγραμμα ενός DPCM διαμορφωτή.

Επίσης από το Σχήμα 4.11 έχουμε $\hat{s}(n) = \tilde{s}(n) + \hat{d}(n)$ άρα με βάση τις παραπάνω εξισώσεις μπορούμε να γράψουμε $\hat{s}(n) = \tilde{s}(n) + d(n) + q_d(n) = s(n) - d(n) + d(n) + q_d(n) = s(n) + q_d(n)$

$\Rightarrow q_d(n) = \hat{s}(n) - s(n)$ όποτε συμπεραίνουμε ότι <u>το σφάλμα κβάντισης του *σφάλματος πρόγνωσης* είναι ίσο με το σφάλμα κβάντισης του αρχικού σήματος της φωνής όμως έχει πολύ μικρότερη ενέργεια (δυναμικό εύρος) από αυτό του αρχικού σήματος. Αυτό έχει ως αποτέλεσμα το σφάλμα κβάντισης του *DPCM* να είναι πολύ μικρότερο από αυτό του *PCM.*</u> Η βελτίωση τώρα που εισάγεται στον σηματοθορυβικό λόγο δίνεται από την σχέση $SNR_{DPCM}(dB) = SNR_{PCM}(dB) + 10\log G_p$ όπου G_p είναι το *κέρδος πρόγνωσης* και δίνεται μέσα από τον τύπο $G_p = \frac{\sigma_s^2}{\sigma_d^2}$ με σ_s^2, σ_d^2 τις ενέργειες του σήματος της ομιλίας και του *σφάλματος πρόγνωσης* αντίστοιχα. Είναι φανερό ότι καλύτερη πρόγνωση μπορεί να βελτιώσει την απόδοση του DPCM γιατί μειώνεται ο παράγοντας σ_d^2 άρα αυξάνεται το SNR_{DPCM} .Για τον λόγο αυτό και έχει μεγάλη σημασία ο προγνώστης. Συνήθως αυξάνοντας την τάξη πρόγνωσης του, δηλαδή αυξάνοντας των αριθμό των προηγούμενων δειγμάτων αυξάνεται και ο σηματοθορυβικός λόγος όμως η αναλογία αυτή διατηρείται μέχρι την $4^η$ ή $5^η$ τάξη ενώ πάνω από αυτές η βελτίωση είναι οριακή.

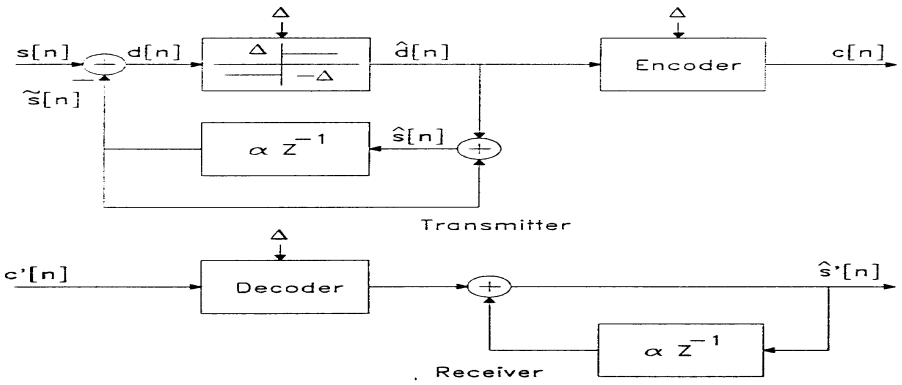
Βέβαια στην περίπτωση του βασικού DPCM που εξετάζουμε εδώ ο κβαντιστής είναι σταθερός, ομοιόμορφης κβάντισης, μπορεί όμως να μορφοποιηθεί σε όλες τις κατηγορίες που εξετάσαμε στις προηγούμενες παραγράφους. Μια τέτοια περίπτωση θα δούμε στην συνέχεια στην *προσαρμοστική διαφορική παλμοκωδική διαμόρφωση (ADPCM).*



Σχήμα 4.12:Μπλοκ διάγραμμα ενός DPCM αποδιαμορφωτή.

4.6 Γραμμική Διαμόρφωση Δέλτα (Linear DM)

Η γραμμική διαμόρφωση δέλτα είναι μια υποπερίπτωση του DPCM όπου ο προγνώστης είναι πρώτης τάξης (P=1) και ο κβαντιστής που χρησιμοποιείται είναι ενός bit. Εδώ βέβαια πρέπει για την διατήρηση της ποιότητας του κωδικοποιημένου σήματος η συχνότητα δειγματοληψίας να είναι αρκετές φορές μεγαλύτερη από την συχνότητα δειγματοληψίας του Nyquist δηλαδή: $f_{DM} = 2Rf_N$ όπου R είναι ο *δείκτης δειγματοληψίας* και f_N η συχνότητα Nyquist. Κατά αυτό τον τρόπο το σήμα που προκύπτει είναι το ¨υπερδειγματοληπτημένο¨ σήμα εισόδου. Στην ουσία <u>ο *δείκτης δειγματοληψίας* $R = F_{DM} / 2F_N$ παίζει το ρόλο του αριθμού των bits/δείγμα για τον κβαντιστή</u> ενώ ο διπλασιασμός της τιμής του οδηγεί σε αύξηση του SNR περίπου κατά 9 dB. Αποτέλεσμα αυτού του γεγονότος είναι να παρατηρείται μεγάλη συσχέτιση μεταξύ των δειγμάτων κάτι το οποίο μας οδηγεί σε ένα βελτιωμένο *σφάλμα πρόγνωσης* d(n) έτσι ώστε αυτό να μπορεί να κωδικοποιηθεί αποτελεσματικά με 1 bit. Το *σφάλμα πρόγνωσης* εδώ δίνεται μέσα από την εξίσωση $d(n) = s(n) - \tilde{s}(n) = s(n) - a \cdot \hat{s}(n-1)$ (Σχήμα 4.13) ενώ ο προγνώστης από την $p(z) = az^{-1}$ με $a < 1$.

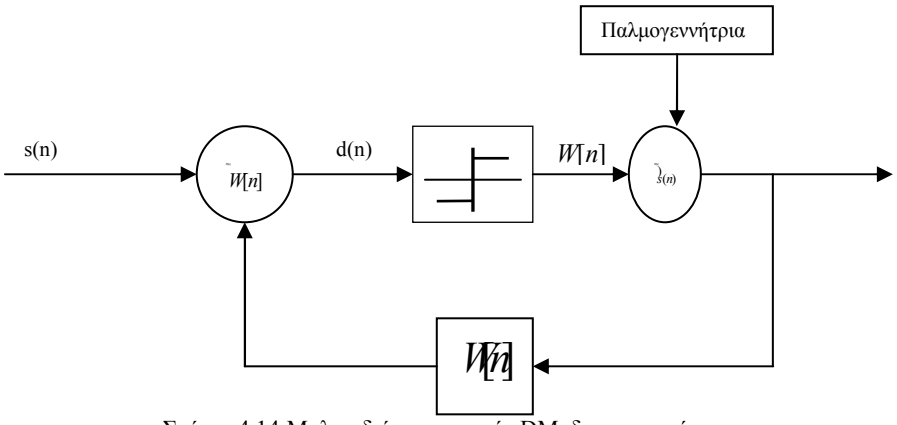


Σχήμα 4.13:Μπλοκ διάγραμμα ενός DM α) διαμορφωτή β) αποδιαμορφωτή.

Η διάταξη του Σχήματος 4.13 τώρα μπορεί να τροποποιηθεί σε αυτή του Σχήματος 4.14 όπου ο προγνώστης ο οποίος εισάγει καθυστέρηση καθώς και ο αθροιστής έχουν αντικατασταθεί από έναν ολοκληρωτή. Η διαδικασία εδώ έχει ως εξής, το σήμα εισόδου $s(n)$ συγκρίνεται με μια κλιμακωτή προσέγγιση $\tilde{s}(n)$ και το *σφάλμα πρόγνωσης* $d(n)$ κβαντίζεται σε δύο στάθμες ανάλογα με το πρόσημο της διαφοράς (αν είναι θετικό ή αρνητικό) σύμφωνα με την σχέση

$$\hat{d}(n) = \begin{cases} \Delta \alpha \nu \, \mathrm{d(n)} \geq 0 \\ -\Delta \alpha \nu \, \mathrm{d(n)} \leq 0 \end{cases}$$

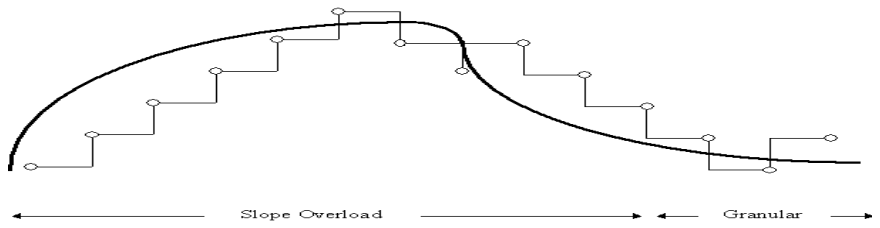
Στη συνέχεια η έξοδος αυτή του κβαντιστή (συγκριτή) πολλαπλασιάζεται με την έξοδο μιας παλμογεννήτριας και το αποτέλεσμα τροφοδοτεί έναν ολοκληρωτή που σε κάθε παλμό αποκρίνεται με μια στάθμη θετική ή αρνητική η οποία είναι η κλιμακωτή προσέγγιση $\tilde{s}(n)$ του σήματος. Επειδή στο αποτέλεσμα του γινομένου το οποίο οδηγείται στην είσοδο του ολοκληρωτή επειδή υπάρχουν μόνο δύο δυνατές στάθμες αυτό μπορεί να μεταδοθεί με την χρήση μιας δυαδικής κυματομορφής. Ο δέκτης τώρα αποτελείται από έναν ολοκληρωτή και ένα βαθυπερατό φίλτρο και επειδή σε ένα πραγματικό σύστημα το LPF δίνει από μόνο του ένα μέτρο ολοκλήρωσης μπορούμε να καταργήσουμε τον ολοκληρωτή του δέκτη. Επίσης για τον ολοκληρωτή του πομπού δεν υπάρχει απαίτηση για την ύπαρξη ενός ιδανικού εξαρτήματος οπότε μπορεί να υλοποιηθεί πολύ απλά με ένα φίλτρο RC χαμηλών συχνοτήτων. Όλα αυτά αποτελούν ένα από τα βασικά πλεονεκτήματα της DM σε σχέση με το PCM γιατί οδηγούν σε απλούστευση των απαιτούμενων κυκλωμάτων.



Σχήμα 4.14:Μπλοκ διάγραμμα ενός DM διαμορφωτή.

Ένα άλλο βασικό χαρακτηριστικό της LDM είναι η κλίση της η οποία και δίνεται από το λόγο Δ/T όπου T είναι η περίοδος δειγματοληψίας. Κατά αυτό τον τρόπο μπορεί να ελεγχθεί με βάση τις τιμές του Δ και της T. Το γεγονός όμως ότι η γραμμική διαμόρφωση δέλτα τείνει να ακολουθεί το

σήμα ομιλίας με γραμμικό τρόπο οδηγεί σε ένα πρόβλημα το οποίο είναι γνωστό σαν *¨υπερφόρτωση κλίσης¨* και η οποία παρουσιάζεται λόγω της κλίσης $d[s(n)]/dt$. Έτσι όταν το σήμα εισόδου μεταβάλλεται κατά τέτοιο τρόπο ώστε τα διαδοχικά δείγματα να μην διαφέρουν μεταξύ τους περισσότερο από εύρος Δ το σήμα προσέγγισης $\tilde{s}(n)$ ακολουθεί με έναν ικανοποιητικό τρόπο. Όταν όμως αυτή η διαφορά γίνει μεγαλύτερη από Δ το σήμα προσέγγισης αδυνατεί να παρακολουθήσει το σήμα εισόδου και έχουμε *¨υπερφόρτωση κλίσης¨*. Το πρόβλημα αυτό μπορεί να αντιμετωπιστεί φιλτράροντας το σήμα για να μειώσουμε την μέγιστη ταχύτητα μεταβολής ή αυξάνοντας το εύρος βαθμίδας Δ και/ή το ρυθμό δειγματοληψίας. Όμως το φιλτράρισμα του σήματος και η αύξηση του Δ έχουν σαν αποτέλεσμα χαμηλή διακριτική ικανότητα του σήματος. Η αλλοίωση αυτή ονομάζεται ονομάζετε *granular noise* (Σχήμα 4.15) και οφείλετε στην ¨ολίσθηση¨ του σήματος εξόδου της διαμόρφωσης σε σχέση με την κυματομορφή του σήματος εισόδου. Επίσης η αύξηση του ρυθμού δειγματοληψίας εισάγει μια αύξηση στο απαιτούμενο εύρος ζώνης. Η λύση στα παραπάνω προβλήματα είναι η ανίχνευση της κατάστασης υπερφόρτωσης και η ανάλογη προσαρμογή του Δ. Οι κωδικοποιητές που λειτουργούν με αυτό τον τρόπο ονομάζονται κωδικοποιητές *προσαρμοστικής διαμόρφωσης δέλτα (Adaptive DM)* και εξετάζονται στην επόμενη παράγραφο.



Σχήμα 4.15:Σφάλματα granular noise και ¨υπερφόρτωση κλίσης¨ της DM.

4.7 Προσαρμοστική Διαμόρφωση Δέλτα (Adaptive DM)

Όπως είδαμε και παραπάνω για την αντιμετώπιση των διαφόρων προβλημάτων που παρουσιάζονται στην DM απαιτείται ένα προσαρμοζόμενο βήμα Δ. Η εφαρμογή αυτής της ιδέας γίνεται στην *ADM*. Εδώ το βήμα Δ προσαρμόζεται κατά τέτοιο τρόπο ώστε να ακολουθεί τις μεταβολές του σήματος εισόδου και ισχύουν σε γενικές γραμμές οι βασικές αρχές της προσαρμοστικής κβάντισης που συζητήθηκαν στις Παράγραφο 4.2.4. Συνήθως στην ADM χρησιμοποιείται backward estimation (Παράγραφος 4.3.2) έτσι ώστε να πετυχαίνουμε διατήρηση της απλότητας στην υλοποίηση και αποφυγή μετάδοσης πλευρικής πληροφορίας. Περισσότερα για την ADM θα δούμε στην επόμενη παράγραφο οπού εξετάζεται η πιο διαδεδομένη μορφή της, η *διαμόρφωση δέλτα συνεχώς μεταβαλλόμενης κλίσης*.

4.8 Διαμόρφωση Δέλτα Συνεχούς Μεταβαλλόμενης Κλίσης (Continuously Variable Slope DM)

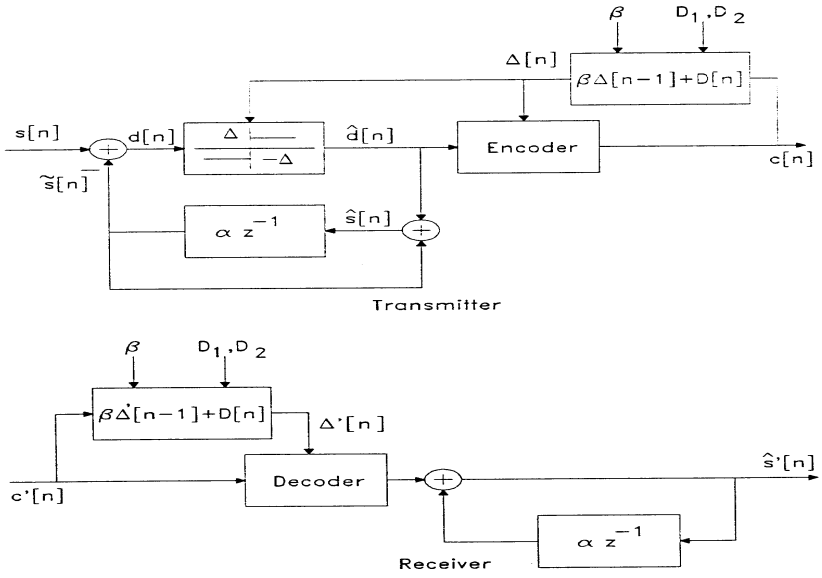
Παρά τα πλεονεκτήματα που εισάγει η ADM παρουσιάζει το μειονέκτημα ότι η ομιλία μπορεί να υποστεί σημαντική αλλοίωση εάν υπάρχουν σφάλματα κατά την μετάδοση. Τα σφάλματα αυτά μπορεί να πολλαπλασιαστούν και να είναι ¨υπολογίσιμα¨ με τον χρόνο για τον λόγο αυτό χρησιμοποιούμε την CVSD. Εδώ το μέγεθος του βήματος προσαρμογής Δ εξαρτάται από *δύο* προηγούμενες τιμές του σήματος στην έξοδο του κωδικοποιητή και δίνεται μέσα από την εξίσωση: $\Delta(n) = \beta\Delta(n-1) + D(n)$. Το $D(n)$ δίνεται μέσα από την

$$D(n) = \begin{cases} D_1 \text{ εαν } c(n-1) = c(n-2) = c(n-3) \\ D_2 \text{ αλλιώς} \end{cases}$$ όπου $D_1 >> D_2 > 0$, $1 > \beta > 0$ και $c(n)$ το σήμα εξόδου

(Σχήμα 4.16). Με βάση τα παραπάνω μπορούμε να πάρουμε την μέγιστη και ελάχιστη τιμή του Δ σύμφωνα με τις $\Delta_{\max} = \frac{D_2}{1-\beta}$, $\Delta_{\min} = \frac{D_1}{1-\beta}$.

Οπότε βασιζόμενοι στις εξισώσεις αυτές βλέπουμε ότι όταν έχουμε αύξηση ή μείωση του σήματος εξόδου για τρία συνεχόμενα δείγματα, το βήμα προσαρμογής αυξάνεται και κατά αυτόν τον τρόπο έχουμε καλύτερη παρακολούθηση του σήματος εισόδου. Σε όλες τις άλλες περιπτώσεις το Δ μειώνεται με ρυθμό ο οποίος καθορίζεται από την τιμή του β. Έτσι για τιμές του β κοντά στο 1 έχουμε πιο γρήγορη προσαρμογή ενώ το αντίθετο γίνεται για τιμές του β κοντά στο 0. Εδώ θα πρέπει να αναφέρουμε ότι πρώτα επιλέγεται το β μετά τα D_1, D_2 και τέλος τα $\Delta_{\min}$, $\Delta_{\max}$.

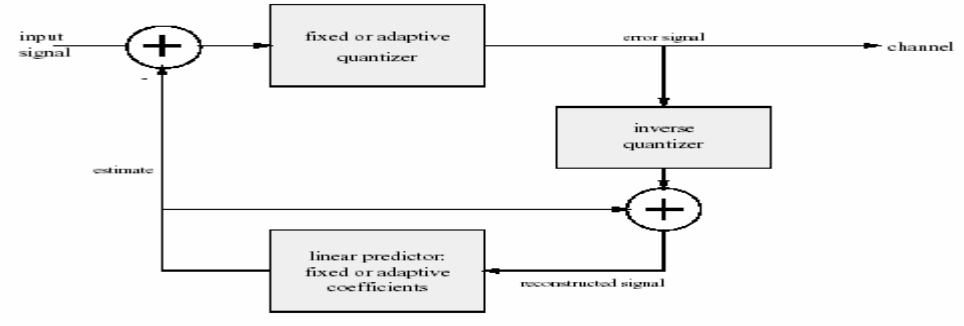
Σαν συμπέρασμα τώρα για την CVSD μπορούμε να πούμε ότι είναι πολύ λιγότερο ευαίσθητη στα σφάλματα που παρουσιάζονται κατά την μετάδοση και αυτό είχε ως αποτέλεσμα να χρησιμοποιείται ευρέως για αρκετό διάστημα. Στην συνέχεια βέβαια αναπτύχθηκαν πιο προηγμένες τεχνικές όπως οι ADPCM ή APC οι οποίες την αντικατέστησαν και τις οποίες θα εξετάσουμε παρακάτω.



Σχήμα 4.16: Μπλοκ διάγραμμα ενός CVSD α) διαμορφωτή β) αποδιαμορφωτή.

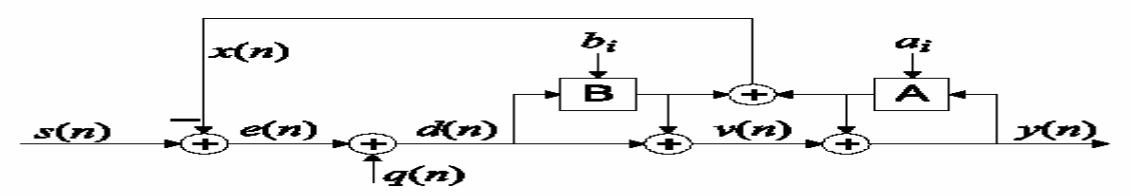
4.9 Προσαρμοστική Διαφορική Παλμοκωδική Διαμόρφωση (Adaptive DPCM)

Στην περίπτωση της ADPCM ο χαρακτηρισμός ¨προσαρμοστική¨ αναφέρεται στο ότι υπάρχει η δυνατότητα να έχουμε *προσαρμοστικούς ή μη κβαντιστές αλλά και προσαρμοστικούς ή μη προγνώστες* (Σχήμα 4.17).



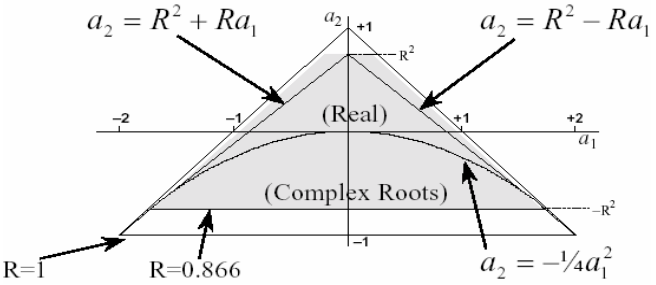
Σχήμα 4.17: Μπλοκ διάγραμμα ενός ADPCM διαμορφωτή.

Έτσι μπορούμε να διακρίνουμε τεσσάρων συνδυασμών ADPCM κωδικοποιητές. Σαν παράδειγμα ενός διαδεδομένου συστήματος ADPCM μπορούμε να αναφέρουμε το στάνταρτ ITU G.721. Στην περίπτωση αυτού του στάνταρτ (Σχήμα 4.18) έχουμε ένα *backward adaptation* προσαρμοστικό κβαντιστή 15 επιπέδων (οπότε έχουμε 4 bits x 8 kHz = 32 kbit/s) και έναν *backward adaptation* προσαρμοστικό προγνώστη πόλων και μηδενικών (2 πόλων, 6 μηδενικών). Ο προγνώστης αυτός στην ουσία είναι ο συνδυασμός δύο διαφορετικών φίλτρων, ενός δεύτερης τάξης $A(z) = a_1 z^{-1} + a_2 z^{-2}$ και ενός έκτης τάξης $B(z) = b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + b_4 z^{-4} + b_5 z^{-5} + b_6 z^{-6}$ τα οποία όπως βλέπουμε βασίζονται μόνο σε προηγούμενα δείγματα. Η προσαρμογή τώρα του προγνώστη γίνεται τροποποιώντας τους συντελεστές των δύο φίλτρων $A(z)$ και $B(z)$.



Σχήμα 4.18: Μπλοκ διάγραμμα ενός ADPCM G.721 διαμορφωτή.

Η σταθερότητα του προγνώστη βέβαια, καθορίζεται με βάση της δύο ρίζες του $A(z)$. Θέλουμε λοιπόν οι πόλοι (ρίζες) του να είναι μικρότερες από την ακτίνα R (Σχήμα 4.19), γεγονός το οποίο έχει ως αποτέλεσμα μεγάλο κέρδος και γρήγορη εξασθένηση των ταλαντώσεων.



Σχήμα 4.19: Διάταξη των πόλων του φίλτρου A(z).

Πάντως περά από το παράδειγμα του G.721 το οποίο είναι ένα μεγάλης πολυπλοκότητας σύστημα με προσαρμοστικό κβαντιστή και προσαρμοστικό προγνώστη υπάρχουν και τα μεσαίας πολυπλοκότητας συστήματα τα οποία χρησιμοποιούν προσαρμοστικούς κβαντιστές και σταθερούς προγνώστες.

4.9.1 Διαφορική Παλμοκωδική Διαμόρφωση με Προσαρμοστική Κβάντιση (DPCM with Adaptive Quantization)

Η πρόγνωση σε τέτοιου είδους κωδικοποιητές με *προσαρμοστική κβάντιση* εμπίπτει στην περίπτωση των *forward adaptation* ή *backward adaptation* και ισχύουν όσα έχουμε αναφέρει σε πιο πάνω παραγράφους (Παράγραφοι 4.3.1 & 4.2.2). Με την χρήση τους πετυχαίνουμε περίπου 10 – 12 dB βελτίωση σε σχέση με ένα σταθερής ομοιόμορφης κβάντισης κωδικοποιητή που χρησιμοποιεί τον ίδιο αριθμό bits.

4.9.2 Διαφορική Παλμοκωδική Διαμόρφωση με Προσαρμοστική Πρόγνωση (DPCM with Adaptive Prediction)

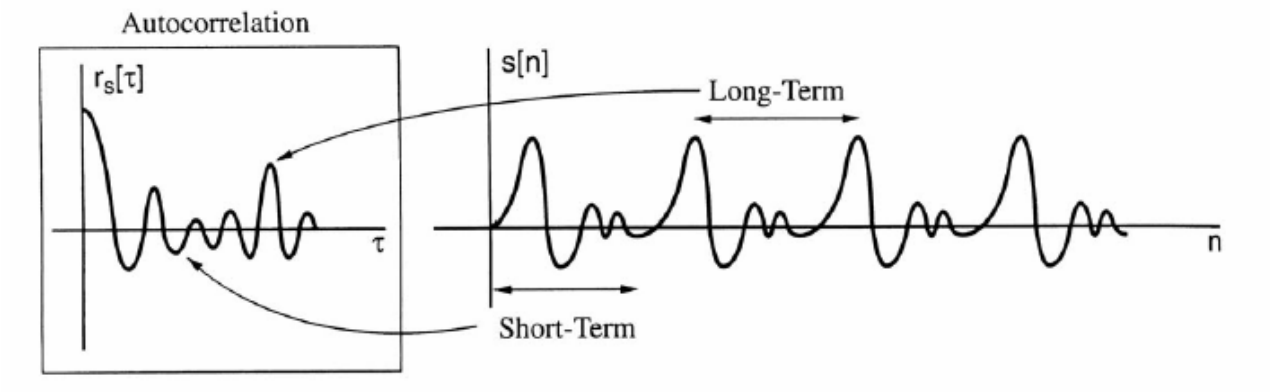
Η περίπτωση της *προσαρμοστικής πρόγνωσης* αναλύεται εκτενώς στην παράγραφο που ακολουθεί.

4.10 Προσαρμοστική Προγνωστική Κωδικοποίηση (Adaptive Predictive Coding)

Οι APC κωδικοποιητές είναι ο συνδυασμός τόσο της κωδικοποίησης κυματομορφής όσο και της παραμετρικής κωδικοποίησης (στην ουσία έχουμε κωδικοποιητές κυματομορφής με προσαρμοστική πρόγνωση). Ακριβώς λοιπόν επειδή υπεισέρχεται και η έννοια της κυματομορφής θα τους μελετήσουμε εδώ.

Η βασική λειτουργία τους τώρα βασίζεται στην *γραμμική πρόγνωση* (Παράγραφος 3.7) και με βάση αυτήν τους χωρίζουμε σε δύο βασικές κατηγορίες. Στους *feed – forward adaptive predictive* κωδικοποιητές στους οποίους, οι παράμετροι του προγνώστη προκύπτουν με βάση το σήμα εισόδου και μεταδίδονται στον δέκτη σαν πλευρική πληροφορία και στους *feedback* κωδικοποιητές στους οποίους, οι παράμετροι προκύπτουν με βάση το σήμα εξόδου οπότε δεν χρειάζεται να μεταδίδονται σαν πλευρική πληροφορία εφόσον είναι διαθέσιμοι στον δέκτη.

Εκτός όμως από αυτή την ταξινόμηση, μπορούμε να διακρίνουμε και τους *adaptive predictive coders with pitch prediction (APC – PP)*. Αυτοί είναι στην ουσία κωδικοποιητές DPCM, οι οποίοι εκμεταλλεύονται τόσο την συσχέτιση μεταξύ δειγμάτων, όσο και την πλεονασμό μεταξύ των *θεμελιωδών συχνοτήτων (pitch)* του σήματος της ομιλίας. Για αυτόν ακριβώς τον λόγο και περιέχουν τόσο *βραχύς διάρκειας προγνώστες (short term predictor)* όσο και *μακράς διάρκειας προγνώστες (long term predictor)* (Σχήμα 4.20).



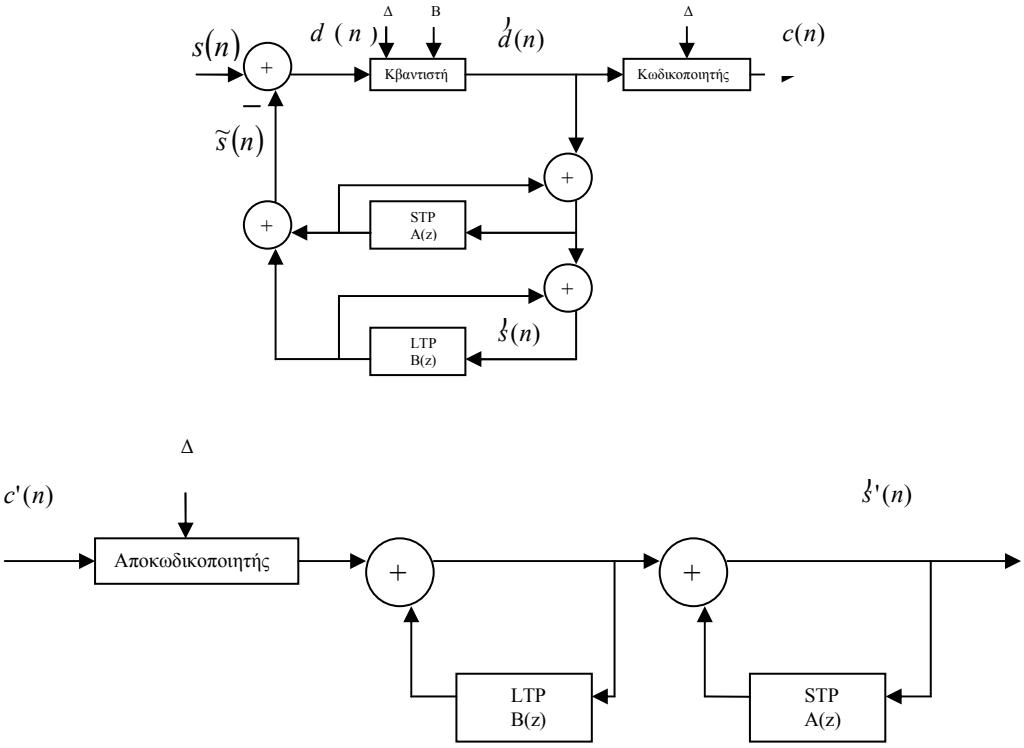
Σχήμα 4.20: Βραχύς διάρκειας και μακράς διάρκειας πρόγνωση βασιζόμενη στην αυτοσυσχέτιση του σήματος της ομιλίας.

Επίσης μια άλλη κατηγορία APC κωδικοποιητών είναι οι *noise – feedback* κωδικοποιητές στους οποίους γίνεται χρήση των ιδιοτήτων της *επικάλυψης (masking)*, η οποία παρουσιάζετε στο ανθρώπινο ακουστικό σύστημα. Με βάση λοιπόν αυτήν την ιδιότητα ελαχιστοποιείτε ο αντιληπτός θόρυβος μορφοποιώντας το φάσμα του θορύβου κωδικοποίησης έτσι ώστε να είναι παρόμοιο με το φάσμα του σήματος ομιλίας. Επειδή όμως σε αυτά τα συστήματα, η ενέργεια του θορύβου είναι μικρότερη από αυτή του σήματος ομιλίας έχουμε ως αποτέλεσμα την επικάλυψη.

4.10.1 Προσαρμοστικοί Προγνωστικοί Κωδικοποιητές με Πρόγνωση Θεμελιώδους Συχνότητας (APC - Pitch Prediction)

Οι κωδικοποιητές αυτοί (Σχήμα 4.21) όπως έχουμε πει περιέχουν τόσο βραχύς όσο και μακράς διάρκειας προγνώστες και βασίζονται στην περιοδικότητα την οποία παρουσιάζει το σήμα της ομιλίας. Ο βραχύς διάρκειας προγνώστης όπως έχουμε δει και σε άλλα συστήματα υλοποιείτε μέσα

από ένα *FIR* φίλτρο $A(z)=\sum_{i=1}^{P} a_i z^{-i}$ με P την τάξη μεγέθους του ενώ ο μακράς διάρκειας προγνώστης είναι της μορφής $B(z)=\sum_{i=-\lambda}^{\lambda} \beta_i z^{-\gamma-i}$ όπου τα β_i είναι οι συντελεστές του, $2\times\lambda+1$ η τάξη μεγέθους του και γ είναι η καθυστέρηση. Η καθυστέρηση αυτή είναι συνήθως ανάλογη μιας περιόδου μεταξύ των *θεμελιωδών συχνοτήτων (pitch)* ή ενός ακέραιου πολλαπλάσιου της και συνήθως επιλέγεται έτσι ώστε να είναι μικρότερη από το μέγεθος της μνήμης του LTP. Ωστόσο επειδή το φίλτρο μπορεί να γίνει ασταθές το $B(z)$ συνήθως παίρνει την μορφή $B(z)=\beta z^{-\gamma}$ με $\beta<1$.



Σχήμα 4.21: Μπλοκ διάγραμμα ενός pitch – predictive AP α)κωδικοποιητή β)αποκωδικοποιητή.

Η διαδικασία τώρα που εφαρμόζεται για την εύρεση των παραμέτρων του LTP σε τέτοιου είδους κωδικοποιητές ονομάζεται *open – loop*[11] (ενώ στους υβριδικούς κωδικοποιητές εφαρμόζεται η διαδικασία του *closed – loop*[12]). Σε αυτή την διαδικασία η εύρεση της *θεμελιώδους συχνότητας (pitch)* γίνεται είτε με βάση την αυτοσυσχέτιση στο σήμα της ομιλίας είτε με βάση το υπόλοιπο του *sort – term predictor*, ενώ το κέρδος β δίνεται από την σχέση,

$$\beta=\frac{R(\gamma)}{R(0)}=\frac{\sum_{n=0}^{N-1} s(n)s(n-\gamma)}{\sum_{n=0}^{N-1} s^2(n)}$$

[11] Στη διαδικασία *open loop* οι παράμετροι προς μετάδοση εξάγονται και κωδικοποιούνται χωρίς να λαμβάνεται υπόψη η διαφορά μεταξύ αρχικού και συντεθειμένου σήματος.

[12] Στη διαδικασία *closed loop* οι παράμετροι προς μετάδοση εξάγονται και κωδικοποιούνται έτσι ώστε η διαφορά μεταξύ αρχικού και συντεθειμένου σήματος να είναι η μικρότερη δυνατή. Παρουσιάζουν συνήθως μεγάλη πολυπλοκότητα.

με N το μέγεθος του μπλοκ ανάλυσης για τον LTP.

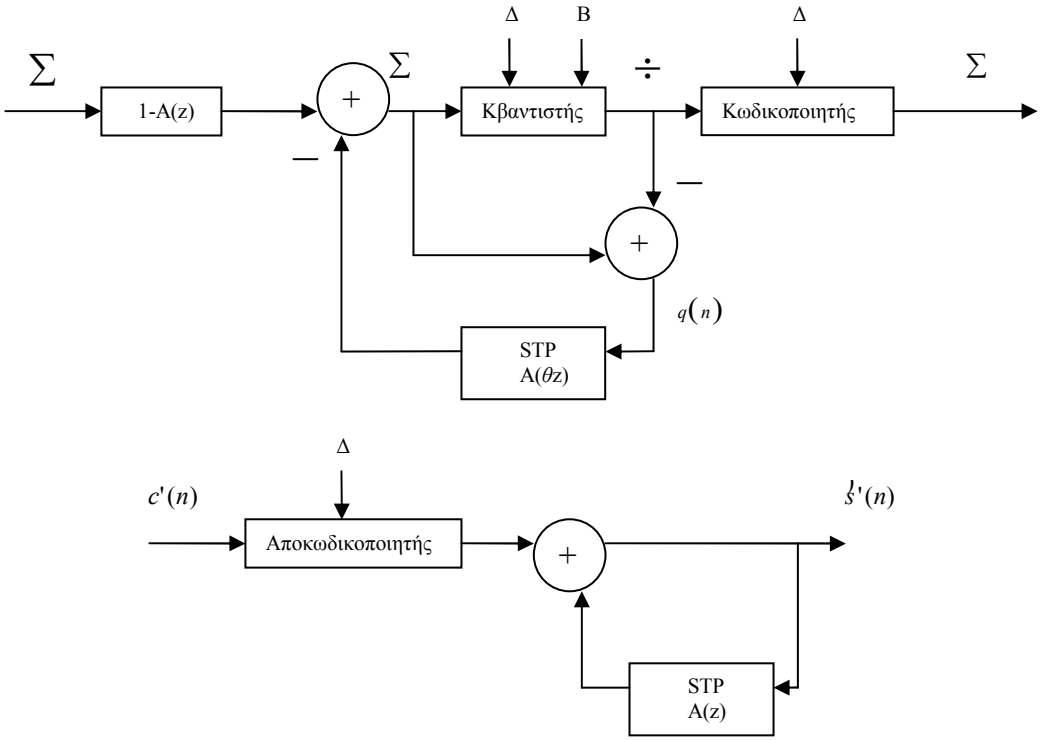
Βέβαια το συνολικό κέρδος το οποίο προκύπτει από τον συνδυασμό των δύο προγνωστών, των οποίων η σειρά δεν παίζει κανένα ρόλο, είναι σε γενικές γραμμές μικρότερο από άθροισμα των ανεξάρτητων κερδών.

4.10.2 Προσαρμοστικοί Προγνωστικοί Κωδικοποιητές με Πρόγνωση Θεμελιώδους Συχνότητας και Noise Feedback

Οι κωδικοποιητές αυτοί όπως είπαμε και πιο πάνω βασίζονται στο φαινόμενο της *επικάλυψης* και προκύπτουν από την κλασσική μορφή του DPCM στο οποίο έχει προστεθεί και ένα φίλτρο το οποίο ονομάζεται *noise shaping filter* και έχει την μορφή $W(z)=\dfrac{1-A(z)}{1-A(\theta z)}$ με $0<\theta<1$.

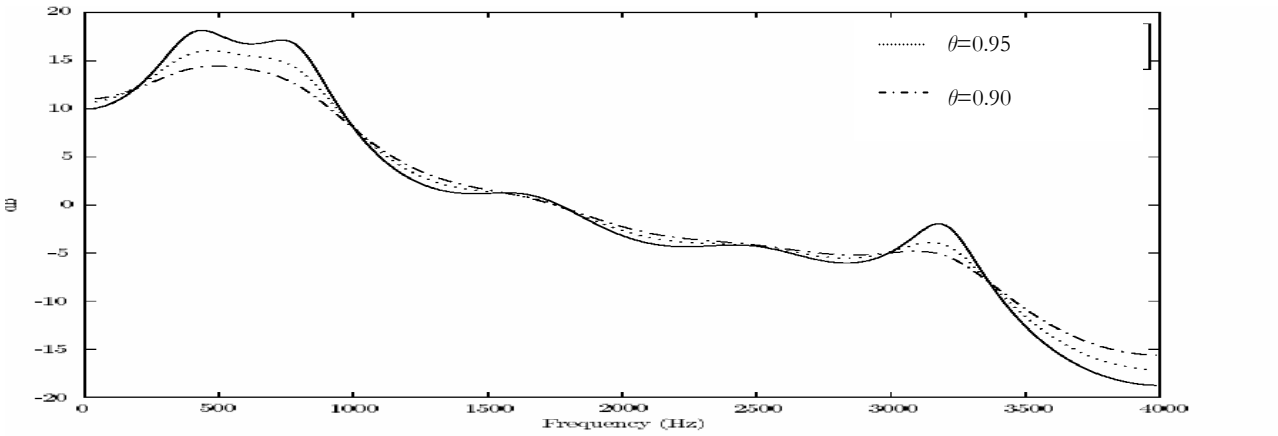
Στο τροποποιημένο αυτό DPCM σύστημα (Σχήμα 4.22) ο *μετασχηματισμός Z* του *σήματος σφάλματος* $y(n)=s(n)-\hat{s}(n)$ στην έξοδο, δίνεται από την εξίσωση, $Y(z)=Q(z)\dfrac{1-A(\theta z)}{1-A(z)}=\dfrac{Q(z)}{W(z)}$

και όπως βλέπουμε <u>είναι στην ουσία ο *θόρυβος κβάντισης* $q(n)$ φιλτραρισμένος από το αντίστροφο φίλτρο $W(z)$.</u>



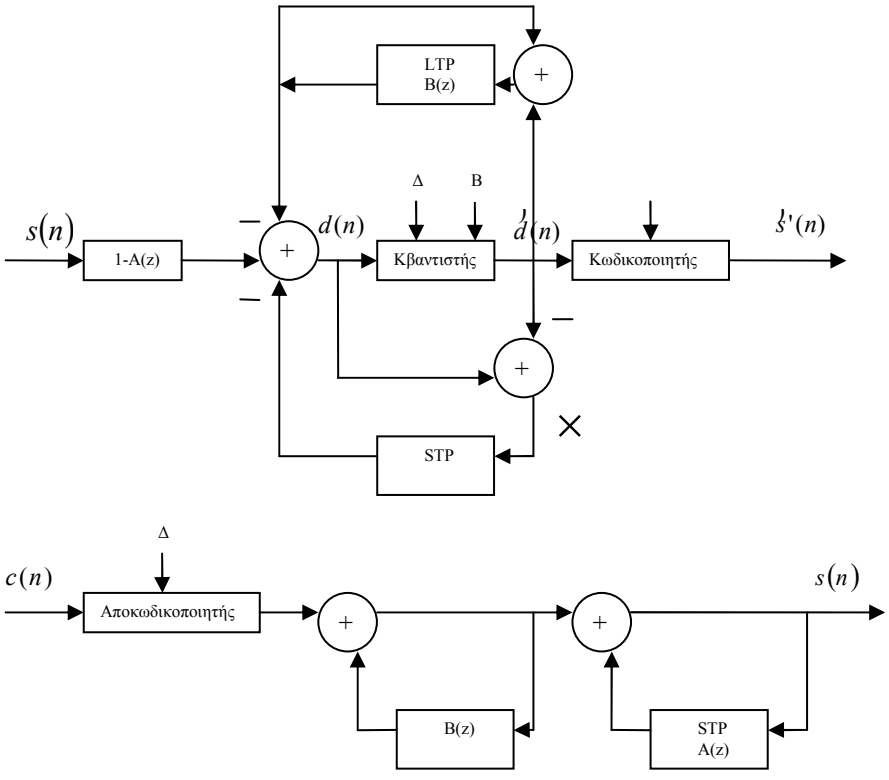
Σχήμα 4.22: Μπλοκ διάγραμμα ενός noise feedback DPCM α)κωδικοποιητή β)αποκωδικοποιητή.

Από το Σχήμα 4.23 τώρα στο οποίο φαίνεται η απόκριση του $(W(z))^{-1}$ για $\theta=0.90$ και $\theta=0.95$ παρατηρούμε ότι μπορούμε, <u>με κατάλληλη επιλογή του θ να ενισχύσουμε το *θόρυβο κβάντισης* στα σημεία εκείνα στα οποία το σήμα έχει μεγάλη ενέργεια (το $W(z)$ είναι μικρό με αποτέλεσμα ο λόγος $Q(z)/W(z)$ να είναι μεγάλος) και να ελαττώσουμε το θόρυβο, στα σημεία όπου το σήμα ομιλίας έχει μικρή ενέργεια, πετυχαίνοντας έτσι επικάλυψη</u>.



Σχήμα 4.23: Αποκρίσεις των φίλτρων $1/A(z)$ και $1/A(\theta z)$.

Ο ολοκληρωμένος τώρα *APC with Noise Feedback and Pitch Prediction (APC – PPNF)* φαίνεται στο Σχήμα 4.24 όπου σε σχέση με το *Noise Feedback DPCM* έχει προστεθεί και ο *μακράς διάρκειας προγνώστης (LTP)*. Έτσι σε αυτόν συνδυάζονται μια σειρά από λειτουργίες όπως είναι η προσαρμοστική πρόγνωση του φάσματος, προσαρμοστική πρόγνωση κορυφής και προσαρμοστική επικάλυψη του θορύβου γεγονός που μας οδηγεί σε ένα αποδοτικό σύστημα με bit rate από 9.6 έως 24 kbps.



Σχήμα 4.24: Μπλοκ διάγραμμα ενός APC – PPNF α)κωδικοποιητή β)αποκωδικοποιητή.

Τέλος να πούμε ότι με κατάλληλη επιλογή των παραμέτρων του *APC – PPNF* μπορούμε να πάρουμε ένα πλήθος παραλλαγών. Έτσι εάν η τιμή του συντελεστή του LTP είναι $\beta = 0$, δηλαδή στην ουσία δεν έχουμε μακράς διάρκειας πρόγνωση, μπορούμε να διακρίνουμε τις εξής υποπεριπτώσεις:

α. για $\theta = 1$ προκύπτει ένα απλό DPCM.

β. για $\theta = 0$ προκύπτει μια υποπερίπτωση του DPCM η οποία συμβολίζεται *D*PCM* και ονομάζεται *open – loop DPCM*.
γ. και για $0 < \theta < 1$ έχουμε την υποπερίπτωση του DPCM *with Noise Feedback*.

ΚΕΦΑΛΑΙΟ 5

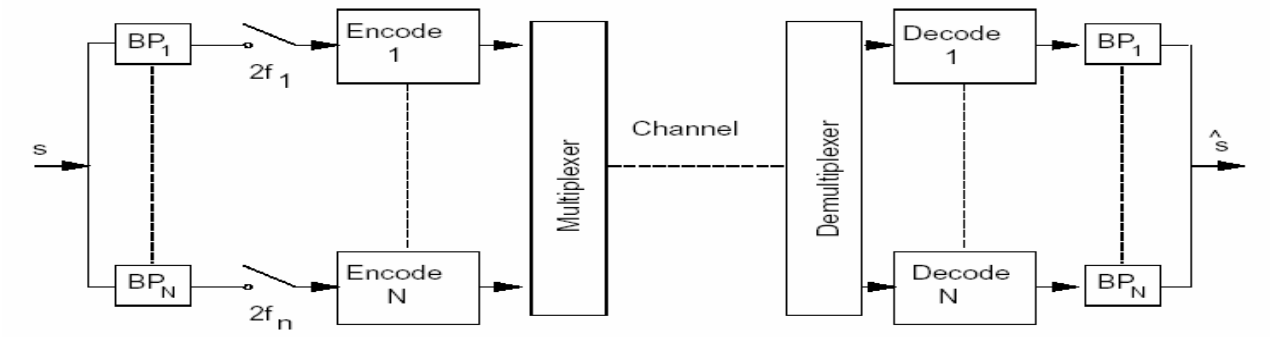
5.1 Εισαγωγή

Ως γνωστό στους κωδικοποιητές κυματομορφής που μελετήσαμε μέχρι τώρα η επεξεργασία του σήματος της ομιλίας γινόταν στο πεδίο του χρόνου. Εκτός όμως από αυτή την τεχνική υπάρχει και η επεξεργασία του σήματος της ομιλίας στο πεδίο των συχνοτήτων ή γενικότερα σε ένα πεδίο μετασχηματισμού. Οι κωδικοποιητές οι οποίοι λειτουργούν με βάση αυτήν την τεχνική ονομάζονται κωδικοποιητές στο *πεδίο των συχνοτήτων*. Σε αυτούς το σήμα της ομιλίας πρώτα μετασχηματίζεται και μετέπειτα γίνεται η επεξεργασία του. Η αποδοτικότητα τους βασίζεται τόσο στις ιδιότητες του σήματος ομιλίας στο πεδίο των συχνοτήτων, όπου παρουσιάζονται πλεονασμοί οι οποίοι είναι ανεξάρτητοι από τους πλεονασμούς στο πεδίο του χρόνου, όσο και στις ιδιότητες του ακουστικού συστήματος του ανθρώπου. Όλα αυτά έχουν ως αποτέλεσμα αποτελεσματικούς κωδικοποιητές από τους οποίους οι κυριότεροι είναι οι *κωδικοποιητές με διαχωρισμό υπο – ζωνών* και οι *προσαρμοστικοί κωδικοποιητές μετασχηματισμού* που αναλύουμε παρακάτω.

5.2 Κωδικοποιητές Με Διαχωρισμό Υπο – Ζωνών (Sub – Band Coders)

5.5.1 Γενική Λειτουργία

Σε αυτή την κλάση των κωδικοποιητών το ψηφιοποιημένο σήμα της φωνής $x(n)$ αρχικά χωρίζεται σε ένα σύνολο από K συχνοτικές συνιστώσες (ζώνες συχνοτήτων) $x_K(n)$ με την χρήση είτε *filterbanks*, είτε φίλτρων διέλευσης ζώνης και η κάθε μια δειγματοληπτείται με την αντίστοιχη συχνότητα Nyquist. Στην συνέχεια οι συνιστώσες αυτές κβαντίζονται και κωδικοποιούνται ξεχωριστά και τελικά πολυπλέκονται και μεταδίδονται από τον πομπό. Στον δέκτη τώρα ακολουθείται η αντίστροφη διαδικασία δηλαδή οι ζώνες συχνοτήτων αποπολυπλέκονται, αποκωδικοποιούνται και συντίθενται έτσι ώστε να μας δώσουν το αρχικό σήμα. Όπως γίνεται αντιληπτό με αυτό τον τρόπο μπορούν να χρησιμοποιηθούν διαφορετικά κριτήρια για τη κωδικοποίηση διαφορετικών ζωνών. Εξάλλου επειδή ο θόρυβος (σφάλμα) κβάντισης της κάθε υπο – ζώνης περιορίζεται σε αυτή, μας δίνεται η δυνατότητα ελέγχου της κατανομής του θορύβου που προκύπτει από την κβάντιση, κατά μήκος όλου του φάσματος του σήματος. Εξαιτίας λοιπόν αυτής της λειτουργίας και βασιζόμενη τόσο στην ιδιότητα της *διαφορετικής αντίληψης για κάθε συχνότητα* της παραμόρφωσης λόγω του θορύβου, όσο και της ιδιότητας της *επικάλυψης* του θορύβου που παρουσιάζει το ακουστικό σύστημα, οδηγούμαστε σε ουσιαστική βελτίωση της ποιότητας της ομιλίας.



Σχήμα 5.1: Τυπική μορφής ενός sub – band κωδικοποιητή.

Βέβαια στους κωδικοποιητές αυτούς πρέπει να πάρουμε υπόψη μας μια σειρά από παραμέτρους για να επιτύχουμε το επιθυμητό αποτέλεσμα. Οι παράμετροι αυτοί είναι ο αριθμός

των υπο – ζωνών (ο οποίος και είναι πρωταρχικής σημασίας γιατί καθορίζει την απόδοση του κωδικοποιητή), ο διαχωρισμός τους καθώς και ο διαμερισμός των διαθέσιμων bits. Συνήθως ο αριθμός των bits που θα χρησιμοποιηθούν προκύπτει με βάση το γεγονός ότι στις ζώνες χαμηλών συχνοτήτων πρέπει να περιγραφούν όσο το δυνατόν καλύτερα η *θεμελιώδης συχνότητα (pitch)* και τα *formants* του σήματος ομιλίας. Αυτό είναι απαραίτητο γιατί όπως έχουμε αναφέρει και στην Παράγραφο 2.3.6 στις χαμηλές συχνότητες περιέχεται σημαντικότερη φωνημική πληροφορία. Βέβαια και στις ζώνες υψηλών συχνοτήτων απαιτείται μεγάλος αριθμός bits εάν θέλουμε να έχουμε καλή ποιότητα ομιλίας. Την λύση συνήθως σε αυτό το πρόβλημα μας την δίνει η χρήση ενός προσαρμοστικού αλγόριθμου όποτε προκύπτει ένας *κωδικοποιητής με διαχωρισμό υπο – ζωνών και προσαρμοστική κατανομή bit*. Ο κωδικοποιητής αυτός μπορεί να προσφέρει υψηλή απόδοση όμως είναι κατά πολύ περισσότερο πολύπλοκος από ένα σύστημα με *μη – προσαρμοστική κατανομή bit* και αυτό είναι μια παράμετρος η οποία πρέπει επίσης να ληφθεί υπόψη.

Επιπρόσθετα μια άλλη παράμετρος που πρέπει να ληφθεί υπόψη είναι και αν οι φασματικές ζώνες που θα χρησιμοποιήσουμε θα έχουν όλες το *ίδιο εύρος* ή θα έχουν *μεταβαλλόμενο εύρος* καθώς και αν θα υπάρχουν ή όχι κενά μεταξύ τους. Επίσης σημαντικό ρόλο στους κωδικοποιητές αυτούς παίζει και τύπος των φίλτρων αν θα είναι *πεπερασμένης κρουστικής απόκρισης (Finite Impulse Response – FIR)* ή *μη πεπερασμένης κρουστικής απόκρισης (Infinite Impulse Response – IIR)* καθώς και αν τα filter banks θα είναι διαρθρωμένα με *δομή δέντρου (tree – structure)* ή με *παράλληλη δομή (parallel – structure)*. Τέλος πολύ σημαντικό ρόλο έχει και η τεχνική η οποία θα ακολουθηθεί στα filter banks η οποία συνήθως είναι η *Quadrature Mirror Filterbank (QMF)* γιατί μας δίνει πολύ μικρή αλλοίωση.

Επειδή λοιπόν βλέπουμε ένα μεγάλο πλήθος παραμέτρων χωρίζουμε την διαδικασία σχεδιασμού ενός τέτοιου κωδικοποιητή σε τρία στάδια. Στο πρώτο στάδιο γίνεται ο σχεδιασμός των filter banks ανάλυσης/σύνθεσης, στο δεύτερο στάδιο καθορίζονται οι κβαντιστές των συχνοτικών συνιστωσών και στο τρίτο στάδιο ορίζεται η διαδικασία της προσαρμοστικής κατανομής των bits.

5.5.2 Αριθμός Υπο – Ζωνών

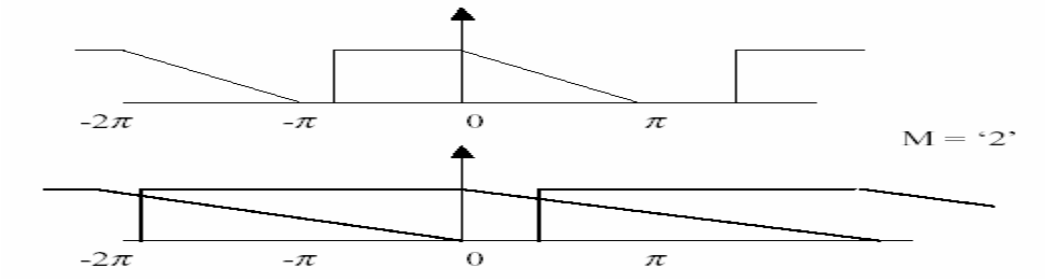
Ο αριθμός των υπο – ζωνών είναι μια παράμετρος από την οποία εξαρτάται άμεσα η βελτίωση του σηματοθορυβικού λόγου. Πιο συγκεκριμένα όσο περισσότερες ζώνες συχνοτήτων έχουμε τόσο μεγαλύτερο κέρδος έχουμε άρα και τόσο μεγαλύτερο σηματοθορυβικό λόγο. Η σχέση που μας δίνει το SNR στην περίπτωση των υπο – ζωνών όταν σε κάθε φασματική συνιστώσα έχουμε κωδικοποίηση PCM είναι η: $SNR_{SBC}(dB) = SNR_{PCM}(dB) + 10\log G_{SBC}$, με G_{SBC} το κέρδος. Παρόλο όμως που με την αύξηση του αριθμού των ζωνών αυξάνεται και το κέρδος, αυτό δεν είναι δυνατόν να γίνει απεριόριστα γιατί υπάρχουν διάφορα προβλήματα υλοποίησης. Το σημαντικότερο από αυτά είναι η επιπρόσθετη καθυστέρηση. Αυτό συμβαίνει γιατί είναι με την χρήση ζωνών με μικρότερο εύρος είναι απαραίτητο τα φίλτρα μας να παρουσιάζουν όσο το δυνατόν πιο απότομη αποκοπή. Έτσι πρέπει αυτά τα φίλτρα να είναι μεγαλύτερης τάξεων (τουλάχιστον στην περίπτωση των *FIR*) γεγονός το οποίο εισάγει και την επιπρόσθετη καθυστέρηση. Υπάρχει λοιπόν μια σχέση αντιστρόφως ανάλογη μεταξύ της καθυστέρησης και του αριθμού των συχνοτικών συνιστωσών και για αυτό τον λόγο συνήθως ο μέγιστος αριθμός υπο – ζωνών που χρησιμοποιείται είναι 16.

5.5.3 Filter Banks (Ομάδες Φίλτρων)

Όπως έχουμε πει και στην εισαγωγή, το σήμα της ομιλίας $x(n)$ διαιρείται σε K συνεχόμενες φασματικές ζώνες με την χρήση ενός *filter bank αναλυτή*, K - καναλιών. Για την διαδικασία αυτή χρησιμοποιείται ένα φίλτρο $h_k(n)$ οπότε και απομονώνεται η k φασματική συνιστώσα του ενδιαφέροντος μας η οποία δίνεται από την εξίσωση $x_k(n) = h_k(n) * x(n)$ ή αντίστοιχα στο πεδίο των συχνοτήτων από την $X_k(\omega) = H_k(\omega)X(\omega)$.Στην συνέχεια ακολουθεί ένας *διαιρέτης (αποδεκατιστής)* συχνότητας ο οποίος έχει ως αποτέλεσμα οι φασματικές συνιστώσες

$x_k(n)$ να υπο – δειγματοληπτούνται με ένα παράγοντα M και έτσι τα σήματα εξόδου $d_k(n)$ να προκύπτουν με βάση μια συχνότητα δειγματοληψίας f_s / M . Η χρήση μιας τέτοιας συχνότητας δειγματοληψίας ισοδυναμεί με πολλαπλασιασμό του εύρους ζώνης του αρχικού σήματος με ένα παράγοντα M . Αυτό έχει ως αποτέλεσμα οι πλευρικές ζώνες του δειγματοληπτούμενου σήματος να παρουσιάζουν μια επικάλυψη και έτσι το αρχικό σήμα να μην μπορεί να ανακτηθεί χωρίς παραμόρφωση (Σχήμα 5.2) .Το φαινόμενο αυτό ονομάζεται *αλλοίωση* ή παραποίηση (aliasing) και μπορεί να μειωθεί σημαντικά με τον περιορισμό του εύρους ζώνης του σήματος εισόδου, ο οποίος πετυχαίνεται με την εφαρμογή του φίλτρου $h_k(n)$ πριν από τον *διαιρέτη*.

Τα σήματα εξόδου που προκύπτουν με βάση αυτή τη συχνότητα δειγματοληψίας $d_k(n)$ δίνονται από την σχέση $d_k(n) = x_k(Mn)$ η οποία στο πεδίο των συχνοτήτων γίνεται $D_k(\omega) = \frac{1}{M}\sum_{i=0}^{M-1} X_k\left(\frac{\omega - 2\pi i}{M}\right)$. Στην συνέχεια αυτά κωδικοποιούνται και μεταδίδονται.

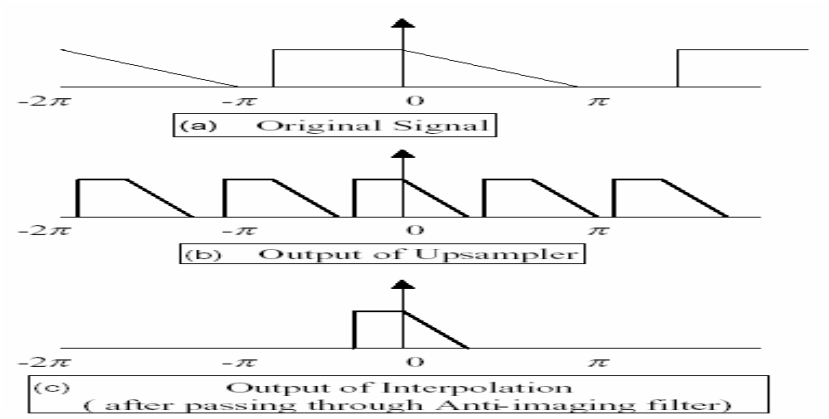


Σχήμα 5.2: Περίπτωση *αλλοίωσης* για M=2.

Αν υποθέσουμε τώρα ότι μετά την μετάδοση, ένα από τα σήματα που φτάνει στο *filterbank σύνθεσης* είναι το $\hat{d}_k(n)$, με την χρήση *πολλαπλασιαστή (παρεμβολέα)* συχνότητας, ο οποίος στην ουσία παρεμβάλει $M-1$ μηδενικά μεταξύ των γειτονικών δειγμάτων παίρνουμε το σήμα εξόδου $u_k(n)$. Σε αυτή την περίπτωση ισχύει ο τύπος

$$u_k(n) = \begin{cases} \hat{d}_k(n/M) & \text{αν } n = \text{πολλαπλάσιο } M \\ 0 & \text{αλλού} \end{cases}$$

Το μειονέκτημα όμως που εμφανίζετε με αυτή τη διαδικασία είναι η εμφάνιση $M-1$ αντιγράφων του βασικού φάσματος του σήματος εισόδου (Σχήμα 5.3b). Το φαινόμενο αυτό ονομάζεται *imaging effect* και η αντίστοιχη εξίσωση για το σήμα εξόδου στο πεδίο των συχνοτήτων είναι $V_k(\omega) = \hat{D}_k(M\omega)$. Στην πράξη κατά την υλοποίηση του συνολικού συστήματος για την αποφυγή τέτοιων φαινομένων μετά τον πολλαπλασιαστή ακολουθεί ένα φίλτρο $g_k(n)$ το οποίο εξαλείφει τα *αντίγραφα* αυτά (Σχήμα 5.3c).



Σχήμα 5.3: Το φαινόμενο του *imaging effect* και ο περιορισμός του με την χρήση κατάλληλου φίλτρου**.**

Το σήμα λοιπόν που προκύπτει μετά και από το φίλτρο αυτό, είναι μια όσο το δυνατόν καλύτερη προσέγγιση του σήματος εισόδου και δίνεται από την σχέση, $\hat{x}_k(n) = g_k(n) * u_k(n)$ ή από την $\hat{X}_k(\omega) = G_k(\omega)V_k(\omega)$. Στο τέλος αυτής της διαδικασίας όλα τα $\hat{x}_k(n)$ αθροίζονται για να μας δώσουν το συνολικό σήμα εξόδου $\hat{x}(n) = \sum_{k=0}^{k-1} \hat{x}_k(n)$.

Για να επιτύχουμε λοιπόν ένα ικανοποιητικό αποτέλεσμα σε όλη αυτή την διαδικασία όπως καταλαβαίνουμε, είναι σημαντικό να περιορίσουμε όσο το δυνατόν την *αλλοίωση* και το *imaging effect*. Η απαίτηση όμως αυτή μας οδηγεί στην χρήση των διαδοχικών φίλτρων $h(n)$ κατά τέτοιο τρόπο ώστε να μας προσφέρουν ικανοποιητικό διαχωρισμό των K φασματικών ζωνών. Αλλά επειδή είναι απαραίτητο να μην ¨αμελούνται¨ κάποιες συχνότητες πρέπει τα φίλτρα αυτά να παρουσιάζουν μια επικάλυψη έτσι ώστε να μην παρατηρούνται κενά μεταξύ τους. Δηλαδή στην ουσία κατά το σχεδιασμό του συστήματος συναντούμε δύο αντιφατικές απαιτήσεις.

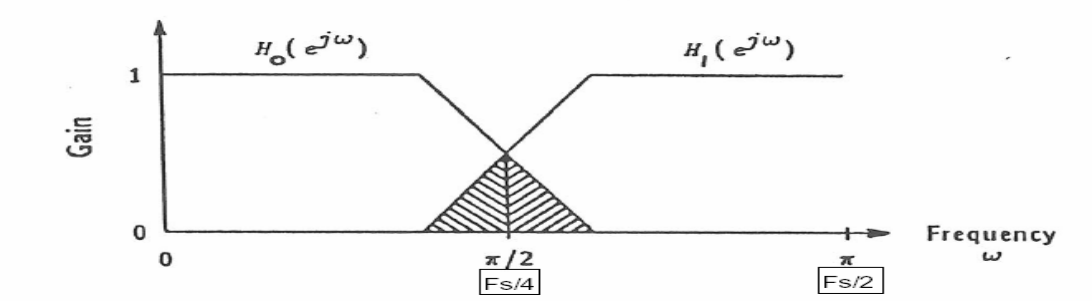
Για την λύση λοιπόν αυτού του προβλήματος έχει αναπτυχθεί μια τεχνική η οποία ονομάζεται Quadrature Mirror Filter (QMF) και η οποία επιτρέπει την ύπαρξη αλλοίωσης κατά την διαδικασία της ανάλυσης η οποία όμως εξουδετερώνεται με την εφαρμογή κατάλληλου φίλτρου κατά την διαδικασία της σύνθεσης. Η προσπάθεια βέβαια αυτή απαιτεί έναν αυστηρό σχεδιασμό τον οποίο και θα δούμε αναλυτικότερα στην επόμενη παράγραφο.

5.3 Quadrature Mirror Filters (QMF)

Η τεχνική του QMF είναι μια πολύπλοκή τεχνική η οποία γνωρίζει πολλές παραλλαγές. Στις παρακάτω παραγράφους θα εξετάσουμε τις πιο βασικές από αυτές.

5.5.1 Δύο Ζωνών Quadrature Mirror Filter Bank

Για να περιγράψουμε την αρχή στην οποία βασίζεται το Quadrature Mirror Filter είναι καλύτερα να δούμε το *δύο ζωνών quadrature mirror filter* το οποίο και αποτελεί την βάση για πιο πολύπλοκα συστήματα. Σε αυτό λοιπόν το σήμα εισόδου $x(n)$ περιορίζεται φασματικά έτσι ώστε η μέγιστη συχνότητα του να είναι $f_s/2$ και δειγματοληπτείται με συχνότητα f_s για να μπορούμε να ανακτήσουμε το αρχικό σήμα. Στην συνέχεια διαιρείται σε δύο φασματικές συνιστώσες με την χρήση δύο φίλτρων, ενός χαμηλοπερατού $h_0(n)$ με μετασχηματισμό – Z $H_0(z)$ και ενός υψηλοπερατού $h_1(n)$ με μετασχηματισμό – Z $H_1(z)$ και εύρος ζώνης $f_s/4$ το κάθε ένα (Σχήμα 5.4).

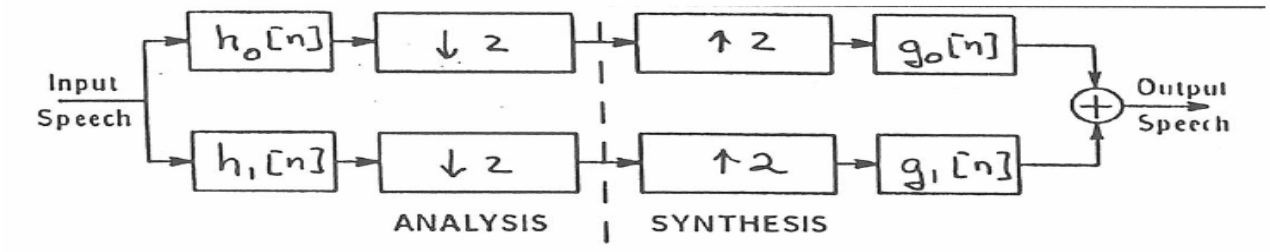


Σχήμα 5.4: Η χρήση των φίλτρων για την δημιουργία δύο φασματικών συνιστωσών.

Επειδή όμως στο σύστημα που υλοποιούμε θέλουμε να διατηρηθεί η συχνότητα δειγματοληψίας του αρχικού σήματος, έτσι η συχνότητα δειγματοληψίας για κάθε μια από τις φασματικές συνιστώσες που έχει δημιουργηθεί είναι υποβιβασμένη κατά το μισό. Κατά αυτό τον τρόπο προκύπτουν τα σήματα εξόδου $Y_0(z)$ και $Y_1(z)$ τα οποία δίνονται μέσα από τις εξισώσεις:

$$Y_0(z)=\frac{1}{2}\left[H_0\left(z^{1/2}\right)X\left(z^{1/2}\right)+H_0\left(-z^{1/2}\right)X\left(-z^{1/2}\right)\right]$$

και $$Y_1(z)=\frac{1}{2}\left[H_1\left(z^{1/2}\right)X\left(z^{1/2}\right)+H_1\left(-z^{1/2}\right)X\left(-z^{1/2}\right)\right]$$ αντίστοιχα (Σχήμα 5.5).

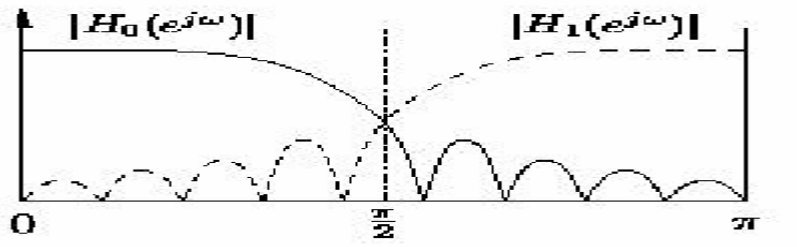


Σχήμα 5.5: Διαδικασία ανάλυσης – σύνθεσης κωδικοποιητή δύο ζωνών.

Στην διαδικασία τώρα της σύνθεσης τα σήματα αυτά συνδυάζονται και έτσι στην έξοδο έχουμε:

$$\hat{X}(z)=\frac{1}{2}[H_0(z)G_0(z)+H_1(z)G_1(z)]X(z)+\frac{1}{2}[H_0(-z)G_0(z)+H_1(-z)G_1(z)]X(-z)$$ όπου ο δεύτερος όρος της είναι η εισερχόμενη αλλοίωση. Αυτή μπορεί να αφαιρεθεί αν κάνουμε την εξής παραδοχή για τα φίλτρα σύνθεσης $G_0(z)=H_1(-z)$, $G_1(z)=-H_0(-z)$. Κατά αυτόν τον τρόπο προκύπτει μηδενική αλλοίωση και η συνάρτηση μεταφοράς $T(z)=\frac{\hat{X}(z)}{X(z)}$ του συστήματος μας γίνεται

$T(z)=\frac{1}{2}H_0(z)H_1(-z)-\frac{1}{2}H_1(z)H_0(-z)=\frac{1}{2}\left[H_0(z)^2-H_1(z)^2\right]=\frac{1}{2}\left[H_0(z)^2-H_0(-z)^2\right]$. Για την επίτευξη τώρα ενός κωδικοποιητή δύο υπο – ζωνών επιλέγουμε να ισχύει για τα δύο φίλτρα $H_1(z)=H_0(-z)$ ή $\left|H_1\left(e^{j\omega}\right)\right|=\left|H_0\left(e^{-(j\omega)}\right)\right|=\left|H_0\left(e^{j(\omega-\pi)}\right)\right|=\left|H_0\left(e^{j(\pi-\omega)}\right)\right|$ δηλαδή αυτά να έχουν μεταξύ τους μια διαφορά φάσης π. Αυτό εισάγει στα πλάτη τους μια συμμετρία γύρω από την συχνότητα $\omega=\pi/2$ (quadrature frequency) όπως φαίνετε και στο Σχήμα 5.6 και για αυτό τον λόγο ονομάζονται και *quadrature mirror filters*.

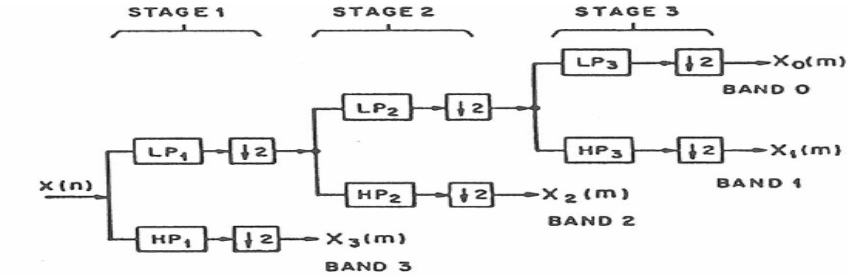


Σχήμα 5.6: Συμμετρία των δύο πλατών γύρω από τα $\pi/2$.

Τα φίλτρα βέβαια στα οποία βασιζόμαστε για την υλοποίηση του κωδικοποιητή στην περίπτωση που είναι ψηφιακά μπορούν να είναι είτε *FIR* είτε *IIR*.Στην περίπτωση της *FIR* υλοποίησης υποθέτουμε ένα *FIR* φίλτρο γραμμική φάσης του οποίου η συνάρτηση μεταφοράς δίνεται μέσα από την σχέση $H_0(z)=\sum_{n=0}^{N-1} h_0(n)z^{-n}$ και το οποίο παρουσιάζει συμμετρική κρουστική απόκριση, δηλαδή $h_0(n)=h_0(N-n)$. Επειδή το κέντρο συμμετρίας είναι ως προς το σημείο $N/2$ η απόκριση συχνότητας του φίλτρου θα δίνεται από την σχέση $H_0(e^{j\omega})=e^{j\omega N/2}\tilde{H}_0(\omega)$ όπου η συνάρτηση πλάτους $\tilde{H}_0(\omega)$ είναι μια πραγματική συνάρτηση του ω. Ως αποτέλεσμα των παραπάνω σχέσεων η απόκριση συχνότητας της συνάρτησης μεταφοράς του συστήματος μας μπορεί να γραφεί ως εξής $\mathrm{T}(e^{j\omega})=\frac{e^{-jN\omega}}{2}\left\{H_0(e^{j\omega})\right|^2-(-1)^N\left|H_0(e^{j(\pi-\omega)})\right|^2\}$. Από την σχέση αυτή παρατηρούμε ότι στην περίπτωση όπου το N είναι άρτιος, τότε έχουμε $T(e^{j\omega})=0$ για $\omega=\pi/2$ γεγονός το οποίο επιφέρει διάφορες αλλοιώσεις πλάτους στην έξοδο. Για τον λόγο αυτό πρέπει το N να είναι περιττό οπότε και έχουμε
$\mathrm{T}(e^{j\omega})=\frac{e^{-jN\omega}}{2}\left\{H_0(e^{j\omega})\right|^2+\left|H_0(e^{j(\pi-\omega)})\right|^2\}=\frac{e^{-jN\omega}}{2}\left\{H_0(e^{j\omega})\right|^2+\left|H_1(e^{j\omega})\right|^2\}$. Από την σχέση αυτή καταλήγουμε στο συμπέρασμα ότι <u>για να έχουμε ακριβή ανασύσταση του σήματος εισόδου πρέπει να ισχύει</u> $\left|H_0(e^{j\omega})\right|^2+\left|H_1(e^{j\omega})\right|^2=1$. Στην περίπτωση αυτή η φασματική αλλοίωση είναι μηδενική όμως η αλλοίωση στο πλάτος εξακολουθεί να υπάρχει. Μπορεί ωστόσο να μηδενιστεί αν το $\left|\mathrm{T}(e^{j\omega})\right|$ είναι σταθερό για όλα τα ω. Κάτι τέτοιο μπορούμε να το επιτύχουμε <u>ρυθμίζοντας από την αρχή τους συντελεστές του φίλτρου</u> $h_0(n)$ <u>ώστε να ισχύει</u> $\left|H_0(e^{j\omega})\right|^2+\left|H_1(e^{j\omega})\right|^2\cong 1$ για κάθε ω. <u>Έτσι πετυχαίνουμε να εξαλείψουμε την αλλοίωση τόσο στο φάσμα όσο και στο πλάτος</u>, κάτι το οποίο είναι και το ζητούμενο από την αρχή.

5.5.2 QMF με Δομή Δέντρου

Στην περίπτωση αυτή η κάθε φασματική ζώνη διαιρείται περαιτέρω με την χρήση *QMF δύο ζωνών* και κατά αυτό τον τρόπο προκύπτουν τέσσερις φασματικές ζώνες με ταυτόχρονη μείωση της συχνότητας δειγματοληψίας για κάθε μια από αυτές σε $f_s/4$.



Σχήμα 5.7: Παράδειγμα ενός κωδικοποιητή υπο – ζωνών με δομή δέντρου.

Η τεχνική αυτή μπορεί να χρησιμοποιηθεί εκτεταμένα έτσι ώστε να προκύψει ο επιθυμητός αριθμός φασματικών ζωνών. Έτσι λοιπόν σε γενικές γραμμές το αρχικό σήμα μπορεί να διαιρεθεί σε 2^p φασματικές συνιστώσες με την κάθε μια από αυτές να δειγματοληπτείται με συχνότητα $f_s/2^p$. Με την αύξηση όμως του αριθμού των ζωνών είναι αναγκαίο τα φίλτρα μας να πληρούν αυστηρότερες προδιαγραφές (πιο απότομη αποκοπή) κάτι το οποίο μεταφράζεται σε φίλτρα μεγαλύτερων τάξεων (τουλάχιστον στην περίπτωση των *FIR*). Το γεγονός αυτό έχει ως αποτέλεσμα σημαντική αύξηση της καθυστέρησης και της πολυπλοκότητας το οποίο αποτελεί ένα βασικό μειονέκτημα. Εκτός όμως από αυτό ένα δεύτερο μειονέκτημα είναι και το γεγονός το εύρος των ζωνών μπορεί να είναι μόνο ίσο με την συχνότητα δειγματοληψίας διαιρεμένη με μια δύναμη του δύο με αποτέλεσμα να μην έχουμε την ευελιξία που απαιτείται σε αρκετές περιπτώσεις. Για την αποφυγή λοιπόν αυτών των προβλημάτων έχουν αναπτυχθεί περισσότερο εξελιγμένες δομές QMF και οι οποίες χρησιμοποιούνται συνήθως σε συστήματα πραγματικού χρόνου μιας και πετυχαίνουν σε σημαντικό βαθμό μείωση της πολυπλοκότητας.

5.4 Κωδικοποίηση στους Sub – Band Κωδικοποιητές

Οι κωδικοποιητές που χρησιμοποιούνται συνήθως στους *sub – band coders* είναι οι APCM και οι ADPCM. Επειδή όμως η χρήση των ADPCM κωδικοποιητών αυξάνει κατά πολύ τον βαθμό πολυπλοκότητας αυτοί χρησιμοποιούνται πολύ αποτελεσματικά σε sub – band coders με μικρό αριθμό φασματικών ζωνών. Από την άλλη πλευρά οι APCM κωδικοποιητές χρησιμοποιούνται σε sub – band coders με μεγάλο αριθμό φασματικών ζωνών. Οι κωδικοποιητές αυτοί κατά κύριο λόγο βασίζονται σε backward estimation για την προσαρμογή του βήματος κβάντισης $\Delta(n)$ το οποίο και δίνεται μέσα από τον τύπο $\Delta(n)=\Delta(n-1)M\{|P(n-1)|\}$, όπου η $M\{|P(n)|\}$ είναι μια χρονικά αμετάβλητη συνάρτηση του πλάτους της n στης κωδικής λέξης. Έτσι ανάλογα αν θέλουμε να αυξήσουμε ή να μειώσουμε το βήμα κβάντισης, η τιμή του $M\{|P(n-1)|\}$ γίνεται μεγαλύτερη ή μικρότερη από το ένα. Επειδή όμως είναι δυνατόν το σήμα εισόδου να έχει αξιοσημείωτες μεταβολές για την αποφυγή εξαιρετικά μεγάλων ή μικρών βημάτων προκαθορίζονται για τον κβαντιστή συγκεκριμένα όρια ως εξής $\Delta_{\min} \le \Delta[n] \le \Delta_{\max}$, ενώ το δυναμικό εύρος του κωδικοποιητή καθορίζεται από τον λόγο $\Delta_{\max}/\Delta_{\min}$. Επίσης είναι δυνατόν να οριστεί μια τιμή του βήματος κβάντισης Δ_{th} σαν διακόπτης με τον οποίο καθορίζεται στον αποκωδικοποιητή αν η κβάντιση θα είναι *mid –rise* ή *mid – tread*. Κατά αυτό τον τρόπο εκμεταλλευόμαστε τα πλεονεκτήματα που πηγάζουν από την χρήση και τον δύο αυτών τεχνικών. Η όλη διαδικασία βέβαια της *προσαρμοστικής κβάντισης με backward estimation* που ακολουθείται εδώ βασίζεται στις γενικές αρχές της Παραγράφου 4.2.2..

5.5 Κατανομή των Bits στους Sub – Band Κωδικοποιητές

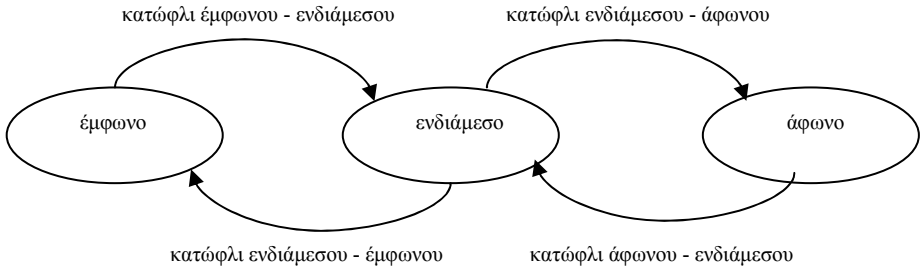
Όπως έχουμε πει και σε προηγούμενη παράγραφο η δυνατότητα κατανομής των bits στις ζώνες συχνοτήτων μας δίνει την ευελιξία να χρησιμοποιήσουμε περισσότερα bit σε εκείνες τις ζώνες οι οποίες είναι πιο σημαντικές από ακουστικής πλευράς πετυχαίνοντας ελαχιστοποίηση του θορύβου. Η κατανομή αυτή βέβαια μπορεί να είναι σταθερή ή καλύτερα προσαρμοστική.

5.5.1 Δυναμική Κατανομή

Στην περίπτωση αυτή τα bits κατανέμονται με βάση μια διαδικασία η οποία στηρίζεται στην τυπική απόκλιση $\sigma_k = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_k(i)\right)^2}$ της κάθε ζώνης συχνοτήτων. Στο τύπο το N δηλώνει τον αριθμό των δειγμάτων της υπο – ζώνης και το $x_k(i)$ το σήμα στη k ζώνη. Για την υλοποίηση λοιπόν της διαδικασίας υπολογίζεται για κάθε υπο – ζώνη η *τυπική απόκλιση* και στην συνέχεια ακολουθείται ένας κανόνας προτεραιότητας. Σύμφωνα με τον κανόνα αυτό η ζώνη με την μεγαλύτερη τυπική απόκλιση παίρνει το πρώτο bit και πριν την κατανομή του επόμενου bit η μεγαλύτερη τυπική απόκλιση μειώνεται κατά ένα παράγοντα γ ως εξής $\sigma_k(i) = \frac{\sigma_k(i)}{\gamma}$. Στην συνέχεια η διαδικασία επαναλαμβάνεται μέχρι να κατανεμηθούν στις φασματικές ζώνες όλα τα διαθέσιμα bits.

5.5.2 Απλή Προσαρμοστική Κατανομή

Η απλή προσαρμοστική κατανομή είναι μια διαδικασία αρκετά πιο πολύπλοκη και πιο αποτελεσματική από την *δυναμική κατανομή*. Βασίζεται στην ταξινόμηση των τμημάτων της ομιλίας σε συνήθως τρεις κατηγορίες: *έμφωνα*, άφωνα και *ενδιάμεσα*[13]. Για την ταξινόμηση αυτή υπάρχουν συγκεκριμένες μέθοδοι όπως είναι εξέταση της ενέργειας του κάθε τμήματος και ο αριθμός τομών με τον άξονα των x (Παράγραφος 3.7.6). Σε κάθε μια από αυτές τις μεθόδους υπάρχουν συγκεκριμένα ¨κατώφλια¨ τα οποία συνδυάζονται μεταξύ τους ώστε να έχουμε επιτυχή αποτελέσματα (Σχήμα 5.8). Αφού ολοκληρωθεί αυτή η διαδικασία ακολουθεί η αντιστοίχηση *προτύπου – ταξινομημένου τμήματος*. Στο κάθε *πρότυπο* έχει αποδοθεί βέβαια ένας συγκεκριμένος αριθμός bit οπότε στην ουσία η αντιστοίχηση είναι *αριθμός bit – ταξινομημένου τμήματος*. Κατά αυτό τον τρόπο λοιπόν καταλήγουμε να έχουμε μια προσαρμοστική κατανομή ανάλογα με το τμήμα της ομιλίας που αντιμετωπίζουμε κάθε φορά.



Σχήμα 5.8: Διάγραμμα κατάστασης της στρατηγικής ταξινόμησης στην απλή προσαρμοστική κατανομή.

Σε αυτή την μέθοδο μια σημαντική παράμετρος που πρέπει να ληφθεί υπόψη είναι προσδιορισμός του αριθμού των bits που έχει αποδοθεί σε κάθε *πρότυπο*. Ο προσδιορισμός αυτός γίνεται με βάση τόσο το τμήμα στο οποίο θα αντιστοιχηθεί το πρότυπο (έμφωνο, άφωνο,

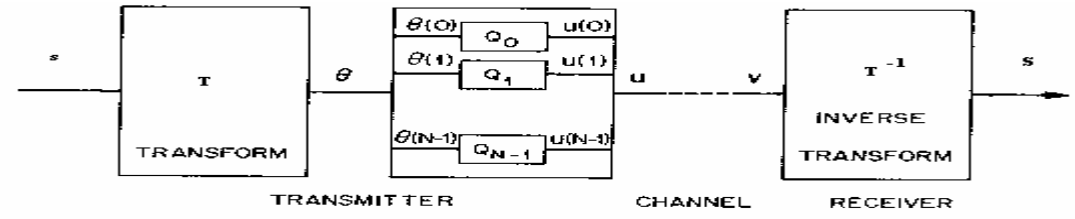
[13] Σαν *ενδιάμεσα* συνήθως ορίζονται εκείνα τα τμήματα ομιλίας τα οποία δεν είναι καθαρά έμφωνα ή άφωνα.

ενδιάμεσο) και είναι συνήθως πειραματικός, όσο και με βάση την ζώνη συχνοτήτων στην οποία συναντάμε το συγκεκριμένο τμήμα. Εδώ ακολουθείτε ο κανόνας που λέει ότι στις ζώνες χαμηλών συχνοτήτων αποδίδονται περισσότερα bits από ότι στις ζώνες υψηλών συχνοτήτων. Η όλη διαδικασία ωστόσο μπορεί να βασίζεται και στην εφαρμογή κάποιον συναρτήσεων οι οποίες κατά κανόνα χρησιμοποιούν τον λόγο θορύβου προς επικάλυψη (noise – to – mask) – NMR. Οι πιο σημαντικές από αυτές τις συναρτήσεις είναι η *average weighted NMR* η οποία και είναι το αποτέλεσμα της άθροισης *επικαλύψεων* σε διάφορες συχνότητες του τμήματος που εξετάζεται. Στις επικαλύψεις αυτές έχουν αποδοθεί βάρη. Ελαχιστοποιώντας αυτό το άθροισμα προκύπτει ένας αριθμός bits, για το εν λόγω τμήμα, ο οποίος είναι ο πλέον αποδοτικός. Μια δεύτερη εξίσου σημαντική συνάρτηση είναι και η *maximal log weighted NMR*. Σε αυτή εφαρμόζεται ο δεκαδικός λογάριθμος αντί για άθροισμα ο οποίος μας δίνει το *log weighted NMR (LWNMR)*. Στη συνέχεια το *maximal log weighted NMR (MLWNMR)* υπολογίζεται από τη σχέση MLWNMR=max(LWNMR). Η ελαχιστοποίηση του προκύπτει αν οδηγήσουμε το *LWNRM* να είναι σταθερό και με βάση αυτή την ελαχιστοποιημένη τιμή μπορούμε να πάρουμε έναν αριθμό bits για το τμήμα του ενδιαφέροντος μας.

Τέλος μια τρίτη συνάρτηση που μπορεί να χρησιμοποιηθεί είναι η *maximal log weighted noise – to – signal ratio*. Η συνάρτηση αυτή είναι μια απλοποιημένη μορφή της MLWNMR και προκύπτει αν αφαιρεθούν οι παραμέτρους εκείνες της MLWNMR που αναφέρονται στην επικάλυψη. Σε αυτή όπως και στις δύο προηγούμενες ακολουθείτε η ίδια διαδικασία ελαχιστοποίησης με βάση την οποία απορρέει και ο ζητούμενος αριθμός bits. Εδώ πρέπει να πούμε σαν γενική παρατήρηση της όλη διαδικασίας ότι, ο αριθμός των bits που παίρνουμε κάθε φορά, στρογγυλοποιείτε σε ακέραια μορφή έτσι ώστε να είναι δυνατή η υλοποίηση του στην πράξη.

5.6 Κωδικοποιητές Μετασχηματισμού

Σε αυτούς τους κωδικοποιητές πραγματοποιείτε στην είσοδο μετασχηματισμός του σήματος της ομιλίας από το πεδίο του χρόνου σε ένα αφηρημένο πεδίο. Αυτό έχει ως αποτέλεσμα να αναπαρίσταται από ένα σύνολο από μετασχηματιστικούς παράγοντες. Κάθε ένας από αυτούς τους παράγοντες τείνει να είναι ασυσχέτιστος με τον γειτονικό του και για αυτό και κβαντίζονται και κωδικοποιούνται ανεξάρτητα ο ένας από τον άλλον. Η κωδικοποίηση γίνεται κατά τέτοιο τρόπο ώστε να αποδίδονται στους πιο σημαντικούς από αυτούς μεγαλύτερος αριθμός bits για να είναι όσο το δυνατόν πιο ακριβή η περιγραφή τους. Στο δέκτη τώρα ακολουθείτε ο αντίστροφος μετασχηματισμός ώστε να αναπαραχθεί το αρχικό σήμα (Σχήμα 5.9).



Σχήμα 5.9: Μπλοκ διάγραμμα ενός κωδικοποιητή μετασχηματισμού.

Για την επεξεργασία του, το σήμα χωρίζεται σε πλαίσια (διανύσματα) στο κάθε ένα από τα οποία εφαρμόζεται ένας διακριτός *unitary* μετασχηματισμός. Η όλη διαδικασία μπορεί να περιγραφεί από την παρακάτω εξίσωση με την μορφή πινάκων ως εξής

$$\begin{bmatrix} S(0) \\ S(1) \\ S(2) \\ \text{M} \\ S(0) \end{bmatrix} = \begin{bmatrix} t_{1,1} & t_{1,2} & t_{1,3} & \Lambda & t_{1,N} \\ t_{2,1} & t_{2,2} & t_{2,3} & \Lambda & t_{2,N} \\ t_{3,1} & t_{3,2} & t_{3,3} & \Lambda & t_{3,N} \\ \text{M} & \text{M} & \text{M} & \text{O} & \text{M} \\ t_{N,1} & t_{N,2} & t_{N,3} & \Lambda & t_{N,N} \end{bmatrix} \begin{bmatrix} s(0) \\ s(1) \\ s(2) \\ \text{M} \\ s(0) \end{bmatrix}$$ ή ως εξής $S = Ts$, ενώ ο

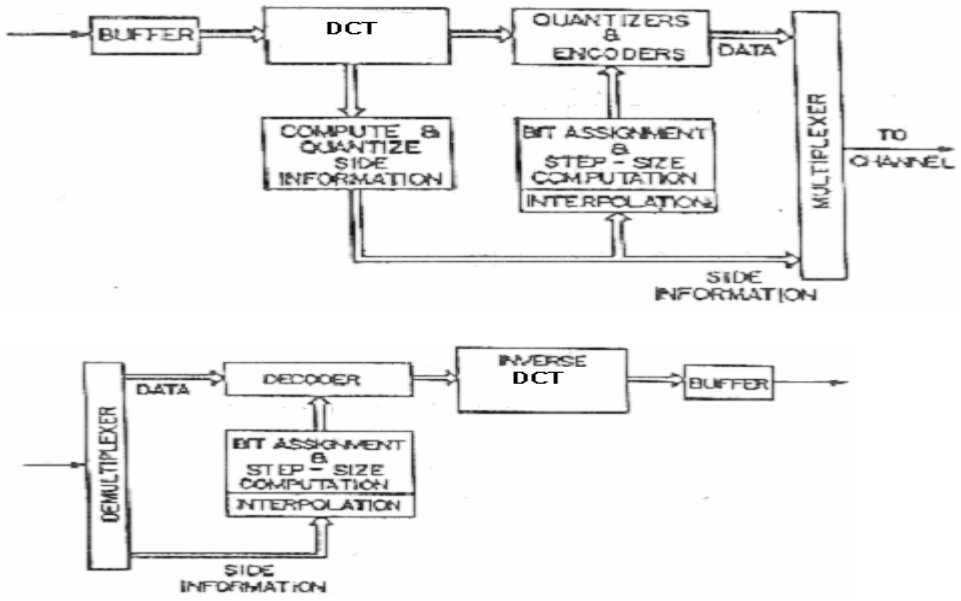
αντίστροφος μετασχηματισμός δίνεται από την σχέση $s = T^{-1}S$. Τα διανύσματα στήλης του πίνακα T^{-1} ονομάζονται *βασικά διανύσματα* και με τον γραμμικό συνδυασμό αυτών γίνεται στην ουσία η ανασύσταση του αρχικού σήματος.

Στους μετασχηματισμούς αυτούς συνήθως χρησιμοποιούνται *ορθογώνιοι* πίνακες γεγονός που εισάγει ότι τα διανύσματα γραμμής του πίνακα T είναι τα *βασικά διανύσματα*. Η χρήση της *ορθογωνιότητας* έχει ως αποτέλεσμα την ανάλυση του σήματος εισόδου σε μη συσχετισμένους παράγοντες και είναι πολύ σημαντική. Υπάρχουν πολλοί μετασχηματισμοί που χρησιμοποιούνται βασιζόμενοι σε αυτή την ιδιότητα και οι πιο σημαντικοί από αυτούς είναι ο Διακριτός Μετασχηματισμός Συνημιτόνου (Discrete Cosine Transform – DCT), ο Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Tansform – DFT) και ο Karhunen Loeve Μετασχηματισμός (KLT).

Από τους πιο πάνω μετασχηματισμούς ο KLT προσφέρει την καλύτερη ανασύσταση του αρχικού σήματος και την μικρότερη αλλοίωση (μικρό MSE) όμως επειδή εξαρτάται από τα δεδομένα εισόδου συνήθως απαιτείται για την υλοποίηση του μεγάλη υπολογιστική ικανότητα. Για τον λόγο αυτό προτιμάται ο DCT μετασχηματισμός ο οποίος έχει περίπου την ίδια απόδοση και είναι ανεξάρτητος από τα δεδομένα εισόδου. Τα *βασικά διανύσματα* του DCT μετασχηματισμού όπως και του DFT έχουν ημιτονοειδή μορφή και μπορούν να υπολογιστούν αποδοτικά με την χρήση του Γρήγορου Μετασχηματισμού Fourier (Fast Fourier Transform – FFT). Βλέπουμε λοιπόν ο DCT παρουσιάζει ένα πλήθος από πλεονεκτήματα συγκρινόμενος με τους άλλους μετασχηματισμούς και για τον λόγο αυτό είναι ο πλέον χρησιμοποιούμενος. Μια από τις πιο διαδεδομένες εφαρμογές που τον συναντούμε είναι ο ATC (Adaptive Transform Coder) τον οποίο θα δούμε αναλυτικότερα παρακάτω ενώ τέλος οι εξισώσεις οι οποίες ορίζουν τον DCT είναι για τον μετασχηματισμό της μορφής $S(k) = \sum_{n=0}^{N-1} s(n)\lambda(k)\cos\left[(2n+1)\pi k / 2N\right]$ με $k = 0,1,2,...,N-1$ και $\lambda(0) = 1, \lambda(k) = \sqrt{2}$ και για τον αντίστροφο μετασχηματισμό της μορφής $s(n) = \frac{1}{N}\sum_{n=0}^{N-1} S(k)\lambda(k)\cos\left[(2n+1)\pi k / 2N\right]$.

5.7 Προσαρμοστικός Κωδικοποιητής Μετασχηματισμού (Adaptive Transform Coder)

Στους κωδικοποιητές αυτούς (Σχήμα 5.10) ακολουθείται η γενική διαδικασία των κωδικοποιητών μετασχηματισμού. Έτσι το σήμα εισόδου διαιρείται σε πλαίσια συγκεκριμένου μήκους (αριθμού δειγμάτων N) τα οποία κανονικοποιούνται διαιρώντας με την συνολική ενέργεια του μπλοκ. Ακολουθεί για κάθε μπλοκ ο DCT μετασχηματισμός με αποτέλεσμα να προκύπτει ένα πλήθος από μετασχηματιστικούς συντελεστές. Στην συνέχεια εφαρμόζεται μια διαδικασία κβάντισης και μια προσαρμοστική κατανομή των bits η οποία βασίζεται σε μια αποτίμηση της ενέργειας του φάσματος του κάθε πλαισίου στο λογαριθμικό πεδίο. Κατά αυτόν τον τρόπο ο αριθμός των bits που αποδίδεται σε κάθε μετασχηματιστικό παράγοντα είναι ανάλογος με την τιμή της ενέργειας. Η προσαρμοστική αυτή κατανομή μεταδίδεται σαν πλευρική πληροφορία και περιλαμβάνει επίσης και τις παραμέτρους της κβάντισης. Στον δέκτη τώρα η πλευρική πληροφορία η οποία συνήθως αποτελείται από L ¨φασματικά σημεία¨ με εύρος του L περίπου 15 – 20 χρησιμοποιείται στους κβαντιστές καθώς επίσης και στην αναδόμηση του αρχικού πλαισίου μήκους N από το L σημείων φάσμα με μια γεωμετρική παρεμβολή.



Σχήμα 5.10: Μπλοκ διάγραμμα ενός προσαρμοστικού α)κωδικοποιητή β)αποκωδικοποιητή μετασχηματισμού.

Οι προσαρμοστικοί κωδικοποιητές μετασχηματισμού πάντως αυτής της μορφής, που έχουν αναπτυχθεί προσφέρουν σαφώς καλύτερες αποδόσεις στα 12 – 32 kbits/s συγκρινόμενοι με το log – PCM κατά 17 έως 23 dB ενώ συγκρινόμενοι με τους ADPCM κωδικοποιητές στα 16 kbits/s προσφέρουν μια βελτίωση κατά 6 dB περίπου. Ωστόσο πέρα από την βελτίωση του σηματοθορυβικού λόγου που εισάγουν το μειονέκτημα τους είναι ότι εμφανίζουν μεγάλη πολυπλοκότητα στην υλοποίηση τους η οποία είναι περίπου δεκαπλάσια της πολυπλοκότητας που παρουσιάζει ένας SBC κωδικοποιητής.

Τέλος να πούμε πως όπως σε κάθε κωδικοποιητή έτσι και εδώ υπάρχει ένα πλήθος παραλλαγών. Μια από αυτές τις παραλλαγές είναι ο ¨vocoder – driven¨ ATC κωδικοποιητής ο οποίος βασίζει την προσαρμοστική στρατηγική του στις κορυφές και στα *formant* της ομιλίας και χρησιμοποιεί ένα μοντέλο γραμμικής πρόγνωσης για την πλευρική πληροφορία. Μια δεύτερη παραλλαγή είναι ο ¨speech – specific¨ κωδικοποιητής ο οποίος και βασίζεται σε ένα ομομορφικό μοντέλο (Παράγραφος 3.6) για την πλευρική πληροφορία. Βέβαια πέρα από την χρήση DCT μετασχηματισμού σε αυτούς του κωδικοποιητές είναι δυνατόν και χρήση άλλων μετασχηματισμών όπως είναι ο Walsh Hadamard μετασχηματισμός καθώς και η χρήση συνδυασμού μετασχηματισμών όπως χρήση WHT και Fourier μετασχηματισμού με αποτέλεσμα να προκύπτει ένα μεγάλο πλήθος παραλλαγών.

5.8 Σύγκριση των Κωδικοποιητών

5.8.1 Σύγκριση ως προς Ποιότητα Ομιλίας/Ρυθμό Μετάδοσης

Όπως έχουμε πει και σε προηγούμενη παράγραφο η πιο διαδεδομένη (υποκειμενική) μέθοδος αξιολόγησης της ποιότητας των κωδικοποιητών είναι η κλίμακα *MOS*. Έτσι η διάκριση τους ανάλογα με την κατηγορία ποιότητας ομιλίας και την κλίμακα MOS έχει ως εξής:

Network ή toll quality: Πετυχαίνεται όταν ισχύει MOS>4 και η ποιότητα της ομιλίας εδώ είναι συγκρίσιμη με την κλασική αναλογική ομιλία.

Communication quality: Πετυχαίνεται όταν ισχύει 3.5<MOS<4 όμως εδώ εισάγεται μια αλλοίωση στην ποιότητα της ομιλίας η οποία όμως είναι αποδεκτή σε τηλεφωνικές εφαρμογές.

Synthetic quality: Πετυχαίνεται όταν ισχύει 2.5<MOS<3.5 και η ομιλία που παίρνουμε είναι κατανοητή αλλά με ¨αφύσικη¨ χροιά. Συνδέεται επίσης με την μη αναγνωρισιμότητα του ομιλητή. Δεν είναι αποδεκτή σε τηλεφωνικές εφαρμογές.

Βέβαια όπως είναι γνωστό η διάκριση των κωδικοποιητών και η απόδοσης τους καθορίζεται από τον ρυθμό μετάδοσης τους. Για την σύγκριση μεταξύ τους χρησιμοποιείται σαν μέτρο αναφοράς ο 56 kb/s *log – PCM* κωδικοποιητής. Τα αποτελέσματα φαίνονται στον Πίνακα 5.1.

Κωδικοποιητής	**Ρυθμός Μετάδοσης για *Network ή toll quality***	**Ρυθμός Μετάδοσης για *Communication quality***
log – PCM	64	36
CVSD	40	20
ADPCM	32	16
SBC	24	14
Adaptive – SBC	16	9.6
APC, ATC	16	8

Πίνακας 5.1: Απαιτούμενος ρυθμός μετάδοσης διαφόρων κωδικοποιητών κυματομορφής για *toll και communication quality*.

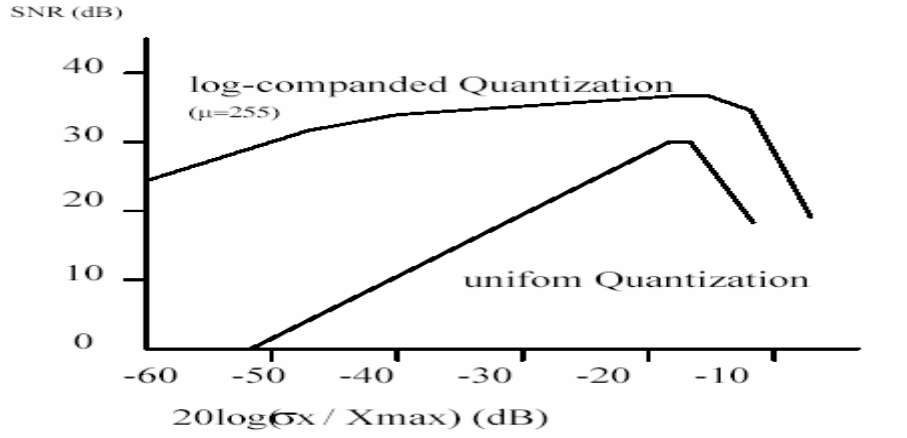
Πέρα όμως από τον Πίνακα 5.1 μπορούμε να χρησιμοποιήσουμε για την σύγκριση των κωδικοποιητών και τον Πίνακα 5.2 ο οποίος μας δείχνει την βαθμολογία που πετυχαίνουν μερικοί από αυτούς με βάση την κλίμακα MOS.

Κωδικοποιητής	**Ρυθμός Μετάδοσης**	**MOS**
PCM	64	4.3
log – PCM	64	4.3
ADPCM	32	4.1
SBC	48/56/64	4.1

Πίνακας 5.2: Βαθμολόγηση ορισμένων κωδικοποιητών κυματομορφής σύμφωνα με την κλίμακα MOS.

Σαν αποτέλεσμα των συγκρίσεων στους δύο παραπάνω πίνακες μπορούμε να πούμε ότι τόσο το log – PCM, όσο και το απλό PCM πετυχαίνουν υψηλή ποιότητα ομιλίας (MOS>4) με αντίστοιχα όμως υψηλά *bit rate* (64 kbps). Τυχόν μείωση του *bit rate* με το οποίο λειτουργούν έχει ως αποτέλεσμα την δραματική μείωση της απόδοσης τους (MOS<4). Αυτό οδηγεί σε αυξημένες απαιτήσεις εύρους ζώνης για την μετάδοση μιας *toll quality* ομιλίας κάτι το οποίο δεν είναι αποδεκτό. Μια λύση στο ¨πρόβλημα¨ αυτό μπορούμε να πούμε ότι είναι ο ADPCM των 32 kbps καθώς και ο SBC των 24 kbps. Και οι δύο προσφέρουν *toll quality* ομιλίας σε χαμηλότερα *bit rate* ενώ ακόμα καλύτερα αποτελέσματα πετυχαίνονται με την χρήση Adaptive – SBC, ATC και APC κωδικοποιητών οι οποίοι μας δίνουν τα ίδια αποτελέσματα στα 16 kbps!!!.

Πέρα όμως από την χρήση υποκειμενικών μετρήσεων υπάρχουν και οι αντικειμενικές μετρήσεις οι οποίες βασίζονται στο σηματοθορυβικό λόγο. Βασιζόμενοι λοιπόν αυτές βλέπουμε για ένα σήμα ομιλίας με $S_{\max} = 8\sigma_x$,δειγματοληπτημένο στην συχνότητα των 8 kHz και με αριθμό bits 2 ότι το απλό PCM δίνει περίπου 4 dB, το PCM με προγνώστη κβάντισης περίπου 8 dB, το DPCM περίπου 13 dB και το ADPCM περίπου 13,5 dB. Επίσης για ένα σήμα με $\sigma_x > 0.1S_{\max}$ το απλό γραμμικό PCM των 11 bits δίνει λίγο καλύτερο λόγο από το αντίστοιχο *νόμου – μ* , μ=500 και 7 bits!!!. Γίνεται λοιπόν φανερό (Σχήμα 5.11) ότι η λογαριθμική κβάντιση προσφέρει (υπό προϋποθέσεις) καλύτερα αποτελέσματα από την ομοιόμορφη κβάντιση αλλά και ότι οι προσαρμοστικοί κωδικοποιητές είναι αυτοί οι οποίοι υπερτερούν και στις αντικειμενικές μετρήσεις.



Σχήμα 5.11: Σύγκριση ομοιόμορφου και νόμου – μ (μ=255), 8 - bits κωδικοποιητή.

5.8.2 Σύγκριση ως προς την Πολυπλοκότητα

Μια ακόμα όμως σημαντική παράμετρος με την οποία μπορούν να συγκριθούν οι κωδικοποιητές είναι με βάση την πολυπλοκότητα που παρουσιάζουν. Μάλιστα όπως είναι γνωστό όταν θέλουμε να μειώσουμε τον ρυθμό μετάδοσης διατηρώντας την ίδια ποιότητα ομιλίας είμαστε αναγκασμένοι να αυξήσουμε την πολυπλοκότητα του κωδικοποιητή μας. Κάτι τέτοιο όμως έχει σαν συνέπεια την ταυτόχρονη αύξηση του κόστους υλοποίησης αλλά και της κατανάλωσης ενέργειας. Η μέτρηση της πολυπλοκότητας τώρα γίνεται με βάση τον αριθμό των πολλαπλασιασμών και των προσθέσεων που απαιτούνται για την υλοποίηση σε επεξεργαστές ψηφιακού σήματος (DSP's). Εκφράζεται σε αριθμό υπολογισμών ανά δευτερόλεπτο (MIP's).

Πιο συγκεκριμένα τώρα για τους κωδικοποιητές κυματομορφής μπορούμε να πούμε σε γενικές γραμμές ότι είναι κωδικοποιητές χαμηλής πολυπλοκότητας και ότι μπορούν υλοποιηθούν σε μικρής υπολογιστής ισχύος επεξεργαστές. Συγκρίνοντας τώρα αυτούς μεταξύ τους παίρνουμε τα αποτελέσματα του Πίνακα 5.3.

Κωδικοποιητής	**Ρυθμός Μετάδοσης**	**MIPS**
PCM	64	0.01
ADPCM	32	2
SBC	48/56/64	5

Πίνακας 5.3: Πολυπλοκότητα ορισμένων κωδικοποιητών κυματομορφής.

Με βάση λοιπόν τον πιο πάνω Πίνακα βλέπουμε ότι ο PCM των 64 kbps είναι εξαιρετικά μικρής πολυπλοκότητας ενώ ο ADPCM και ο SBC είναι μεσαίας πολυπλοκότητας. Μεγάλης πολυπλοκότητας μπορούν να θεωρηθούν ο Adaptive – SBC και οι APC και ATC με MIPS γύρω στα 13 (δεν φαίνονται στον Πίνακα 5.3). Έτσι με βάση και την παραπάνω παράγραφο βλέπουμε ότι αν και αν και ο Adaptive – SBC και οι APC και ATC δίνουν εξαιρετική ποιότητα ομιλίας ακόμα και για χαμηλά bit rate αντισταθμίζουν αυτήν την ¨επιτυχία¨ τους με την αυξημένη πολυπλοκότητα που παρουσιάζουν. Από την άλλη πλευρά το PCM και το log – PCM που δίνουν υψηλής ποιότητας ομιλία με μεγάλα bit rate δεν παρουσιάζουν αυξημένη πολυπλοκότητα. Στην μέση αυτών των δύο κατηγοριών μπορούμε να πούμε ότι είναι οι ADPCM και οι SBC κωδικοποιητές οι οποίοι δίνουν υψηλής ποιότητας ομιλία διατηρώντας χαμηλά τόσο το bit rate όσο και την πολυπλοκότητα τους!!!, χωρίς όμως αυτό να τους κάνει ¨καθολική επιλογή¨. Κάτι τέτοιο όπως καταλαβαίνουμε δεν μπορεί να γίνει αλλά η επιλογή του κωδικοποιητή διαφοροποιείται κάθε φορά ανάλογα με ιδιαιτερότητες του κάθε συστήματος.

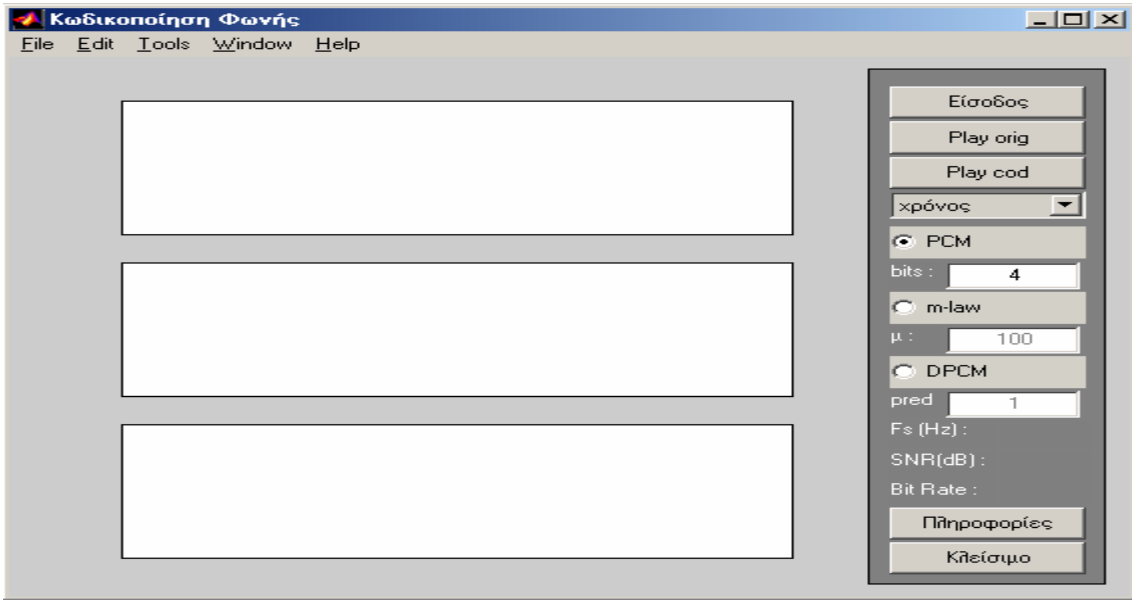
ΚΕΦΑΛΑΙΟ 6

6.1 Εισαγωγή

Το κεφάλαιο αυτό αναφέρεται στην υλοποίηση κωδικοποιητών φωνής μέσα από το περιβάλλον του Matlab. Η εν λόγω υλοποίηση γίνεται με την μορφή κώδικα και την χρήση m files και μας προσφέρει την δυνατότητα μεταβολής των κύριων παραμέτρων των κωδικοποιητών αυτών καθώς και δυνατότητα παρακολούθησης των βασικότερων αποτελεσμάτων όπως είναι ο *σηματοθορυβικός λόγος (SNR)* και το *bit rate* του κωδικοποιητή. Επιπρόσθετα δίνεται η δυνατότητα τόσο να δούμε όσο και να ακούσουμε τα αποτελέσματα της κωδικοποίησης με απλό και εύχρηστο τρόπο.

6.2 Δομή Υλοποίησης

Οι κωδικοποιήσεις οι οποίες επιλέχθηκαν να υλοποιηθούν είναι η ομοιόμορφη παλμοκωδική κωδικοποίηση (PCM), η λογαριθμική κωδικοποίηση νόμου – μ και οι διαφορική παλμοκωδική κωδικοποίηση (DPCM). Για κάθε μια από αυτές αναπτύχθηκε μια συνάρτηση με ονόματα *pcm, mlaw* και *dpcm* αντίστοιχα. Οι συναρτήσεις αυτές καλούνται μέσα από την κύρια συνάρτηση *guip1* η οποία και αντιστοιχεί στο περιβάλλον εργασίας. Έτσι λοιπόν πληκτρολογώντας *guip1* στην prompt γραμμή του Matlab εμφανίζεται η εικόνα του Σχήματος 6.1 λεπτομέρειες χρήσης τις οποίας θα αναφερθούν σε παρακάτω παραγράφους.



Σχήμα 6.1: Γραφικό περιβάλλον των τριών κωδικοποιητών.

6.3 Λεπτομέρειες Υλοποίησης

Όπως είπαμε και πιο πάνω το γραφικό περιβάλλον του Σχήματος 6.1 συνδυάζεται με τρεις συναρτήσεις οι οποίες και αντιστοιχούν στους τρεις κωδικοποιητές.

6.3.1 Λεπτομέρειες Υλοποίησης PCM Κωδικοποιητή

Για τον PCM κωδικοποιητή έχουμε την συνάρτηση *pcm* της οποίας ο κώδικας φαίνεται παρακάτω:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Written by Goumenidis Theodoros  %
% December 2003                     %
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [coded,error,snr]=pcm(original,arithmos_bits_kbantisti)

[sima,fs,number_bits]=wavread(original);

%%%%%%%%%%%%%%%
% Parametroi %
%%%%%%%%%%%%%%%
N=2^arithmos_bits_kbantisti;

%%%%%%%%%%%%%%%%%
% Normalization %
%%%%%%%%%%%%%%%%%
max_sign= max(abs(sima));
norm_sign=sima/max_sign;
max_norm=max(norm_sign);
bima=(2*max_norm)/N;

%%%%%%%%%%%%
% Quantize %
%%%%%%%%%%%%
for z=1:length(norm_sign);
   for i=0:(N-1)
      if norm_sign(z)>=-1+i*bima & norm_sign(z)<=-1+(i+1)*bima
         quant_sign(z)=((-1+(i*bima))+(-1+(i+1)*bima))/2;
      end
   end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
coded=quant_sign;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%
% SNR %
%%%%%%%
quant_sign=quant_sign';
error=norm_sign-quant_sign;
errorvar=var(error);
signvar=var(norm_sign);
VSNR=signvar/errorvar;
VSNRdb=10*log10(VSNR);
snr=VSNRdb;
```

Αρχικά ορίζεται η *pcm* με την δήλωση:

```
function [coded,error,snr]=pcm(original,arithmos_bits_kbantisti)
```

της οποίας ορίσματα εισόδου αποτελούν το *σήμα εισόδου* – `original` και ο *αριθμός bits του κβαντιστή* - `arithmos_bits_kbantisti`. Σαν έξοδο μας δίνει το *κωδικοποιημένο σήμα* – `coded`, *το σφάλμα* – `error` και τον *σηματοθορυβικό λόγο σε dB* – `snr`. Πρέπει να πούμε ότι το *σήμα εισόδου* είναι στην ουσία η διεύθυνση του σήματος την οποία και παίρνει στην αμέσως επόμενη γραμμή σαν όρισμα η – *wavread* και μας επιστρέφει το σήμα με μορφή (πίνακα) κατάλληλη για επεξεργασία – `sima`. Επίσης η *wavread*, μας δίνει και την *συχνότητα* με την οποία έχει γίνει η δειγματοληψία του σήματος – `fs`, καθώς και τον *αριθμό των bits/δείγμα* – `number_bits`. Στην συνέχεια ακολουθεί ο υπολογισμός των σταθμών κβάντισης – `N` και η κανονικοποίηση του σήματος. Η κανονικοποίηση επιταχύνει την διαδικασία και δεν επηρεάζει το αποτέλεσμα αφού ο λόγος *μέγιστη τιμή*

σήματος/τυπική απόκλιση σήματος από τον οποίο εξαρτάται ουσιαστικά η τιμή του SNR παραμένει αναλλοίωτος και ίσος με τον λόγο *μέγιστη τιμή κανονικοποιημένου σήματος/τυπική απόκλιση κανονικοποιημένου σήματος*. Πριν από την κβάντιση τώρα υπολογίζεται και το βήμα κβάντισης – `bima`, με βάση τον τύπο που ισχύει στον *mid – zero* κβαντιστή και μετά ακολουθεί η υλοποίηση του κβαντιστή μέσα από το διπλό loop:

```
for z=1:length(norm_sign);
   for i=0:(N-1)
      if norm_sign(z)>=-1+i*bima & norm_sign(z)<=-1+(i+1)*bima
         quant_sign(z)=((-1+(i*bima))+(-1+(i+1)*bima))/2;
      end
   end
end
```

Εδώ, το πρώτο `for` εξασφαλίζει ότι θα επεξεργαστούμε όλο το σήμα. Το δεύτερο `for` σε συνδυασμό με την παράμετρο `(i+1)` της παρακάτω γραμμής, και το `if` εξασφαλίζει ότι θα ελεγχθεί διαδοχικά για όλες τις στάθμες η κάθε τιμή του κανονικοποιημένου σήματος ανάμεσα σε ποίες τιμές βρίσκεται. Με βάση αυτό και τον τύπο $m_i = \frac{x_{i-1} + x_i}{2}$ προκύπτει η κβαντισμένη τιμή. Πρέπει να πούμε βέβαια ότι η κβάντιση γίνεται μέσα στο διάστημα [-1,1] αφού αυτά είναι και τα όρια του *κανονικοποιημένου* σήματος.

Στις τελευταίες γραμμές του κώδικα τώρα, υπολογίζεται τόσο το σφάλμα το οποίο και είναι η διαφορά μεταξύ κανονικοποιημένου και κβαντισμένου σήματος `error=norm_sign-quant_sign`, όσο και το SNR σύμφωνα με τον τύπο $SNR = \frac{E[x^2(n)]}{E[e^2(n)]} = \frac{\sigma_x^2}{\sigma_e^2}$. Σαν σήμα εισόδου x παίρνεται το κανονικοποιημένο σήμα.

6.3.2 Λεπτομέρειες Υλοποίησης Κωδικοποιητή Νόμου – μ

Για τον κωδικοποιητή *νόμου – μ* έχουμε την συνάρτηση *mlaw* της οποίας ο κώδικας φαίνεται παρακάτω:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Written by Goumenidis Theodoros  %
% December 2003                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [coded,error,snr]=mlaw(original,arithmos_bits_kbantisti,mu)

[sima,fs,number_bits]=wavread(original);
%%%%%%%%%%%%%%%
% Parametroi %
%%%%%%%%%%%%%%%
N=2^arithmos_bits_kbantisti;

%%%%%%%%%%%%%%%%%%
% Normalization %
%%%%%%%%%%%%%%%%%%
max_sign= max(abs(sima));
norm_sign=sima/max_sign;
max_norm=max(abs(norm_sign));
bima=(2*max_norm)/N;

%%%%%%%%%%%%%%%%%%%%%%%%%%
% Compress using mu - law %
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%
denom = log(1 + mu);
prosimo_sign=sign(norm_sign);
comp_sign_temp = (log(1 + mu.*abs((norm_sign)/max_norm))./denom;
comp_sign =max_norm*comp_sign_temp.*norm_sign;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Normalize The Compressed Signal %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
max_comp_sign=max(abs(comp_sign));
norm_comp_sign=comp_sign/max_comp_sign;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Pass Through Uniform Quantizer %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for z=1:length(norm_comp_sign);
   for i=0:(N-1)
      if norm_comp_sign(z)>=-1+i*bima & norm_comp_sign(z)<-1+(i+1)*bima
         quant_sign(z)=((-1+(i*bima))+(-1+(i+1)*bima))/2;
      end
   end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%
coded=quant_sign;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%
% SNR %
%%%%%%%
quant_sign=quant_sign';
error=norm_sign-quant_sign;
errorvar=var(error);
signvar=var(norm_sign);
VSNR=signvar/errorvar;
VSNRdb=10*log10(VSNR);
snr=VSNRdb;
```

Και εδώ αρχικά ορίζεται η *mlaw* η οποία εκτός από τα δύο προηγούμενα ορίσματα εισόδου που έχει η *pcm* έχει και ένα τρίτο την τιμή της παραμέτρου μ – `mu`. Επίσης και εδώ το σήμα παίρνεται σε μορφή κατάλληλη για επεξεργασία με την wavread. Ακολουθεί ο υπολογισμός των σταθμών κβάντισης, η κανονικοποίηση του σήματος και ο υπολογισμός του βήματος κβάντισης. Στην συνέχεια εφαρμόζεται η συνάρτηση συμπίεσης του νόμου – μ και στο συμπιεσμένο σήμα που προκύπτει γίνεται εκ νέου κανονικοποίηση. Το νέο αυτό κανονικοποιημένο σήμα - `norm_comp_sign` περνά μέσα από έναν ομοιόμορφο κβαντιστή ο οποίος λειτουργεί ακριβώς με την ίδια λογική με την οποία λειτουργεί και ο κβαντιστής στην συνάρτηση *pcm*. Μετά τον κβαντιστή υπολογίζεται όπως και παραπάνω το σφάλμα - `error=norm_sign-quant_sign` το οποίο είναι ίσο με την διαφορά του αρχικού κανονικοποιημένου σήματος πλην το σήμα στην έξοδο του κβαντιστή. Επίσης υπολογίζεται και ο λόγος SNR ο οποίος μας δίνεται μέσα από τον τύπο $SNR = \frac{E[x^2(n)]}{E[e^2(n)]} = \frac{\sigma_x^{\,2}}{\sigma_e^{\,2}}$. Σαν σήμα εισόδου x παίρνεται το κανονικοποιημένο σήμα.

6.3.3 Λεπτομέρειες Υλοποίησης DPCM Κωδικοποιητή

Για τον DPCM κωδικοποιητή έχουμε την συνάρτηση *dpcm* της οποίας ο κώδικας φαίνεται παρακάτω:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Written by Goumenidis Theodoros  %
% December 2003                    %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [coded,error,snr]=dpcm(original,arithmos_bits_kbantisti,order)

[sima,fs,number_bits]=wavread(original);

%%%%%%%%%%%%%%%
% Parametroi %
%%%%%%%%%%%%%%%
N=2^arithmos_bits_kbantisti;

%%%%%%%%%%%%%%%%%%
% Normalization %
%%%%%%%%%%%%%%%%%%
max_sign=max(abs(sima));
norm_sign=sima/max_sign;
max_norm=max(norm_sign);
bima=(2*max_norm)/N;
```

Μέχρι το σημείο αυτό γίνεται η ο ορισμός της συνάρτησης *dpcm* η οποία έχει ένα επιπλέον όρισμα εισόδου σε σχέση με την *pcm* το οποίο είναι η τάξη του προγνώστη *A(z)* – `order`. Επίσης γίνεται ο υπολογισμός των παραμέτρων και η κανονικοποίηση του σήματος. Μετά τον υπολογισμό των παραμέτρων ακολουθεί ο υπολογισμός των συντελεστών του προγνώστη $A(z) = \sum_{i=1}^{P} a_i z^{-i}$ ο οποίος εφαρμόζεται στην *DPCM* κωδικοποίηση.

```
order_plus=order+1;
%%%%%%%%%%%%%%%%%%%%
% Filter Parameter %
%%%%%%%%%%%%%%%%%%%%
for l=1:order_plus
   %Calculate the autocorrelations (lagged products) of the signal
   r(l)=sum(norm_sign(1:(length(norm_sign)-l)).*norm_sign(l:(length(norm_sign)-
1)))/length(norm_sign);
end
```

Με το ποιο πάνω `for` υπολογίζουμε τους συντελεστές αυτοσυσχέτισης του κανονικοποιημένου σήματος. Το loop επαναλαμβάνεται τόσες φορές όσες είναι και η τάξη του προγνώστη όμως επειδή η πίνακες στο Matlab ξεκινούν από το ένα έχουμε μια ολίσθηση δεξιά κατά +1. Κατά αυτό τον τρόπο ο πρώτος συντελεστής αυτοσυσχέτισης $R(0)$ αντιστοιχεί στην ουσία με το πρώτο στοιχείο r(1) του πίνακα κ.ο.κ. Για τον υπολογισμό τώρα των συντελεστών αυτών - r(l) εφαρμόζεται ο τύπος $R(i) = \frac{1}{N} \sum_{i=0}^{N-|m|-1} s(n+|m|)s(n)$.

```
%%%%%%%%%%%%%%%%%%%%
% Gain Calculation %
%%%%%%%%%%%%%%%%%%%%
for l=2:order_plus
   temp=sum(r(l));
```

```
end
gain=sqrt(r(1)-temp);

%%%%%%%%%%%%%%%%%%%%%%%%
% Ypologismos Toeplitz %
%%%%%%%%%%%%%%%%%%%%%%%%
normal_order=order_plus-1;
toeplitz=ones(normal_order,normal_order);
for l=1:normal_order
   l1=l;
   k=1;
   while l~=k & l1~=0
      toeplitz(l,k)=r(l1);
      k=k+1;
      l1=l1-1;
   end
   if l==k
      k1=k;
      for z=1:order_plus-k1
         if k~=order_plus
            toeplitz(l,k)=r(z);
            k=k+1;
         end
      end
   end
end
```

Μετά τον υπολογισμό των συντελεστών αυτοσυσχέτισης υπολογίζεται το κέρδος – `gain` σύμφωνα με τον τύπο $G^2 = R(0) - \sum_{k=1}^{p} a_k R(k)$ της Παραγράφου 3.7.5. Η όλη διαδικασία όπως βλέπουμε βασίζεται στην γραμμική πρόγνωση και πιο συγκεκριμένα στη μέθοδο αυτοσυσχέτισης. Για να σχηματίσουμε όμως τις εξισώσεις Yule – Walker της μεθόδου αυτοσυσχέτισης με την μορφή πινάκων, είναι απαραίτητος ο σχηματισμός του πίνακα Toeplitz. Αυτό γίνεται με το for που ακολουθεί αμέσως μετά από το κέρδος. Το γενικό loop βέβαια του πίνακα Toeplitz ενσωματώνει διάφορα αλλά loop καθώς και κάποιες προϋποθέσεις κατά την εκτέλεση του (`if` και `while`).

Ας εξετάσουμε την περίπτωση όπου ισχύει η συνθήκη `while l~=k & l1~=0` τότε από το σημείο εκείνο και στην γραμμή **l** μπαίνουν οι αντίστοιχοι συντελεστές αυτοσυσχέτισης της γραμμής (αφού l1=l και toeplitz(l,k)=r(l1)). Οι συντελεστές αυτοί τοποθετούνται κατά φθίνουσα σειρά αφού, l1=l1-1. Η διαδικασία αυτή συνεχίζεται μέχρις ότου να μην πληρείται ένας από τους παραπάνω όρους οπότε και είμαστε στην περίπτωση του `if l==k`. Από το σημείο εκείνο και μετά οι συντελεστές αυτοσυσχέτισης ξεκινώντας από τον πρώτο $R(0)$ μπαίνουν κατά αύξουσα σειρά, δηλαδή $R(0)$ $R(1)$ $R(2)$ κ.ο.κ. Η διαδικασία αυτή συνεχίζεται μέχρι να φτάσουμε στο τελευταίο στοιχείο της γραμμής στην οποία βρισκόμαστε (`if k~=order_plus`).

Μόλις ολοκληρωθεί και αυτό το loop η διαδικασία επαναλαμβάνεται για την επόμενη γραμμή l με αρχικοποίηση της στήλης, αφού κάθε φορά δηλώνετε k=1. Αυτό συνεχίζεται έως ότου να σχηματιστεί ο ζητούμενος πίνακας Toeplitz.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Sximatismos pinaka [r] kai pinaka sintleston [a] %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear l;
R=ones(1,normal_order); %einai 1xp
a=zeros(1,normal_order); %einai 1xp
for l=1:normal_order
   R(1,l)=r(l+1);
end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Find the optimal predictor solution %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
a = R*inv(toeplitz);
```

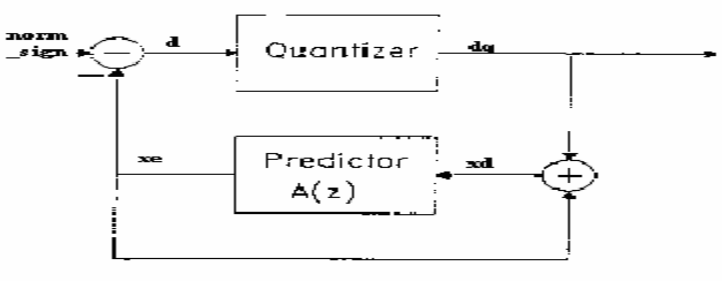
Με τον σχηματισμό του Toeplitz ακολουθεί ο σχηματισμός του πίνακα των συντελεστών του προγνώστη καθώς και του πίνακα των συντελεστών αυτοσυσχέτισης. Στην συνέχεια με πολλαπλασιασμό υπολογίζονται οι τιμές για τους συντελεστές πρόγνωσης a.

```
p= length(a);
xe= 0*norm_sign;  % Exodos Filtrou A(z)
dq= 0*norm_sign;  % Kbantismeno sima
xd1= 0*norm_sign; % Eisodos Filtrou A(z)
xd= zeros(1,p); %arxiki eisodos filtrou einai [0 0 0 0 ...] gia tis proigoumenes
times

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Differencial Quantization %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
temp=1;
for n= temp:length(norm_sign)
   temp=n;
   xe(n)= a*xd'*gain; %exodos filtrou
   d(n)= norm_sign(n)-xe(n); %diafora metaxi kanonikis timis kai exodou filtrou

   %%%%%%%%%%%%%%%%%%%%%%%
   if d(n)<-1
      d(n)=-1
   elseif d(n)>1
      d(n)=1
   end
   %%%%%%%%%%%%%%%%%%%%%%%

   for i=0:(N-1)
      if d(n)>=-1+i*bima & d(n)<=-1+(i+1)*bima
         dq(n)=((-1+(i*bima))+(-1+(i+1)*bima))/2;
      end
   end
   %dq(n)= d;
   xd1(n)= xe(n)+dq(n); %eisodos Filtrou A(z)=exodos Filtrou A(z)+kbantismeni
exodos
   xd= [xd1(n) xd(1:p-1)]; %olisthimeni Filtrou A(z)
end
```

Αφού έχουν υπολογιστεί και οι συντελεστές πρόγνωσης του φίλτρου ακολουθεί η διαδικασία της διαφορικής κβάντισης (Σχήμα 6.2). Αρχικά ορίζεται η ολισθημένη τιμή της εισόδου του φίλτρου ίση με το μηδέν `xd=zeros(1,p)` και ακολουθεί για όλες τις τιμές του κανονικοποιημένου σήματος ο υπολογισμός της εξόδου του φίλτρου *A(z)* σύμφωνα με τον τύπο $xe(n)=\sum_{i=1}^{P} a_i xd(n-i)$.
Το μέγεθος βέβαια της ολισθημένης εισόδου εξαρτάτε κάθε φορά όπως είναι φανερό από την τάξη του προγνώστη.

Μετά από το `xe` υπολογίζεται το *σφάλμα πρόγνωσης* - `d(n)` το οποίο και μπαίνει μέσα σε έναν ομοιόμορφο κβαντιστή όμοιο με αυτόν του *pcm*. Στην έξοδο του κβαντιστή παίρνουμε το σήμα `dq` το οποίο και προστίθεται μαζί με την έξοδο του φίλτρου και μας δίνει το `xd1`. Στην συνέχεια αυτό υφίσταται μια ολίσθηση, `xd= [xd1(n) xd(1:p-1)]`κατά τέτοιο τρόπο ώστε να λαμβάνουμε υπόψη μας τόσες προηγούμενες τιμές όσες και η τάξη του φίλτρου. Το σήμα που προκύπτει

εισάγεται στο *A(z)*. Μόλις ολοκληρωθεί και αυτό το βήμα η διαδικασία επαναλαμβάνεται για την επόμενη τιμή.



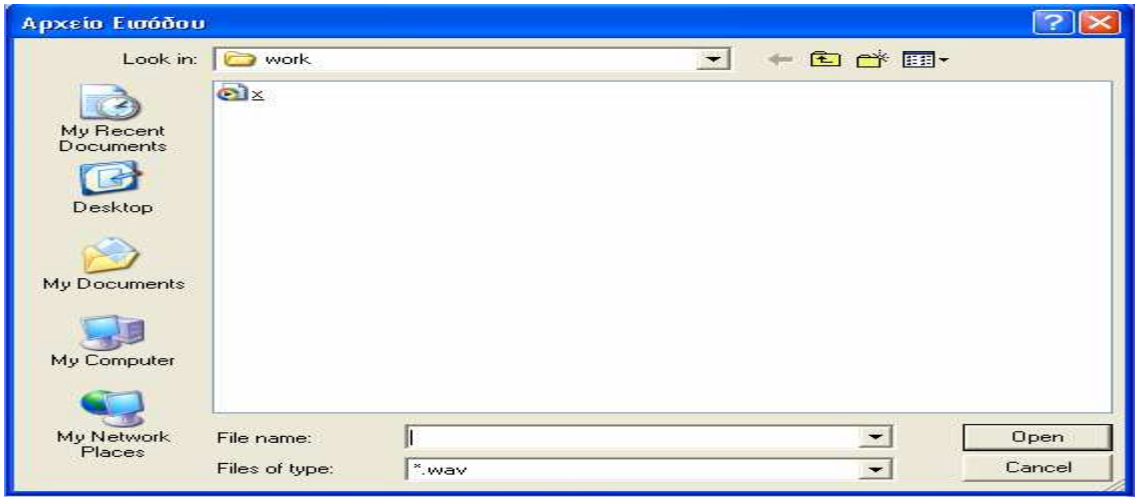
Σχήμα 6.2: Γραφική αναπαράσταση της υλοποίησης του DPCM.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear temp;
coded=dq;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%
% SNR %
%%%%%%%
error=norm_sign-dq;
errorvar=var(error);
signvar=var(norm_sign);
VSNR=signvar/errorvar;
VSNRdb=10*log10(VSNR);
snr=VSNRdb;
```

Στις τελευταίες γραμμές του κώδικα γίνεται υπολογισμός του σφάλματος και του SNR σύμφωνα με τον τύπο $SNR = \frac{E[x^2(n)]}{E[e^2(n)]} = \frac{\sigma_x^2}{\sigma_e^2}$. Σαν σήμα εισόδου x παίρνεται το κανονικοποιημένο σήμα.

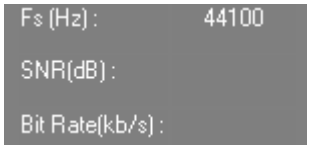
6.4 Λεπτομέρειες Χρήσης

Όπως είπαμε στην Παράγραφο 6.1 το κύριο περιβάλλον αλληλεπίδρασης με το πρόγραμμα το οποίο εμφανίζεται πληκτρολογώντας *guip1* στην prompt γραμμή του Matlab, είναι αυτό του Σχήματος 6.1. Στο δεξί μέρος του, υπάρχει μια σειρά από κουμπιά και επιλογές τα οποία εκτελούν κάποιες λειτουργίες. Με το πάτημα λοιπόν του κουμπιού *¨Είσοδος¨* εμφανίζεται το παράθυρο του Σχήματος 6.3 με το οποίο μπορούμε να επιλέξουμε μέσα από τον υπολογιστή μας το αρχείο *.wav* το οποίο θέλουμε να επεξεργαστούμε. Στην περίπτωση τώρα που δεν επιλέξουμε αρχείο προς επεξεργασία ή που το αρχείο που επιλέξαμε δεν είναι τύπου .wav θα εμφανιστεί ένα μήνυμα λάθους. Η αρχική διεύθυνση στην οποία γίνεται αναζήτηση αρχείων είναι η διεύθυνση c:/MATLABR11/work (αν βέβαια το Matlab είναι εγκατεστημένο στον δίσκο c:), ωστόσο η αναζήτηση μπορεί να γίνει οπουδήποτε αλλού μέσα στον Η/Υ όπως ακριβώς γίνεται και στα Windows.



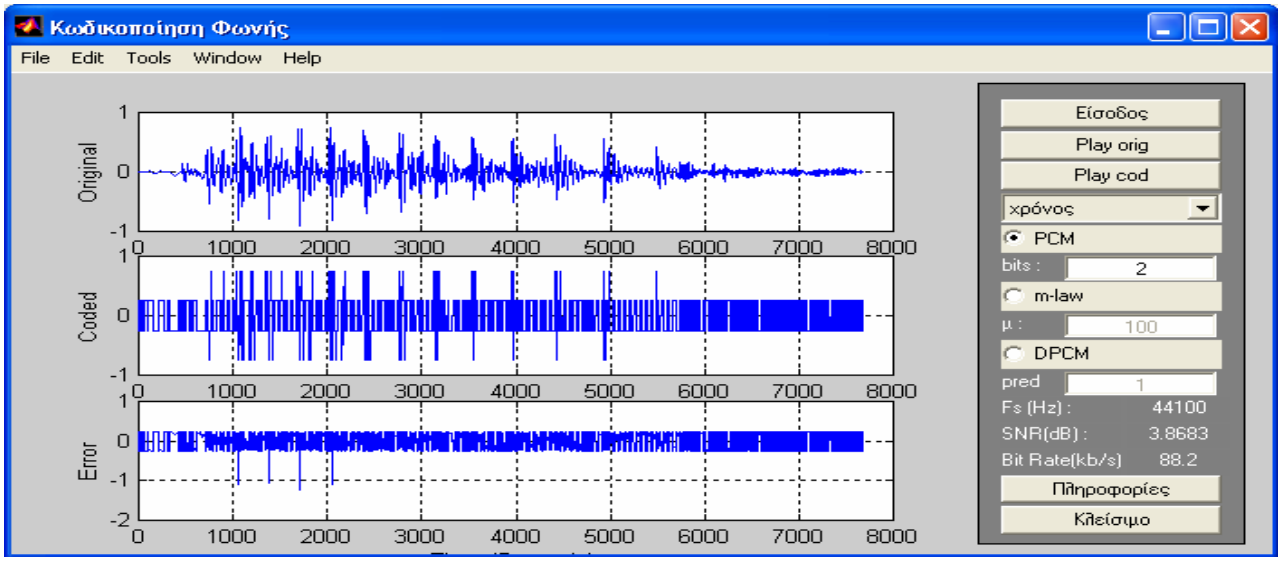
Σχήμα 6.3: Παράθυρο εισαγωγής αρχείων .wav προς επεξεργασία.

Με την εισαγωγή αρχείου και μετά εμφανίζονται στο γραφικό μας περιβάλλον οι πρώτες πληροφορίες στο κάτω δεξί μέρος. Ποίο συγκεκριμένα εμφανίζεται η συχνότητα δειγματοληψίας *Fs* του σήματος εισόδου μας (Σχήμα 6.4) σε Hz.



Σχήμα 6.4: Εμφάνιση της συχνότητας δειγματοληψίας Fs σε Hz.

Επίσης, αφού έχει γίνει η εισαγωγή του αρχείου μπορούμε να εκτελέσουμε μια σειρά από ενέργειες. Πιο συγκεκριμένα μπορούμε να επιλέξουμε το είδος της κωδικοποίησης που θέλουμε να εφαρμόσουμε στο σήμα μας. Η προεπιλεγμένη κωδικοποίηση είναι η *PCM*. Σε αυτή μπορούμε να καθορίσουμε τον αριθμό των bits που θέλουμε να έχει ο κβαντιστής μας δίνοντας έναν αριθμό στο πλαίσιο ¨*bits*¨. Ο περιορισμός εδώ έχει οριστεί μέχρι 32 bits και αν δοθούν παραπάνω τότε εμφανίζεται ένα μήνυμα λάθους. Για να ξεκινήσει η κωδικοποίηση αφού έχουμε δώσει τον αριθμό των bits που θέλουμε πρέπει να κάνουμε κλικ πάνω στο πλήκτρο ¨*PCM*¨. Μόλις η επεξεργασία ολοκληρωθεί στο αριστερό τμήμα της εικόνας εμφανίζονται οι γραφικές παραστάσεις του με σειρά από πάνω προς τα κάτω του *αρχικού σήματος (Original)*, του *κωδικοποιημένου σήματος (Coded)* και του *σφάλματος κωδικοποίησης (Error)* (Σχήμα 6.5).



Σχήμα 6.5: Εμφάνιση των γραφικών παραστάσεων και των αποτελεσμάτων.

Εφόσον η επιλογή στο μενού πάνω από το πλήκτρο του ¨*PCM*¨ είναι στο *χρόνο* τα τρία σήματα θα εμφανιστούν στο πεδίο του χρόνου. Αν η επιλογή τώρα είναι στο *σπεκτρόγραμμα* τα σήματα αυτά θα εμφανιστούν σαν σπεκτρογραφήματα. Η εναλλαγή μεταξύ *χρόνου – σπεκτρόγραμμα* μπορεί να γίνει και εκ τον υστέρων αφού έχουμε κωδικοποίηση το σήμα μας και να δούμε πάλι τις ανάλογες αναπαραστάσεις.
Εκτός όμως από την *pcm* κωδικοποίηση υπάρχει η δυνατότητα για *νόμου – μ* καθώς και για *dpcm*. Κάνοντας λοιπόν τις ανάλογες επιλογές κάτω από την κάθε μια ενεργοποιούνται τα πεδία ¨*μ*¨ και ¨*pred*¨ αντίστοιχα. Στο πεδίο ¨*μ*¨ δηλώνουμε το μέγεθος της παραμέτρου *μ* του τύπου

$$t(n) = S_{\max} \frac{\log(1+\mu\left|\frac{s(n)}{s_{\max}}\right|)}{\log(1+\mu)} \sin g(s(n))$$

ενώ στο πεδίο ¨*pred*¨ δηλώνουμε την τάξη του προγνώστη $A(z) = \sum_{i=1}^{P} a_i z^{-i}$. Με την επιλογή του ¨*μ*¨ ή του ¨*dpcm*¨ ξεκινά ταυτόχρονα με την ενεργοποίηση του αντίστοιχου πεδίου και η κωδικοποίηση του σήματος εισόδου για τις προεπιλεγμένες τιμές αυτών των πεδίων (100 – 1) και για αριθμό bits αυτό που έχουμε δηλώσει. Με το που ολοκληρώνεται αυτή η ¨αρχική¨ επεξεργασία έχουμε την δυνατότητα να τροποποιήσουμε τις παραμέτρους. Και εδώ όμως έχουν τεθεί κάποιοι περιορισμοί. Έτσι η μέγιστη τιμή που μπορούμε να δώσουμε για την παράμετρο *μ* είναι 500 ενώ η μέγιστη τιμή για την τάξη του προγνώστη είναι 10.

Φυσικά μετά το τέλος της κωδικοποίησης εμφανίζονται στο αριστερό μέρος της οθόνης τα τρία σήματα μας. Εκτός όμως από την γραφική αυτή αναπαράσταση των σημάτων στο κάτω δεξί μέρος παίρνουμε μετά την επεξεργασία δύο ακόμα χρήσιμα αποτελέσματα το πρώτο είναι ο *σηματοθορυβικός λόγος SNR* σε dB, ενώ το δεύτερο είναι το *Bit Rate* σε kb/s (Σχήμα 6.5).

Τέλος να πούμε ότι εκτός από την γραφική απεικόνιση υπάρχει η δυνατότητα να ακούσουμε τόσο το σήμα εισόδου πατώντας το πλήκτρο ¨*Play Orig*¨ όσο και το κωδικοποιημένο σήμα πατώντας το πλήκτρο ¨*Play Cod*¨. Πατώντας τα πλήκτρα αυτά, πάλι στο αριστερό μέρος θα εμφανιστούν οι αντίστοιχες απεικονίσεις. Εδώ πρέπει να επισημάνουμε ότι το κωδικοποιημένο σήμα που ακούμε είναι ανάλογα με την επιλογή που έχουμε κάνει κάθε φορά ως προς την κωδικοποίηση μας.

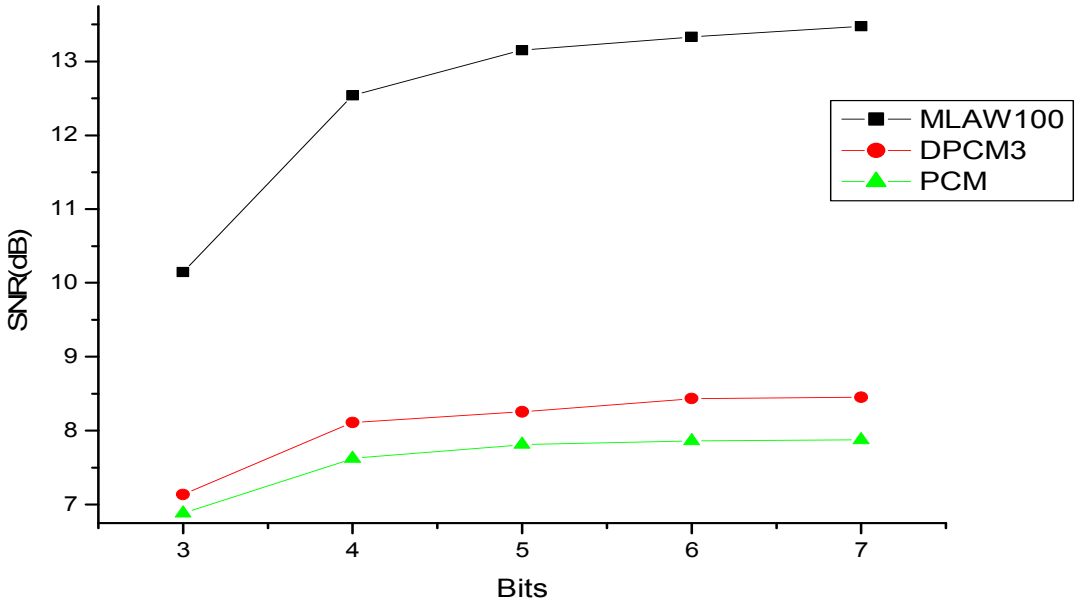
6.5 Συμπεράσματα

Με βάση τις πιο πάνω υλοποιήσεις βλέπουμε ότι εμφανίζονται τα αναμενόμενα αποτελέσματα για τους κωδικοποιητές. Έτσι παρατηρούμε ότι το DPCM παρουσιάζει σαφώς καλύτερη απόδοση από το PCM. Το ίδιο ισχύει και για το m – law. Επίσης βλέπουμε ότι το m – law είναι αυτό που σε γενικές γραμμές υπερτερεί έναντι των άλλων δύο. Βέβαια αυτό δεν είναι κανόνας γιατί η σύγκριση μεταξύ m – law και DPCM εξαρτάται από την τιμή της παραμέτρου μ και από την τάξη του προγνώστη αντίστοιχα. Μεταβάλλοντας τις τιμές αυτών των δύο οι αποδόσεις ποικίλουν.

Πέρα όμως από την σύγκριση των κωδικοποιητών είναι φανερό στο DPCM ότι η αύξηση της τάξης του προγνώστη δεν επιφέρει πάντα υπολογίσιμη βελτίωση στο σηματοθορυβικό λόγο. Αυτό βέβαια δεν πρέπει να μας παραξενεύει μιας και όπως είναι γνωστό η αυτοσυσχέτιση των δειγμάτων εξαρτάται από την δειγματοληψία του σήματος και για απομακρυσμένα δείγματα (δηλαδή μεγάλη τάξη προγνώστη) μπορεί να είναι αρνητική. Κάτι τέτοιο όπως καταλαβαίνουμε έχει άμεσες συνέπειες στον προγνώστη μας και κατ' επέκταση στην ¨ποιότητα¨ του σήματος πρόγνωσης άρα και στην συνολική απόδοση του κωδικοποιητή μας.

Επίσης από πλευράς πολυπλοκότητας αξιολογώντας τον χρόνο που απαιτεί κάθε κωδικοποίηση για να μας παρουσιάσει τα αποτελέσματα βλέπουμε ότι αυτό που παρουσιάζει την μεγαλύτερη είναι το DPCM και ακολουθεί το m – law και τέλος το PCM.

Τέλος με επισταμένες παρατηρήσεις και εξετάζοντας διαφορετικά σήματα, μπορούμε να πούμε σε γενικές γραμμές ότι καταλήγουμε στα συμπεράσματα εκείνα τα οποία έχουμε αναφέρει για τους συγκεκριμένους κωδικοποιητές στην Παράγραφο 5.8.



Σχήμα 6.6: Εμφάνιση των αποτελεσμάτων των κωδικοποιητών pcm, m – law (με μ=100) και dpcm (με τάξη προγνώστη 3) για διαφορετικούς αριθμούς bits.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Atal S. Bishnu, and Nikil S. Jayant, "Speech Codign", *AT&T Bell Laboratories, Murray Hill.*

2. Barnewell Thomas P III., Kambiz Nayebi, and Craig H. Richardon, "Speech Coding: A Computer Labatory Textbook", *John Wiley & Sons, Inc.*, 1996.

3. Batri Nadim, "Robust Spectral Parameter Coding in Speech Processing", *http://www. tsp.ece.mcgill.c /MMSP/Theses/1998/BatriT1998.pdf*, May 1998.

4. Bradbury Jeremy, "Linear Predictive Coding", *http://www.cs.queensu.ca/home/bradbury/pdf/ lpc_paper.pdf*, December 2000.

5. Burnett G. C., Ng L. C., Holzrichter J. F., and Gable T. J., "Denoising Of Human Speech Using Combined Acoustic And EM Sensor Signal Processing", *http://www.icsi.berkeley.edu/~dpwe/ research/etc/ icassp2000/pdf/1914_103.PDF*.

6. Crochiere R. E., "Sub – Band Coding", *The Bell System Technical Journal*, September 1981.

7. Γεωργάκης Θεόδωρος, Κάππας Κωνσταντίνος, "Παραγωγή Και Επεξεργασία Σήματος", *Οργανισμός Εκδόσεως Διδακτικών Βιβλίων.*

8. Goyal K Vivek, Jun Zhuang and Martin Vetterli, "Transform Coding with Backward Adaptive Updates", *http://lcavwww.epfl.ch/publications/00/postscripts/GoyalZhuangVetterli.pdf*, September 1996.

9. Hasegawa – Johnson Mark, "Lecture Notes In Speech Production, Speech Coding, and Speech Recognition", February 2000.

10. Kiviluoto Antti, "*Speech Coding Standards*", *http://mia.ece.uic.edu/~papers/WWW/ MultimediaStandards/ chapter3.pdf*.

11. Leis John, "Speech Coding Lecture Notes & Examples", *http://www.usq.edu.au/users/leis/notes /sigproc/spchcode.pdf*, December 2000.

12. Mitra S.K., "Lecture Notes In Quadrature Mirror Filter Banks".

13. McClellan Stan, and Gibson Jerry D., "Speech Signal Processing: Coding, Transmission and Storage".

14. Mitra S. K., "Quadrature – Mirror Filter Bank", *http://www.ee.nmt.edu/~rison/mitr / Ch10_4.pdf.*

15. Morgan N., and Gold B., "Medium & High Rate Coding – Lecture 26", *http:// www.icsi.berkeley.edu/ eecs225d/spr01/lectures/lect_3_23.pdf*, Spring 1999.

16. Νασίοπουλος Αθανάσιος Δρ, Δημήτρης Χατζόπουλος, "Συστήματα Εκπομπής – Λήψης", *Οργανισμός Εκδόσεως Διδακτικών Βιβλίων,* Φερβουάριος 2000.

17. Naylor A Patrick Dr, "Lecture Notes", *http://www.ee.ic.ac.uk/hp/staff/pnaylor/SpeechProcessing.html*.

18. Ozgu Ozun, Philipp Steurer, and Daniel Thell, "Wideband Speech Coding with Linear Predictive Coding (LPC)", *http://www.ee.ucla.edu/~psteurer/projects/ee214aprojectreport.pdf*, Winter 2002.

19. Peleg Nimrod , "Introduction to Speech Coding", *http://cs.haifa.ac.il/~nimrod/Compression Speech/ S2CODING2004.pdf,* Jan 2004.

20. Rabiner L. R., and Schafer R. W., "Digital Processing Of Speech Signals", *Prentice – Hall Inc*, 1978.

21. Schniter Phil, "Two Branch Quadvalue Mirror Filterbank (QMF)", *Creative Commons Attribution,* 2003.

22. Schussler Marc, "Design And Simulation Of A Speech Doder For Mobile Communication Systems",*Master's Thesis*, 1994.

23. Sheth Amit, "Speech Redundacy", *http://web.njit.edu/~ams7/SpeechRedundancy.ppt*.

24. Shlomot Eyal, Vladimir Cuperman, and Gersho Allen, "Hybrid Coding: Combined Harmonic and Waveform Coding of Speech at 4 kb/s", *IEEE Transactions On Speech And Audio Processing, Vol. 9, No. 6*, September 2001.

25. Smith W.Steven, "The Scientist And Engineer's Guide To Digital Signal Processing", *California Technical Publishing (Second Edition),* 1999.

26. Sturt Christian, "Low Bit Rate Speech Coding", *Master Of Philosophy (Universtiy Of Surrey),* 2001.

27. Tansony R.W., and Kabal P., "A Variable Rate Adaptive Transform Coder For Digital Storage Of Audio Signals", *IEEE Int. Conf. Communications*, pp. 42.1.1-42.1.6, June 1988.

28. Voran Stephen, "Perception Based Bit Allocation Algorithms For Audio Coding", *Proceeding of IEEE ASSP Workshop,* 1997.

29. Warburton – Φιλιππάκη Ειρήνη, "Εισαγωγή Στη Θεωρητική Γλωσσολογία", Εκδόσεις Νεφέλη, 1992.

30. University Purdue, "Labatory 9: Speech Processign", *Purdue University: EE438 – Digital Signal Processign with Applications,* April 2002.

31. http://home.ewha.ac.kr/~ewhaelec/exhibit/work/project/2_SpeechCoding/speech.pdf, "A Study on the Performance Evaluation of a Speech Coder for Wireless System".

32. http://mi.eng.cam.ac.uk/~ajr/SA95/, "Speech Analysis", 1998.

33. http://scs4.tec.tuiasi.ro/t-teodor/Master/lecture%2020.pdf, "Speech Coding Methods Based on Speech Waveform Representations and Speech Models – Model Based Coding and Coding Standarts".

34. http://www.cen.uiuc.edu/~ece320/handouts/speech.pdf., "Speech Analysis and Synthesis", Fall 2001.

35. http://www.cisco.com/warp/public/788/signalling/waveform_coding.pdf, "Waveform Coding Techniques".

36. http://www.csd.uch.gr/~hy431/docs/DSP_Tutorials.pdf., "TMS320 Software Cooperative Resource Guide".

37. http://www.cs.tut.fi/sgn/arg/intro/ , "Audio Signal Processing Basics".

38. http://www.ece.pdx.edu/~ssp/Slides/Autocorrelationx4.pdf, "Autocorrelation".

39. http://www.ee.oulu.fi/~skidi/teaching/internet_multimedia/audio_long.pdf., "Audio and Speech", October 2001".

40. http://www.ind.rwth-aachen.de/research/speech_coding.html, "Digital source coding of speech signals".

41. http://www.lnt.de/LMS/lecture/seminar_ss/seminar_SS03/documents/slides_prediction_in_source_coding_audio_1.pdf, "Prediction in Source Coding – Speech and Audio".

42. http://www-mobile.ecs.soton.ac.uk/speech_codecs/, "Speech Coding".

43. http://www-mobile.ecs.soton.ac.uk/speech_codecs/waveform.html, "Waveform Codecs".

44. http://www.usq.edu.au/users/leis/courses/ELE4607/module7c.pdf, "Advanced Digital Communications. Module 7 – Speech Coding".

45. http://xanthippi.ceid.upatras.gr/courses/mobi_net/notes4.doc, "Κινητά Δίκτυα Επικοινωνιών. Διάλεξη 4: Τεχνικές Κωδικοποίησης Πηγής – Η Περίπτωση της Φωνής ".