

Τ.Ε.Ι. ΚΡΗΤΗΣ – ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΝΙΚΗΣ
ΠΑΡΑΡΤΗΜΑ ΧΑΝΙΩΝ

**ΑΝΑΛΥΣΗ ΤΕΧΝΙΚΩΝ
ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΦΩΝΗΣ
ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

ΣΠΟΥΔΑΣΤΗΣ:
ΓΙΑΝΝΟΥΛΗΣ ΙΩΣΗΦ

ΥΠΕΥΘΥΝΟΣ ΚΑΘΗΓΗΤΗΣ:
Δρ. Α. ΚΩΝΣΤΑΝΤΑΡΑΣ

ΧΑΝΙΑ 29/6/2009

ΠΕΡΙΕΧΟΜΕΝΑ

ΣΥΝΤΟΜΗ ΕΙΣΑΓΩΓΗ	
ΕΙΣΑΓΩΓΗ	4
ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ: Η ΑΝΘΡΩΠΙΝΗ ΦΩΝΗ	
1.1. ΠΑΡΑΓΩΓΗ ΦΩΝΗΣ	8
1.2. ΦΩΝΗΤΙΚΗ ΕΠΙΚΟΙΝΩΝΙΑ	9
1.2.1. ΓΝΩΣΤΙΚΑ ΜΟΝΤΕΛΑ	9
1.2.2. ΓΛΩΣΣΙΚΑ ΜΟΝΤΕΛΑ	10
1.2.3. ΜΟΝΤΕΛΑ ΦΩΝΗΤΙΚΟΥ ΣΥΣΤΗΜΑΤΟΣ	10
1.2.4. ΑΚΟΥΣΤΙΚΑ ΜΟΝΤΕΛΑ	11
1.3. ΚΩΔΙΚΟΠΟΙΗΣΗ ΦΩΝΗΣ	12
1.3.1. ΓΕΝΙΚΑ ΓΙΑ ΤΗΝ ΚΩΔΙΚΟΠΟΙΗΣΗ ΦΩΝΗΣ	12
1.3.2. ΧΡΗΣΙΜΟΤΗΤΑ ΤΗΣ ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΦΩΝΗΣ	12
1.3.3. ΚΩΔΙΚΟΠΟΙΗΣΗ & ΚΒΑΝΤΙΣΗ	12
1.3.4. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΟΣ ΚΩΔΙΚΟΠΟΙΗΣΗΣ	13
1.4. ΚΩΔΙΚΟΠΟΙΗΤΕΣ ΦΩΝΗΣ	14
1.4.1. ΟΡΙΣΜΟΣ ΚΩΔΙΚΟΠΟΙΗΤΗ ΦΩΝΗΣ	14
1.4.2. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΚΩΔΙΚΟΠΟΙΗΤΩΝ ΦΩΝΗΣ	15
1.5. ΠΡΟΤΥΠΑ ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΦΩΝΗΣ	15
ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ: WAVEFORM CODERS	
2.1. ΕΙΣΑΓΩΓΗ	17
2.1.1. ΣΥΝΤΟΜΗ ΑΝΑΛΥΣΗ PCM	18
2.1.2. SIGNAL TO NOISE RATIO (SNR)	22
2.2. ΟΜΟΙΟΜΟΡΦΗ ΚΒΑΝΤΙΣΗ	22
2.3. ΜΗ ΟΜΟΙΟΜΟΡΦΗ ΚΒΑΝΤΙΣΗ	24
2.3.1. ΛΟΓΑΡΙΘΜΙΚΟΙ ΚΒΑΝΤΙΣΤΕΣ	26
2.3.2. ΒΕΛΤΙΣΤΟΙ ΚΒΑΝΤΙΣΤΕΣ	27
2.4. ΠΡΟΣΑΡΜΟΣΤΙΚΗ ΚΒΑΝΤΙΣΗ	28
2.4.1. FEED FORWARD ΠΡΟΣΑΡΜΟΓΗ	29
2.4.1.1. ΠΡΟΣΑΡΜΟΓΗ ΜΕΓΕΘΟΥΣ ΒΗΜΑΤΟΣ	31
2.4.1.2. ΚΕΡΔΟΣ ΠΡΟΣΑΡΜΟΓΗΣ	32
2.4.2. FEEDBACK ΠΡΟΣΑΡΜΟΓΗ	34
2.5. RETURN TO ZERO Ή NON RETURN TO ZERO PCM	34
2.6. ΙΣΤΟΡΙΚΑ ΣΤΟΙΧΕΙΑ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥ PCM	35
2.6.1. ΣΥΣΤΗΜΑ PCM	36
2.6.1.1. FAILSAFE	37
2.7. ΔΙΑΜΟΡΦΩΣΗ ΘΕΣΗΣ ΠΑΛΜΟΥ (PPM)	37
2.7.1. ΥΠΟΒΑΘΡΟ	37
2.7.2. ΔΟΚΙΜΗ ΕΥΡΟΥΣ	38
2.7.3. ΠΑΡΕΜΒΟΛΕΣ ΣΤΑ PPM ΚΑΙ PCM	38
2.7.4. ΑΣΦΑΛΕΙΑ	38
ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ - ΚΩΔ. WAVEFORM ΜΕ ΣΤΑΘΕΡΗ ΠΡΟΒΛΕΨΗ	
3.1. ΕΙΣΑΓΩΓΗ	39
3.2. ΒΑΣΙΚΟ DPCM	40
3.3. DPCM ΜΕ ΠΡΟΣΑΡΜΟΣΤΙΚΗ ΚΒΑΝΤΙΣΗ (ADPCM)	41
3.4. ΔΙΑΜΟΡΦΩΣΗ ΔΕΛΤΑ (DM)	42
3.4.1. ΓΡΑΜΜΙΚΗ ΔΙΑΜΟΡΦΩΣΗ ΔΕΛΤΑ (LDM)	43
3.4.2. ΠΡΟΣΑΡΜΟΣΤΙΚΗ ΔΙΑΜΟΡΦΩΣΗ ΔΕΛΤΑ (ADM)	44
3.4.3. ΣΥΝΕΧΟΥΣ ΜΕΤΑΒΛΗΤΗΣ ΚΛΙΣΗΣ ΔΙΑΜΟΡΦΩΣΗ ΔΕΛΤΑ (CVSD)	45
ΚΕΦΑΛΑΙΟ ΤΕΤΑΡΤΟ - ΚΩΔ. WAVEFORM ΜΕ ΠΡΟΣΑΡΜ. ΠΡΟΒΛΕΨΗ	
4.1. ΕΙΣΑΓΩΓΗ	47
4.2. ΚΩΔΙΚΟΠΟΙΗΣΗ ΠΡΟΣΑΡΜΟΣΤΙΚΗΣ ΠΡΟΒΛΕΨΗΣ	47
4.3. APC ΜΕ ΠΡΟΒΛΕΨΗ ΚΛΙΣΗΣ (APC-PP)	50
4.4. ΚΩΔΙΚΟΠΟΙΗΣΗ ΘΟΡΥΒΟΥ-ΑΝΑΔΡΑΣΗΣ	51
4.4.1. ΑΝΟΙΚΤΟΥ ΒΡΟΓΧΟΥ DPCM (D*PCM)	53

4.4.2. DPCM ΜΕ ΑΝΑΔΡΑΣΗ ΘΟΡΥΒΟΥ	53
4.4.3. APC ΜΕ ΑΝΑΔΡΑΣΗ ΘΟΡΥΒΟΥ ΚΑΙ ΠΡΟΒΛΕΨΗ PITCH	54

ΚΕΦΑΛΑΙΟ ΠΕΜΠΤΟ: VOCODERS

5.1. ΙΣΤΟΡΙΑ ΤΩΝ VOCODERS	56
5.2. ΕΙΣΑΓΩΓΗ ΣΤΟΥΣ VOCODERS	56
5.2.1. ΜΟΝΤΕΛΟ ΦΩΝΗΣ	56
5.2.2. ΧΡΗΣΕΙΣ VOCODER	57
5.3. ΚΑΤΗΓΟΡΙΕΣ VOCODER	57
5.3.1. CHANNEL VOCODERS (ΚΩΔΙΚΟΠΟΙΗΤΕΣ ΚΑΝΑΛΙΩΝ)	57
5.3.2. FORMANT VOCODERS	59
5.3.3. ΚΩΔΙΚΟΠΟΙΗΤΕΣ ΓΡΑΜΜΙΚΗΣ ΠΡΟΒΛΕΨΗΣ (LPC)	60
5.3.3.1. ΑΝΑΛΥΣΗ LPC	61
5.3.3.2. ΠΡΟΤΥΠΟ LPC-10	62
5.3.3.3. ΛΕΙΤΟΥΡΓΙΑ LPC	64
5.3.3.4. ΕΝΑΛΛ. ΜΕΘ. ΔΗΜ. ΤΟΥ ΣΗΜ. ΔΙΕΓΕΡΣΗΣ ΣΤΟΝ ΑΠΟΚΩΔ. LPC	65

ΕΠΙΛΟΓΟΣ - ΣΥΜΠΕΡΑΣΜΑΤΑ

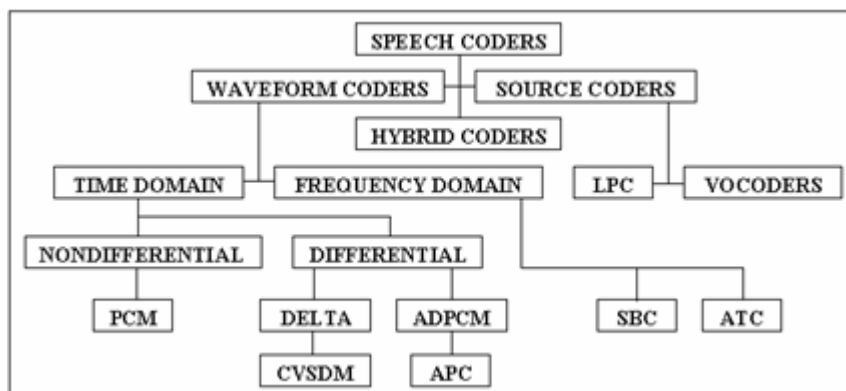
6.1. Η ΣΥΝΕΧΗΣ ΑΝΑΓΚΗ ΓΙΑ ΤΗ ΣΥΜΠΙΕΣΗ ΦΩΝΗΣ	68
6.2. ΟΙ ΣΤΟΧΟΙ ΤΗΣ ΤΕΧΝΟΛΟΓΙΑΣ	68
6.3. ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ	69
6.3.1. ΚΩΔΙΚΟΠΟΙΗΣΗ, ΕΠΙΚΟΙΝΩΝΙΑ ΚΑΙ ΔΙΚΤΥΩΣΗ	69
6.3.2. ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ	69

ΠΑΡΑΡΤΗΜΑ

1. ΠΙΝΑΚΑΣ ΒΑΣΙΚΟΤΕΡΩΝ ΑΥΔΙΟ CODERS	70
-------------------------------------	----

ΣΥΝΤΟΜΗ ΕΙΣΑΓΩΓΗ

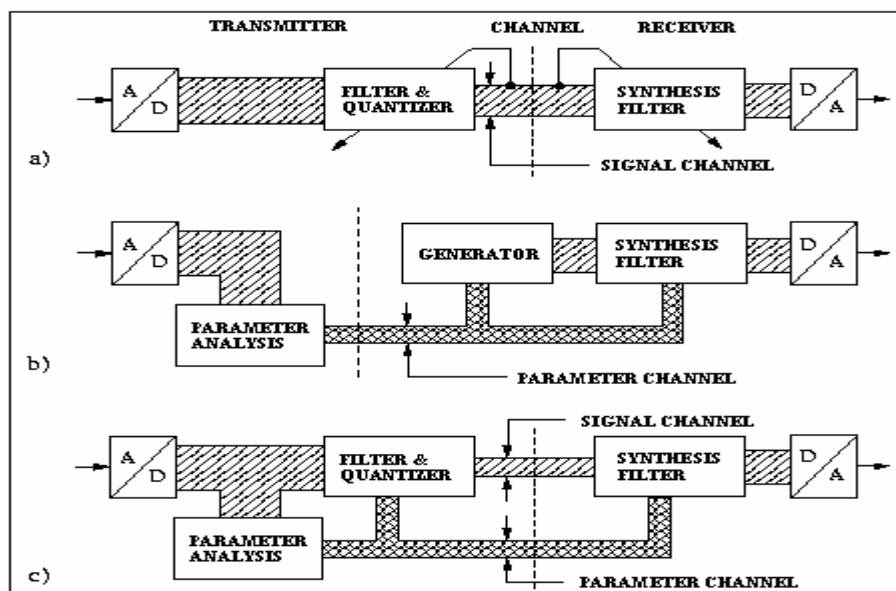
Η παρούσα πτυχιακή εργασία επιχειρεί να περιγράψει τα βασικά χαρακτηριστικά των κυριότερων κωδικοποιητών φωνής. Υπάρχουν διάφορες τεχνικές για κωδικοποίηση. Μια ιεραρχική κατηγοριοποίηση των διάφορων τεχνικών κωδικοποίησης παρουσιάζεται στο παρακάτω σχήμα:



ΣΧΗΜΑ 1: Τεχνικές κωδικοποίησης

Οι αλγόριθμοι κωδικοποίησης φωνής μπορούν, με βάση τον τρόπο με τον οποίο επιτυγχάνουν συμπίεση, να ταξινομηθούν σε τρεις βασικές κατηγορίες:

- a) Waveform coders (Κωδικοποιητές κυματομορφής)
- b) Vocoders (**Voice Coders** - Κωδικοποιητές φωνής)
- c) Hybrid coders (Υβριδικοί κωδικοποιητές)



ΣΧΗΜΑ 2: Αλγόριθμοι κωδικοποίησης φωνής

a) Waveform coders: (Περιγράφονται εκτενέστερα στα κεφάλαια 2, 3 και 4). Οι Waveform Coders προσπαθούν να αναπαράγουν την κυματομορφή του σήματος της φωνής όσο το δυνατόν πιστότερα. Είναι σχεδιασμένοι ώστε να μην εξαρτώνται από το σήμα εισόδου και γι' αυτό μπορούν να κωδικοποιήσουν εξίσου καλά μια ποικιλία σημάτων. Στον κωδικοποιητή, μια μείωση της δυναμικής των σημάτων μπορεί να επιτευχθεί από μια σταθερή ή προσαρμοστική κβάντιση. Καλύτερα αποτελέσματα επιτυγχάνονται εάν υιοθετείται ένα (σταθερό ή προσαρμοστικό) φιλτράρισμα πρόβλεψης, σύμφωνα με τις

ιδιότητες συσχετισμού του σήματος. Κάτω από ορισμένες συνθήκες, ένα κέρδος πρόβλεψης μπορεί να χρησιμοποιηθεί για να μειώσει το bit rate, εάν το (υπόλοιπο) σήμα λάθους πρόβλεψης κβαντιστεί αντί του αρχικού σήματος. Οι παράμετροι των φίλτρων πρόβλεψης μπορούν να προσαρμοστούν χρησιμοποιώντας το αναδημιουργημένο σήμα. Έχουν το πλεονέκτημα ότι είναι ανθεκτικοί σε ένα ευρύ φάσμα χαρακτηριστικών της φωνής καθώς και σε θορυβώδη περιβάλλοντα. Όλα αυτά τα πλεονεκτήματα επιτυγχάνονται με ελάχιστη πολυπλοκότητα. Ένα μειονέκτημα είναι ότι επιτυγχάνουν μόνο μέση οικονομία στο ρυθμό μετάδοσης δεδομένων.

Μερικά παραδείγματα Waveform Coders είναι τα PCM (Pulse Code Modulation), DPCM (Differential Pulse Code Modulation), ADPCM (Adaptive Differential Pulse Code Modulation), DM (Delta Modulation), CVSDM (Continuously Variable Slope Delta Modulation) και APC (Adaptive Predictive Coding).

Οι κωδικοποιητές κυματομορφής προσπαθούν, χωρίς τη χρησιμοποίηση οποιασδήποτε γνώσης για το πως δημιουργήθηκε το σήμα προς κωδικοποίηση, να παράγουν ένα αναδομημένο σήμα του οποίου η κυματομορφή να πλησιάζει όσο το δυνατόν περισσότερο την αρχική. Αυτό σημαίνει ότι θεωρητικά πρέπει να είναι ανεξάρτητοι σημάτων και να δουλεύουν καλά με μη λεκτικά σήματα. Γενικά είναι χαμηλής πολυπλοκότητας κωδικοποιητές που παράγουν υψηλής ποιότητας ήχο σε ποσοστό περίπου πάνω από 16 kbits/s. Όταν το ποσοστό στοιχείων πέσει κάτω από αυτό το επίπεδο, η αναδομημένη ποιότητα ήχου που μπορεί να ληφθεί υποβαθμίζεται γρήγορα.

Οι waveform coders περιγράφουν την ομιλία κώδικα με μια εξ' ολοκλήρου προσέγγιση χρονικών περιοχών. Οι προσεγγίσεις περιοχών συχνότητας είναι επίσης δυνατές, και έχουν ορισμένα πλεονεκτήματα. Παραδείγματος χάριν στην κωδικοποίηση υποζωνών (Sub-Band Coding - SBC) ο εισαγωγικός ήχος είναι χωρισμένος σε διάφορες ζώνες συχνότητας, ή σε υποζώνες, και κάθε μια κωδικοποιείται ανεξάρτητα χρησιμοποιώντας παραδείγματος χάριν έναν ADPCM κωδικοποιητή. Στο δέκτη τα σήματα υποζωνών αποκωδικοποιούνται και επανασυνδυάζονται για να δώσουν το αναδημιουργημένο λεκτικό σήμα. Τα πλεονεκτήματα αυτά προέρχονται από το γεγονός ότι ο θόρυβος σε κάθε υποζώνη εξαρτάται μόνο από την κωδικοποίηση που χρησιμοποιείται σε εκείνη την υποζώνη. Επομένως μπορούμε να διανείμουμε περισσότερα bit στις σημαντικές υποζώνες έτσι ώστε ο θόρυβος σε αυτές τις περιοχές συχνότητας να είναι χαμηλός, ενώ σε άλλες υποζώνες μπορούμε να επιτρέψουμε έναν υψηλό θόρυβο κωδικοποίησης επειδή ο θόρυβος σε αυτές τις συχνότητες είναι λιγότερο αντιληπτός. Τα προσαρμοστικά σχέδια κατανομής bit μπορούν να χρησιμοποιηθούν στην περαιτέρω εκμετάλλευση αυτών των ιδεών. Οι κωδικοποιητές SBC τείνουν να παράγουν στις επικοινωνίες ποιότητα φωνής που θα περιορίζεται στο φάσμα των 16 ως 32Kbit/s. Λόγω του ότι κατά το φιλτράρισμα η ομιλία πρέπει να χωριστεί σε υποζώνες, οι κωδικοποιητές SBC είναι πιο σύνθετοι από τους απλούς κωδικοποιητές DPCM, και εισάγουν περισσότερη καθυστέρηση κωδικοποίησης. Εντούτοις η πολυπλοκότητα και η καθυστέρηση είναι σχετικά χαμηλές συγκρινόμενες με αυτές των περισσότερων hybrid coders.

Η απλούστερη και πιο συνηθισμένη τεχνική κωδικοποίησης φωνής, η γραμμική διαμόρφωση παλμού-κώδικα (PCM), δεν κάνει ουσιαστικά καμία υπόθεση όσον αφορά τα χαρακτηριστικά του προς κωδικοποίηση σήματος ή την τελική χρήση του. Αυτό έχει ως αποτέλεσμα, τα γραμμικά PCM συστήματα να απαιτούν υψηλότερο bit-rate για να παράγουν δελεαστική ποιότητα φωνής. Έχουν όμως το πλεονέκτημα ότι επειδή δεν έχουν προτιμήσεις σε κάποια συγκεκριμένη κατηγορία σημάτων, μπορούν να χρησιμοποιηθούν για την κωδικοποίηση άλλων κατηγοριών σημάτων όπως σημάτων πληροφοριών ή μουσικής. Τα συμπίεσης/αποσυμπίεσης (companding - compressing/expanding) συστήματα PCM, στην πιο βασική του μορφή, κάνουν απευθείας χρήση των χαρακτηριστικών επικάλυψης του αυτιού. Στα τελικά κωδικοποιημένα σήματα φωνής που δημιουργούνται, ο θόρυβος συσχετίζεται με το σήμα, το οποίο είναι καλό, αλλά ο θόρυβος επίσης εξαπλώνεται σε όλο το εύρος συχνότητων με κανένα σεβασμό στην παρουσία ενέργειας σήματος, το οποίο είναι κακό. Τα companded συστήματα PCM είναι εξαιρετικά απλά και με χαμηλό κόστος υλοποίησης, αλλά ο ρυθμός στον οποίο πρέπει να λειτουργήσουν για να πετύχουν δελεαστική ποιότητα φωνής είναι ακόμα σχετικά υψηλός.

Συστήματα όπως τα διαφορικής διαμόρφωσης κώδικα παλμού (DPCM) και διαμόρφωσης δέλτα (Delta Modulation - DM) κάνουν απευθείας χρήση ενός μακροχρόνιου σταθερού στατιστικού μοντέλου για την παραγωγή φωνής. Η προσαρμοστική διαφορική διαμόρφωση κώδικα παλμού (ADPCM) και η προσαρμοστική διαμόρφωση δέλτα (Adaptive

Delta Modulation – ADM) επίσης χρησιμοποιούν την, με αργούς ρυθμούς διαφοροποιούμενη, φύση των βραχυχρόνιων ενεργειών, προκαλώντας ισχυρή συσχέτιση του θορύβου και του σήματος φωνής, και προκαλώντας μια δραματική μείωση του άσκοπου θορύβου. Διάφορες μορφές των κωδικοποιητών προσαρμοστικής μετατροπής (Adaptive Transform Coders – ATC) και των κωδικοποιητών προσαρμοστικής πρόβλεψης (Adaptive Predictive Coders – APC) κάνουν χρήση όλων των φωνητικών και ακουστικών χαρακτηριστικών, και τέτοια συστήματα είναι ικανά να παράγουν τέλεια ποιότητα φωνής στα μεσαία και χαμηλά bit-rates. Βασικά χαρακτηριστικά των Waveform Coders είναι τα εξής:

- *Απλή υλοποίηση*
- *Μικρή καθυστέρηση*
- *Είναι γενικοί, δηλαδή είναι χρήσιμοι και για άλλα σήματα*
- *Ρυθμοί μετάδοσης $\geq 32\text{kbps}$ (kbits/sec)*

b) Vocoders (voice coders): (Περιγράφονται εκτενέστερα στο κεφάλαιο 5). Οι vocoders μπορούν να επιτύχουν μεγάλη οικονομία στο ρυθμό μετάδοσης δεδομένων αλλά είναι γενικά περισσότερο πολύπλοκοι. Βασίζονται στη χρησιμοποίηση της πρότερης (priori) γνώσης του σήματος που πρόκειται να κωδικοποιηθεί και γι' αυτό εξειδικεύονται σε συγκεκριμένα σήματα (signal specific). Στους Vocoders, οι παράμετροι ενός μοντέλου πηγής-φίλτρου φωνής και όχι τα σήματα, κβαντίζονται και μεταδίδονται. Αυτή η αντιπροσώπευση σύνθεσης πηγής-φίλτρου ακολουθεί στενά το πρότυπο της λεκτικής παραγωγής. Το φίλτρο σύνθεσης διαφοροποίησης χρόνου (time varying) αντιστοιχεί στο φωνητικό κομμάτι και μπορεί να περιλάβει ένα πρότυπο του ακουστικού σωλήνα. Η χρήση αυτού του φίλτρου αντιστοιχεί στην αρχή της κωδικοποίησης γραμμικής πρόβλεψης (Linear Predictive Coding - LPC). Οι vocoders χωρίζονται σε τρεις κατηγορίες, τους channel vocoders, τους formant vocoders και τους LPC. Βασικά χαρακτηριστικά των vocoders είναι τα εξής:

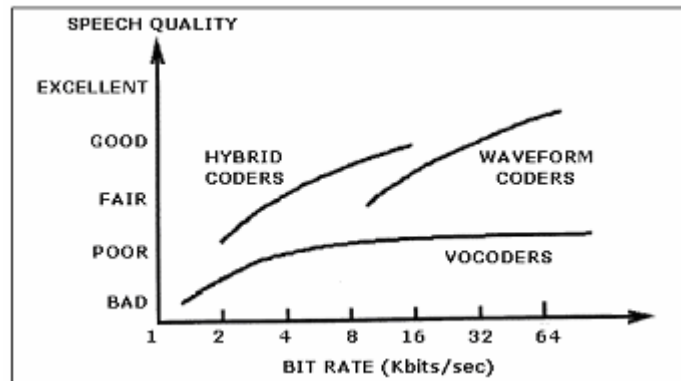
- *Μέση πολυπλοκότητα*
- *Πολύ χαμηλοί ρυθμοί μετάδοσης*
- *Χαμηλή ποιότητα*

c) Hybrid coders: (Επειδή οι υβριδικοί κωδικοποιητές δεν είναι μια ευκρινώς διαχωρισμένη κατηγορία κωδικοποιητών, διότι περιλαμβάνει στοιχεία και από τις δύο προηγούμενες κατηγορίες, γενική περιγραφή τους γίνεται σε όλα τα κεφάλαια).

Αυτή η ενδιάμεση κατηγορία μεταξύ των waveform coders και vocoders κυριαρχεί στις λύσεις υψηλής τεχνικής των κωδικοποιητών φωνής για τα μεσαία bit rate (0,5 ως 2 bit ανά δείγμα) και τα υψηλής ποιότητας ψηφιακά ενσύρματα και ασύρματα συστήματα επικοινωνιών. Όπως και στους vocoders, οι παράμετροι ενός φίλτρου σύνθεσης LPC κβαντίζονται και μεταδίδονται ως δευτερεύουσες πληροφορίες (κανάλι παραμέτρων). Επιπλέον, τα περιοδικά (φωνητικά) μέρη του λεκτικού σήματος μοντελοποιούνται από ένα δεύτερο φίλτρο, το οποίο καλείται μακροπρόθεσμο φίλτρο πρόβλεψης (Long Term Predictive - LTP). Η πλειοψηφία των σύγχρονων υβριδικών κωδικοποιητών φωνής είναι βασισμένη στην αρχή της κωδικοποίησης γραμμικής ανάλυσης-από-σύνθεση πρόβλεψης γνωστής επίσης ως CELP (Code-Excited Linear Prediction). Οι παραλλαγές των κωδικοποιητών ανάλυσης-από-σύνθεση είναι δομές όπως ACELP (Algebraic CELP), RPE (Regular Pulse Excitation), MPE (Multi Pulse CELP), VSELP (Vector-Sum Excited Linear Prediction). Βασικά χαρακτηριστικά των Hybrid coders είναι τα εξής:

- *Υψηλή πολυπλοκότητα*
- *Μεταβλητή καθυστέρηση*
- *Χαμηλοί μεσαίοι ρυθμοί μετάδοσης*

Συγκριτικά, η ποιότητα της κωδικοποιημένης φωνής ως προς το ρυθμό μετάδοσης φαίνεται στο παρακάτω σχήμα:



ΣΧΗΜΑ 3

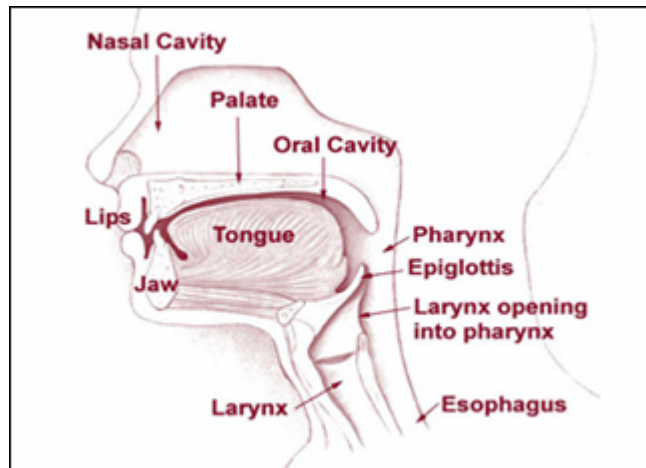
Coding Technique	Bit Rate (kbps)	Aural Noise Masking	Aural Frequency	Aural Phase	Syllabic Energy	Vocal Tract Model	Short Time Stationarity	Pitch
Linear PCM	80-120	No	No	No	No	No	No	No
Companded PCM	50-100	Yes	No	No	No	No	No	No
Delta Modulation	50-80	Yes	No	No	No	Yes	No	No
DPCM	40-80	Yes	No	No	No	Yes	No	No
ADM	16-40	Yes	No	No	Yes	Yes	No	Maybe
Subband Coder	10-32	Yes	Yes	Maybe	Yes	No	No	Maybe
ATC	8-32	Yes	Yes	Maybe	Yes	Yes	Yes	Yes
APC	8-32	Yes	No	No	Yes	Yes	Yes	Maybe
MPLPC	8-16	Yes	No	Yes	Yes	Yes	Yes	Yes
CELP	4-16	Yes	No	Yes	Yes	Yes	Yes	Yes
SEV	4-16	Yes	No	Yes	Yes	Yes	Yes	Yes
LPC Vocoder	0,6-2,4	Yes	No	Yes	Yes	Yes	Yes	Yes

ΠΙΝΑΚΑΣ 1: Σύγχρονες τεχνικές κωδικοποίησης φωνής.

ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ Η ΑΝΘΡΩΠΙΝΗ ΦΩΝΗ

1.1. ΠΑΡΑΓΩΓΗ ΦΩΝΗΣ

Για να μπορέσει κάποιος να καταλάβει πώς λειτουργούν οι μέθοδοι επεξεργασίας φωνής, μια συνοπτική εξήγηση της παραγωγής φωνής απαιτείται. Το σχήμα 1.1 παρουσιάζει το ανθρώπινο φωνητικό σύστημα. Για να παραγάγει την ομιλία, ο πνεύμονας αντλεί αέρα μέσω της τραχείας. Για μερικούς ήχους, αυτό το ρεύμα του αέρα διακόπτεται περιοδικά από τις φωνητικές χορδές.



ΣΧΗΜΑ 1.1: Το ανθρώπινο φωνητικό σύστημα: Ρινική κοιλότητα (nasal cavity), ουρανίσκος (palate), στοματική κοιλότητα (oral cavity), χείλη (lips), γλώσσα (tongue), σαγόι (jaw), φάρυγγας (pharynx), λάρυγγας (larynx), επιγλωττίδα (epiglottis), οισοφάγος (esophagus).

Η ροή του αέρα διασχίζει τη φωνητική περιοχή (σύστημα οργάνων). Η φωνητική περιοχή εκτείνεται από το άνοιγμα στις φωνητικές χορδές ως το στόμα. Ένα μέρος της ροής ταξιδεύει μέσω της ρινικής κοιλότητας. Η φωνητική περιοχή έχει ορισμένα ηχητικά χαρακτηριστικά, τα οποία μπορούν να αλλάξουν διαφοροποιώντας τη μορφή της φωνητικής περιοχής, για παράδειγμα με την κίνηση της γλώσσας. Αυτά τα ηχητικά χαρακτηριστικά μετασχηματίζουν τη ροή του αέρα που προέρχεται από τις φωνητικές χορδές για να δημιουργήσουν έναν συγκεκριμένο ήχο. Οι συχνότητες αντήχησης καλούνται formants. Βασικά, υπάρχουν τρεις κατηγορίες λεκτικών ήχων που μπορούν να παραχθούν. Αυτές οι κατηγορίες είναι:

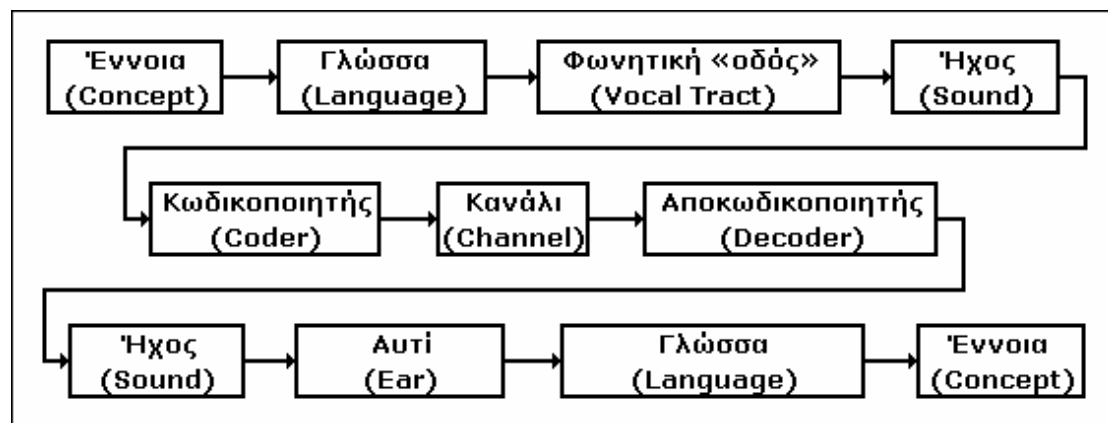
- Οι **φωνούμενοι ήχοι** (voiced) δημιουργούνται όταν δονούνται οι φωνητικές χορδές ανοικτές και κλειστές. Με αυτό το τρόπο, περιοδικοί παλμοί αέρα βγαίνουν από το άνοιγμα των φωνητικών χορδών. Το ποσοστό στο οποίο το άνοιγμα και το κλείσιμο εμφανίζονται καθορίζει το βάθος της φωνής.
- Για να παραγάγουν τους **μη φωνούμενους ήχους** (unvoiced), οι φωνητικές χορδές δεν δονούνται, κρατιούνται ανοικτές. Ο αέρας στέλνεται έπειτα με υψηλές ταχύτητες μέσω μιας συστολής στη φωνητική περιοχή, δημιουργώντας ένα θόρυβο, όπως αναταραχή.
- Οι **έκκροτοι ήχοι** προκύπτουν από την ενίσχυση της πίεσης αέρα πίσω από μια περάτωση στη φωνητική περιοχή και έπειτα την ξαφνική απελευθέρωση αυτού του αέρα.

Ένα σημαντικό γεγονός είναι ότι η μορφή της φωνητικής περιοχής και ο τύπος διέγερσης (η ροή του αέρα που βγαίνει από τις φωνητικές χορδές) αλλάζουν σχετικά αργά. Αυτό σημαίνει ότι για τα σύντομα χρονικά διαστήματα, π.χ. 20ms, το σύστημα λεκτικής παραγωγής μπορεί να θεωρηθεί σχεδόν στάσιμο. Μια άλλη σημαντική παρατήρηση είναι ότι τα λεκτικά σήματα παρουσιάζουν υψηλό βαθμό προβλεψιμότητας. Μερικές φορές εξαιτίας

του περιοδικού σήματος που δημιουργείται από τις φωνητικές χορδές και επίσης εξαιτίας των ηχητικών χαρακτηριστικών του φωνητικού κομματιού.

1.2. ΦΩΝΗΤΙΚΗ ΕΠΙΚΟΙΝΩΝΙΑ

Στην πιο γενική της μορφή η μετάδοση φωνής είναι η διαδικασία προφορικής μετάδοσης μιας ιδέας από έναν άνθρωπο σε έναν άλλο. Ο στόχος του όλου συστήματος είναι να μεταδώσει σωστά μια ιδέα από τον ένα ομιλητή στον άλλο μέσω καναλιών επικοινωνίας. Σε αυτή τη διαδικασία η ιδέα ή σκέψη μετατρέπεται πρώτα σε πρόταση η οποία έπειτα μετατρέπεται σε μυϊκές συσπάσεις των φωνητικών οργάνων, του λάρυγγα και των πνευμόνων. Αυτά μετατρέπουν την πρόταση σε ακουστικό κύμα πεπιεσμένου αέρα (πυκνώματα και αραιώματα των μορίων του αέρα) το οποίο λαμβάνεται από το μικρόφωνο (είσοδος) του συστήματος κωδικοποίησης φωνής. Η δουλειά του κωδικοποιητή φωνής είναι να ψηφιοποιήσει το σήμα φωνής και να το αναπαραστήσει με μία ροή ψηφιακών bit. Ο ρυθμός bit (bit rate) της ροής αυτής πρέπει να είναι σύμφωνος με την δυνατότητα μετάδοσης του καναλιού. Στον δέκτη, ο αποκωδικοποιητής φωνής λαμβάνει την ψηφιακή ροή bit και προσπαθεί να δημιουργήσει ένα νέο σήμα φωνής το οποίο να προσεγγίζει όσο το δυνατόν περισσότερο το αρχικό. Το σήμα αυτό έπειτα μεταδίδεται ακουστικά μέσα στο αυτί του ακροατή.



ΣΧΗΜΑ 1.2: Ολικό περιβάλλον λειτουργίας του κωδικοποιητή φωνής.

Στόχος της σχεδίασης των κωδικοποιητών φωνής είναι να παράγουν την υψηλότερη δυνατή ποιότητα φωνής με το χαμηλότερο δυνατό bit rate. Η πρόκληση για έναν κωδικοποιητή φωνής είναι να επιτύχει έναν ουσιαστικό ρυθμό δεδομένων φωνής ενόσω διατηρεί την διακρινόμενη ποιότητα του συστήματος υψηλής ροής δεδομένων. Το πλεονέκτημα των κωδικοποιητών φωνής είναι ότι δεν χρειάζεται να κωδικοποιήσουν οποιοδήποτε ηχητικό σήμα αλλά μόνο σήματα τα οποία προέρχονται από ακουστικά σήματα που παράγονται από τα ανθρώπινα φωνητικά όργανα, το οποίο αντιπροσωπεύει έναν κατανοήσιμο λόγο στην ανθρώπινη γλώσσα και το οποίο θα ληφθεί από έναν ακροατή. Καθένα από αυτά τα στοιχεία προσδίδει δομή στο σήμα και ο κωδικοποιητής μπορεί να χρησιμοποιήσει απόψεις του καθενός για να αφαιρέσει τον πλεονασμό και να μειώσει το bit rate. Ωστόσο χρησιμοποιώντας αυτά τα στοιχεία για τη μείωση του bit rate έχει και μειονεκτήματα. Όσο χαμηλότερο είναι το bit rate σε έναν υψηλής ποιότητας κωδικοποιητή φωνής, τόσο πιο αδύναμος είναι ο κωδικοποιητής σε σήματα που δεν είναι "καλά". Τέτοιοι κωδικοποιητές (χαμηλού bit rate) συχνά υποβαθμίζονται δραματικά για μη φωνητικά σήματα, για σήματα με θόρυβο και για πολλαπλούς ομιλητές. Επιπλέον τέτοια συστήματα τείνουν να είναι πιο αδύναμα όταν διαφθείρονται από καναλικά λάθη και ίσως να έχουν αρκετά μεγάλη καθυστέρηση στην επεξεργασία.

1.2.1. ΓΝΩΣΤΙΚΑ ΜΟΝΤΕΛΑ

Το αρχικό σημείο μιας έκφρασης φωνής είναι η ιδέα στον εγκέφαλο ενός ομιλητή. Το μυαλό χειρίζεται και οργανώνει τις ιδέες συνδυαστικά με έναν τρόπο ο οποίος δεν είναι απολύτως κατανοητός. Θα ήταν ιδανικό αν μια ιδέα μπορούσε με κάποιο τρόπο να μεταδίδεται απ' ευθείας σε έναν ακροατή χρησιμοποιώντας έναν μηχανισμό ο οποίος δεν χρησιμοποιεί καθόλου ομιλία.

1.2.2. ΓΛΩΣΣΙΚΑ ΜΟΝΤΕΛΑ

Η προφορική πρόταση έχει αισθητά περισσότερες πληροφορίες παρά στη γραπτή μορφή της. Η γραπτή πρόταση είναι ουσιαστικά ένας σωρός λέξεων και σημείων στίξεως και εξαρτάται ισχυρά από το περιεχόμενο και τη γνώση του αναγνώστη να αφαιρέσει την αμφιβολία στο νόημα της. Η προφορική πρόταση έχει λιγότερες αμφιβολίες διότι ο ομιλητής έχει γνώση του επιθυμητού νοήματος.

Παρόλο που τα στοιχεία που σχετίζονται με τη γλώσσα στους κωδικοποιητές φωνής αναπτύσσονται, δεν είναι ακόμη πολύ συνηθισμένα. Τα γλωσσικά μοντέλα έχουν χρησιμοποιηθεί επιτυχώς στα πολύ χαμηλού bit rate αναγνώρισης/σύνθεσης συστήματα στα οποία ο κωδικοποιητής φωνής είναι παρόμοιος με ένα σύστημα αναγνώρισης φωνής και ο αποκωδικοποιητής φωνής είναι όμοιος με ένα σύστημα σύνθεσης κατά κανόνα (synthesis-by-rule system). Ωστόσο τα γλωσσικά μοντέλα έχουν παίξει άλλο ένα σημαντικό ρόλο στο σχεδιασμό των σύγχρονων κωδικοποιητών φωνής. Τα γλωσσικά μοντέλα παρέχουν μια απόκλιση από τη διαδικασία επικοινωνίας το οποίο μας επιτρέπει τα συγκεκριμένα στοιχεία της εργασίας να περιγραφούν και αναλυθούν αποτελεσματικά. Μελετώντας τη σχέση μεταξύ των στοιχείων απόκλισης της γλώσσας και πως αυτά επενεργούν μεταξύ τους πάνω στην πραγματική κυματομορφή φωνής, οι ερευνητές κατάφεραν να συνθέσουν μια καθαρότερη εικόνα των χαρακτηριστικών της ακουστικής κυματομορφής τα οποία πρέπει να διατηρηθούν στη διαδικασία κωδικοποίησης φωνής.

1.2.3. ΜΟΝΤΕΛΑ ΦΩΝΗΤΙΚΟΥ ΣΥΣΤΗΜΑΤΟΣ

Το ανθρώπινο φωνητικό σύστημα είναι κυρίως ένας μη ομοιόμορφος σωλήνας του οποίου το σχήμα αλλάζει με το χρόνο. Ήχος παράγεται όταν ένα ακουστικής διέγερσης σήμα εγχέεται κάποια στιγμή στο φωνητικό σύστημα. Για φωνούμενους ήχους, όπως τα φωνήεντα, το σήμα διέγερσης είναι ένα ψευδο-περιοδικό σήμα το οποίο παράγεται από τις δονούμενες φωνητικές χορδές στη γλωττίδα. Για μη φωνούμενους ήχους το διεγερμένο σήμα είναι γενικά μια πηγή οξέος θορύβου που παράγεται από μια στροβιλιζόμενη ροή αέρα σε κάποιο στενό σημείο του φωνητικού συστήματος.

Σε μια πρώτη προσέγγιση, οι επιδράσεις του ακουστικού κύματος μέσα στο φωνητικό σύστημα κατά τη διάρκεια της δημιουργίας ενός ακουστικού σήματος είναι γραμμικές. Για το λόγο αυτό, είναι δυνατό να μοντελοποιήσουμε αποτελεσματικά το φωνητικό σύστημα σαν ένα αργό, χρονικά μεταβαλλόμενο ακουστικό φίλτρο το οποίο διεγείρεται από ένα ή περισσότερα σήματα διέγερσης. Περισσότερο σημαντικό είναι το γεγονός ότι το φωνητικό σύστημα είναι ένα μηχανικό σύστημα. Αυτό σημαίνει ότι οι κινήσεις του φωνητικού συστήματος είναι σχετικά αργές διότι παρεμποδίζονται από τη μάζα των φωνητικών μερών (γλώσσα, σαγόι, χείλη, δόντια). Αυτό καταλήγει σε ένα μοντέλο φωνητικού συστήματος όπου το γραμμικό φίλτρο διαφοροποιείται σχετικά αργά με το χρόνο. Αν το φίλτρο του φωνητικού συστήματος μπορεί να επιτευχθεί χρησιμοποιώντας ένα μοντέλο παραμετρικού φίλτρου, οι παράμετροι αυτοί θα αλλάζουν αργά με το χρόνο. Συνεπώς, τα θεμελιώδη bit rate που έχουν σχέση με τις παραμέτρους του φίλτρου θα είναι χαμηλότερα από ότι το ίδιο το φωνητικό σήμα.

Το σήμα διέγερσης έχει επίσης ιδιότητες οι οποίες μπορούν να αντιπροσωπευθούν παραμετρικά. Κατά τη διάρκεια των φωνητικών σημάτων, η αρχική διέγερση για το φίλτρο του φωνητικού συστήματος προέρχεται από δονήσεις της γλωττίδας. Η συχνότητα της δόνησης στη γλωττίδα ονομάζεται θεμελιώδης συχνότητα και την αντιλαμβανόμαστε σαν τον τόνο (pitch) της φωνής. Ο πρότυπος τόνος του φωνητικού σήματος είναι εύκολα

αντιλαμβανόμενος από τους ακροατές και φέρει μεγάλη ποσότητα πληροφορίας. Ωστόσο ο τόνος αυτός καθαυτός είναι σημαντικός και όχι οι λεπτομέρειες της δόνησης. Λόγω της φυσιολογίας της γλωττίδας, η θεμελιώδης συχνότητα δόνησης δεν διαφοροποιείται σημαντικά με το χρόνο. Έτσι, η διέγερση φωνής μπορεί συχνά να μοντελοποιηθεί επαρκώς χρησιμοποιώντας ένα αργά μεταβαλλόμενο μοντέλο της θεμελιώδους συχνότητας. Αυτό το pitch σήμα μπορεί να αναπαρασταθεί με πολύ χαμηλότερο bit rate από ότι το φωνητικό σήμα διέγερσης.

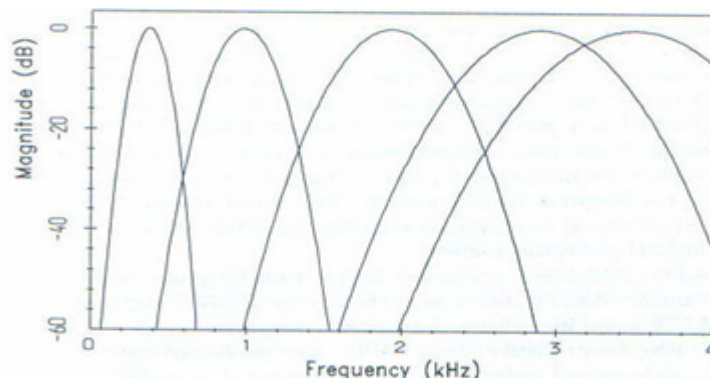
1.2.4. ΑΚΟΥΣΤΙΚΑ ΜΟΝΤΕΛΑ

Κοινό χαρακτηριστικό όλων των συστημάτων κωδικοποίησης φωνής είναι ότι παράγουν σήματα τα οποία αντιλαμβάνεται το ανθρώπινο αυτί. Αν υπάρχει καμιά πληροφορία στο αρχικό σήμα φωνής που φιλτράρεται έξω από το αυτί, τότε αυτή η πληροφορία μπορεί να μείνει έξω από την κωδικοποιημένη αναπαράσταση και το απαιτούμενο bit rate μειώνεται αναλόγως.

Το μοντέλο του αυτιού που έχει επιτυχώς χρησιμοποιηθεί από τα συστήματα κωδικοποίησης φωνής είναι το γνωστό κρίσιμης ζώνης (critical band) μοντέλο για ακουστική παρατήρηση. Σε αυτό το μοντέλο το αυτί αναπαρίσταται από συνεχόμενα γραμμικά φίλτρα μέσα στα οποία το εύρος ζώνης κάθε συγκεκριμένου φίλτρου διαφοροποιείται με τη συχνότητα. Ο πίνακας 1.1 παρουσιάζει μια λίστα από πειραματικά δεδομένα όσον αφορά τα φίλτρα αυτά. Το εύρος ζώνης και η κεντρική συχνότητα είναι μη ομοιόμορφα και δύσκολα ανταποκρίνονται σε μια διάταξη φίλτρων 1/6 της οκτάβας. Μια απλούστερη διάταξη φίλτρων 5 ζωνών παρουσιάζεται στο σχήμα 1.3.

Filter number	Center freq. (Hz)	Bandwidth (Hz)	Filter number	Center freq. (Hz)	Bandwidth (Hz)
1	50	70	13	1020	127
2	120	70	14	1148	140
3	190	70	15	1288	153
4	260	70	16	1442	168
5	330	70	17	1610	183
6	400	70	18	1794	199
7	470	70	19	1993	217
8	540	70	20	2221	235
9	617	86	21	2446	255
10	703	95	22	2701	276
11	798	105	23	2978	298
12	904	116	24	3276	321
			25	3497	346

ΠΙΝΑΚΑΣ 1.1: Κεντρικές συχνότητες και εύροι κρίσιμης ζώνης



ΣΧΗΜΑ 1.3: Φίλτρα κρίσιμης ζώνης

Το μοντέλο critical band προσπαθεί να αιχμαλωτίσει έναν αριθμό από σχετικές απόψεις της ακουστικής αντίληψης. Το πρώτο είναι η ακουστική ανάλυση συχνότητας (aural frequency resolution). Το εύρος ζώνης μιας κρίσιμης ζώνης συχνοτήτων σε συγκεκριμένη συχνότητα είναι η μέτρηση του πόσο μακριά, όσον αφορά τη συχνότητα, πρέπει να είναι δύο τόνοι, έτσι ώστε να ξεχωρίζει ο ένας από τον άλλο. Απλά η ικανότητα του ανθρώπινου αυτιού να αναλύει τόνους είναι δύσκολη αναλογικά με το εύρος των critical bands στην περιοχή των τόνων. Το δεύτερο είναι η ακουστική επικάλυψη θορύβου (aural noise-masking). Με απλά λόγια ένα φωνητικό σήμα σε μια συγκεκριμένη κρίσιμη ζώνη θα επικαλύψει ένα άλλο σήμα (θόρυβος) μέσα στην ίδια ζώνη. Αυτό το μοντέλο πρακτικά καθορίζει πως το φάσμα του θορύβου πρέπει να σχηματίζεται σαν λειτουργία στο φάσμα του ήχου έτσι ώστε να επιτύχουμε την ελάχιστη αντιληπτή σύγκρουση.

1.3. ΚΩΔΙΚΟΠΟΙΗΣΗ ΦΩΝΗΣ

1.3.1. ΓΕΝΙΚΑ ΓΙΑ ΤΗΝ ΚΩΔΙΚΟΠΟΙΗΣΗ ΦΩΝΗΣ

Γενικά η κωδικοποίηση φωνής μπορεί να θεωρηθεί ως μια συγκεκριμένη ειδικότητα στη γενική κατηγορία επεξεργασία φωνής, η οποία επίσης περιλαμβάνει ανάλυση και αναγνώριση φωνής. Ολόκληρο το πεδίο της επεξεργασίας φωνής βιώνει προσφάτως πρόοδο που έχουν επιφέρει οι τεχνικές και τα συστήματα της ψηφιακής επεξεργασίας σήματος (Digital Signal Processing – DSP). Αντιστρόφως είναι δίκαιο να πούμε ότι η επεξεργασία φωνής υπήρξε το λίκνο για τα DSP με την έννοια ότι οι περισσότεροι από τους ευρέως διαδεδομένους αλγόριθμους στο DSP αναπτύχθηκαν και πρωτοεμφανίστηκαν στην πράξη από ανθρώπους που δούλευαν στα συστήματα επεξεργασίας φωνής. Αυτοί οι αλγόριθμοι περιλαμβάνουν τεχνικές ψηφιακού φιλτραρίσματος, διαμόρφωση παλμού κώδικα (Pulse Code Modulation – PCM), κωδικοποίηση γραμμικής πρόβλεψης (Linear Predictive Coding – LPC), βραχυχρόνιο μετασχηματισμό Fourier (Short-time Fourier Transform – STFT), γενικές χρόνου-συχνότητας αναπαραστάσεις, προσαρμοστικές τεχνικές φιλτραρίσματος και πολλές ακόμη. Από όλα τα πεδία που αφορούν την επεξεργασία φωνής, η κωδικοποίηση φωνής είναι η πιο κατανοητή και πιο ώριμη. Υπάρχουν τρεις λόγοι που συμβαίνει αυτό:

1. Το γεγονός ότι η φωνή είναι ένα σήμα χαμηλού εύρους ζώνης
2. Η επανάσταση που έφερε η ταχεία ανάπτυξη της τεχνολογίας κυκλωμάτων VLSI (Very Large-Scale Integrated circuits).
3. Η αποτελεσματικότητα των διαφόρων αλγορίθμων DSP στη λύση πολλών κύριων προβλημάτων που σχετίζονται με τα συστήματα κωδικοποίησης φωνής.

1.3.2. ΧΡΗΣΙΜΟΤΗΤΑ ΤΗΣ ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΦΩΝΗΣ

Η κωδικοποίηση της φωνής αποτελεί μια ιδιαίτερα χρήσιμη διαδικασία αν αναλογιστούμε ότι μεγάλο μέρος της διακινούμενης πληροφορίας είναι φωνή και ότι μέσω της συμπίεσης της μπορεί να επιτευχθεί σημαντική μείωση του απαιτούμενου εύρους ζώνης (μέχρι και 4-5 φορές χωρίς αισθητή υποβάθμιση της ποιότητας). Στα κινητά τηλεπικοινωνιακά συστήματα, το διαθέσιμο εύρος ζώνης αποτελεί σημαντικό πόρο καθώς πραγματοποιείται προσπάθεια για εξυπηρέτηση ολοένα και περισσότερων χρηστών μέσα σε αυτό. Για το λόγο αυτό σύνθετοι αλγόριθμοι κωδικοποίησης φωνής χρησιμοποιούνται για να μειώσουν την κατοχή εύρους ζώνης ενός σήματος χωρίς υποβάθμιση της ποιότητας. Οι περισσότεροι από τους γνωστούς αλγορίθμους κωδικοποίησης φωνής είναι ρητά βασισμένοι σε ένα πρότυπο παραγωγής φωνής. Στους χαμηλούς και μέσους ρυθμούς bit (περίπου 0,5... 2 bit ανά δείγμα, δηλ. 4... 16 KBIT/S σε ένα ποσοστό δειγματοληψίας 8 kHz), οι ιδιότητες του ανθρώπινου αυτιού επίσης χρησιμοποιούνται. Είναι αναγκαία, λοιπόν, η κατασκευή κωδικοποιητών φωνής που θα συμπιέζουν τη φωνή σε περιορισμένο εύρος ζώνης με όσο το δυνατόν καλύτερη ποιότητα. Δεδομένου ότι η πολυπλοκότητα αυξάνεται, η καθυστέρηση φωνής αυξάνεται εξαιτίας του γεγονότος ότι παίρνει περισσότερο για να κωδικοποιήσει και να αποκωδικοποιήσει ένα τμήμα της ομιλίας. Οι καθυστερήσεις που είναι μεγαλύτερες από το 100msec μπορούν να γίνουν ενοχλητικές στο χρήστη.

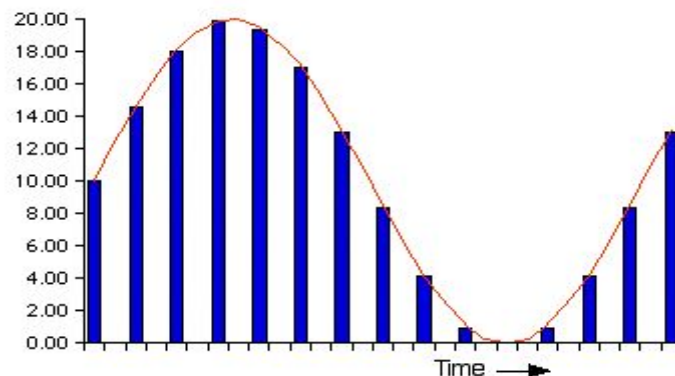
1.3.3. ΚΩΔΙΚΟΠΟΙΗΣΗ & ΚΒΑΝΤΙΣΗ

Η κωδικοποίηση είναι η διαδικασία αναπαράστασης των δειγματοληπτημένων τιμών με n bit ψηφιακές λέξεις. Το n επιλέγεται ως δύναμη του 2 ανάλογα με το επίπεδο ακρίβειας (ποιότητα) που απαιτείται. Αυξάνοντας το n θα μειωθεί το μέγεθος βημάτων μεταξύ των γειτονικών επιπέδων κβάντισης και αυτό θα οδηγήσει σε μια μείωση του θορύβου κβάντισης καθώς επίσης και μια αύξηση του όγκου των ψηφιακών στοιχείων.

Κβάντιση είναι η διαδικασία προσδιορισμού ενός από τα πιθανά διακριτά επίπεδα κάθε αποκτηθέντος δείγματος. Ο συνολικός αριθμός επιπέδων εξαρτάται από τον αριθμό των bit που χρησιμοποιούνται για να αντιπροσωπευθεί κάθε δείγμα. Η κβάντιση μπορεί να επιτευχθεί είτε με τη στρογγυλοποίηση (που ορίζει το πλησιέστερο επίπεδο στο σήμα που μπορεί να είναι μεγαλύτερο ή μικρότερο από το πραγματικό σήμα) είτε την περικοπή (που ορίζει το πλησιέστερο επίπεδο που είναι μικρότερο από το πραγματικό σήμα). Αυτή η διαδικασία δίνει μια κλιμακωτή κυματομορφή που μοιάζει με το σήμα της πηγής. Η διαφορά μεταξύ του πραγματικού δείγματος και της αξίας που του ορίζεται είναι γνωστό ως θόρυβος κβάντισης. Ο θόρυβος κβάντισης (e) μπορεί να πάρει τις τιμές που προσδιορίζονται από τον τύπο παρακάτω.

$$-\frac{\delta}{2} \leq e < \frac{\delta}{2}$$

Το παρακάτω σχήμα παρουσιάζει ένα αναλογικό σήμα (κόκκινη γραμμή) και δείγματα κώδικα παλμού (μπλε μπάρες).



Η ακρίβεια του δείγματος καθορίζεται από τον αριθμό των bit που χρησιμοποιούνται για να αναπαρασταθούν τα δειγματοληπτημένα δεδομένα. Για παράδειγμα αν χρησιμοποιήσουμε 7 bit θα έχουμε 128 πιθανές τιμές. Αφού το αναλογικό σήμα είναι συνεχόμενο κάθε τιμή PCM θα προσεγγίζει το αναλογικό σήμα. Παρακάτω παρουσιάζονται οι μέθοδοι που χρησιμοποιούμε:

Full 7-bit Sample: Κάθε δείγμα ερμηνεύεται με μια τιμή μεταξύ 0 και 127. Επτά bit ανά δείγμα σημαίνει ότι επτά bit απαιτούνται για να αντιπροσωπεύσουν αυτήν την τιμή. Εάν υποθέσουμε ότι το αναλογικό σήμα αλλάζει σχετικά αργά έναντι του ποσοστού δειγματοληψίας μπορούμε να μειώσουμε τα bit που μεταδίδονται. Αυτό είναι η βάση των άλλων δύο μεθόδων που παρουσιάζονται παρακάτω.

4-bit Differential: Αν το αναλογικό σήμα αλλάζει περισσότερο από όσο μπορεί να προσαρμοστεί, η λογική κωδικοποίησης μπορεί να απαιτήσει αρκετές περιόδους δειγματοληψίας για "να προφθάσει" την αλλαγή.

1-bit Delta Modulation (Διαμόρφωση Δέλτα): Η διαμόρφωση δέλτα είναι η έκδοση DPCM για ένα bit. Εάν ένα σήμα δειγματοληπτηθεί σε ένα ποσοστό πολύ υψηλότερο από το ποσοστό Nyquist ο συσχετισμός των παρακείμενων δειγμάτων αυξάνεται. Αυτό απλοποιεί την κβάντιση. Ένα μόνο bit μεταδίδεται δείχνοντας εάν το σήμα έχει αυξηθεί ή μειωθεί. Το δείγμα πρέπει να αλλάξει από την προηγούμενη τιμή και μπορεί να αλλάξει το πολύ κατά ένα επίπεδο.

1.3.4. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΟΣ ΚΩΔΙΚΟΠΟΙΗΣΗΣ

α) Ρυθμός μετάδοσης: Η ψηφιακή κωδικοποίηση χαρακτηρίζεται από απώλεια πληροφορίας (lossy compression). Στόχος είναι η επίτευξη της ελάχιστης απώλειας για ένα δοσμένο ρυθμό μετάδοσης.

β) Ποιότητα φωνής: Εκτιμάται με βάση τη φυσικότητα και την ευκολία κατανόησης (intelligibility) της κωδικοποιημένης φωνής. Υπάρχουν αντικειμενικά και υποκειμενικά μέτρα παραμόρφωσης.

➤ Αντικειμενικά μέτρα παραμόρφωσης (Objective distortion measures): Αυτά πρέπει:

1. Να αντιστοιχούν σε ψυχολογικά χαρακτηριστικά της ανθρώπινης αντίληψης
2. Να υπολογίζονται εύκολα και αποτελεσματικά
3. $d(x_1, x_2) \geq 0$

➤ Υποκειμενικά μέτρα παραμόρφωσης (Subjective distortion measures):

Mean-opinion score (MOS): Η ποιότητα φωνής αξιολογείται συνήθως σε μια κλίμακα πέντε επιπέδων, γνωστή ως κλίμακα MOS, η οποία προέρχεται από τον μέσο όρο των δεδομένων φωνής, των ομιλητών και των ακροατών. Τα πέντε επίπεδα της ποιότητας είναι: *Κακό (bad)*, *ανεπαρκές (poor)*, *μέτριο (fair)*, *καλό (good)* και *άριστο (excellent)*. Η ποιότητα θεωρείται ικανοποιητική για βαθμό 3,5 ή υψηλότερο που γενικά υπονοεί υψηλά επίπεδα σαφήνειας, αναγνώρισης ομιλητών και φυσικότητας (Βαθμολογία: 1-Bad, 2-Poor, 3-Fair, 4-Good, 5-Excellent).

γ) Καθυστερήση κωδικοποιητή (Encoding Delay): Δημιουργείται από την πολυπλοκότητα του κωδικοποιητή και προστίθεται στις καθυστερήσεις μετάδοσης (π.χ. δορυφορικά κανάλια) και την κωδικοποίηση πηγής. Προξενεί σημαντικά προβλήματα στην επικοινωνία καθώς:

1. Μπορεί να δημιουργηθεί ηχώ (σε μετασχηματιστές δισύρματων σε τετρασυρμάτινες ζεύξεις)

2. Παρεμποδίζει την επικοινωνία

Επιτρεπτές τιμές: $\leq 1\text{ms}$ στην τηλεφωνία, $\leq 500\text{ms}$ video telephony, αυθαίρετη για αποθήκευση (π.χ. για voice mail)

δ) Ευαισθησία στα σφάλματα μετάδοσης:

1. Μπορεί να έχουμε bursts (mobile communications)

2. Οι κωδικοποιητές αφαιρούν περιττή πληροφορία σε μικρότερους ρυθμούς μετάδοσης και συνήθως

αυξάνεται η ευαισθησία στα σφάλματα μετάδοσης

3. Υποκειμενική εκτίμηση των σφαλμάτων μετάδοσης.

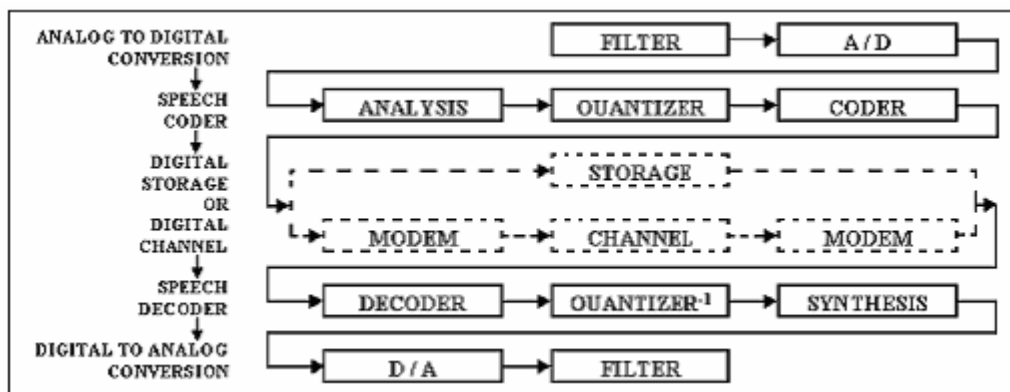
ε) Πολυπλοκότητα: Η πολυπλοκότητα ενός αλγορίθμου κωδικοποίησης είναι η προσπάθεια επεξεργασίας που απαιτείται για να αναπαρασταθεί ο αλγόριθμος και τυπικά μετριέται σε σχέση με την αριθμητική δυνατότητα και τις απαιτήσεις μνήμης ή ισοδύναμα από την άποψη του κόστους. Μια μεγάλη πολυπλοκότητα μπορεί να οδηγήσει στη μεγάλη κατανάλωση ενέργειας στο hardware.

1.4. ΚΩΔΙΚΟΠΟΙΗΤΕΣ ΦΩΝΗΣ

1.4.1. ΟΡΙΣΜΟΣ ΚΩΔΙΚΟΠΟΙΗΤΗ ΦΩΝΗΣ

Όλα τα αρχικά στοιχεία ενός ψηφιακού συστήματος κωδικοποίησης φωνής παρουσιάζονται στο σχήμα 3. Στο αριστερό μέρος του σχήματος είναι μια λίστα των υποσυστημάτων τα οποία εξηγούνται αναλυτικά στα κουτιά του σχήματος. Η είσοδος του συστήματος είναι μια συνεχής κυματομορφή ήχου $s(t)$. Αυτό το σήμα φιλτράρεται σε χαμηλές συχνότητες χρησιμοποιώντας ένα anti-aliasing φίλτρο και δειγματοληπείται από ένα μετατροπέα αναλογικού σε ψηφιακό (A/D converter) δίνοντας το ψηφιακό σήμα φωνής $s(n)$. Αυτή είναι η είσοδος του κωδικοποιητή.

Οι κωδικοποιητές φωνής γενικά αποτελούνται από τρία μέρη: Ανάλυση φωνής, κβάντιση παραμέτρων και κωδικοποίηση παραμέτρων. Η είσοδος για το επίπεδο της ανάλυσης είναι το ψηφιακό σήμα φωνής, ενώ η έξοδος είναι η νέα αναπαράσταση του σήματος φωνής η οποία θα κβαντιστεί και θα κωδικοποιηθεί. Η έξοδος για το επίπεδο της ανάλυσης μπορεί να διαφέρει δραματικά αναλόγως από το πώς το σήμα φωνής είναι μοντελοποιημένο. Μετά την ανάλυση αυτοί οι παράμετροι είναι μη κβαντισμένοι και αναπαριστούν το σήμα όσο καλύτερα γίνεται όσον αφορά την ανάλυση.



ΣΧΗΜΑ 1.4: Στοιχεία συστήματος κωδικοποίησης φωνής

Μετά την ανάλυση οι παράμετροι πρέπει να κβαντιστούν για να μειωθεί ο αριθμός των bit που χρειάζονται. Με άλλα λόγια η κβάντιση χρησιμοποιείται για να μειωθεί η φυσική ροή πληροφορίας του σήματος φωνής που αναπαραστήθηκε. Η έξοδος του κβαντιστή παρέχεται στον κωδικοποιητή ο οποίος αντιστοιχεί ένα δυαδικό κώδικα σε κάθε πιθανή κβαντισμένη αναπαράσταση. Αυτοί οι δυαδικοί κώδικες τοποθετούνται μαζί για αποδοτικότερη μετάδοση ή αποθήκευση.

Η ψηφιακά κωδικοποιημένη φωνή χρησιμοποιείται συχνά στις εφαρμογές επικοινωνιών και στις εφαρμογές αποθήκευσης και αναπαραγωγής (store-and-playback).

Οι εφαρμογές αποθήκευσης φωνής συνήθως δεν έχουν περιορισμούς στην καθυστέρηση κωδικοποίησης και συχνά ο κωδικοποιητής δεν χρειάζεται καν να γίνει αντιληπτός σε πραγματικό χρόνο. Συνήθως δεν αντιμετωπίζουν λάθη όσον αφορά τα bit. Έτσι η εφαρμογές αυτές συχνά έχουν υψηλότερη ποιότητα, για το ίδιο bit rate, όταν συγκρίνονται με εφαρμογές μετάδοσης φωνής αλλά μεγαλύτερους περιορισμούς όσον αφορά το κόστος.

Ο αποκωδικοποιητής φωνής λειτουργεί αντίστροφα από τον κωδικοποιητή. Αφότου αποκωδικοποιηθεί η ψηφιακή ροή bit, μετατρέπεται σε κβαντισμένες εκδοχές των παραμέτρων φωνής μέσω του αντίστροφου κβαντιστή. Αυτές οι παράμετροι είναι ταυτόσημες με τις παραμέτρους στην έξοδο του επιπέδου κβάντισης του κωδικοποιητή, λόγω του ότι δεν υπάρχουν λάθη bit. Αυτές οι παράμετροι χρησιμοποιούνται έπειτα για να συντεθεί το κωδικοποιημένο σήμα φωνής $s[n]$.

1.4.2. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΚΩΔΙΚΟΠΟΙΗΤΩΝ ΦΩΝΗΣ

Όπως αναφέραμε παραπάνω οι κωδικοποιητές φωνής κατασκευάστηκαν με σκοπό τη συμπίεση φωνής σε περιορισμένο εύρος ζώνης με όσο δυνατόν καλύτερη ποιότητα. Επιθυμητά χαρακτηριστικά που πρέπει να έχει ένας κωδικοποιητής φωνής είναι:

- Σημαντική συμπίεση δεδομένων
- Ικανοποιητικό ρυθμό συμπίεσης
- Υψηλή ποιότητα φωνής
- Χρησιμοποίηση όσο το δυνατό μικρότερου εύρους ζώνης από το κανάλι
- Χαμηλή κατανάλωση ενέργειας
- Μικρή καθυστέρηση επικοινωνίας
- Χαμηλό κόστος

Κατά την υλοποίηση ενός κωδικοποιητή, η προσπάθεια αύξησης της δυνατότητας του για μεγαλύτερους ρυθμούς δεδομένων οδηγεί στην αύξηση της πολυπλοκότητας του. Όσο πιο πολύπλοκος όμως είναι ο κωδικοποιητής, τόσο μεγαλώνει η καθυστέρηση επεξεργασίας της φωνής καθώς και το κόστος της υλοποίησης. Για μια συγκεκριμένη υλοποίηση θα πρέπει ανάλογα με τις εκάστοτε ανάγκες μας να ισορροπήσουμε ή να δώσουμε βάρος σε κάποιον ή κάποιους από τους παραπάνω συγκρουόμενους παράγοντες.

1.5. ΠΡΟΤΥΠΑ ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΦΩΝΗΣ

Τα πρότυπα κωδικοποίησης φωνής παίζουν σημαντικό ρόλο στην ανάπτυξη και χρήση των κωδικοποιητών φωνής. Υπάρχουν μερικές εφαρμογές κωδικοποίησης φωνής στις οποίες η διαλειτουργικότητα δεν είναι θέμα. Για τέτοιου είδους εφαρμογές ο κωδικοποιητής φωνής της επιλογής μας μπορεί να είναι ο καλύτερος και ο πιο οικονομικός που είναι διαθέσιμος τη στιγμή που σχεδιάζεται το σύστημα, χωρίς να θεωρείται σημαντική η διαλειτουργικότητα. Για την πλειοψηφία των εφαρμογών, ωστόσο, η διαλειτουργικότητα είναι σημαντικό θέμα. Όλες οι εφαρμογές τηλεπικοινωνιών ανήκουν καθαρά σε αυτή την κατηγορία όπως επίσης και εφαρμογές μεταφοράς μέσου (carry-away media) όπως είναι τα CD. Για να επιτευχθεί η διαλειτουργικότητα, τα πρότυπα πρέπει να οριστούν και να υλοποιηθούν.

Τα πρότυπα αναπτύχθηκαν από διάφορους οργανισμούς. Ίσως το απλούστερο πρότυπο που πρέπει να αναφερθεί, αν και δεν είναι συγκεκριμένο πρότυπο κωδικοποίησης φωνής, είναι το format που χρησιμοποιείται στα CD. Τέτοια συστήματα χρησιμοποιούν 16 bit γραμμική κωδικοποίηση PCM. Αυτό το format δεν χρησιμοποιεί χαρακτηριστικά που είναι συγκεκριμένα στη φωνή και έτσι είναι κατάλληλο για την κωδικοποίηση φωνούμενων και μη σημάτων.

Πολλά από τα πρώτα πρότυπα κωδικοποίησης φωνής δημιουργήθηκαν από το υπουργείο άμυνας (Department of Defense - DoD). Αυτό έγινε αρχικά διότι οι ψηφιακές τεχνικές κρυπτογράφησης είναι πιο αποτελεσματικές από τις αναλογικές. Ένα από τα πρώτα πρότυπα που χρησιμοποιείται ακόμα είναι μια μορφή προσαρμοστικής διαμόρφωσης δέλτα που ονομάζεται διαμόρφωση συνεχούς μεταβλητής κλίσης δέλτα (Continuously Variable Slope Delta Modulation-CVSD), το οποίο μπορεί να χρησιμοποιηθεί στα 16 ή στα 32 kbps.

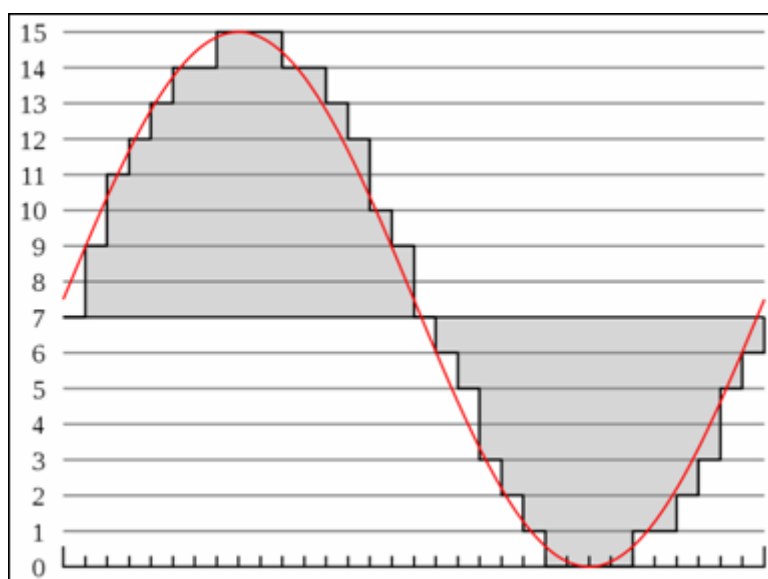
Δύο άλλα DoD πρότυπα κωδικοποίησης φωνής είναι τα αμερικάνικα ομοσπονδιακά πρότυπα 1015 1016. Το πρώτο από αυτά αναφέρεται συχνά ως LPC10e το οποίο είναι ένας pitch excited γραμμικής πρόβλεψης κωδικοποιητής ο οποίος λειτουργεί στα 2.4 kbps. Το δεύτερο και πιο πρόσφατο DoD πρότυπο είναι μιας μορφής code-excited γραμμικής πρόβλεψης κωδικοποιητής (CELP) ο οποίος λειτουργεί στα 4.8 kbps.

Μια άλλη σημαντική σύσταση προτύπων είναι η Διεθνής Ένωση Τηλεπικοινωνιών (International Telecommunications Union - ITU) η οποία είναι διάδοχος της Διεθνούς Τηλεφωνικής και Τηλεγραφικής Συμβουλευτικής Επιτροπής (International Telephone and Telegraph Consultative Committee - CCITT). Η επιτροπή αυτή προσδιορίζει πρότυπα για το παγκόσμιο δίκτυο τηλεφωνίας. Υπάρχουν κάποια πρότυπα ITU που χρησιμοποιούνται ακόμη. Τα πιο ευρέως χρησιμοποιούμενα πρότυπα είναι οι compressed PCM κωδικοποιητές στα 64 kbps τα οποία συναντώνται σε εφαρμογές ψηφιακών διακοπών. Περιλαμβάνουν το μ-law compressed PCM για τη Βόρεια Αμερική και A-law compressed PCM για την Ευρώπη (ITU G.711). Άλλα σημαντικά πρότυπα τηλεφωνίας στα 3.4 kHz σήματα φωνής περιλαμβάνουν ένα ADPCM (G.726) που λειτουργεί στα 16, 24, 32, 40 kbps και ένα χαμηλής καθυστέρησης CELP (G.728) που λειτουργεί στα 16 kbps. Υπάρχουν επίσης ακουστικά πρότυπα όπως το ο κωδικοποιητής two-band subband (G.722) ο οποίος λειτουργεί πάνω σε σήματα των 7 kHz σε bit rate των 48, 56 και 64 kbps.

ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ PCM ΚΑΙ ARPCM ΚΒΑΝΤΙΣΗ

2.1. ΕΙΣΑΓΩΓΗ

Η απλούστερη μορφή κωδικοποιητή φωνής είναι αυτή η οποία δειγματοληπτεί απ'ευθείας το σήμα φωνής και έπειτα κβαντίζει και κωδικοποιεί κάθε ξεχωριστό δείγμα. Τέτοια συστήματα δεν εκμεταλλεύονται το γεγονός ότι κωδικοποιούν σήματα φωνής και η λειτουργία τους γενικά δεν περιορίζεται σε τέτοια σήματα. Τέτοιοι κωδικοποιητές έχουν ωστόσο τη δυνατότητα να κάνουν κάποια μικρή χρήση των φωνητικών οργάνων και των ακουστικών μοντέλων για να επιτύχουν καλύτερη απόδοση.

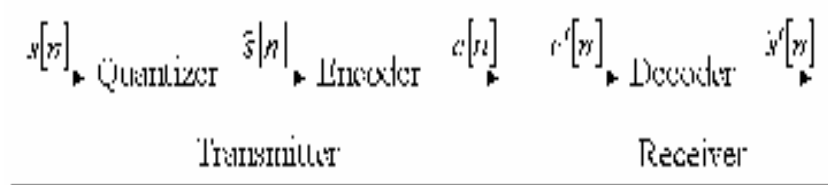


ΣΧΗΜΑ 2.1: Σχηματική απεικόνιση ενός PCM σήματος

Γενικά ένα ψηφιακό σήμα φωνής λαμβάνεται δειγματοληπώντας και κβαντίζοντας μία low-pass-filtered εκδοχή ενός συνεχούς σήματος φωνής. Ιδανική δειγματοληψία ενός περιορισμένης ζώνης σήματος είναι η διαδικασία κατά την οποία δε χάνεται πληροφορία. Η ψηφιακή φωνή λαμβάνεται κβαντίζοντας και κωδικοποιώντας το δειγματοληπτημένο σήμα φωνής, αναπαριστώντας έτσι κάθε δείγμα με ακρίβεια. Η κβάντιση είναι πάντα μια διαδικασία με απώλειες και παραμορφώσεις. Η κβαντισμένη φωνή μπορεί γενικά να μοντελοποιηθεί σαν ένα τέλειο, άπειρης ανάλυσης δειγματοληπτημένο σήμα έχοντας επιπλέον ένα λάθος κβάντισης. Το κβαντισμένο σήμα πάντα παραμορφώνεται ως κάποιο βαθμό και αυτή η παραμόρφωση αν είναι αισθητή, οπωσδήποτε δεν είναι επιθυμητή στις εφαρμογές επικοινωνιών. Στόχος κατά τη σχεδίαση των κβαντιστών είναι η σχεδίαση κβαντιστών οι οποίοι καταλήγουν στην ελάχιστη δυνατή παρατηρούμενη παραμόρφωση στην κωδικοποίηση φωνής στον επιθυμητό ρυθμό δεδομένων.

Οι κβαντιστές διακρίνονται σε ομοιόμορφους (uniform) ή μη ομοιόμορφους (non-uniform) και σε σταθερούς (fixed) ή προσαρμοστικούς (adaptive). Οι uniform και οι fixed κβαντιστές είναι οι απλούστεροι τύποι κβαντιστών. Οι non-uniform και adaptive κβαντιστές είναι πιο πολύπλοκοι αλλά η πρόσθετη πολυπλοκότητα συχνά αξίζει τον κόπο γιατί μπορεί να χρησιμοποιηθεί για να μειωθούν οι αντιληπτικές επιδράσεις κατά την κβάντιση. Ένας uniform κβαντιστής ακολουθούμενος από έναν κωδικοποιητή ο οποίος αντιστοιχίζει έναν κώδικα (συνήθως δυαδικό) σε κάθε κβαντισμένο επίπεδο, ονομάζεται γραμμικός κωδικοποιητής παλμού κώδικα (linear pulse code modulator-linear PCM). Σε ένα linear PCM ο ρυθμός μετάδοσης των bit είναι $I=f_s B$ bits per second, όπου b είναι ο αριθμός bits ανά δείγμα και f_s ο ρυθμός δειγματοληψίας. Για παράδειγμα για ένα ρυθμό δειγματοληψίας 8 kHz και 16 bit ανά δείγμα, το bit rate είναι $8 \cdot 16 = 128 \text{ kbps}$. Στο σχήμα 2.2 φαίνεται ένα block διάγραμμα ενός συστήματος PCM. Το αρχικό σήμα εισόδου $s[n]$ κβαντίζεται

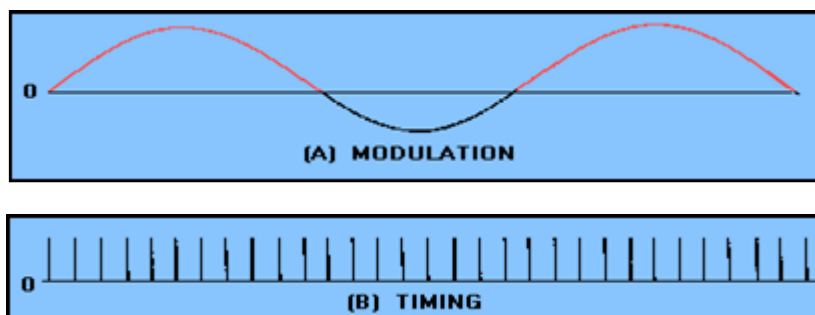
ομοιόμορφα και το κβαντισμένο σήμα $\hat{s}[n]$ κωδικοποιείται. Ο κωδικοποιητής αντιστοιχεί ένα δυαδικό κώδικα $c[n]$ σε κάθε επίπεδο κβάντισης. Το κωδικοποιημένο σήμα μεταδίδεται και ο λαμβανόμενος κώδικας $c'[n]$ αποκωδικοποιείται στο δέκτη για να πετύχει το σήμα $\hat{s}'[n]$. Υποθέτοντας ένα τέλειο κανάλι μετάδοσης τότε $c'[n]=c[n]$ και το κβαντισμένο σήμα $s(n)$ μπορεί να αναδημιουργηθεί επακριβώς στο δέκτη. Δηλαδή $\hat{s}'[n]=s(n)$.

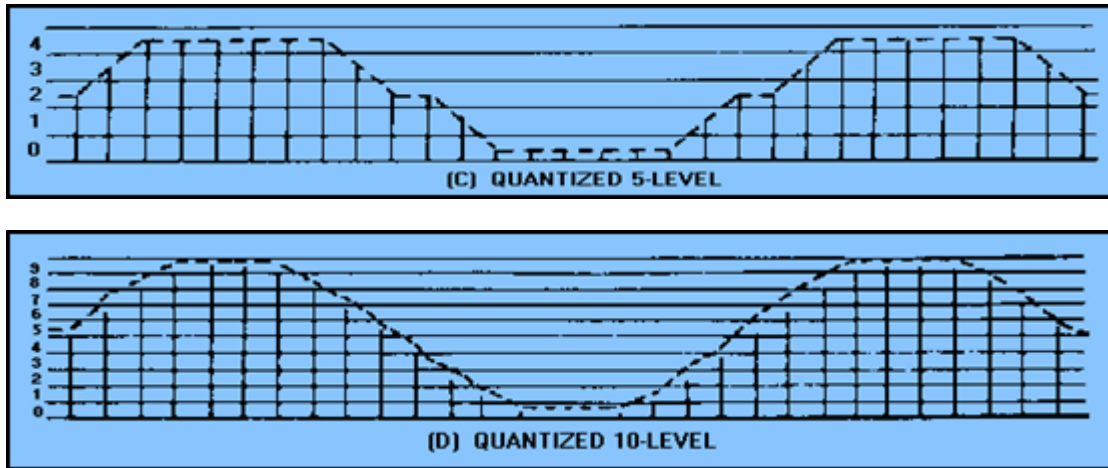


ΣΧΗΜΑ 2.2: Block διάγραμμα ενός PCM

2.1.1. ΣΥΝΤΟΜΗ ΑΝΑΛΥΣΗ PCM

Η διαμόρφωση κώδικα παλμού (PCM) αναφέρεται σε ένα σύστημα στο οποίο οι τυποποιημένες τιμές ενός κβαντισμένου σήματος υποδεικνύονται από μια σειρά κωδικοποιημένων παλμών. Όταν αυτοί οι παλμοί αποκωδικοποιούνται, δείχνουν τις τυποποιημένες τιμές του αρχικού κβαντισμένου σήματος. Αυτοί οι κώδικες μπορούν να είναι δυαδικοί, στους οποίους το σύμβολο για κάθε κβαντισμένο στοιχείο συνίσταται σε παλμούς και διαστήματα (δυαδικοί "0", "1"), τριαδικοί, όπου ο κώδικας για κάθε στοιχείο αποτελείται από οποιοδήποτε από τρία ευδιάκριτα είδη τιμών (όπως οι θετικοί παλμοί, οι αρνητικοί παλμοί, και τα διαστήματα) ή n-καταστάσεων, στην οποία ο κώδικας για κάθε στοιχείο αποτελείται από τον αριθμό άρνησης (n) ευδιάκριτων τιμών. Όλα τα συστήματα PCM που αναφέρθηκαν προηγουμένως παρέχουν τις μεθόδους μετατροπής αναλογικού σήματος σε ψηφιακό (παλμοί που εμφανίζονται σε διακριτά διαστήματα, κάποια χαρακτηριστικά των οποίων ποικίλουν ως συνεχής συνάρτηση του αναλογικού σήματος). Ολόκληρο το εύρος τιμών (συχνότητα ή φάση) του αναλογικού σήματος μπορεί να διαιρεθεί αυθαίρετα σε σειρά τυποποιημένων τιμών. Κάθε παλμός μιας ακολουθίας παλμών [σχήμα 2.3, άποψη (B)] παίρνει την τυποποιημένη τιμή που πλησιάζει την πραγματική κατά τη διαμόρφωση. Το διαμορφωμένο σήμα μπορεί να αναπαραχθεί πιστά, όπως φαίνεται στις απόψεις (C) και (D). Το εύρος πλάτους έχει διαιρεθεί σε 5 τυποποιημένες τιμές κατά την άποψη (C). Σε κάθε παλμό δίνεται οποιαδήποτε τυποποιημένη τιμή είναι πλησιέστερη στην πραγματική στιγμιαία τιμή της. Κατά την άποψη (D), το ίδιο εύρος πλάτους έχει διαιρεθεί σε 10 τυποποιημένα επίπεδα. Η καμπύλη της άποψης (D) (10 επίπεδα κβάντισης) είναι μια πολύ πιο στενή προσέγγιση της καμπύλης της άποψης (A), από ότι είναι η καμπύλη της άποψης (C) (5 επίπεδα κβάντισης). Από αυτό καταλαβαίνουμε ότι όσο μεγαλύτερος ο αριθμός χρησιμοποιούμενων τυποποιημένων επιπέδων τόσο περισσότερο το κβαντισμένο σήμα προσεγγίζει το αρχικό. Αυτό γίνεται επίσης εμφανές από το γεγονός ότι ένας άπειρος αριθμός τυποποιημένων επιπέδων αναπαράγει ακριβώς το αρχικό μη κβαντισμένο σήμα (την αρχική αναλογική κυματομορφή).





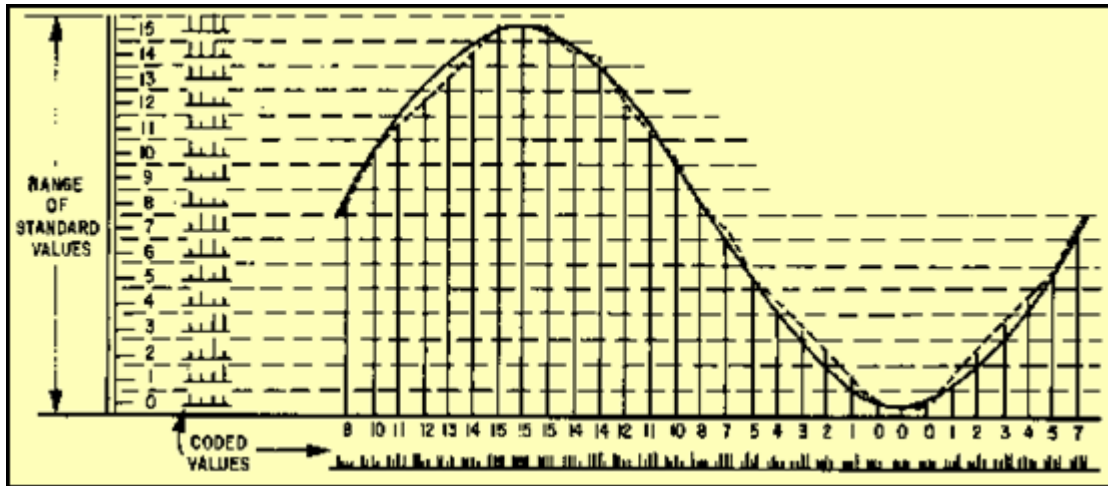
Σχήμα 2.3: Επίπεδα κβάντισης του αναλογικού σήματος.

Αν και οι κβαντισμένες καμπύλες του σχήματος 2.3 είναι βασισμένες σε 5 και 10 επίπεδα κβάντισης, στην πραγματικότητα τα επίπεδα καθιερώνονται συνήθως με κάποια εκθετική αξία σε δύναμη του 2, όπως $4(2^2)$, $8(2^3)$, $16(2^4)$, $32(2^5)$... $N(2^n)$. Η κβάντιση FM είναι όμοια με την κβάντιση AM. Δηλαδή το εύρος της απόκλισης συχνότητας διαιρείται σε πεπερασμένο αριθμό τυποποιημένων τιμών. Κάθε παλμός δειγματοληψίας οδηγεί σε μια απόκλιση ίση με την τυποποιημένη τιμή που είναι πλησιέστερη στην πραγματική απόκλιση της συγκεκριμένης στιγμής δειγματοληψίας. Ομοίως, για τη διαμόρφωση φάσης (PM), η κβάντιση καθορίζει ένα σύνολο τυποποιημένων τιμών. Η κβάντιση χρησιμοποιείται συνήθως στα "εύρος - συχνότητα διαμορφωμένα" συστήματα παλμού.

Το σχήμα 2.4 παρουσιάζει τη σχέση μεταξύ των δεκαδικών αριθμών, των δυαδικών αριθμών, και ενός κώδικα σήματος-παλμού που αντιπροσωπεύει τους αριθμούς. Ο πίνακας είναι για έναν 16 επιπέδων κώδικα δηλαδή 16 τυποποιημένες τιμές ενός κβαντισμένου σήματος θα μπορούσαν να αντιπροσωπευθούν από αυτές τις ομάδες παλμών. Μόνο η παρουσία ή η απουσία των παλμών είναι σημαντική. Η επόμενη βαθμίδα θα ήταν ένας 32 επιπέδων κώδικας, με κάθε δεκαδικό αριθμό να αντιπροσωπεύεται από μια σειρά από πέντε δυαδικά ψηφία, παρά τα τέσσερα ψηφία του σχήματος 2.4. Έξι ομάδες ψηφίων θα παρείχαν έναν 64 επιπέδων κώδικα, επτά ψηφία έναν κώδικα 128 επιπέδων, και ούτω καθ'εξής. Το σχήμα 2.5 παρουσιάζει την εφαρμογή των ομάδων των παλμών στις τυποποιημένες τιμές ενός κβαντισμένου σήματος.

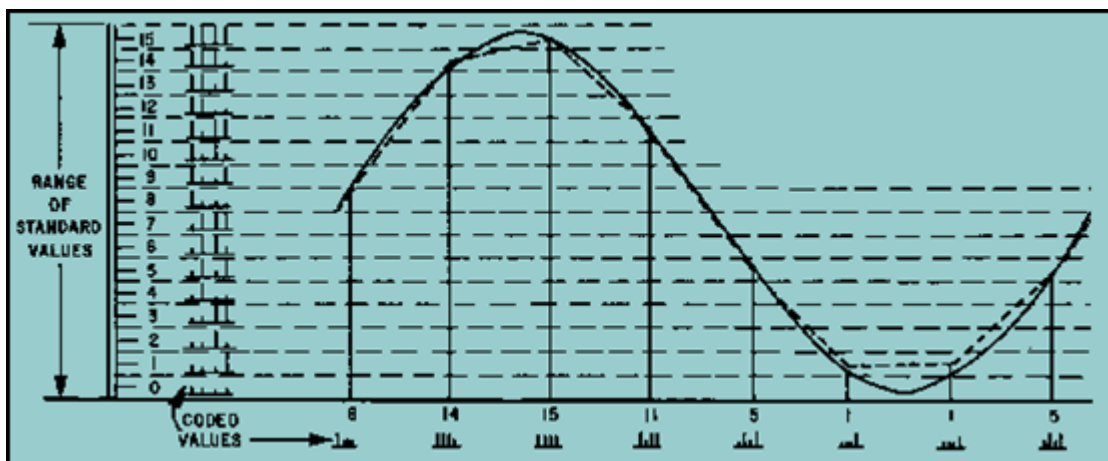
DECIMAL NUMBER	BINARY EQUIVALENT				PULSE-CODE WAVEFORMS			
	2^3	2^2	2^1	2^0	2^3	2^2	2^1	2^0
0	0	0	0	0				
1	0	0	0	1				↑
2	0	0	1	0			↑	
3	0	0	1	1			↑	↑
4	0	1	0	0		↑		
5	0	1	0	1		↑		↑
6	0	1	1	0		↑	↑	
7	0	1	1	1		↑	↑	↑
8	1	0	0	0	↑			
9	1	0	0	1	↑			↑
10	1	0	1	0	↑		↑	
11	1	0	1	1	↑		↑	↑
12	1	1	0	0	↑	↑		
13	1	1	0	1	↑	↑		↑
14	1	1	1	0	↑	↑	↑	
15	1	1	1	1	↑	↑	↑	↑

Σχήμα 2.4: Δυαδικοί αριθμοί και αντίτιμα παλμού-κώδικα.



Σχήμα 2.5: Εφαρμογή των ομάδων παλμών στο κβαντισμένο σήμα.

Στο σχήμα 2.5 η σταθερή καμπύλη αντιπροσωπεύει τις μη κβαντισμένες τιμές μιας διαμορφωμένης ημιτονοειδούς καμπύλης. Η διακεκομμένη καμπύλη που αναδημιουργείται από τις κβαντισμένες τιμές που λαμβάνονται στο διάστημα δειγματοληψίας παρουσιάζει μια καλή προσέγγιση της αρχικής. Το σχήμα 2.6 είναι πανομοιότυπο με το σχήμα 2.5 εκτός από το ότι το διάστημα δειγματοληψίας είναι τέσσερις φορές μεγαλύτερο και η αναδημιουργημένη καμπύλη δεν είναι πιστή στην αρχική. Όπως αναφέρθηκε νωρίτερα, το ποσοστό δειγματοληψίας ενός συστήματος παλμών πρέπει να είναι τουλάχιστον δύο φορές η μέγιστη διαμορφωμένη συχνότητα για να πάρει μια εφαρμόσιμη αναδημιουργημένη καμπύλη διαμόρφωσης. Στο ποσοστό δειγματοληψίας του σχήματος 2.5 και με έναν δυαδικό κώδικα 4 στοιχείων, 128 bit (παρουσία ή απουσία παλμών) πρέπει να μεταδοθούν για κάθε κύκλο της διαμορφωμένης συχνότητας. Στο ποσοστό δειγματοληψίας του σχήματος 2.6 απαιτούνται μόνο 32 bit, ενώ στο ελάχιστο ποσοστό δειγματοληψίας μόνο 8 bit.

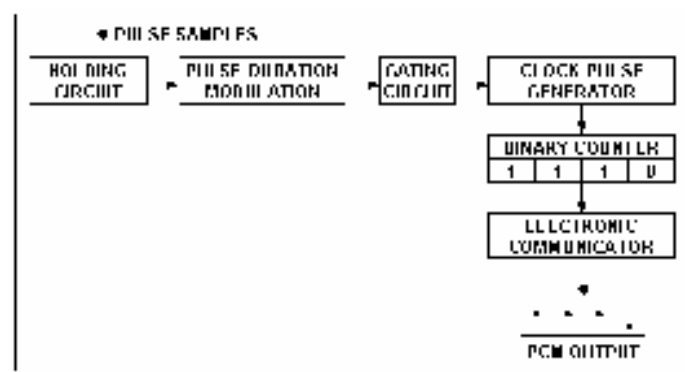


Σχήμα 2.6: Σήμα με τέσσερις φορές μεγαλύτερη περίοδο.

Για την απλούστευση των πραγμάτων, ειδικά για να απλοποιηθεί η αποδιαμόρφωση του PCM, οι ακολουθίες παλμών που πραγματικά μεταδίδονται είναι αντίστροφοι από εκείνους που παρουσιάζονται στα σχήματα 2.4, 2.5 και 2.6 δηλαδή ο παλμός με τη χαμηλότερη δυαδική αξία (λιγότερο σημαντικό ψηφίο - less significant bit) μεταδίδεται πρώτα και όσο περνούν οι παλμοί πηγαίνουμε σε περισσότερο σημαντικά ψηφία έως ότου φτάσουμε στο όριο του κώδικα το οποίο είναι το πιο σημαντικό ψηφίο (most significant bit). Η κωδικοποίηση παλμού μπορεί να εκτελεστεί με διάφορους τρόπους χρησιμοποιώντας συμβατικά στοιχεία κυκλώματος ή με τη κωδικοποίηση που συντελείται με τη βοήθεια ακτινών ειδικών καθοδικών σωλήνων.

Μια μορφή κυκλώματος κωδικοποίησης παρουσιάζεται στο σχήμα 2.7. Σε αυτήν την περίπτωση, τα δείγματα παλμού εφαρμόζονται σε ένα κύκλωμα αποθήκευσης (ένας

πυκνωτής που αποθηκεύει τις πληροφορίες εύρους παλμού) και ο διαμορφωτής μετατρέπει PAM σε PDM. Οι παλμοί PDM χρησιμοποιούνται έπειτα για να διέρχονται μέσω πυλών στην έξοδο μιας γεννήτριας παλμών ακριβείας που ελέγχει τον αριθμό παλμών που εφαρμόζονται σε έναν δυαδικό μετρητή. Η διάρκεια των παλμών των πυλών δεν είναι απαραίτητα ένας ολοκληρωμένος αριθμός των επαναληπτικών παλμών από την γεννήτρια παλμών ρολογιού. Επομένως, οι παλμοί ρολογιού που διέρχονται στο δυαδικό μετρητή από το PDM μπορούν να είναι διάφοροι παλμοί συν την αιχμή ενός πρόσθετου παλμού. Αυτός ο "μερικός" παλμός είτε μπορεί να έχει την επαρκή διάρκεια για να σκανδαλίσει το μετρητή, είτε όχι. Ο μετρητής αποκρίνεται έτσι μόνο στους ακέραιους αριθμούς κβαντίζοντας αποτελεσματικά το σήμα, ενώ συγχρόνως το κωδικοποιεί. Κάθε δισταθές στάδιο του μετρητή αποθηκεύει μηδέν ή ένα για κάθε δυαδικό ψηφίο που αντιπροσωπεύει (το δυαδικό 1110 ή το δεκαδικό 14 παρουσιάζεται στο σχήμα 2.7). Μια γεννήτρια δειγματοληπτεί τις 2^0 , 2^1 , 2^2 , και 2^3 θέσεις ψηφίων σε σειρά και μεταδίδει ένα bit (παλμός ή κενό) σε συμφωνία με την κατάσταση κάθε επιπέδου του μετρητή. Το κύκλωμα αποθήκευσης είναι πάντα κενό και μηδενίζεται πριν από την έναρξη της ακολουθίας για το επόμενο παλμό.



ΣΧΗΜΑ 2.7: Block διάγραμμα κυκλώματος κωδικοποίησης.

Ο αποδιαμορφωτής PCM θα αναπαράγει το σωστό τυποποιημένο πλάτος που αντιπροσωπεύεται από την ομάδα κώδικα παλμών. Εντούτοις, θα αναπαράγει τα σωστά πρότυπα μόνο εάν είναι σε θέση να αναγνωρίσει σωστά την παρουσία ή την απουσία παλμών σε κάθε θέση. Ο θόρυβος δεν εισάγει κανένα λάθος εάν ο λόγος σήματος προς θόρυβο (Signal to Noise Ratio - SNR) είναι τέτοιος ώστε οι υψηλότερες αιχμές του θορύβου να μην αναγνωρίζονται ως παλμοί. Όταν ο θόρυβος είναι τυχαίος, η πιθανότητα της εμφάνισης μιας αιχμής θορύβου συγκρίσιμης στο πλάτος με τους παλμούς μπορεί να καθοριστεί. Αυτή η πιθανότητα μπορεί να καθοριστεί από μαθηματική άποψη για οποιοδήποτε λόγο σήματος προς μέση ισχύς θορύβου. Όταν αυτό γίνεται για 10^5 παλμούς ανά δευτερόλεπτο, το κατά προσέγγιση ποσοστό λάθους για τρεις τιμές της ισχύος σημάτων ως προς τη μέση ισχύ θορύβου είναι:

- 17 DB - 10 λάθη ανά δευτερόλεπτο
- 20 DB - 1 λάθος κάθε 20 λεπτά
- 22 DB - 1 λάθος κάθε 2.000 ώρες

Πάνω από ένα όριο του λόγου σήματος προς θόρυβο περίπου 20 DB, ουσιαστικά κανένα λάθος δεν εμφανίζεται. Σε όλα τα άλλα συστήματα διαμόρφωσης, ακόμη και με το λόγο σήματος προς θόρυβο τόσο υψηλό όπως 60 DB, ο θόρυβος θα έχει κάποια επίδραση. Επιπλέον, το σήμα PCM μπορεί να αναμεταδοθεί, όπως σε ένα πιο πολύπλοκο σύστημα συνδέσεων ηλεκτρονικών, όσες φορές επιθυμούμε, χωρίς την εισαγωγή των πρόσθετων παρεμβολών θορύβου. Ο θόρυβος, δηλαδή, δεν είναι αθροιστικός στους σταθμούς ηλεκτρονικών όπως είναι με άλλα συστήματα διαμόρφωσης.

Το σύστημα βέβαια εισάγει κάποια παραμόρφωση με το να κβαντίζει το σήμα. Οι δύο τυποποιημένες τιμές που επιλέγονται και το διάστημα δειγματοληψίας τείνουν να κάνουν το αναδημιουργημένο σήμα να διαφοροποιηθεί από το αρχικό. Αυτή η παραμόρφωση, ο λεγόμενος κβαντικός θόρυβος, εισάγεται αρχικά κατά την κβάντιση και κατά την κωδικοποίηση διαμορφωτή και παραμένει σταθερή σε όλες τις διαδικασίες

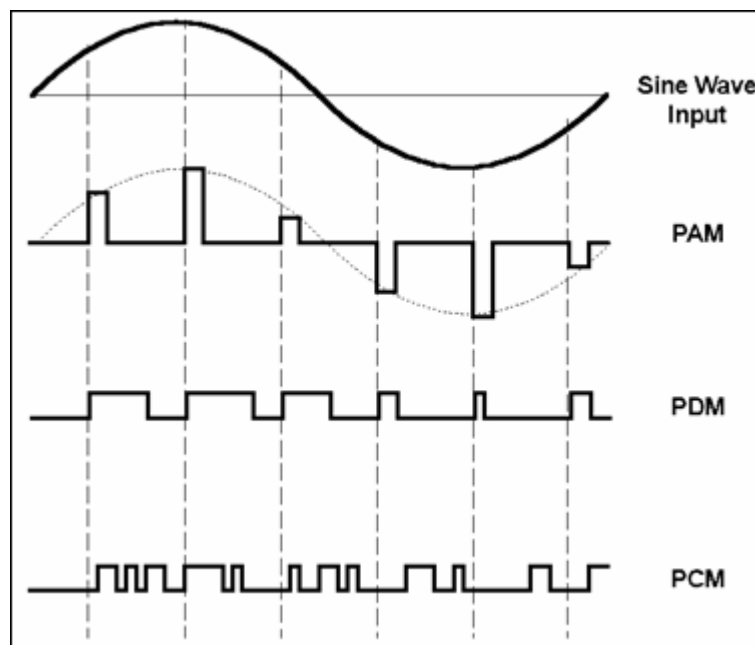
μετάδοσης και αναμετάδοσης. Το μέγεθός του μπορεί να μειωθεί φέρνοντας τα τυπικά κβαντικά επίπεδα πιο κοντά. Η σχέση του κβαντισμένου θορύβου προς τον αριθμό των ψηφίων στο δυαδικό κώδικα δίνεται από την ακόλουθη σχέση:

$$\frac{\text{peak signal power}}{\text{average quantizing noise power}} = (10.8 + 6n) \text{ dB}$$

Όπου: το n είναι ο αριθμός ψηφίων στο δυαδικό κώδικα.

Έτσι, με τον τετραψήφιο κώδικα των σχημάτων 2.5 και 2.6, ο κβαντισμένος θόρυβος θα είναι περίπου 35 DB πιο αδύναμος από το σήμα κορυφής που το κανάλι θα προσαρμόσει.

Τα πλεονεκτήματα του PCM είναι δύο. Πρώτον, η παρεμβολή θορύβου σχεδόν πλήρως αποβάλλεται όταν υπερβαίνουν τα σήματα παλμού τα επίπεδα θορύβου κατά μία τιμή 20 DB ή περισσότερο. Δεύτερον, το σήμα μπορεί να παραληφθεί και να αναμεταδοθεί όσες φορές επιθυμούμε χωρίς να παραμορφωθεί.



Σχήμα 2.8: Τεχνικές "παλμού-κώδικα"

2.1.2. SIGNAL TO NOISE RATIO (SNR)

Οι τεχνικές κβάντισης non-uniform και adaptive χρησιμοποιούν τις ιδιότητες της φωνής και του ανθρώπινου ακουστικού συστήματος στην απλούστερη μορφή τους για να βελτιώσουν την παρατηρούμενη ποιότητα του κωδικοποιητή. Και οι δύο τεχνικές προσπαθούν να διατηρήσουν υψηλό ρυθμό σήματος προς θόρυβο (Signal to Noise Ratio - SNR) για την κωδικοποιημένη φωνή. Το SNR προσδιορίζεται ως ο λόγος της ισχύς σήματος προς την ισχύ του θορύβου κβάντισης, όπου η ισχύς σήματος ορίζεται ως:

$$\sigma_s^2 = \frac{1}{M} \sum_{n=1}^M (s[n])^2$$

Και η ισχύς του θορύβου κβάντισης ορίζεται ως:

$$\sigma_q^2 = \frac{1}{M} \sum_{n=1}^M (\hat{s}[n] - s[n])^2$$

Όπου το M είναι ο αριθμός των σημείων στο σήμα. Χρησιμοποιώντας την κλίμακα dB το SNR ορίζεται ως:

$$SNR = 10 \log_{10} \frac{\sigma_s^2}{\sigma_q^2}$$

Το SNR είναι μια μέση μέτρηση του λάθους πάνω σε ολόκληρο το σήμα και συχνά καλύπτεται από τα υψηλής ενέργειας τμήματα του σήματος. Οι non-uniform κβαντιστές τυπικά χρησιμοποιούνται για να παράγουν υψηλότερο SNR για σήματα μεγάλης δυναμικής εμβέλειας.

2.2. ΟΜΟΙΟΜΟΡΦΗ ΚΒΑΝΤΙΣΗ

Το σχήμα 2.9a δείχνει τη συνάρτηση εισόδου-εξόδου ενός μεσαίου-βήματος (mid-tread) ομοιόμορφου κβαντιστή με 8 επίπεδα κβάντισης. Στο σχήμα αυτό, $s[n]$ είναι το σήμα εισόδου και $\hat{s}[n]$ η έξοδος, δηλαδή το κβαντισμένο σήμα. Όπως φαίνεται στο σχήμα, κάθε επίπεδο κβάντισης προσδιορίζεται από ένα δυαδικό κώδικα 3 bit για κωδικοποίηση. Γενικά $B = \lceil \log_2 L \rceil$ bits απαιτούνται για να κωδικοποιηθούν L επίπεδα κβάντισης όπου $[x]$ αντιπροσωπεύει τον μικρότερο ακέραιο ο οποίος είναι μεγαλύτερος του x . Για παράδειγμα για να κωδικοποιηθούν $L=256$ επίπεδα απαιτούνται 8 bit.

Η διαδικασία μια ομοιόμορφης κβάντισης ορίζεται με το να αναπαραστήσει όλες τις τιμές εισόδου μεταξύ s_{i-1} και s_i με \hat{s}_i . ως:

$$s_i - s_{i-1} = \Delta$$

και

$$\hat{s}_i - \hat{s}_{i-1} = \Delta$$

όπου Δ είναι το μέγεθος του βήματος κβάντισης, και του λάθους κβάντισης, $q[n]$ ορίζεται ως:

$$q[n] = s(n) - \hat{s}[n]$$

όπου

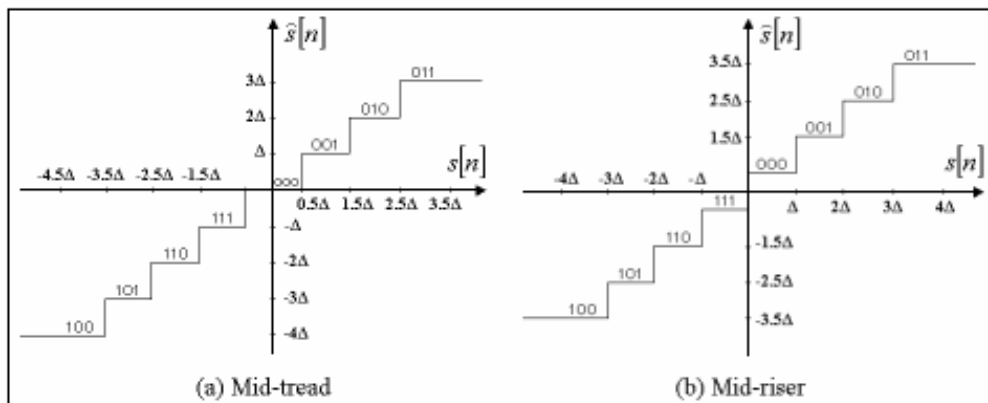
$$-\frac{\Delta}{2} \leq q[n] \leq \frac{\Delta}{2}$$

Ένας δεύτερος τύπος ομοιόμορφου κβαντιστή είναι γνωστός και ως mid-riser και ο οποίος φαίνεται στα σχήμα 2.9b. Η διαφορά μεταξύ των δύο προαναφερθέντων κβαντιστών είναι ότι ο mid-riser δεν έχει ένα μηδενικό επίπεδο εξόδου αλλά είναι συμμετρικός ως προς το μηδέν. Σε αντίθεση, ο mid-tread έχει ένα μηδενικό επίπεδο εξόδου αλλά δεν έχει τον ίδιο αριθμό θετικών και αρνητικών επιπέδων. Οι παράμετροι το ομοιόμορφου κβαντιστή, οι οποίοι φαίνονται στο σχήμα 2.10, είναι ο αριθμός των επιπέδων κβάντισης (αριθμός των bits, B), το μέγεθος του βήματος κβάντισης, Δ , το μέγιστο εύρος σήματος S_{\max} . Οι παράμετροι κβάντισης σχετίζονται από την εξίσωση:

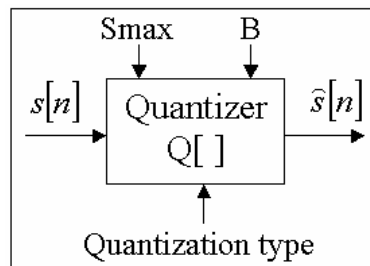
$$2S_{\max} = \Delta 2^B$$

όπου S_{\max} είναι το μέγιστο ασφαλιστικό σήμα, το οποίο είναι ίσο με τη μέγιστη τιμή του σήματος. Αν το S_{\max} επιλεγεί μικρό, τότε το σήμα θα ψαλιδιστεί έντονα. Από την άλλη, μεγάλο S_{\max} οδηγεί σε ένα μεγάλο μέγεθος βήματος άρα και σε μια κανονική κβάντιση. Το μέγεθος του βήματος μπορεί να μειωθεί αυξάνοντας το B . Αυξάνοντας το B οδηγεί σε ένα μικρό λάθος κβάντισης όπως επίσης σε ένα υψηλό bit rate για τον κωδικοποιητή. Ο πίνακας 2.1 δείχνει τις παραμέτρους του ομοιόμορφου κβαντιστή οι οποίες είναι διαθέσιμες σε ένα

πρόγραμμα ομοιόμορφης κβάντισης. Αυτός ο πίνακας παρέχει επίσης το εύρος και μερικές τυπικές τιμές για τις παραμέτρους αυτές.



ΣΧΗΜΑ 2.9: Mid-tread και mid-riser ομοιόμορφοι κβαντιστές με οκτώ επίπεδα κβάντισης. Σε κάθε επίπεδο αντιστοιχεί ένας 3-bit δυαδικός κώδικας, $c[n]$.



ΣΧΗΜΑ 2.10: Block διάγραμμα ενός ομοιόμορφου κβαντιστή με τις παραμέτρους εισόδου του.

Parameters	Name	Range	Typical values
Quantizer max level	S_{max}	0 - 32767	32767
Number of bits	B	1 - 16	7
Quantization type	-	-	mid-treat or mid-riser

ΠΙΝΑΚΑΣ 2.1: Παράμετροι ενός ομοιόμορφου PCM κβαντιστή.

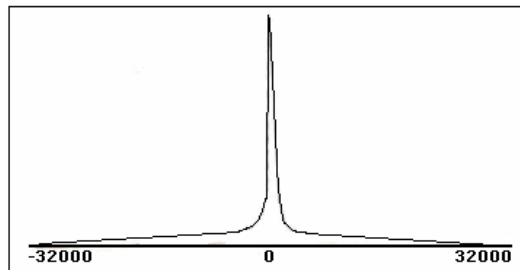
2.3. ΜΗ ΟΜΟΙΟΜΟΡΦΗ ΚΒΑΝΤΙΣΗ

Οι ομοιόμορφοι κβαντιστές παρουσιάζονται πιο ευνοϊκοί για σήματα με μια ομοιόμορφη κατανομή πάνω στο δυναμικό εύρος του σήματος. Όπως φαίνεται στη συνάρτηση πυκνότητας πιθανότητας ενός σήματος φωνής στο σχήμα 2.11, τα σήματα φωνής δεν έχουν ομοιόμορφη κατανομή. Όπως φαίνεται από την συνάρτηση δεν κατανέμονται ομοιόμορφα και η πλειοψηφία των δειγμάτων του σήματος έχει σχετικά μικρά μεγέθη.

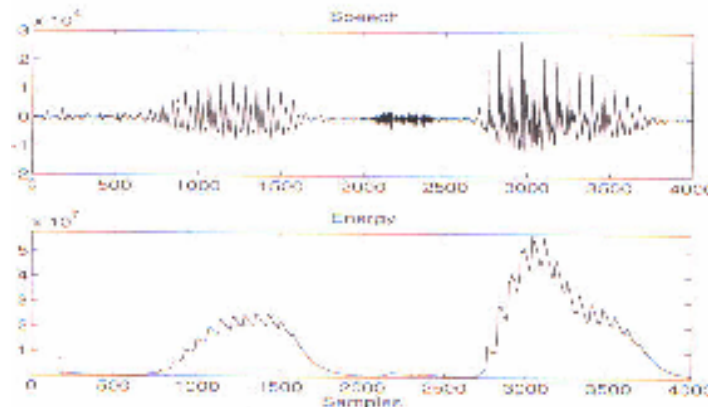
Ένας άλλος τρόπος για να κατανοήσουμε το βασικό πρόβλημα το οποίο συνδέεται με έναν ομοιόμορφο κβαντιστή είναι να κατανοήσουμε τις ιδιότητες των σημάτων φωνής που σχετίζονται με τη διαφοροποίηση τους στο χρόνο (time-varying). Τα σήματα φωνής έχουν συλλαβικά ενεργειακά περιγράμματα το οποίο σημαίνει ότι η ενέργεια των σημάτων φωνής ποικίλει από μεγάλη σε μικρή, πάνω σε συλλαβικά διαστήματα μέσα σε δέκατα του δευτερολέπτου. Το σχήμα 2.12 δείχνει την ποικιλία της ενέργειας σε τμήμα φωνής των 500 msec. Ένας "καλός" κβαντιστής θα παράγει ομοιόμορφο λευκό θόρυβο σε ένα σταθερό επίπεδο, το οποίο σημαίνει ότι ο λόγος SNR της κωδικοποιημένης φωνής θα ποικίλει στο συλλαβικό ρυθμό, το οποίο είναι χειρότερο κατά την διάρκεια "ήσυχων" περιόδων παρά "θορυβωδών". Από την ιδιότητα της ακουστική επικάλυψης γνωρίζουμε ακόμη ότι, κατά προσέγγιση, τα σήματα επικάλυπτουν τον θόρυβο. Έτσι ένα ειδικό επίπεδο θορύβου θα είναι πιο εύκολο να ακουστεί σε "ήσυχες" παρά σε "θορυβώδεις" περιόδους. Οι ομοιόμορφοι

κβαντιστές οδηγούν σε σταθερό προσθετικό θόρυβο όταν αυτό που χρειάζεται είναι ένας κβαντιστής ο οποίος οδηγεί σε πολύ μεγαλύτερο θόρυβο. Δηλαδή χαμηλός θόρυβος κατά τη διάρκεια χαμηλών πλατών σήματος και υψηλός θόρυβος κατά τη διάρκεια υψηλών πλατών θορύβου. Οι μη ομοιόμορφοι κβαντιστές των οποίων το μέγεθος των βημάτων αυξάνει την ποσότητα σήματος είναι έτσι περισσότερο αποτελεσματικοί και οδηγούν σε ένα σχετικά σταθερό SNR πάνω σε υψηλής και χαμηλής ενέργειας περιόδους.

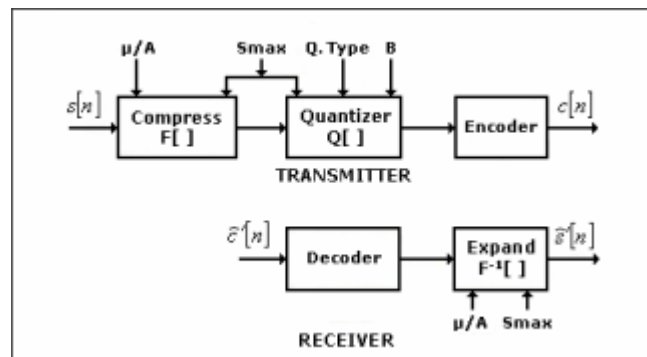
Το σχήμα 2.13 δείχνει ένα block διάγραμμα ενός γενικού συστήματος συμπίεσης-αποσυμπίεσης (companding) PCM. Σε αυτό το PCM σύστημα, το σήμα περνάει μέσω μιας μη γραμμικής συνάρτησης η οποία "δεν έχει μνήμη" που ονομάζεται συμπίεστης. Αυτή η συνάρτηση F , επιλέγεται έτσι ώστε το συμπίεσμένο σήμα να έχει μια σχετική κατανομή. Το συμπίεσμένο σήμα κβαντίζεται ομοιόμορφα και κωδικοποιείται με ομοιόμορφη κατανομή. Στο δέκτη, μετά την αποκωδικοποίηση το σήμα εκτείνεται χρησιμοποιώντας την αντίστροφη συνάρτηση F^{-1} .



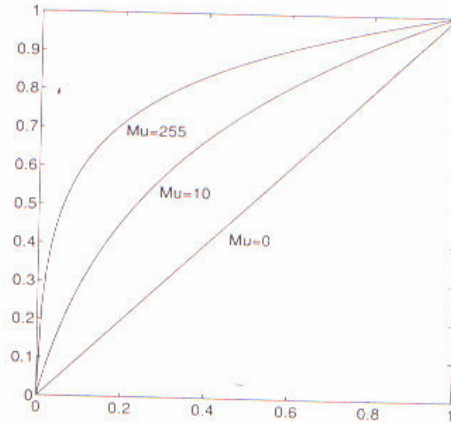
ΣΧΗΜΑ 2.11: Μια τυπική συνάρτηση πυκνότητας της φωνής. Αυτό είναι ένα ιστόγραμμα 3 sec ομιλίας.



ΣΧΗΜΑ 2.12: Μια περίοδος 500msec ομιλίας και η διακύμανση ενέργειας της.



ΣΧΗΜΑ 2.13: Block διάγραμμα ενός companding PCM



ΣΧΗΜΑ 2.14: Η μ -law συνάρτηση συμπίεσης για τρεις διαφορετικές τιμές του μ .

2.3.1. ΛΟΓΑΡΙΘΜΙΚΟΙ ΚΒΑΝΤΙΣΤΕΣ

Η λειτουργία λογαρίθμου ταιριάζει καλά στη συμπίεση φωνής επειδή οδηγεί ρητά σε πολλαπλασιαστικό θόρυβο κβάντισης. Δυστυχώς, η λογαριθμική λειτουργία δεν είναι τέλεια διότι έχει άπειρο εύρος δηλαδή οι αλγόριθμοι των αριθμών που πλησιάζουν το μηδέν προσεγγίζουν αρνητική απειρία. Για να αποφύγουμε αυτά τα προβλήματα, τροποποιημένες λογαριθμικές λειτουργίες χρησιμοποιούνται ευρέως στους companding κβαντιστές. Δύο συνηθισμένοι τύποι συμπίεστων-αποσυμπίεστων (companders) που χρησιμοποιούνται στην πράξη είναι οι μ -law και A-law. Και οι δύο λειτουργίες έχουν λογαριθμική συμπεριφορά και πεπερασμένο εύρος. Ο μ -law συμπίεστής καθορίζεται ως:

$$t[n] = F(s[n]) = S_{\max} \frac{\log\left(1 + \mu \frac{|s[n]|}{S_{\max}}\right)}{\log(1 + \mu)} \text{sign}(s[n])$$

όπου μ είναι η παράμετρος συμπίεσης, το $\text{sign}(s[n])$ παίρνει μια τιμή μεταξύ +1 και -1 και $\log(\)$ είναι η φυσική αλγοριθμική λειτουργία. Το σχήμα 2.14 δείχνει τη λειτουργία συμπίεσης για κάποιες τυπικές τιμές του μ . Για μικρές τιμές του $s[n]$ αυτή η λειτουργία είναι σχεδόν γραμμική διότι:

$$\log(1 + \mu s[n]) \approx \mu s[n]$$

Ένας άλλος λογαριθμικός compander είναι ο A-law ο οποίος καθορίζεται ως:

$$t[n] = \begin{cases} \frac{As[n]}{1 + \log A} & 0 \leq s[n] \leq S_{\max}/A \\ S_{\max} \frac{1 + \log(A|s[n]|/S_{\max})}{1 + \log A} \text{sign}(s[n]) & S_{\max}/A \leq s[n] \leq S_{\max} \end{cases}$$

όπου A είναι ο παράγοντας συμπίεσης. Στην συμπίεση A-law, τα χαρακτηριστικά του κβαντιστή είναι γραμμικά για μεγέθη μικρότερα του S_{\max}/A και λογαριθμικά πάνω από αυτό. Παρατηρούμε ότι και οι δύο companders χαρτογραφούν το S_{\max} στο S_{\max} . Γι' αυτό το λόγο, όπως και στους ομοιόμορφους κβαντιστές, το S_{\max} είναι το κοινό στοιχείο των δύο companding κβαντιστών. Ο πίνακας 2.2 εμφανίζει τις παραμέτρους για μ -law και A-law για μη ομοιόμορφους κβαντιστές. Αυτοί οι κβαντιστές σχηματίζουν τη βάση για το φωνητικό πρότυπο ITU G.711 για τη συμπίεση/αποσυμπίεση γραμμικών PCM δειγμάτων. Οι κβαντιστές μ -law χρησιμοποιούνται στη Βόρεια Αμερική με $\mu=255$ και οι A-law χρησιμοποιούνται στην Ευρώπη με $A=87.56$.

Για τις ελάχιστες τιμές των παραγόντων συμπίεσης, $\mu=0$ ή $A=1$, η συνάρτηση F είναι η ενωτική συνάρτηση και δεν πραγματοποιείται συμπίεση. Οι μεγαλύτερες τιμές του μ

ή του A καταλήγουν στη συμπίεση και σε ένα σταθερό SNR πάνω από μεγαλύτερο εύρος της φωνής εισόδου. Παρόλο που μεγάλες τιμές των παραγόντων συμπίεσης, όπως $\mu > 500$ ή $A > 100$, οδηγούν σε ένα σταθερό SNR για ένα μεγαλύτερο εύρος, παράγουν ένα γενικά χαμηλότερο SNR για την κβαντισμένη φωνή.

parameters	name	range	typical value
compression factor	μ	0-1000	255
	A	1-100	87.56
quantizer max level	S_{max}	0-32767	32767
number of bits	B	1-16	5
quantization type	--	--	mid-tread or mid-riser

ΠΙΝΑΚΑΣ 2.2: Οι παράμετροι ενός λογαριθμικού companding PCM (log-PCM)

2.3.2. ΒΕΛΤΙΣΤΟΙ ΚΒΑΝΤΙΣΤΕΣ

Μια άλλη προσέγγιση των μη ομοιόμορφων κβαντιστών είναι να χρησιμοποιήσουμε βέλτιστους κβαντιστές (optimum quantizers) για να μεγιστοποιήσουμε το μέσο SNR. Ο στόχος ενός βέλτιστου κβαντιστή είναι να επιλέξει τη μη γραμμική συνάρτηση συμπίεσης έτσι ώστε για το δοθέντα αριθμό επιπέδων κβάντισης, η τελική κωδικοποιημένη φωνή να έχει το μεγαλύτερο δυνατό SNR. Ο σκοπός ενός βέλτιστου κβαντιστή είναι να αντιστοιχίσει τα μεγέθη των βημάτων του κβαντιστή σε μια κατανομή του κωδικοποιημένου σήματος. Η σχεδίαση ενός βέλτιστου κβαντιστή είναι ισοδύναμη με το να λάβουμε μια συνάρτηση F_{opt} που καταλήγει σε συμπιεσμένο σήμα ομοιόμορφης κατανομής. Οι λογαριθμικοί κβαντιστές προσεγγίζουν μια τέτοια συνάρτηση χρησιμοποιώντας μ -law και A-law companders αλλά δεν είναι βέλτιστοι αν η κατανομή του σήματος είναι γνωστή. Στις περιπτώσεις όπου η κατανομή του σήματος είναι γνωστή, είναι δυνατόν να λάβουμε τα επίπεδα του κβαντιστή (ή F_{opt}) τα οποία ελαχιστοποιούν το λάθος κβάντισης και μεγιστοποιούν το SNR. Για να λάβουμε τον βέλτιστο κβαντιστή, το λάθος κβάντισης δίνεται από:

$$\sigma_e^2 = \sum_{i=-\frac{\Delta t}{2}+1}^{\frac{\Delta t}{2}} \int_{s_{i-1}}^{s_i} (\hat{s}_i - s)^2 p(s) ds$$

όπου το $p(s)$ είναι συνάρτηση πυκνότητας πιθανότητας, η οποία πρέπει να ελαχιστοποιηθεί. Αυτό μπορεί να επιτευχθεί παραγωγίζοντας το σ_e^2 σύμφωνα με τις παραμέτρους του κβαντιστή, s_i και \hat{s}_i , και θέτοντας το αποτέλεσμα στο 0. Ο βέλτιστος κβαντιστής στον οποίο καταλήγουμε έχει τις παρακάτω ιδιότητες:

1. Η βέλτιστη περιοχή του επιπέδου κβάντισης \hat{s}_i είναι στο κέντρο της μάζας πυκνότητας πιθανότητας πάνω στο διάστημα s_{i-1} και s_i .
2. Τα βέλτιστα οριακά σημεία κβάντισης βρίσκονται στη μισή απόσταση μεταξύ των $M/2$ επιπέδων του κβαντιστή.

Η φωνή έχει μια συνάρτηση πυκνότητας η οποία πλησιάζει σχετικά την Λαπλασιανή συνάρτηση πυκνότητας. Ο πίνακας 2.3 παρουσιάζει τις βέλτιστες παραμέτρους του κβαντιστή για τη Λαπλασιανή κατανομή για κβαντιστές 1, 2, 3, 4 και 5 bit οι οποίοι μπορούν να χρησιμοποιηθούν για να υλοποιήσουν το βέλτιστο Λαπλασιανό κβαντιστή. Σαν παράδειγμα θεωρούμε την κατάσταση 2 bit του πίνακα 2.3. Για σήματα με τυπική απόκλιση, πλάτος μεταξύ 0 και 1.102 κβαντίζονται στο 0.395, και τα πλάτη πάνω από 1.102 κβαντίζονται στο 1.810. Το αρνητικό αυτών των αποτελεσμάτων έχει εφαρμογή σε πλάτη σημάτων που είναι μικρότερα του μηδενός. Τα βέλτιστα χαρακτηριστικά του συμπιεστή για ένα σήμα με τη Λαπλασιανή συνάρτηση κατανομής είναι γνωστό σαν m-law συμπιεστής και ορίζεται ως:

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

όπου η βέλτιστη αξία του m είναι:

$$m_{opt} = \sqrt{2} S \max 3\sigma_s$$

1 bit		2 bits		3 bits		4 bits		5 bits	
s_i	\hat{s}_i	s_i	\hat{s}_i	s_i	\hat{s}_i	s_i	\hat{s}_i	s_i	\hat{s}_i
∞	0.707	1.102	0.395	0.504	0.222	0.266	0.126	0.147	0.072
		∞	1.810	1.818	0.785	0.566	0.407	0.302	0.222
				2.285	1.576	0.910	0.726	0.467	0.382
				∞	2.994	1.317	1.095	0.642	0.551
						1.821	1.504	0.829	0.732
						2.499	2.103	1.031	0.926
						3.605	2.895	1.250	1.136
						∞	4.316	1.490	1.365
								1.756	1.616
								2.055	1.896
								2.398	2.214
								2.804	2.583
								3.305	3.025
								3.978	3.586
								5.069	4.371
								∞	5.768

ΠΙΝΑΚΑΣ 2.3: Βέλτιστες παράμετροι κβαντιστή με Λαπλασιανή κατανομή για 1, 2, 3, 4 και 5 bit κβάντιση.

2.4. ΠΡΟΣΑΡΜΟΣΤΙΚΗ ΚΒΑΝΤΙΣΗ

Στο προηγούμενο τμήμα είδαμε ότι η μη ομοιόμορφοι κβαντιστές σχηματίζουν ένα σταθερό σήμα με ένα μεγάλο δυναμικό εύρος. Το δυναμικό εύρος πρέπει να είναι μεγάλο για να περιλάβει και τα "ήσυχα" και τα "θορυβώδη" τμήματα φωνής. Λόγω του ότι το σήμα φωνής δεν είναι σταθερό των οποίων τα χαρακτηριστικά μεταβάλλονται με το χρόνο, ένας καλύτερος τρόπος να σχηματιστεί ένα σήμα είναι όπως ένα μη σταθερό σήμα με σχεδόν παρόμοιο δυναμικό εύρος, αλλά ένα σήμα του οποίου οι ενεργειακές μεταβολές γίνονται σχετικά αργά με το χρόνο. Ένας καλός τρόπος να διευθυνοδοτήσουμε το μη σταθερό σήμα και το υψηλό εύρος δυναμικού του σήματος είναι να χρησιμοποιήσουμε προσαρμοστικούς κβαντιστές όπου τα χαρακτηριστικά τους είναι δυναμικά αλλαγμένα για να ταιριάξουν τις ιδιότητες του φωνητικού σήματος.

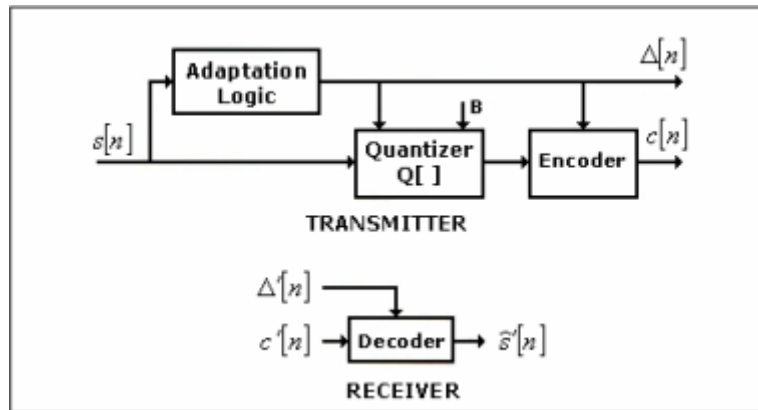
Η βασική ιδέα στην προσαρμοστική κβάντιση είναι να ποικίλει το μέγεθος βήματος του κβαντιστή για να ταιριάξει τη διαφορά του σήματος εισόδου. Εναλλακτικά και ισοδύναμα το σήμα φωνής μπορεί προσαρμοστικά να κλιμακωθεί για να ταιριάξει τις ιδιότητες ενός ειδικού κβαντιστή. Οι στρατηγικές προσαρμογής μπορούν να βασιστούν είτε στο φωνητικό σήμα εισόδου είτε στο κβαντισμένο σήμα εξόδου.

Το σχήμα 2.15 δείχνει ένα block διάγραμμα ενός συστήματος το οποίο προσαρμόζει τον κβαντιστή βασισμένο στο σήμα εξόδου. Η παράμετρος προσαρμογής $\Delta(n)$ εκτιμάται από το σήμα εισόδου το οποίο έπειτα χρησιμοποιείται για να ελέγξει τον κβαντιστή. Επειδή ο δέκτης δε μπορεί να υπολογίσει αυτή την αξία, το $\Delta(n)$ επίσης μεταδίδεται στο δέκτη το οποίο χρησιμοποιείται για να ελέγξει τη διαδικασία αποκωδικοποίησης. Το σύστημα αυτό καλείται feed-forward προσαρμοστικό PCM (PCM-AQF) διότι ο κβαντιστής ελέγχει παραμέτρους οι οποίοι πρέπει να μεταδοθούν κατά μήκος με τα κωδικοποιημένα δείγματα.

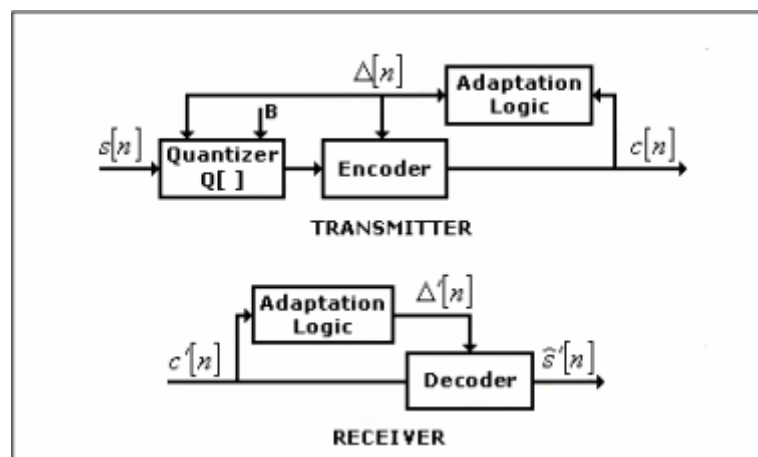
Το σχήμα 2.16 δείχνει μια εναλλακτική προσέγγιση το οποίο προσαρμόζει και το οποίο βασίζεται στην κβαντισμένη φωνή εξόδου. Από τη στιγμή που η κωδικοποιημένη φωνή είναι διαθέσιμη στο δέκτη και τον πομπό, το σύστημα δεν χρειάζεται να μεταδώσει τις παραμέτρους ελέγχου του κβαντιστή. Για το λόγο αυτό ονομάζεται προσαρμοστικό PCM ανάδρασης (feedback adaptive PCM, PCM-AQB)

Ανάλογα με το πόσο συχνά παρουσιάζεται η προσαρμογή, τα προσαρμοστικά συστήματα μπορούν να χωριστούν σε "στιγμαία" και "συλλαβικά". Στη στιγμαία

προσαρμογή, οι παράμετροι του κβαντιστή ενημερώνονται σε κάθε δείγμα του σήματος εισόδου, δεδομένου ότι στη συλλαβική προσαρμογή οι αλλαγές είναι πιο αργές και είναι σε ένα ρυθμό ο οποίος είναι σε ίδιας τάξης μεγέθους με το συλλαβικό ρυθμό της φωνής (16-20 msec). Στα συστήματα feed forward, η προσαρμογή συνήθως παρουσιάζεται με έναν αργό ρυθμό διότι οι προσαρμοστικές παράμετροι χρειάζονται για να μεταδοθούν έπειτα από κάθε ενημέρωση παραμέτρου. Αντίθετα, από τη στιγμή που η ανάδραση προσαρμογής δεν απαιτεί την μετάδοση κανενός από τις παραμέτρους, οι προσαρμογές ανάδρασης συνήθως είναι στιγμιαίες.



ΣΧΗΜΑ 2.15: Block διάγραμμα ενός PCM-AQF



ΣΧΗΜΑ 2.16: Block διάγραμμα ενός PCM-AQB

2.4.1. FEED FORWARD ΠΡΟΣΑΡΜΟΓΗ

Στην feed forward προσαρμογή, το σήμα εισόδου χρησιμοποιείται για να υπολογίσει τις παραμέτρους προσαρμογής. Μια κοινή προσέγγιση είναι να χρησιμοποιηθεί η φωνητική διακύμανση στην διαδικασία προσαρμογής. Η διακύμανση μπορεί να χρησιμοποιηθεί με δύο διαφορετικούς τρόπους. Πρώτα, μπορεί να χρησιμοποιηθεί σε μία στρατηγική προσαρμοστικού μεγέθους βήματος όπου το μέγεθος βήματος του κβαντιστή αλλάζει με το χρόνο. Σε μια δεύτερη προσέγγιση η διακύμανση σήματος μπορεί να χρησιμοποιηθεί σε μια στρατηγική προσαρμοστικού κέρδους όπου το σήμα κλιμακώνεται με έναν παράγοντα χρονικά μεταβαλλόμενου κέρδους πριν περάσει μέσα στο σταθερό κβαντιστή. Αυτή η διαδικασία συχνά αναφέρεται ως προσαρμοστικό κέρδος ελέγχου (adaptive gain control-AGC).

Η διακύμανση του σήματος φωνής σε ένα feed forward σχήμα μπορεί να υπολογιστεί μετρώντας τη βραχυχρόνια ενέργεια του σήματος φωνής η οποία μπορεί να οριστεί ως:

$$\sigma^2[n] = \sum_{m=-\infty}^{\infty} s^2[m]w[n-m]$$

Όπου το $w[n]$ είναι μια συνάρτηση παράθυρου (window function) η οποία χρησιμοποιείται για να περιορίσει τον όγκο των δεδομένων ή τη διάρκεια της ομιλίας που χρησιμοποιείται στη μέτρηση της βραχυχρόνιας ενέργειας. Μια εκθετική συνάρτηση προσθήκης ορίζεται ως:

$$w[n] = \begin{cases} \alpha^{n-1} & n \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

με $0 < \alpha < 1$ να χρησιμοποιείται συχνά. Αυτό έχει ως αποτέλεσμα την εξίσωση συμπίεσης διαφοράς.

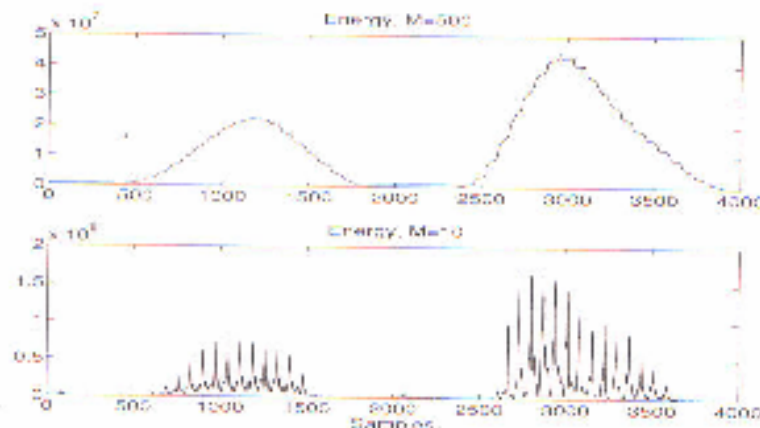
$$\sigma^2[n] = \alpha\sigma^2[n-1] + (1-\alpha)s^2[n]$$

η οποία υπολογίζει τη διακύμανση. Μια άλλη γνωστή συνάρτηση είναι η συνάρτηση ορθογώνιου παράθυρου (rectangular window) η οποία καταλήγει σε έναν υπολογισμό της διακύμανσης της μορφής:

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

όπου M είναι το μέγεθος του ορθογώνιου παράθυρου. Ενώ η εξίσωση $w[n]$ δεν απαιτεί καθόλου αποθηκευμένα δείγματα φωνής και έτσι δεν εμφανίζει καθόλου διαδικαστική καθυστέρηση, ο υπολογισμός της εξίσωσης $\sigma^2[n]$ απαιτεί αποθηκευμένα δείγματα σημάτων M και σαν αποτέλεσμα εμφανίζει M δείγματα καθυστέρησης κωδικοποίησης. Ο υπολογισμός της εξίσωσης $\sigma^2[n]$ μετράται και εφαρμόζεται για δείγματα φωνής μεταξύ των χρόνων n και $n+M-1$.

Η επιλογή του α στο εκθετικό παράθυρο και του M στο ορθογώνιο παράθυρο ελέγχει το αποτελεσματικό μήκος του διαστήματος το οποίο εμποδίζει τον υπολογισμό της διακύμανσης. Ενώ μεγάλες τιμές του α και του M καταλήγουν σε ένα υπολογισμό συλλαβικής διακύμανσης, μικρές τιμές καταλήγουν σε ένα στιγμιαίο υπολογισμό της διακύμανσης. Το σχήμα 2.17 παρουσιάζει μια συλλαβική και μια στιγμιαία διακύμανση του τμήματος φωνής του σχήματος 2.12. Οι διακυμάνσεις λαμβάνονται χρησιμοποιώντας ορθογώνιο παράθυρο με $M=500$ και $M=10$. Τα ορθογώνια παράθυρα συνήθως χρησιμοποιούνται στα συστήματα feed forward ενώ τα εκθετικά παράθυρα προτιμούνται στα συστήματα ανάδρασης. Με μια συλλαβική προσαρμογή, η πλευρική πληροφορία αποτελεί συνήθως ένα μικρό ποσοστό (περίπου 1%) του συνολικού bit rate.



ΣΧΗΜΑ 2.17: Υπολογισμοί διακύμανσης ενός τμήματος φωνής με ορθογώνιο παράθυρο μήκους 500 (κορυφή) και 10 (βάση).

2.4.1.1. ΠΡΟΣΑΡΜΟΓΗ ΜΕΓΕΘΟΥΣ ΒΗΜΑΤΟΣ

Το σχήμα 2.18 παρουσιάζει ένα block διάγραμμα ενός τυπικού προσαρμοστικού παλμού κώδικα διαμορφωτή (Adaptive Pulse Code Modulator – APCM) με προσαρμοστικό feed forward μέγεθος βήματος (PCM-ASF). Στην προσαρμογή μεγέθους βήματος η τιμή του μεγέθους βήματος είναι ανάλογη της διακύμανσης του σήματος φωνής. Ένας καλός κανόνας της προσαρμογής μεγέθους βήματος συνδέει το μέγεθος βήματος με την υπολογιζόμενη τυπική απόκλιση

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

όπου Δ_0 είναι ένας σταθερός παράγοντας κλιμάκωσης ο οποίος ονομάζεται κλιμακωτό βήμα μεγέθους. Χρησιμοποιώντας αυτή τη στρατηγική προσαρμογής, τα μέρη του σήματος φωνής με υψηλή διακύμανση θα έχουν μεγάλο μέγεθος βήματος κβάντισης, και το μέγεθος θα μειώνεται όσο θα μειώνεται η ενέργεια του σήματος. Αυτό έχει ως αποτέλεσμα την αποδοτική χρήση των διαθέσιμων bit για διαφορετικά επίπεδα ενέργειας του σήματος.

Ο πίνακας 2.4 εμφανίζει τις παραμέτρους ενός PCM-ASF. Οι παράμετροι περιλαμβάνουν:

- Τον αριθμό των bit, B
- Τον τύπο κβάντισης
- Το μέγιστο και ελάχιστο μέγεθος βήματος, Δ_{min} , Δ_{max}
- Το μέγεθος βήματος του κλιμακωτού κβαντιστή, Δ_0
- Το μέγεθος του ορθογώνιου παραθύρου για τον υπολογισμό της διακύμανσης

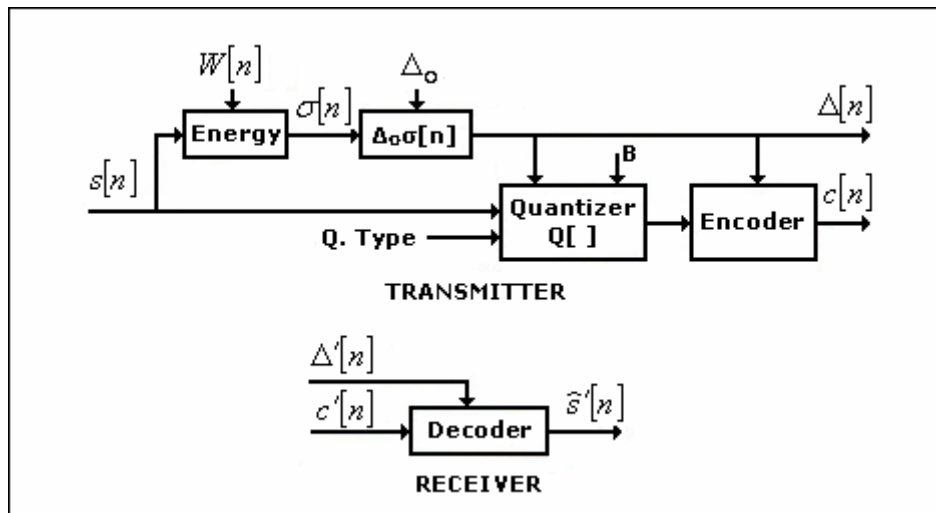
Από τη στιγμή που η διακύμανση του δυναμικού εύρους του σήματος φωνής μπορεί να είναι πολύ υψηλή, η προσαρμογή της εξίσωσης 2.16 μπορεί να αποδώσει παράλογα μικρά ή μεγάλα μεγέθη βήματος. Για να το αποφύγουμε αυτό τα Δ_{min} και Δ_{max} είναι προκαθορισμένα για τον κβαντιστή:

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

Η τιμή του Δ_{min} επιλέγεται συχνά να είναι χαμηλή, αρκετά, έτσι ώστε να ελαχιστοποιηθεί ο άχρηστος θόρυβος καναλιού, και ο Δ_{max} επιλέγεται να είναι μεγάλος αρκετά έτσι ώστε να ελαχιστοποιεί τον ψαλιδισμό του κβαντιστή. Ο λόγος $\Delta_{max} / \Delta_{min}$ καθορίζει το δυναμικό εύρος του κβαντιστή (με μια τυπική τιμή 100) και το Δ_0 καθορίζει το μέγεθος βήματος του κβαντιστή για την διακύμανση σήματος.

parameters	name	range	typical value
number of bits	B	1 - 16	5
scaling step size	Δ_0	0.01 - 1	0.1
min. step size	Δ_{min}	0 - 500	50
max. step size	Δ_{max}	50 - 10000	500
window length (FF)	M	1 - 500	100
exponential window (FB)	α	0.1 - 0.99	0.95
quantization type	--	--	mid-tread or mid-riser

ΠΙΝΑΚΑΣ 2.4: Παράμετροι των PCM-ASF και PCM-ASB.



ΣΧΗΜΑ 2.18: Block διάγραμμα ενός PCM-ASF.

2.4.1.2. ΚΕΡΔΟΣ ΠΡΟΣΑΡΜΟΓΗΣ

Το σχήμα 2.19 δείχνει το block ενός feed forward προσαρμοστικό APCM (PCM-AGF). Στους κβαντιστές προσαρμοστικού κέρδους (gain adaptive quantizers), το φωνητικό σήμα κλιμακώνεται από έναν παράγοντα κέρδους $G[n]$, ο οποίος μεταβάλλεται με το χρόνο, για να λάβει ένα σήμα εισόδου του κβαντιστή με ένα ομοιόμορφο εύρος. Ο κβαντιστής είναι ειδικά τροποποιημένος και τα χαρακτηριστικά του είναι επιλεγμένα για να ταιριάζουν τα χαρακτηριστικά του σήματος. Το κέρδος επιλέγεται να είναι αντιστρόφως ανάλογο με την απόκλιση σήματος. Το κέρδος αυτό ορίζεται από τον τύπο:

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

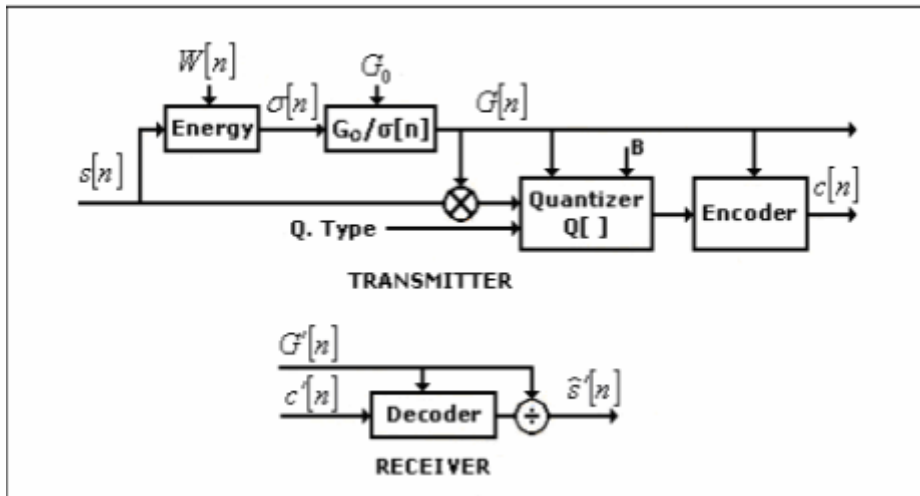
όπου G_0 είναι μια σταθερά ίση με το κέρδος για την διακύμανση εισόδου. Σε αυτή την εξίσωση, ένα σήμα χαμηλής ενέργειας αποδίδει μεγάλο κέρδος και ένα υψηλής ενέργειας σήμα αποδίδει μικρό κέρδος. Αυτό οδηγεί σε ένα σήμα με σχετικά ομοιόμορφο εύρος το οποίο είναι καταλληλότερο για ένα τροποποιημένο κβαντιστή παρά για το αρχικό σήμα. Για να αποφύγουμε την κλιμάκωση του σήματος πέραν κάποιου ορίου η διακύμανση από την συνάρτηση κέρδους συνήθως περιορίζεται από κάποιες μέγιστες και ελάχιστες τιμές όπως:

$$G_{\min} \leq G[n] \leq G_{\max}$$

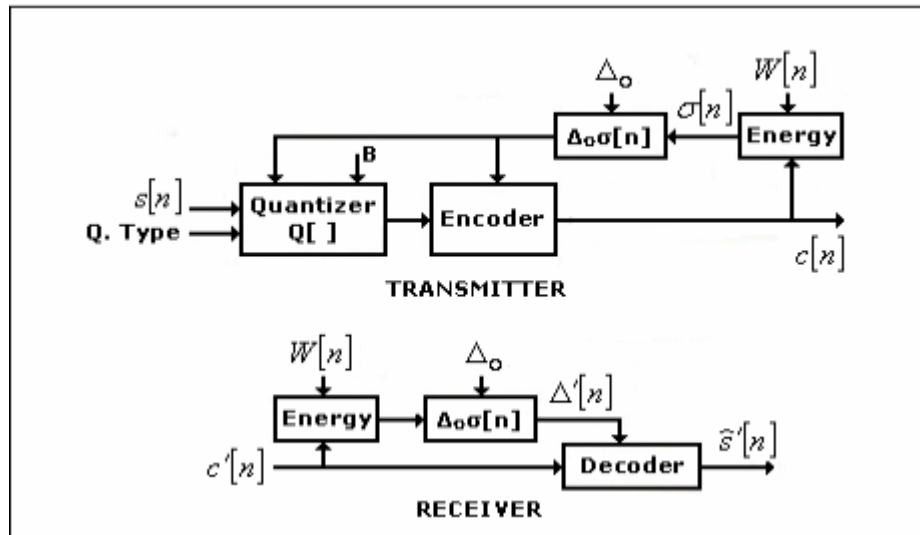
Ο λόγος G_{\max} / G_{\min} ελέγχει το δυναμικό εύρος του κλιμακωτού σήματος. Αλλάζοντας το G_0 , το σήμα μπορεί να κλιμακωθεί ομοιόμορφα σε οποιοδήποτε επιθυμητό επίπεδο. Οι παράμετροι του PCM-AGF δίνονται στον πίνακα 2.5.

parameters	name	range	typical value
number of bits	B	1 - 16	5
quantizer max. level	Smax	0 - 032767	32767
scaling gain	G_0	--	--
min. gain	G_{\min}	0 - 100	0.1
max. gain	G_{\max}	10 - 10000	100
rectangular window length (FF)	M	1 - 500	100
exponential window parameter (FB)	α	0.1 - 0.9999	0.95
quantization type	--	--	mid-tread or mid-riser

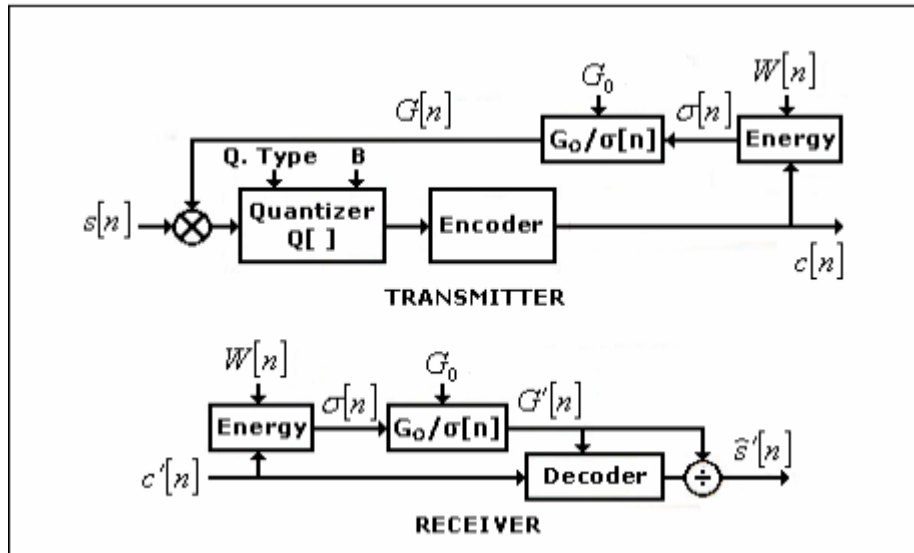
ΠΙΝΑΚΑΣ 2.5: Παράμετροι ενός PCM-AGF και ενός PCM-AGB.



ΣΧΗΜΑ 2.19: Block διάγραμμα ενός PCM-AGF



ΣΧΗΜΑ 2.20: Block διάγραμμα ενός PCM-ASB.



ΣΧΗΜΑ 2.21: Block διάγραμμα ενός PCM-AGB

2.4.2. FEEDBACK ΠΡΟΣΑΡΜΟΓΗ

Στις στρατηγικές προσαρμοστικής ανάδρασης (feedback) οι παράμετροι προσαρμογής υπολογίζονται από το κωδικοποιημένο σήμα φωνής. Επειδή το σήμα είναι διαθέσιμο στο δέκτη, η μετάδοση της πλευρικής πληροφορίας, όπως το μέγεθος βήματος ή το κέρδος, δεν είναι απαραίτητη. Τα σχήματα 2.20 και 2.21 δείχνουν block διαγράμματα APCM με κβαντιστή ανάδρασης προσαρμοστικού μεγέθους βήματος (PCM-ASB) και προσαρμοστικό κέρδος ανάδρασης (PCM-AGB) αντίστοιχα. Και τα δύο αυτά συστήματα χρησιμοποιούν τους προσαρμοστικούς κανόνες των εξισώσεων 2.16 και 2.18.

Στα συστήματα ανάδρασης ο υπολογισμός της διακύμανσης του σήματος φωνής βασίζεται μόνο πάνω στις προηγούμενες τιμές του κωδικοποιημένου σήματος. Στη μέθοδο ανάδρασης ένα εκθετικό παράθυρο χρησιμοποιείται συνήθως το οποίο οδηγεί στην εξίσωση διαφοροποίησης:

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

για να υπολογιστεί η διακύμανση. Οι παράμετροι του PCM ανάδρασης δίνονται στους πίνακες 2.4 και 2.5.

Στα συστήματα ανάδρασης, για να επιτευχθεί ένας σχετικά καλός υπολογισμός της απόκλισης του σήματος, προτιμάται συνήθως στιγμιαία προσαρμογή. Αυτό έρχεται σε αντίθεση με την περίπτωση feed forward όπου πιο αργή προσαρμογή απαιτείται για να μειωθεί το ποσό της πλευρικής πληροφορίας που πρέπει να μεταδοθεί στον αποκωδικοποιητή.

2.5. RETURN TO ZERO Ή NON RETURN TO ZERO PCM

Το PCM μπορεί να είναι είτε Return to Zero (RZ) είτε Non Return to Zero (NRZ). Για ένα σύστημα NRZ που για να συγχρονιστεί χρησιμοποιεί τις πληροφορίες εντός ζώνης, δεν πρέπει να υπάρξουν μακροχρόνιες ακολουθίες ταυτόσημων συμβόλων, όπως οι άσσοι ή τα μηδενικά. Για τα δυαδικά συστήματα PCM, η πυκνότητα των 1-συμβόλων καλούνται "πυκνότητα-ένα" (ones-density).

Η ones-density συχνά ελέγχεται χρησιμοποιώντας τεχνικές προ-κωδικοποίησης, όπου ο κώδικας PCM επεκτείνεται σε έναν ελαφρώς εκτενέστερο κώδικα που έχει σαφή σχέση με την ones-density πριν από τη διαμόρφωση της στο κανάλι. Σε άλλες περιπτώσεις,

επιπλέον framing bits προστίθενται στη ροή τα οποία εγγυώνται τουλάχιστον περιστασιακές μεταβάσεις συμβόλων.

Μια άλλη τεχνική που χρησιμοποιείται για να ελέγξει την ones-density είναι η χρήση μιας συσκευής παρεμβολών "πολυώνυμο" πάνω στα ακατέργαστα στοιχεία που θα τείνει να μετατρέψει την ακατέργαστη ροή στοιχείων σε ροή που μοιάζει ψευδο-τυχαία, αλλά όπου η ακατέργαστη ροή μπορεί να ανακτηθεί ακριβώς με την αντιστροφή της επίδρασης του πολυωνύμου. Σε αυτήν την περίπτωση, μακροσκελείς ακολουθίες μηδενικών ή άσων είναι ακόμα πιθανές στην έξοδο, αλλά θεωρείται αρκετά απίθανο να είναι μέσα στην κανονική ανοχή της εφαρμοσμένης μηχανικής.

Σε άλλες περιπτώσεις, η μακροχρόνια τιμή του συνεχούς ρεύματος του διαμορφωμένου σήματος είναι σημαντική, όπως ενισχύοντας μια μετατόπιση συνεχούς ρεύματος θα τείνει να αποκλίνει τα κυκλώματα ανιχνευτών από το εύρος λειτουργίας τους. Σε αυτήν την περίπτωση ειδικά μέτρα λαμβάνονται για να κρατήσουν μια αρίθμηση της αθροιστικής μετατόπισης συνεχούς ρεύματος, και για να τροποποιήσουν τους κώδικες για να κάνουν, εάν είναι απαραίτητο, το συνεχές ρεύμα να τείνει πάντα στο μηδέν.

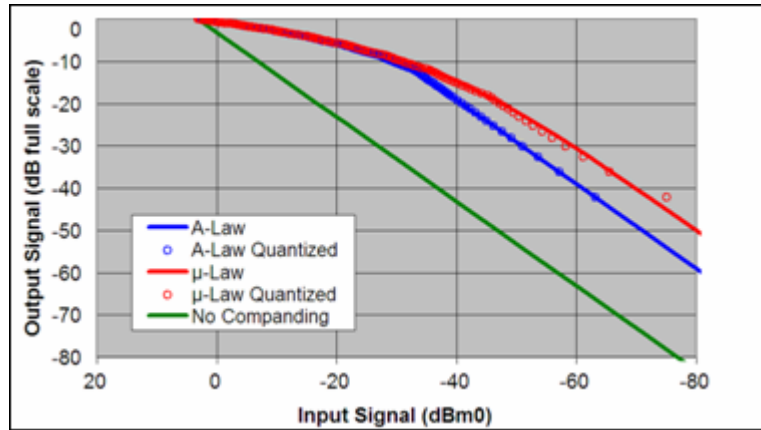
Πολλοί από αυτούς τους κώδικες είναι διπολικοί που οι παλμοί μπορούν να είναι θετικοί, αρνητικοί ή ανύπαρκτοι. Συνήθως οι μη μηδενικοί παλμοί εναλλάσσονται μεταξύ θετικού και αρνητικού. Αυτοί οι κανόνες μπορούν να παραβιαστούν για να παράγουν ειδικά σύμβολα που χρησιμοποιούνται για framing ή άλλους ειδικούς σκοπούς.

2.6. ΙΣΤΟΡΙΚΑ ΣΤΟΙΧΕΙΑ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥ PCM

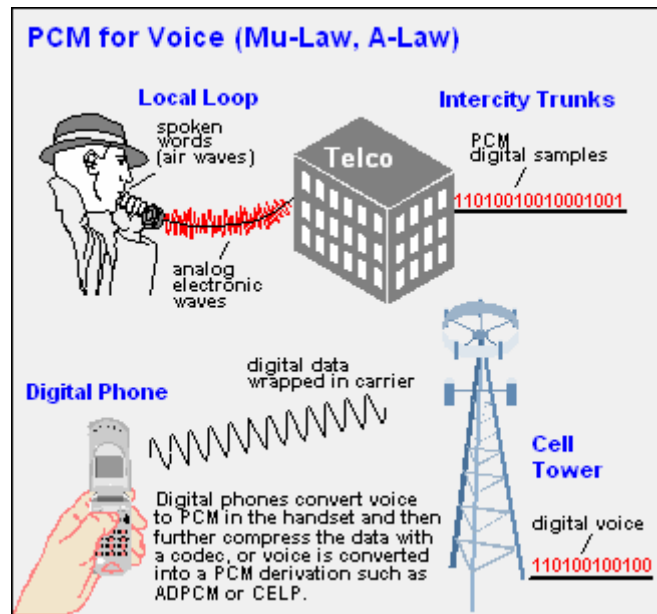
Στη δεκαετία του '30 η ανάγκη να μετατραπούν τα αναλογικά στοιχεία σε ψηφιακή μορφή για υπολογιστή ήταν άγνωστη διότι δεν υπήρχε κανένας υπολογιστής ικανός για να την χρησιμοποιήσει. Εντούτοις ο Alec Reeves χρησιμοποίησε την ιδέα αυτή ως μέσο επικοινωνίας τέλειας πιστότητας. Κανένα λάθος δεν θα μπορούσε να προκύψει σε ένα σύστημα βασισμένο στα δυαδικά "0" και "1". Το 1938 κατέθεσε ένα δίπλωμα ευρεσιτεχνίας στη Γαλλία που εισήγαγε την έννοια της ψηφιακής επικοινωνίας που ονόμασε διαμόρφωση κώδικα παλμού (PCM) το οποίο είναι τώρα σύνηθες για τους υπολογιστές, το ραδιόφωνο, την τηλεόραση, τα CD, recording και τα λοιπά.

Το PCM παρουσιάστηκε στις ΗΠΑ στις αρχές του 1960 όταν οι τηλεφωνικές εταιρίες άρχισαν να μετατρέπουν τη φωνή σε ψηφιακή μορφή για να την μεταφέρουν μεταξύ των πόλεων. Στην πραγματικότητα ο όρος "κώδικας-παλμός" προέρχεται από την αποστολή ή όχι παλμού στην γραμμή. Στην τηλεφωνία, διάφορες ροές PCM μπορούν να πολυπλεχθούν σε ένα μεγαλύτερο σύνολο ροών δεδομένων.

Mu-Law ονομάζεται το Βόρειο-Αμερικάνικο πρότυπο μετατροπής δεδομένων από αναλογική σε ψηφιακή μορφή, ενώ το αντίστοιχο ITU πρότυπο, ονομάζεται A-Law. Και τα δύο πρότυπα χρησιμοποιούν μια companding τεχνική η οποία παρέχει περισσότερα κβαντικά επίπεδα στα χαμηλότερα εύρη παρά στα υψηλότερα εύρη. Η Βόρεια Αμερική και η Ιαπωνία χρησιμοποιούν το mu-Law πρότυπο, ενώ το A-Law πρότυπο χρησιμοποιείται στην Ευρώπη. Το mu-Law προέρχεται από το μ-Law στο οποίο χρησιμοποιείται το Ελληνικό γράμμα μ.

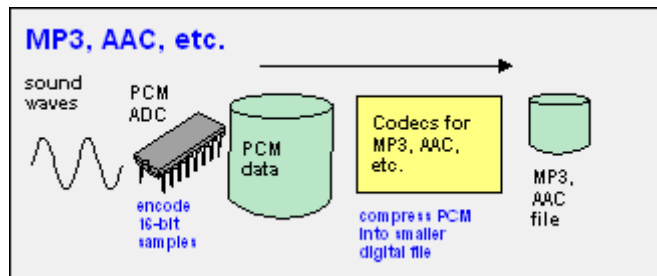


ΣΧΗΜΑ 2.22: Διάγραμμα μ-law και A-law αλγορίθμων

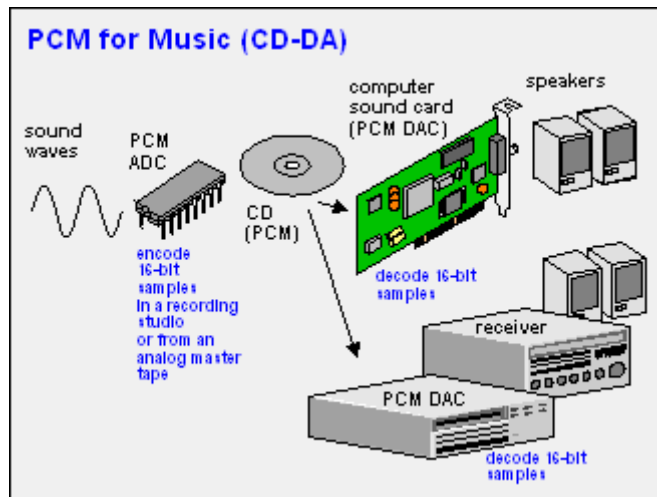


ΣΧΗΜΑ 2.23

Τα μικρόφωνα και τα κυκλώματα των καρτών ήχου παράγουν PCM δείγματα, και όλες οι κάρτες ήχου απαιτούν PCM για έξοδο. Συμπίεσμένες φόρμες ήχου όπως τα MP3 και τα AAC προέρχονται από PCM και οι κάρτες ήχου μετατρέπουν το PCM σήμα σε αναλογικό για τα ηχεία.



ΣΧΗΜΑ 2.24

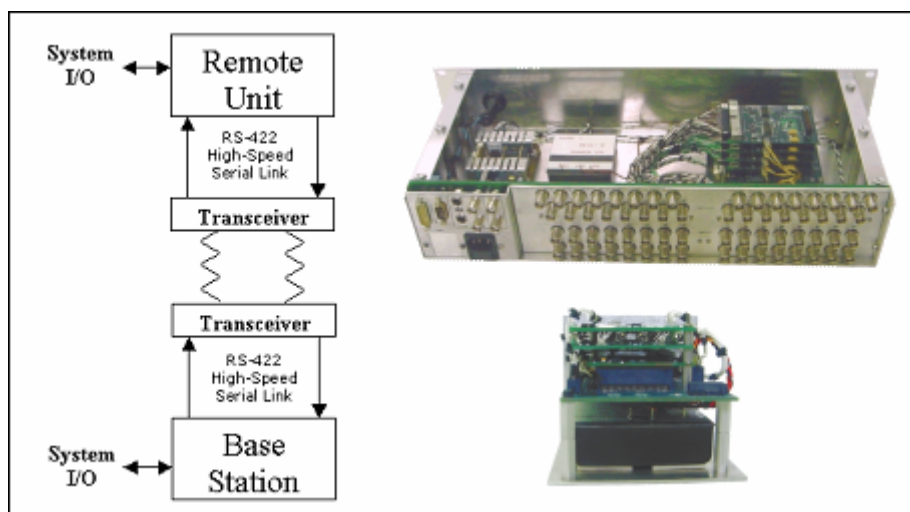


ΣΧΗΜΑ 2.25

PCM είναι ο βασικότερος τρόπος μετατροπής αναλογικών κυμάτων σε ψηφιακά στις συνομιλίες, όπως επίσης και στην μουσική. Στο συμβατικό PCM, πριν ψηφιοποιηθεί, τα αναλογικά στοιχεία μπορούν να υποβληθούν σε επεξεργασία (π.χ. συμπίεση), αλλά μόλις ψηφιοποιηθεί, το σήμα PCM δεν υποβάλλεται σε περαιτέρω επεξεργασία (π.χ. ψηφιακή συμπίεση). Codecs όπως τα MP3 και τα AAC τα οποία συμπιέζουν περαιτέρω τα δεδομένα, εφαρμόζουν αλγόριθμους πάνω στα PCM δείγματα έτσι ώστε να ελαχιστοποιήσουν τυχών αλληλοκαλυπτόμενες συχνότητες όπως επίσης και ήχους τους οποίους δεν αντιλαμβάνεται το ανθρώπινο αφτί.

2.6.1. ΣΥΣΤΗΜΑ PCM

Το σύστημα PCM ένα σύστημα απόκτησης στοιχείων και τηλεμετρίας. Στην τυποποιημένη διαμόρφωσή του, αποτελείται από μια εξ' αποστάσεως μονάδα (remote unit) και έναν σταθμό-βάση (base station) συνδεδεμένα μεταξύ τους με μια υψηλής ταχύτητας σειριακή σύνδεση, από την οποία, η βάση λαμβάνει τηλεμετρία από τη μονάδα. Δεδομένου ότι η τμηματική σύνδεση είναι πλήρως αμφίδρομη, το σύστημα PCM μπορεί επιλεκτικά να χρησιμοποιηθεί σαν σύστημα τηλεμετρίας ή ελέγχου. Είναι σε θέση να μεταφέρει τις παραμέτρους από τη μονάδα πίσω στη βάση, και αντιστρόφως, που χρησιμοποιεί το IRIG συμβατό PCM. Το ποσοστό δεδομένων είναι σημαντικά υψηλότερο από αυτό που παρέχεται με τα συμβατικά ασύρματα modem (**MOD**ulator **DEM**odulator).



ΣΧΗΜΑ 2.26: Block διάγραμμα συστήματος PCM, Base station, Remote Unit.

2.6.1.1. FAILSAFE

Πολλοί άνθρωποι όταν αρχικά ακούνε για failsafe (προστασία), νομίζουν ότι είναι ένα σύστημα το οποίο σχεδιάστηκε για να σώζει πρότυπα, αν και μπορεί, εάν προγραμματιστεί κατάλληλα. Failsafe είναι μια λειτουργία των περισσότερων συστημάτων PCM που επιτρέπει στο σύστημα να κινήσει τα servos προς μια προκαθορισμένη θέση σύμφωνα με την απώλεια σήματος. Συνήθως, υπάρχει μια επιλογή για κάθε κανάλι είτε να μετακινηθεί σε μια προκαθορισμένη θέση είτε να παραμείνει στην τελευταία καθορισμένη κατάσταση. Λόγω της φύσης του failsafe, δεν μπορεί ποτέ να τεθεί εκτός λειτουργίας. Το σύστημα PCM αναγνωρίζει πότε το δεδομένο που έλαβε είναι αβάσιμο, και απλά δεν μπορούμε να το κάνουμε να ενεργήσει επάνω του. Πρέπει να τονίσουμε ότι αυτή η failsafe διαδικασία δεν είναι προεπιλογή για οποιοδήποτε γνωστό PCM. Πρέπει να τεθεί ρητά από το χρήστη.

2.7. ΔΙΑΜΟΡΦΩΣΗ ΘΕΣΗΣ ΠΑΛΜΟΥ (PPM)

Υπάρχουν δύο σημαντικές μέθοδοι που τα σύγχρονα ράδιο συστήματα ελέγχου FM χρησιμοποιούν για να κωδικοποιήσουν σήματα: η διαμόρφωση κώδικα παλμού (Pulse Code Modulation - PCM) και η διαμόρφωση θέσης παλμού (Pulse Position Modulation - PPM). Η περιγραφή της διαμόρφωσης PPM σ' αυτή την ενότητα θα γίνει με τη μορφή σύγκρισης ως προς την διαμόρφωση PCM. Η διαμόρφωση PCM είναι παρόμοια με τη διαμόρφωση PPM δεδομένου ότι οι λαμβανόμενες πληροφορίες δεν καθορίζονται από τη μορφή του παλμού, αλλά έχει το πρόσθετο πλεονέκτημα ότι η ακριβής θέση του παλμού επίσης δεν είναι σημαντική. Η PCM είναι μια πιο προηγμένη μέθοδος κωδικοποίησης που προσφέρει την απόρριψη θορύβου και πρόσθετα χαρακτηριστικά γνωρίσματα. Εάν χρησιμοποιείται κατάλληλα, το μόνο μειονέκτημα της PCM είναι το υψηλό κόστος της.

2.7.1. ΥΠΟΒΑΘΡΟ

Υπάρχει μεγάλη σύγχυση που περιβάλλει την διαμόρφωση PCM για το ποια ακριβώς είναι τα πλεονεκτήματα και ποια τα μειονεκτήματά της. Οι PCM και PPM είναι και οι δύο μέθοδοι κωδικοποίησης που χρησιμοποιούνται για τα σήματα FM, και το PCM δεν είναι μια τελειώς διαφορετική μέθοδος μετάδοσης όπως η AM. Επομένως, το σήμα PCM επηρεάζεται από την ίδια παρεμβολή με ένα σήμα PPM. Ο λόγος που η PCM είναι καλύτερη μέθοδος δεν είναι το γεγονός ότι έχει λιγότερες παρεμβολές αλλά το ότι μπορεί να τις αντιμετωπίσει καλύτερα.

Υπάρχει μια θεμελιώδης διαφορά μεταξύ των σημάτων PCM και PPM. Ένα σήμα PCM είναι ψηφιακό, ενώ ένα σήμα PPM είναι αναλογικό. Σε ένα ψηφιακό σήμα, υπάρχουν δύο ευδιάκριτες καταστάσεις, ανοικτό και κλειστό (on-off), και για να τις μπερδέψει κανείς απαιτείται ένα σημαντικό ποσό παρεμβολής. Στο αναλογικό σήμα, υπάρχει ένας συνεχής αριθμός καταστάσεων που αντιπροσωπεύουν το σήμα, και κάποια παρεμβολή μπορεί να αναγκάσει κάθε κατάσταση να μετατοπιστεί σε μια παρακείμενη. Επομένως, ένας δέκτης PCM θα λειτουργήσει τέλεια, χωρίς βλάβες ή κομμένα πλαίσια, κάτω από πολύ πιο υψηλά επίπεδα παρεμβολών από ένα PPM.

Το ψηφιακό σήμα επιτρέπει, επίσης, στο σύστημα PCM να εκτελέσει πρόσθετες λειτουργίες, επειδή δεδομένα μπορούν να προστεθούν στο σήμα. Αρχικά, οι PCM ραδιοπομποί στέλνουν αυτό που είναι γνωστό ως "checksum" (άθροισμα ελέγχου μνήμης), το οποίο προσθέτει τον πλεονασμό στο σύστημα με το να επιτρέπει στον δέκτη να ελέγχει αν τα δεδομένα που λαμβάνει είναι αλλοιωμένα. Στην πραγματικότητα, το checksum είναι τόσο αποτελεσματικό που είναι σχεδόν αδύνατο για έναν δέκτη να δεχτεί κατά λάθος αβάσιμα στοιχεία. Εάν ο δέκτης καθορίσει ότι το στοιχείο είναι αβάσιμο, θα συγκρατήσει την τελευταία κατάσταση. Σχεδόν όλα τα ραδιοσυστήματα PCM μεταδίδουν επίσης το "failsafe" (ασφαλές σε περίπτωση βλάβης) δεδομένο, το οποίο χρησιμοποιείται για να θέσει τις σερβοθέσεις όταν κρίνεται το σήμα αλλοιωμένο για περισσότερο από περίπου μισό δευτερόλεπτο.

2.7.2. ΔΟΚΙΜΗ ΕΥΡΟΥΣ

Η δοκιμή εύρους ενός συστήματος PCM είναι διαφορετική από ένα σύστημα PPM, επειδή τα συστήματα PCM δεν παρουσιάζουν οποιαδήποτε "glitch" (παρασιτικό σήμα) επάνω στην απώλεια σήματος. Όταν ένας δέκτης PCM πλησιάζει στο όριο του εύρους σταματά αρχικά τα "frames", ή τις σερβο εντολές θέσης. Εάν αρκετά frames πέσουν σε μια σειρά, ο ραδιοπομπός πηγαίνει σε failsafe έως ότου επανακτάται το σήμα. Ποτέ τα servos δεν τρέμουν ή αναπηδούν όπως τα συστήματα PPM.

2.7.3. ΠΑΡΕΜΒΟΛΕΣ ΣΤΑ PPM ΚΑΙ PCM

Το PCM καλύπτει την παρεμβολή για τρεις λόγους: failsafe, checksum και ποιότητα σημάτων. Επειδή το σήμα PCM είναι ψηφιακό, είναι λιγότερο ευαίσθητο στην παρεμβολή και η παρεμβολή που οδηγεί σε παρασιτικά σήματα στο PPM δεν θα έχει καμία επίδραση στο PCM. Η ενεργοποίηση failsafe είναι ένα σίγουρο σημάδι ότι η παρέμβαση είναι αρκετά υψηλή ώστε ένας δέκτης PPM θα είχε κατασταθεί πολύ καιρό πριν άχρηστος.

Το PPM αποκρίνεται περισσότερο και είναι ακριβέστερο από το PCM. Αυτό είναι ένα κατάλοιπο από τα παλαιότερα συστήματα PCM, τα οποία ήταν κατώτερα από τα συστήματα που χρησιμοποιούμε σήμερα. Τα σύγχρονα συστήματα PCM έχουν υψηλότερη ανάλυση. Στη χειρότερη περίπτωση, τα σύγχρονα συστήματα PCM ανταποκρίνονται τόσο όσο τα αντίστοιχα PPM τους.

2.7.4. ΑΣΦΑΛΕΙΑ

Το PCM έχει διάφορα πλεονεκτήματα όσον αφορά την ασφάλεια σε σύγκριση με τα PPM, όλα αυτά σχετικά με την προβλεψιμότητα της διαφορετικής συμπεριφοράς σημάτων των PCM. Όταν ένας δέκτης PCM χάνει το σήμα, ή παίρνει ένα με σημαντική παρεμβολή, είτε θα συγκρατήσει τις σερβο θέσεις είτε θα τις κινήσει προς τις προκαθορισμένες ασφαλείς θέσεις. Αυτό έρχεται σε αντίθεση με το PPM, που θα μπορούσε να κάνει οτιδήποτε, από το κράτημα των servos μέχρι την μετακίνηση τους στο όριο της διαδρομής τους.

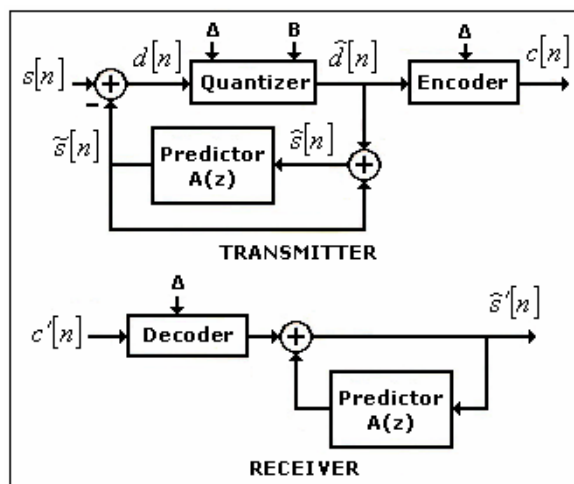
ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ ΚΩΔΙΚΟΠΟΙΗΣΗ WAVEFORM ΜΕ ΣΤΑΘΕΡΗ ΠΡΟΒΛΕΨΗ

3.1. ΕΙΣΑΓΩΓΗ

Οι προσαρμοστικές και συμπίεσης/αποσυμπίεσης τεχνικές κβάντισης αρχίζουν να εκμεταλλεύονται την ιδιότητα επικάλυψης θορύβου της ακουστικής προοπτικής και την διακύμανση συλλαβικής ενέργειας των σημάτων ανθρώπινης φωνής. Δεν εκμεταλλεύονται όμως το γεγονός ότι τα σήματα φωνής παράγονται από το ανθρώπινο φωνητικό σύστημα

Για να εκμεταλλευτούμε τις επιδράσεις του φωνητικού συστήματος, το σήμα φωνής μπορεί να παρατηρηθεί μέσα από δύο διαφορετικές οπτικές, την μακροχρόνια και τη βραχυχρόνια. Η μακροχρόνια προοπτική λαμβάνει υπόψη τις μέσες χρονικά εξαρτώμενες ιδιότητες της φωνής και οδηγεί φυσικά σε μη προσαρμοστικές στρατηγικές κωδικοποίησης. Η βραχυχρόνια προοπτική λαμβάνει υπόψη τις αργές χρονικά εξαρτώμενες ιδιότητες της φωνής οι οποίες παράγονται από τις μηχανικές ιδιότητες του φωνητικού συστήματος. Αυτό οδηγεί σε ένα βραχυχρόνιο σταθερό στατιστικό μοντέλο το οποίο μπορούμε να εκμεταλλευθούμε καλύτερα μέσω προσαρμοστικών στρατηγικών κωδικοποίησης φωνής.

Η διαφορική κωδικοποίηση αναφέρεται στην κωδικοποίηση των διαφορών μεταξύ των δύο σημάτων παρά των ίδιων των σημάτων. Η αρχική μακροχρόνια ιδιότητα της φωνής που μπορούμε να εκμεταλλευτούμε με ένα μη προσαρμοστικό διαφορικό κωδικοποιητή είναι η σχετικά υψηλή συσχέτιση γειτονικών δειγμάτων. Αυτό είναι ένα αποτέλεσμα των ολικών χαμηλοπερατών χαρακτηριστικών της φωνής λόγω των συνδυασμένων επιδράσεων του σχήματος παλμού της γλωττίδας, της ακουστική ιδιότητα φιλτραρίσματος του φωνητικού συστήματος και της ακουστικής επίδρασης απόσβεσης ακτινοβολίας λόγω της ακτινοβολίας από το στόμα. Λόγω της υψηλής συσχέτισης μεταξύ των παρακείμενων δειγμάτων του σήματος φωνής (ειδικά της φωνούμενης ομιλίας), κάθε δείγμα φωνής μπορεί συνήθως να υπολογιστεί από τα προηγούμενα δείγματα φωνής είτε τα κωδικοποιημένα δείγματα φωνής. Επί πρόσθετα λόγω των προσεγγιστικών γραμμικών ιδιοτήτων του ακουστικού φίλτρου του φωνητικού συστήματος, ο απαιτούμενος υπολογισμός μπορεί να γίνει με ένα γραμμικό φίλτρο.



ΣΧΗΜΑ 3.1: Block διάγραμμα ενός DPCM.

Στην διαφορική διαμόρφωση, ο μακροχρόνιος πλεονασμός της κυματομορφής της φωνής απομακρύνεται όσο περισσότερο γίνεται και το εναπομείναν σήμα κβαντίζεται και κωδικοποιείται χρησιμοποιώντας τεχνικές παρόμοιες με αυτές που χρησιμοποιήθηκαν στο Κεφάλαιο 2. Αυτό πραγματοποιείται μορφοποιώντας ένα λάθος σήματος ή ένα σήμα διαφοράς αφαιρώντας έναν υπολογισμό του σήματος από το αρχικό σήμα. Ο υπολογισμός γενικά επιτυγχάνεται από μια γραμμική πρόβλεψη η οποία υπολογίζει τα τρέχοντα δείγματα σήματος από ένα γραμμικό συνδυασμό ενός οι περισσότερων δειγμάτων. Το σχήμα 3.1

δείχνει το block διάγραμμα ενός DPCM. Η κύρια πηγή της βελτίωσης παρουσίασης για τους διαφορικούς κωδικοποιητές είναι το μειωμένο δυναμικό εύρος του σήματος εισόδου στο οποίο η διαφορά σήματος $d[n]$ είναι γενικά μικρότερη σε πλάτος από το αρχικό σήμα. Από τη στιγμή που ο θόρυβος κβάντισης είναι ανάλογος του μεγέθους βήματος, ένα σήμα με μικρότερο δυναμικό εύρος μπορεί να κωδικοποιηθεί ακριβέστερα με δοσμένο αριθμό επιπέδων κβάντισης.

Τα βασικά θέματα στη διαφορική κωδικοποίηση είναι οι ιδιότητες του συστήματος πρόβλεψης (predictor) και η κβάντιση του σήματος λάθους πρόβλεψης. Η απλούστερη μορφή ενός διαφορικού κωδικοποιητή είναι αυτή η οποία χρησιμοποιεί έναν τροποποιημένο predictor και ένα ομοιόμορφα τροποποιημένο predictor. Τα περισσότερα πολύπλοκα DPCM αποτελούν συστήματα με προσαρμοστικούς κβαντιστές (ADPCM), προσαρμοστικούς predictors (APC) ή και τα δύο.

3.2. ΒΑΣΙΚΟ DPCM

Το σχήμα 3.1 δείχνει ένα block διάγραμμα ενός βασικού DPCM με ένα σταθερό predictor και ένα σταθερό ομοιόμορφο predictor. Σε αυτό το σχήμα, το $\hat{s}[n]$ δηλώνει τον υπολογισμό του $s[n]$ το οποίο εξασφαλίζεται από τα προηγούμενα κωδικοποιημένα δείγματα ως:

$$\tilde{s}[n] = \sum_{i=1}^P \alpha_i \hat{s}[n-1]$$

όπου $\alpha_i, i=1\dots,P$ είναι οι συντελεστές του γραμμικού predictor $A(z)$ και $\hat{s}[n]$ είναι τα προηγούμενα κωδικοποιημένα δείγματα. Ο predictor, $A(z)$, ορίζεται ως:

$$A(z) = \sum_{i=1}^P \alpha_i z^{-i}$$

όπου P ο βαθμός του predictor. Οι συντελεστές του predictor συνήθως υπολογίζονται μέσω μιας γραμμικής προσαρμοστικής ανάλυσης της φωνής, όπου οι α_i υπολογίζονται ελαχιστοποιώντας το λάθος πρόβλεψης το οποίο ορίζεται από:

$$E = \sum_n (s[n] - \tilde{s}[n])^2$$

Το σήμα λάθους πρόβλεψης, $d[n]$, το οποίο δίδεται από:

$$d[n] = s[n] - \tilde{s}[n]$$

είναι η ποσότητα η οποία κβαντίζεται και κωδικοποιείται. Το κβαντισμένο σήμα λάθους μπορεί να αντιπροσωπευθεί ως:

$$\tilde{d}[n] = Q(d[n]) = d[n] + q_d[n]$$

όπου $q_d[n]$ είναι το λάθος κβάντισης. Όπως δείχνει το σχήμα 4.1 η είσοδος είναι:

$$\hat{s}[n] = s[n] + \tilde{d}[n]$$

Συνδυάζοντας τις εξισώσεις 3.4, 3.5 και 3.6, το $\hat{s}[n]$, μπορεί να γραφεί ως:

$$\hat{s}[n] = s[n] + q_d[n]$$

το οποίο είναι το κβαντισμένο σήμα. Αυτό δείχνει ότι το σήμα λάθους εξόδου σε αυτό το σύστημα είναι ίδιο όπως στο λάθος κβάντισης του σήματος πρόβλεψης λάθους, $d[n]$. Αυτό είναι σημαντικό διότι το $d[n]$ έχει μικρότερο δυναμικό εύρος (διακύμανση) απ' ό,τι το σήμα φωνής, και συνεπώς το ανταποκρινόμενο λάθος κβάντισης θα είναι μικρότερο. Η βελτίωση του SNR για τα DPCM συγκρινόμενη με το PCM είναι ανάλογη του κέρδους πρόβλεψης και ορίζεται ως:

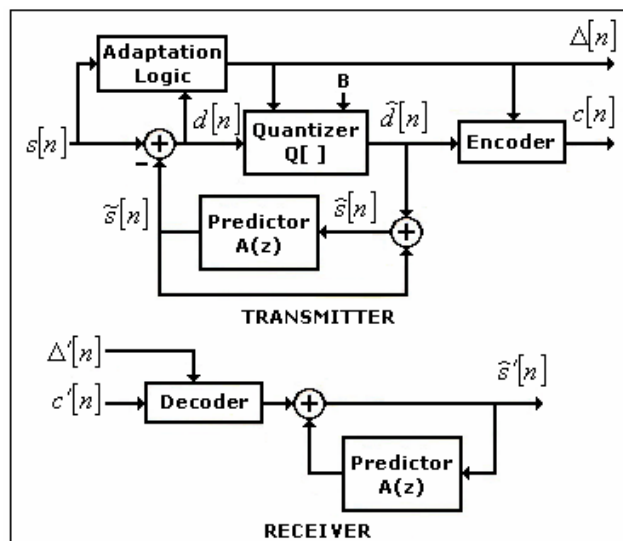
$$G_p = \frac{\sigma_s^2}{\sigma_d^2}$$

όπου το σ_s^2 είναι η διακύμανση του σήματος φωνής και το σ_d^2 δηλώνει την διακύμανση του σήματος λάθους πρόβλεψης $d[n]$. Είναι προφανές ότι καλύτερη πρόβλεψη μπορεί να βελτιώσει τη λειτουργία της DPCM διότι μειώνει την διακύμανση του $d[n]$. Έτσι μια σημαντική παράμετρος της DPCM είναι ο predictor, η οποία καθορίζεται από το βαθμό και τους συντελεστές του predictor. Παρόλο που αυξάνοντας το βαθμό του predictor συνήθως καταλήγουμε σε υψηλότερο SNR, ο predictor διευθετεί υψηλότερα από 4 ή 5 βελτιώσεις απόδοσης οριακού SNR.

Στο βασικό DPCM χρησιμοποιείται ένας σταθερός ομοιόμορφος κβαντιστής. Οι παράμετροι του κβαντιστή επιλέγονται έτσι ώστε να ταιριάζουν το δυναμικό εύρος με την κατανομή του $d[n]$. Ο πίνακας 3.1 δείχνει τις παραμέτρους του βασικού DPCM οι οποίες συμπεριλαμβάνουν τις παραμέτρους των ομοιόμορφων κβαντιστών και ενός σταθερού predictor. Αν ο σταθερός DPCM predictor είναι κατάλληλα σχεδιασμένος και υλοποιημένος, το DPCM μπορεί να έχει ως αποτέλεσμα αισθητά καλύτερη απόδοση ως προς το SNR απ' ό τι το PCM. Καλύτερη απόδοση είναι δυνατή χρησιμοποιώντας προσαρμοστική πρόβλεψη.

parameters	name	range	typical value
predictor order	P	1 - 10	1
quantizer max. level	Smax	0 - 32767	32767
number of bits	B	1 - 16	6
quantization type	--	--	mid-tread or mid-riser

ΠΙΝΑΚΑΣ 3.1: Παράμετροι ενός DPCM.

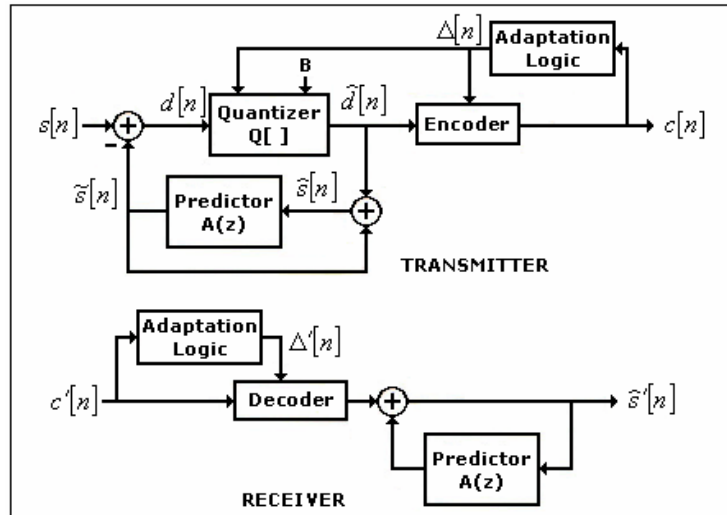


ΣΧΗΜΑ 3.2: Block διάγραμμα ενός DPCM-AQF

3.3. DPCM ΜΕ ΠΡΟΣΑΡΜΟΣΤΙΚΗ ΚΒΑΝΤΙΣΗ (ADPCM)

Τα DPCM με σταθερή κβάντιση παρέχουν μια μέση, περίπου στα 6 dB, βελτίωση στο SNR σε σχέση με το PCM με τον ίδιο αριθμό bit ανά δείγμα. Όπως και στους APCM, οι προσαρμοστικοί κβαντιστές μπορούν να χρησιμοποιηθούν για να πετύχουν περαιτέρω βελτίωση στις διαδικασίες κωδικοποίησης. Τα σχήματα 3.2 και 3.3 παρουσιάζουν τα block διαγράμματα ενός προσαρμοστικού DPCM (ADPCM) με feed-forward προσαρμοστικό κβαντιστή (DPCM-AQF) και feedback (ανάδραση) προσαρμοστικό κβαντιστή (DPCM-AQB). Οι ADPCM μπορούν να χρησιμοποιήσουν οποιονδήποτε από τους προσαρμοστικούς αλγορίθμους. Για παράδειγμα η προσαρμογή μεγέθους βήματος (step size adaptation) της

εξίσωσης 3.16 μπορεί να χρησιμοποιηθεί και στα δυο, feed-forward και feed back, συστήματα. Μια δυνατή προσέγγιση στα DPCM-AQF είναι να χρησιμοποιηθεί η διακύμανση του $d[n]$ για να προσαρμοστεί το μέγεθος βήματος του κβαντιστή. Αφού η διακύμανση του $d[n]$ είναι ανάλογη της διακύμανσης του αρχικού σήματος εισόδου, η προσαρμογή μπορεί επίσης να βασιστεί στο σήμα φωνής της εισόδου. Στα DPCM-AQB, η διακύμανση υπολογίζεται από το $d[n]$, δηλαδή τα κβαντισμένα δείγματα του $d[n]$.



ΣΧΗΜΑ 3.3: Block διάγραμμα ενός DPCM-AQB

parameters	name	range	typical value
predictor order	P	1 - 10	5
rectangular window size	M (FF only)	1 - 500	50
exponential window	α (FB only)	0.1 - 0.99	0.95
number of bits	B	1 - 16	5
scaling step size	Δ_0	--	--
min. step size	Δ_{min}	50 - 500	100
max. step size	Δ_{max}	100 - 10000	5000

ΠΙΝΑΚΑΣ 3.2: Παράμετροι ενός ADPCM. Όπου FF σημαίνει feed forward και FB σημαίνει feedback.

Όπως φαίνεται στον Πίνακα 3.2, οι παράμετροι του DPCM περιλαμβάνουν τις παραμέτρους πρόβλεψης και τις παραμέτρους του κβαντιστή. Όπως και στον PCM-AQF, ο DPCM-AQF χρησιμοποιεί ένα ορθογώνιο παράθυρο μεγέθους M για τον υπολογισμό της διακύμανσης. Αυτή η διακύμανση εκσυγχρονίζεται ανά M δείγματα. Στα προσαρμοστικά συστήματα ADPCM με ανάδραση (DPCM-AQB), ένα εκθετικό παράθυρο με παράμετρο α χρησιμοποιείται.

Για να επιτύχουμε περαιτέρω βελτίωση του SNR και της υποκειμενικής ποιότητας, το κέρδος πρόβλεψης των συστημάτων DPCM μπορεί να βελτιωθεί χρησιμοποιώντας προσαρμοστικές τεχνικές πρόβλεψης.

3.4. ΔΙΑΜΟΡΦΩΣΗ ΔΕΛΤΑ (DM)

Η διαμόρφωση δέλτα (Delta modulation - DM) είναι ένας συγκεκριμένος τύπος DPCM με ένα πρώτης τάξης προβλεπτή ($P=1$) και ένα 1-bit κβαντιστή ($B=1$). Η διαμόρφωση δέλτα, για να αντισταθμίσει τη χονδροειδή κβάντιση και να διατηρήσει την ποιότητα του κωδικοποιημένου σήματος, ο ρυθμός δειγματοληψίας του DM σήματος εισόδου, f_{DM} , επιλέγεται να είναι αρκετές φορές μεγαλύτερος από την συχνότητα Nyquist έτσι ώστε:

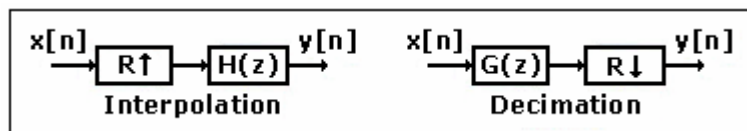
$$F_{DM} = 2Rf_N$$

όπου R είναι ο λόγος υπερδειγματοληψίας (oversampling ratio) και f_N ο ρυθμός Nyquist. Τέτοιος υψηλός ρυθμός δειγματοληψίας μπορεί να επιτευχθεί δειγματοληπώντας το αναλογικό σήμα στον επιθυμητό ρυθμό ή παρεμβάλλοντας το σήμα φωνής να αποκτήσει τον επιθυμητό ρυθμό δειγματοληψίας. Η λειτουργία παρεμβολής (interpolation), η οποία φαίνεται στο σχήμα 3.4, αποτελείται από δειγματοληπιμένο μέρος (upsample) του σήματος εισόδου ακολουθούμενο από το φίλτρο παρεμβολής. Το σήμα εξόδου μπορεί να θεωρηθεί ότι είναι μια υπερδειγματοληπιμένη εκδοχή του σήματος εισόδου. Τέτοιοι υψηλοί ρυθμοί δειγματοληψίας καταλήγουν σε μια υψηλή συσχέτιση των δειγμάτων του δειγματοληπιμένου σήματος φωνής. Αυτό, αντιστρόφως, καταλήγει σε μια πολύ μικρή διακύμανση του σήματος πρόβλεψης λάθους $d[n]$, το οποίο οδηγεί σε υψηλή πρόβλεψη κέρδους G_p . Το σχήμα 3.5 παρουσιάζει το block διάγραμμα μιας DM. Στην DM, η πρόβλεψη είναι της μορφής

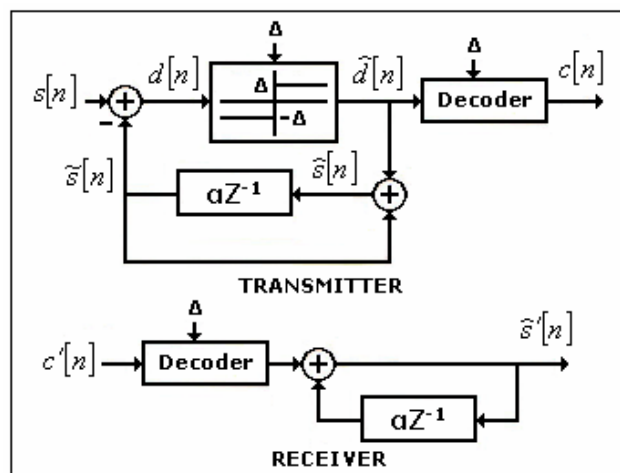
$$p(z) = az^{-1}$$

όπου το a επιλέγεται να είναι μικρότερο του 1. Η τιμή πρόβλεψης ακριβώς 1 συνήθως παραλείπεται διότι καταλήγει σε ένα φίλτρο πλήρους ολοκληρωτή σύνθεσης, το οποίο τείνει να συσσωρεύει λάθη κωδικοποίησης.

Το κύριο πλεονέκτημα της DM είναι η απλότητα της. Στη DM, το bit rate είναι ίσο με το ρυθμό δειγματοληψίας του DM σήματος εισόδου F_{DM} . Προφανώς, οι σημαντικές παράμετροι της λειτουργίας της DM είναι ο λόγος υπερδειγματοληψίας και ο τύπος κβάντισης (σταθερός ή προσαρμοστικός).



ΣΧΗΜΑ 3.4



ΣΧΗΜΑ 3.5: Block διάγραμμα ενός DM

3.4.1. ΓΡΑΜΜΙΚΗ ΔΙΑΜΟΡΦΩΣΗ ΔΕΛΤΑ (LDM)

Η Γραμμική Διαμόρφωση Δέλτα (Linear Delta Modulation - LDM) είναι η απλούστερη μορφή DM όπου ένας σταθερός, δύο επιπέδων, κβαντιστής χρησιμοποιείται. Το σχήμα 3.6 παρουσιάζει αυτόν τον 1 bit κβαντιστή με μέγεθος βήματος Δ του οποίου η έξοδος ορίζεται ως:

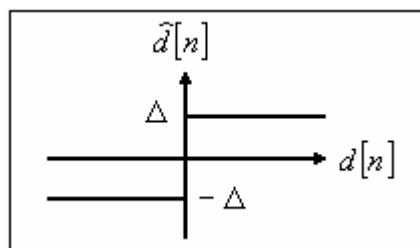
$$\tilde{d}[n] = \begin{cases} \Delta & \text{if } d[n] \geq 0 \\ -\Delta & \text{if } d[n] < 0 \end{cases}$$

Το LDM προσπαθεί να ακολουθήσει την κυματομορφή του σήματος φωνής εισόδου σε μια γραμμική μορφή. Η κλίση του LDM δίνεται από το λόγο Δ/T όπου T είναι η περίοδος δειγματοληψίας. Επομένως η κλίση μπορεί να ελεγχθεί αλλάζοντας το Δ και το T . Επειδή η μείωση του T έχει ως αποτέλεσμα υψηλότερη, δείγμα προς δείγμα, συσχέτιση, το κέρδος πρόβλεψης μπορεί να αυξηθεί μειώνοντας το T . Το SNR μπορεί να αυξηθεί περίπου 9 dB για κάθε διπλασιασμό του λόγου υπερδειγματοληψίας R , το οποίο ανταποκρίνεται στη μείωση του T σε υποδιπλασιάζοντας και διπλασιάζοντας το bit rate.

Υπάρχουν δύο τύποι παραμόρφωσης οι οποίοι αναμειγνύονται στην διαμόρφωση δέλτα (DM): κοκκώδης θόρυβος και θόρυβος υπερφόρτωσης κλίσης. Ο κοκκώδης θόρυβος είναι το αποτέλεσμα των ταλαντώσεων της DM εξόδου γύρω από την τιμή που στοχεύουμε. Η τιμή του μεγέθους βήματος Δ χρειάζεται να είναι μικρή για να μειώσει τον κοκκώδη θόρυβο. Από την άλλη μεριά, το Δ χρειάζεται να είναι μεγάλο για να εμποδιστεί η υπερφόρτωση της κλίσης η οποία απαντάται όταν η κλίση του σήματος είναι μεγαλύτερη από την κλίση της LDM. Έτσι η σωστή επιλογή του Δ είναι ένας συμβιβασμός και είναι κρίσιμη για την παρουσίαση του LDM. Οι παράμετροι της LDM περιλαμβάνουν ένα μέγεθος βήματος κβαντιστή, Δ , τον συντελεστή πρόβλεψης, α , και του λόγου υπερδειγματοληψίας, R . Ο πίνακας 3.3 δίνει μερικές τυπικές τιμές γι' αυτές τις παραμέτρους.

parameters	name	range	typical value
oversampling ratio	R	1 - 8	4
step size	Δ	100 - 1000	500
predictor parameter	α	0.1 - 0.99	0.85

ΠΙΝΑΚΑΣ 3.3: Παράμετροι ενός LDM



ΣΧΗΜΑ 3.6: Ένας 1bit κβαντιστής χρησιμοποιούμενος σε ένα LDM

3.4.2. ΠΡΟΣΑΡΜΟΣΤΙΚΗ ΔΙΑΜΟΡΦΩΣΗ ΔΕΛΤΑ (ADM)

Από τη στιγμή που οι απαιτήσεις για το Δ ώστε να μειωθούν και ο κοκκώδης θόρυβος και ο θόρυβος υπερφόρτωσης κλίσης είναι αντιφατικοί, είναι επιθυμητό ένα προσαρμοστικό μέγεθος βήματος. Στην προσαρμοστική διαμόρφωση δέλτα (ADM), το μέγεθος βήματος αλλάζει δυναμικά για να ικανοποιήσει και τις δύο απαιτήσεις. Τέτοιο βήμα μεταβαλλόμενο με το χρόνο μπορεί να χρησιμοποιηθεί για να εμποδίσει και τις δύο παραμορφώσεις θορύβων ταιριάζοντας την κλίση της DM στην κλίση της κυματομορφής εισόδου. Για να διατηρήσουμε την απλότητα του κωδικοποιητή και να αποφύγουμε την μετάδοση της πληροφορίας κάθε πλευράς, προτιμάται η προσαρμοστική ανάδραση. Το σύστημα το οποίο αναφέρεται ως ADM έχει ένα Jayant τύπο κβαντιστή του οποίου ο κανόνας του μεγέθους βήματος δίδεται από την:

$$\Delta[n] = M[n]\Delta[n-1]$$

όπου το $M[n]$ παίρνει μια από τις δύο τιμές σύμφωνα με τον παρακάτω κανόνα

$$M[n] = \begin{cases} M_1 & \text{if } c[n] = c[n-1] = c[n-2] \\ M_2 & \text{otherwise} \end{cases}$$

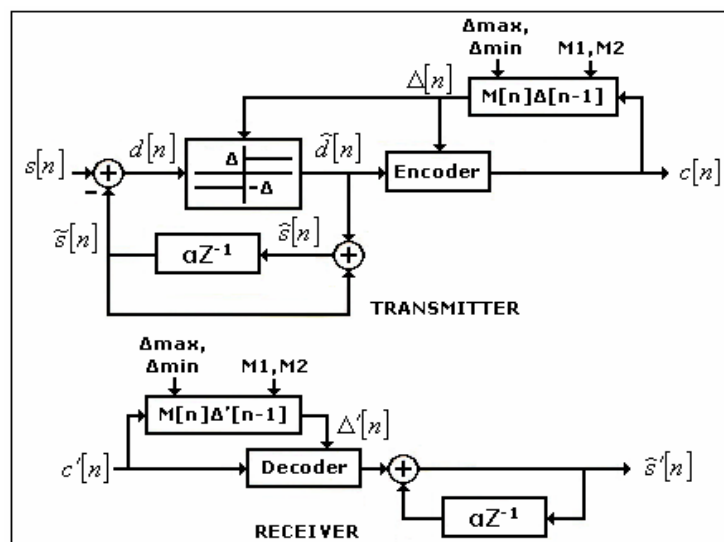
όπου $c[n]$ είναι ο κώδικας εξόδου την στιγμή n . Η κατάσταση $c[n]=c[n-1]=c[n-2]$ δείχνει τρεις συνεχείς αυξομειώσεις του σήματος. Αυτό απαιτεί ένα μεγαλύτερο μέγεθος σήματος για να εμποδίσει μια δύσκολη υπερφόρτωση κλίσης. Γενικά, τα M_1 και M_2 επιλέγονται ώστε $M_1 > 1$ $M_2 < 1$ και

$$M_1 M_2 \leq 1$$

για σταθερότητα. Όπως σε άλλα προσαρμοστικού μεγέθους βήματος συστήματα, ελάχιστες και μέγιστες τιμές μεγέθους βήματος, τα Δ_{min} και Δ_{max} χρησιμοποιούνται στο ADM. Το σχήμα 3.7 δείχνει ένα block διάγραμμα ενός συστήματος ADM. Όπως φαίνεται στον πίνακα 3.4 το ADM χαρακτηρίζεται από έξι παραμέτρους: τον λόγο υπερδειγματοληψίας, R , την τιμή της παραμέτρου του predictor, α , τις τιμές των παραμέτρων ελέγχου του κβαντιστή M_1 και M_2 και τις ελάχιστες και μέγιστες τιμές του μεγέθους βήματος Δ_{min} και Δ_{max} .

parameters	name	range	typical value
oversampling ratio	R	1 - 8	4
predictor parameter	α	0.1 - 0.99	0.85
max. step size multiplier	M_1	1 - 10	1.2
min. step size multiplier	M_2	0 - 0.99	0.7
min. step size	Δ_{min}	10 - 500	20
max. step size	Δ_{max}	100 - 5000	2000

ΠΙΝΑΚΑΣ 3.4: Παράμετροι ενός ADM



ΣΧΗΜΑ 3.7: Block διάγραμμα ενός ADM

4.4.3. ΣΥΝΕΧΗΣ ΜΕΤΑΒΛΗΤΗΣ ΚΛΙΣΗΣ ΔΙΑΜΟΡΦΩΣΗ ΔΕΛΤΑ (CVSD)

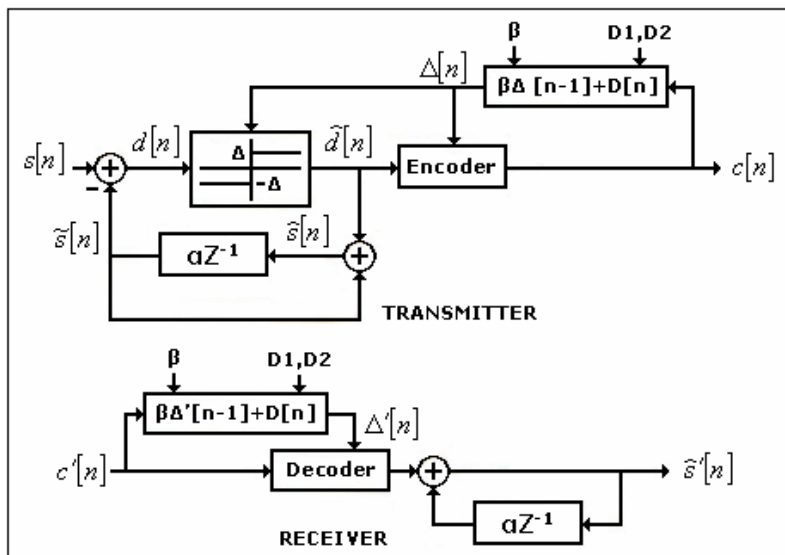
Μια άλλη γνωστή διαμόρφωση δέλτα είναι η συνεχής μεταβλητής κλίσης διαμόρφωση δέλτα (Continuously Variable Slope Delta modulation - CVSD), η οποία έχει ένα διαφορετικό κανόνα για το μέγεθος βήματος κβάντισης από ότι το ADM. Στο CVSD το $\Delta[n]$ υπολογίζεται ως:

$$\Delta[n] = \beta \Delta[n-1] + D[n]$$

όπου $D[n]$ ισούται με μια από τις δύο σταθερές, ανάλογα με το αν οι τρεις τελευταίες τιμές του $c[n]$ είναι ίδιες. Η τιμή του $D[n]$ προσδιορίζεται από

$$D[n] = \begin{cases} D_1 & \text{if } c[n-1] = c[n-2] = c[n-3] \\ D_2 & \text{otherwise} \end{cases}$$

όπου $D_1 \gg D_2 > 0$. Αν η έξοδος αυξομειώνεται για τρία συνεχόμενα δείγματα, το μέγεθος βήματος αυξάνεται. Διαφορετικά, το μέγεθος βήματος μειώνεται κατά ένα λόγο ο οποίος καθορίζεται από το β . Τιμές του β οι οποίες βρίσκονται κοντά στο 1 οδηγούν σε πιο αργές διακυμάνσεις του μεγέθους βήματος. Από την άλλη, μια μικρότερη τιμή του β και μια μεγάλη του D_2 μπορεί να χρησιμοποιηθεί για στιγμιαία προσαρμογή.



ΣΧΗΜΑ 3.8: Block διάγραμμα ενός CVSD

Το σχήμα 3.8 δείχνει το block διάγραμμα ενός CVSD. Στο CVSD, το μέγεθος βήματος είναι έμφυτο το οποίο οριοθετείται από την εξίσωση προσαρμογής 3.16. Το ελάχιστο μέγεθος βήματος για το CVSD είναι $D_2/(1-b)$, και το ανταποκρινόμενο μέγιστο μέγεθος βήματος δίδεται ως $D_1/(1-b)$. Το CVSD γι' αυτό το λόγο χαρακτηρίζεται από 5 παραμέτρους: το λόγο της δειγματοληψίας του σήματος εισόδου, την τιμή της παραμέτρου του predictor, α , την τιμή της παραμέτρου του βήματος ολοκληρωτή β , το ελάχιστο μέγεθος βήματος παραμέτρου, D_2 και το ανταποκρινόμενο μέγιστο μέγεθος βήματος, D_1 . D_1/D_2 είναι γνωστό ως ο λόγος διαστολής ο οποίος προσδιορίζει το δυναμικό εύρος του κβαντιστή. Ο πίνακας 3.5 περιέχει το εύρος των παραμέτρων και μερικές τυπικές τιμές τους.

parameters	name	range	typical value
oversampling ratio	R	1 - 8	4
predictor parameter	α	0.1 - 0.99	0.85
step integrator	β	0.1 - 0.99	0.95
step size parameter	D_1	1 - 1000	100
step size parameter	D_2	0 - 100	10

ΠΙΝΑΚΑΣ 3.5: Παράμετροι ενός CVSD

ΚΕΦΑΛΑΙΟ ΤΕΤΑΡΤΟ ΚΩΔΙΚΟΠΟΙΗΣΗ WAVEFORM ΜΕ ΠΡΟΣΑΡΜΟΣΤΙΚΗ ΠΡΟΒΛΕΨΗ

4.1. ΕΙΣΑΓΩΓΗ

Αυτό το κεφάλαιο αναφέρεται στην κατηγορία των διαφορικών κωδικοποιητών με προσαρμοστική πρόβλεψη. Τέτοιοι κωδικοποιητές εμφανίζονται σε δύο μορφές. Πρώτον, εμφανίζονται ως διαφορικά συστήματα DPCM που χρησιμοποιούν έναν προσαρμοστικό γραμμικό προβλεπτή για να εντοπίσουν τα βραχυπρόθεσμα στατικά στατιστικά στοιχεία της κυματομορφής της φωνής. Δεύτερον, εμφανίζονται σαν κωδικοποιητές φωνής που διεγείρονται από την κυματομορφή της φωνής (waveform-excited vocoders). Τέτοια συστήματα, τα οποία παραδοσιακά ονομάζονται κωδικοποιητές προσαρμοστικής πρόβλεψης (adaptive-predictive coders – APCs), μπορούν να λειτουργήσουν με υψηλότερη ποιότητα από τους κωδικοποιητές διεγερσης κλίσης (pitch-excited coders). Τα APCs είναι σημαντικά διότι είναι ο ζωτικός συνδετικός κρίκος μεταξύ των κωδικοποιητών κυματομορφής και των παραμετρικών κωδικοποιητών.

Οι διαφορικοί κωδικοποιητές προσπαθούν να εκμεταλλευτούν την βραχυπρόθεσμη προβλεψιμότητα του σήματος φωνής. Επειδή τα σήματα φωνής είναι μη στατικά, οι διαφορικοί PCM με σταθερή πρόβλεψη μπορούν να επιτύχουν μόνο περιορισμένο κέρδος πρόβλεψης. Η προσαρμοστική πρόβλεψη και η προσαρμοστική κβάντιση μπορούν να συνδυαστούν για να παραχθεί μια πιο εκλεπτυσμένη κατηγορία κωδικοποιητών κυματομορφής φωνής. Οι κωδικοποιητές προσαρμοστικής πρόβλεψης (adaptive-predictive coders) είναι ένα σημαντικό μέλος της κατηγορίας αυτής. Αυτοί οι κωδικοποιητές χρησιμοποιούν τη θεωρία γραμμικής πρόβλεψης (linear prediction – LP) για να αυξήσουν το κέρδος πρόβλεψης. Στους APC, οι παράμετροι πρόβλεψης λαμβάνονται είτε από τη φωνή εισόδου είτε από την αναδομημένη φωνή εξόδου. Οι κωδικοποιητές φωνής που λαμβάνουν παραμέτρους πρόβλεψης από τη φωνή εισόδου πρέπει να μεταδώσουν τις LPC παραμέτρους στο δέκτη, και ονομάζονται feed forward κωδικοποιητές. Οι κωδικοποιητές που λαμβάνουν παραμέτρους πρόβλεψης από το σήμα εξόδου δεν είναι ανάγκη να τις μεταδώσουν διότι το σήμα εξόδου είναι διαθέσιμο στο δέκτη. Τέτοια συστήματα ονομάζονται feedback (ανάδρασης) κωδικοποιητές. Οι κατασκευές feed forward είναι συνήθως πιο αποτελεσματικές και πιο σθεναρές στα λάθη καναλιών, αλλά απαιτούν τη μετάδοση της απαραίτητης πλευρικής πληροφορίας για να ελέγξουν τον προσαρμοστικό προβλεπτή. Από την άλλη πλευρά, τα συστήματα feedback δεν απαιτούν την μετάδοση ουδεμίας πλευρικής πληροφορίας, αλλά γενικά λειτουργούν με χαμηλότερη ποιότητα και είναι πιο ευάλωτα στα λάθη καναλιών.

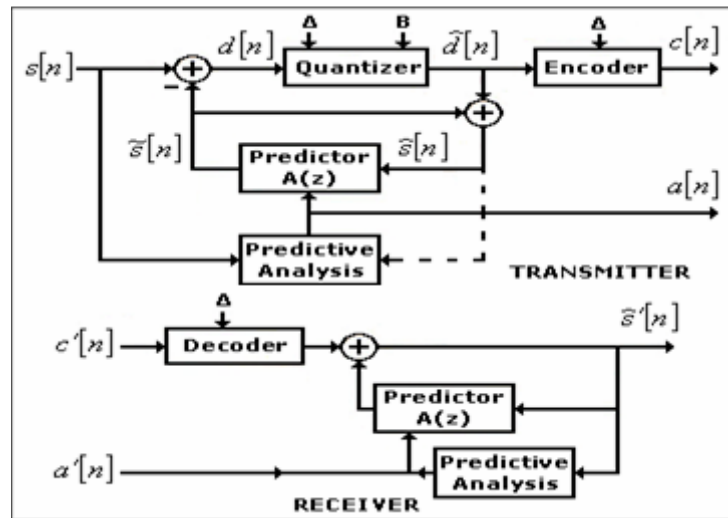
Άλλο ένα μέλος της κατηγορίας των κωδικοποιητών κυματομορφής με προσαρμοστική πρόβλεψη είναι οι κωδικοποιητές προσαρμοστικής πρόβλεψης με διεγερση κλίσης (adaptive-predictive coder with pitch prediction, APC-PP). Οι APC-PP είναι DPCM βασιζόμενοι κωδικοποιητές που εκμεταλλεύονται τον πλεονασμό κλίσης του σήματος φωνής όπως επίσης και την βραχυπρόθεσμη συσχέτιση των δειγμάτων φωνής. Έτσι ο APC-PP βασίζεται πάνω σε δύο προσαρμοστικούς προβλεπτές, έναν βραχυπρόθεσμο προβλεπτή (short-term predictor – STP) για φασματική εκτίμηση και έναν μακροπρόθεσμο προβλεπτή (long-term predictor – LTP) για πρόβλεψη κλίσης. Η προσθήκη του LTP βελτιώνει σημαντικά την απόδοση APC.

Οι θορύβου-ανάδρασης κωδικοποιητές (Noise-feedback coders) είναι ένα άλλο μέλος της κατηγορίας των κωδικοποιητών προσαρμοστικής πρόβλεψης. Στην κωδικοποίηση θορύβου-ανάδρασης, χρησιμοποιούνται οι ιδιότητες επικάλυψης του ανθρώπινου ακουστικού συστήματος για να ελαχιστοποιηθεί ο όγκος του ακουστικού θορύβου στο αναδημιουργημένο σήμα φωνής. Για να επιτευχθεί αυτή η επικάλυψη θορύβου, το φάσμα του κωδικοποιημένου θορύβου σχηματίζεται έτσι ώστε να είναι παρόμοιο με το φάσμα της φωνής. Σε τέτοια συστήματα, η ενέργεια του φάσματος θορύβου θα είναι κανονικά λιγότερη από την ενέργεια του φάσματος ήχου και ο θόρυβος θα επικαλύπτεται.

4.2. ΚΩΔΙΚΟΠΟΙΗΣΗ ΠΡΟΣΑΡΜΟΣΤΙΚΗΣ ΠΡΟΒΛΕΨΗΣ

Το σχήμα 4.1 παρουσιάζει το block διάγραμμα ενός APC. Σ' αυτό το σχήμα, οι παράμετροι του κβαντιστή θεωρούνται κανονικοποιημένοι. Στα APC με προσαρμοστική feed forward πρόβλεψη (APC with adaptive feed forward prediction, APC-APF), οι συντελεστές πρόβλεψης λαμβάνονται από το σήμα φωνής και μεταδίδονται σαν πλευρική πληροφορία. Αντίθετα, οι APC με προσαρμοστική feedback πρόβλεψη (APC with adaptive feedback

prediction, APC-APB) χρησιμοποιούν το κωδικοποιημένο σήμα για να καθορίσουν τις παραμέτρους πρόβλεψης. Στο σχήμα 4.1, οι έντονες μαύρες γραμμές δείχνουν την feed forward διαδρομή και οι διακεκομμένες γραμμές την feedback διαδρομή του προσαρμοστικού προβλεπτή.



ΣΧΗΜΑ 4.1: Block διάγραμμα ενός APC

Στην feed forward στρατηγική, οι παράμετροι του προβλεπτή καθορίζονται από μια γραμμική πρόβλεψη ανάλυση. Οι παράμετροι που πρέπει να εξαχθούν και να κωδικοποιηθούν για τον APC είναι ακριβώς οι ίδιοι με αυτούς ενός φίλτρου φωνητικής περιοχής με LPC διέγερσης κλίσης. Τα μειονεκτήματα ενός συστήματος feed forward περιλαμβάνουν καθυστέρηση κωδικοποίησης που απαιτείται για τον υπολογισμό του προβλεπτή (όπως στα LPC) και τα παραπάνω bit που χρειάζονται για να μεταδοθεί η πλευρική πληροφορία. Τα πλεονεκτήματα του συστήματος αυτού είναι ότι είναι λιγότερο ευαίσθητο στα λάθη μετάδοσης bit. Όπως και στα LPC διέγερσης κλίσης, οι παράμετροι του προβλεπτή ενημερώνονται πάνω σε μια frame-by-frame βάση για να ελαχιστοποιηθεί ο όγκος της πλευρικής πληροφορίας.

Στην feedback μορφή ενός APM συστήματος, ο προβλεπτής λαμβάνεται κατευθείαν από τα κωδικοποιημένα δείγματα φωνής χρησιμοποιώντας επίσης γραμμική προσαρμοστική ανάλυση. Λόγω του ότι ένα δείγμα πρέπει να είναι διαθέσιμο ώστε να χρησιμοποιηθεί, μόνο οι περασμένες τιμές του σήματος φωνής αξιοποιούνται για να υπολογίσουν τους συντελεστές του predictor. Επίσης, επειδή δεν απαιτείται πλευρική πληροφορία, ο predictor μπορεί να ενημερωθεί σε κάθε λόγο χωρίς να επηρεάζει το bit rate. Έτσι οι συντελεστές του predictor ενημερώνονται σε έναν υψηλό λόγο μερικές φορές σε κάθε δείγμα. Αυτό προσθέτει στην υπολογιστική πολυπλοκότητα του συστήματος feedback. Ένα μεγάλο μειονέκτημα του συστήματος αυτού είναι ότι με την κοινή κβάντιση, ο predictor συνήθως δεν είναι τόσο καλός, το οποίο μπορεί να επηρεάσει την επίδοση του συστήματος σε χαμηλά bit rate. Στη μορφή feedback, οι παράμετροι του προβλεπτή συνήθως λαμβάνονται χρησιμοποιώντας τεχνικές ανάλυσης περιοδικά επαναλαμβανόμενες όπως είναι η μέθοδος περιοδικά επαναλαμβανόμενης αυτοσυσχέτισης, η μέθοδος ελάχιστου τετραγωνικού μέσου (least mean square-LMS) ή η μέθοδος απότομης κατάβασης (steepest descent). Σε αυτό το κείμενο δίνουμε έμφαση στα feed forward συστήματα λόγω της ευρωστίας και της εύκολης υλοποίησης τους. Οι παράμετροι ενός APC-APF (το οποίο θα ονομάζουμε APC από δω και πέρα) με ένα σταθερό ομοιόμορφο κβαντιστή φαίνονται στον πίνακα 4.1. Οι παράμετροι του predictor λαμβάνονται μέσω μιας LPC ανάλυσης του σήματος φωνής. Επιπλέον στην κατηγορία του predictor, P, και στο LPC μέγεθος frame, I, οι παράμετροι κβάντισης πρέπει να καθοριστούν.

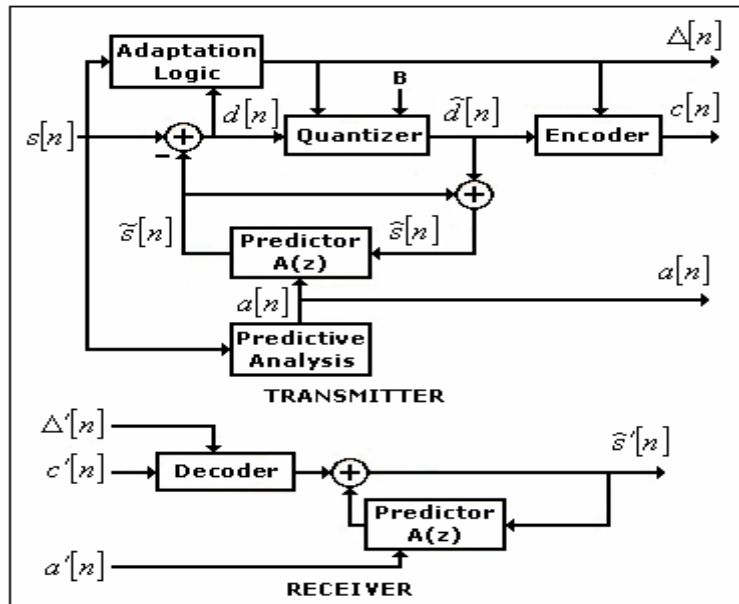
parameters	name	range	typical value
predictor order	P	1 - 10	1
LPC frame size	I	80 - 360	160
quantizer max.	Smax	1 - 32767	32767
number of bits	B	1 - 16	5

ΠΙΝΑΚΑΣ 4.1: Παράμετροι ενός APC με feed forward πρόβλεψη και ομοιόμορφο κβαντιστή.

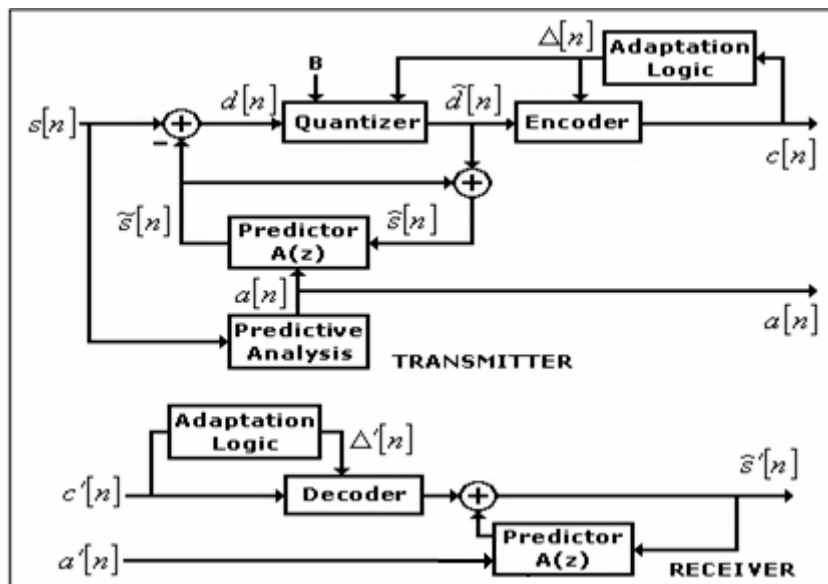
parameters	name	range	typical value
predictor order	P	1 - 10	4
LPC frame size	l	80 - 360	160
adaptive quantizer parameters (see quantizer parameter tables)	--	--	--

ΠΙΝΑΚΑΣ 4.2: Παράμετροι ενός APC με feedback προσαρμοστικό κβαντιστή.

Σε αυτό το κείμενο, το APC θεωρείται DPCM με προσαρμοστικό predictor και προσαρμοστικό κβαντιστή. Στα APC, ο κβαντιστής μπορεί να είναι είτε προσαρμοστικός feed forward είτε προσαρμοστικός feedback. Τα σχήματα 4.2 και 4.3 παρουσιάζουν τα block διαγράμματα ενός APC συστήματος με feed forward και feedback προσαρμοστικούς κβαντιστές. Στο πλάνο feed forward, οι παράμετροι κβάντισης μεταδίδονται σαν πλευρική πληροφορία. Ο πίνακας 4.2 παρουσιάζει τις παραμέτρους ενός APC με feedback προσαρμοστική κβάντιση.



ΣΧΗΜΑ 4.2: Block διάγραμμα ενός APC με feed forward προσαρμοστικό κβαντιστή.



ΣΧΗΜΑ 4.3: Block διάγραμμα ενός APC με feedback προσαρμοστικό κβαντιστή.

4.3. APC ΜΕ ΠΡΟΒΛΕΨΗ ΚΛΙΣΗΣ (APC-PP)

Ένας άλλος τρόπος πρόβλεψης του σήματος φωνής και αύξησης του κέρδους κωδικοποίησης είναι η χρήση του πλεονασμού κλίσης του φωνούμενου ήχου. Κατά τη διάρκεια του φωνούμενου ήχου, η κυματομορφή είναι σχεδόν περιοδική, και αυτή η ψευδο-περιοδικότητα μπορεί να χρησιμοποιηθεί για να βοηθήσει στην πρόβλεψη της κυματομορφής φωνής. Αυτό φυσιολογικά επιτυγχάνεται χρησιμοποιώντας έναν μακροπρόθεσμο προβλεπτή (LTP). Όπως ο βραχυπρόθεσμος προβλεπτής, ο LTP είναι επίσης γραμμικός προβλεπτής, όμως αντίθετα με το βραχυπρόθεσμο προβλεπτή του οποίου η πρόβλεψη βασίζεται σε προσαρμοσμένα δείγματα, η πρόβλεψη του LTP βασίζεται σε δείγματα μίας ή περισσότερων περιόδων κλίσης (pitch periods) στο παρελθόν. Οι μακροπρόθεσμοι προβλεπτές μπορούν να χρησιμοποιηθούν για να αφαιρεθεί ο πλεονασμός κλίσης του σήματος φωνής. Όπως ο STP παράγει έναν υπολογισμό του φασματικού φακέλου του σήματος φωνής, ο LTP αναπαριστά την αρμονική δομή του φάσματος φωνής. Οι κωδικοποιητές προσαρμοστικής πρόβλεψης με διέγερση κλίσης προσπαθούν να εκμεταλλευθούν βραχυπρόθεσμους και μακροπρόθεσμους πλεονασμούς του σήματος φωνής για να παράγουν έναν αποτελεσματικό κωδικοποιητή κυματομορφής σε μεσαία bit-rates.

Το σχήμα 4.4 παρουσιάζει το block διάγραμμα ενός APC-PP. Το STP είναι ένα φίλτρο πρόβλεψης τάξης P που ορίζεται ως:

$$A(z) = \sum_{k=1}^P a_k z^{-k}$$

και οι LTP είναι συνήθως της μορφής:

$$B(z) = \sum_{k=-l}^l \beta_k z^{-\gamma-k}$$

όπου τα β είναι οι συντελεστές του LTP και $2 \times \ell + 1$ είναι η τάξη του predictor. Το γ είναι η μακροπρόθεσμη καθυστέρηση που συνήθως ανταποκρίνεται σε μια περίοδο κλίσης ή ένα ακέραιο πολλαπλάσιο της περιόδου αυτής. Επειδή το ανταποκρινόμενο φίλτρο σύνθεσης μπορεί να γίνει ασταθές, προτιμάται ένα μονής αντλίας LTP (single-tap LTP) της μορφής:

$$B(z) = \beta z^{-1}$$

όπου $\beta < 1$. Σε χρόνο n , η μακροπρόθεσμη καθυστέρηση λαμβάνεται ψάχνοντας την τιμή του γ που μεγιστοποιεί την ομαλοποιημένη συσχέτιση φωνής που δίνεται από:

$$R_n[\gamma] = \frac{\sum_{n=0}^{N-1} s[n]s[n-\gamma]}{\sum_{n=0}^{N-1} s^2[n] \sum_{n=0}^{N-1} s^2[n-\gamma]}$$

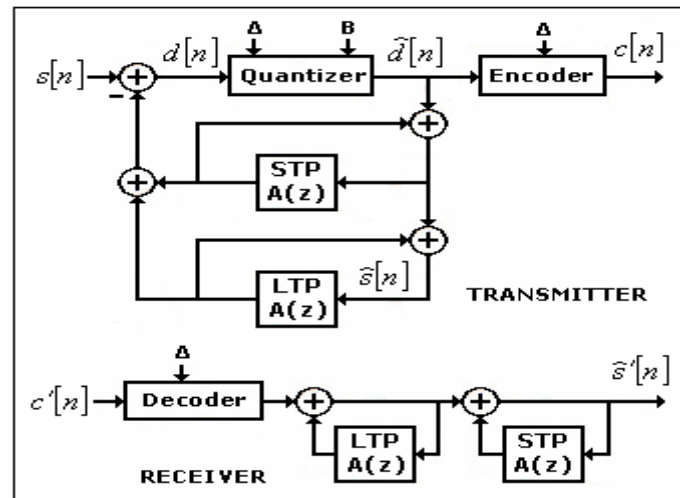
όπου N το μέγεθος του τμήματος μακροχρόνιας ανάλυσης. Η καθυστέρηση LTP, γ , συνήθως επιλέγεται να είναι μικρότερη από το L , όπου L είναι το μέγεθος μνήμης του LTP. Το κέρδος LTP, β , δίνεται από το λόγο της αυτοσυσχέτισης φωνής στις τιμές γ και 0,

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

Η παραπάνω διαδικασία απολαβής των παραμέτρων LTP ονομάζεται μέθοδος ανοιχτού βρόχου (open-loop method). Η μέθοδος κλειστού βρόχου (closed-loop method) βασίζεται στην ανάλυση-από-σύνθεση (analysis-by-synthesis) της φωνής. Οι μακροπρόθεσμοι προβλεπτές είναι πολύ αποτελεσματικοί στη βελτίωση της απόδοσης των διαφορικών κωδικοποιητών.

Παρόλο που η κατηγορία στην οποία συγκαταλέγονται οι προβλεπτές, δεν είναι σημαντική, οι STP συνήθως αναφέρονται πριν τους LTP. Το κέρδος στους SNR που

παρέχεται από τον συνδυασμό των δύο προβλεπτών είναι γενικότερα μικρότερο από το άθροισμα των ξεχωριστών κερδών. Όπως φαίνεται στο σχήμα 4.4, αντιστροφές του μακροπρόθεσμου φίλτρου, $1-B(z)$, και βραχυπρόθεσμου φίλτρου, $1-A(z)$, χρησιμοποιούνται για να παραχθεί το κωδικοποιημένο σήμα φωνής στο δέκτη. Οι παράμετροι του APC-PP περιλαμβάνουν τις παραμέτρους LTP, STP και κβάντισης, που δίνονται στον πίνακα 4.3.



ΣΧΗΜΑ 4.4: Block διάγραμμα ενός APC-PP.

parameters	name	range	typical value
predictor order	P	1 - 10	1
LPC frame size	I	80 - 360	160
LTP frame size	N	10 - 80	40
long-term memory size	L	128 - 512	256
number of bits	B	1 - 5	3
quantizer parameters	--	--	--

ΠΙΝΑΚΑΣ 5.3: Παράμετροι ενός APC-PP.

4.4. ΚΩΔΙΚΟΠΟΙΗΣΗ ΘΟΡΥΒΟΥ-ΑΝΑΔΡΑΣΗΣ

Η κωδικοποίηση θορύβου-ανάδρασης (noise-feedback coding, NFC) αναφέρεται σε μια τροποποιημένη εκδοχή του κωδικοποιητή DPCM ο οποίος προσπαθεί να εκμεταλλευτεί τις ιδιότητες επικάλυψης του ανθρώπινου φωνητικού συστήματος. Στο NFC η υποκειμενική ένταση ήχου του κωδικοποιημένου θορύβου μειώνεται σχηματίζοντας το φάσμα του κωδικοποιημένου θορύβου για να ταιριάζει με το φάσμα του σήματος φωνής εισόδου. Μια γενική διαμόρφωση κωδικοποιητή DPCM αποδίδει περίπου λευκό κωδικοποιημένο θόρυβο (αν υποθέσουμε καλή κβάντιση). Ο θόρυβος δεν θα ήταν αισθητός σε περιοχές του φάσματος φωνής που περιέχουν ουσιαστική ενέργεια σήματος αλλά σε συχνότητες που υπάρχει λίγη ή καθόλου ενέργεια ήχου. Για να μειώσουμε αυτό το θόρυβο, το φάσμα θορύβου πρέπει να σχηματιστεί έτσι ώστε να ταιριάζει με το φάσμα ήχου. Αυτό έχει ως αποτέλεσμα μεγαλύτερη συνολική ενέργεια θορύβου αλλά λόγω της επικάλυψης του κωδικοποιημένου θορύβου από τον ήχο, λαμβάνεται λιγότερος θόρυβος.

Το σύστημα DPCM του σχήματος 4.1 μπορεί να αναπαρασταθεί παρόμοια όπως φαίνεται στο σχήμα 4.5. αυτή η εναλλακτική αναπαράσταση του DPCM μπορεί να χρησιμοποιηθεί για την απολαβή κωδικοποιητών κυματομορφής που έχουν τη δυνατότητα να σχηματίζουν το θόρυβο. Επειδή η είσοδος του κβαντιστή στο DPCM, $d[n]$, μπορεί να εκφραστεί ως:

$$d[n] = s[n] - \tilde{s}[n] = s[n] - \sum_{i=1}^P a_i \hat{s}[n-i]$$

και επειδή:

$$\hat{s}[n] = s[n] + q[n]$$

όπου $q[n]$ είναι ο κωδικοποιημένος θόρυβος, το $d[n]$ μπορεί να εκφραστεί ως:

$$d[n] = s[n] - \sum_{i=1}^P a_i s[n-i] - \sum_{i=1}^P a_i q[n-i]$$

Η παραπάνω εξίσωση μπορεί να εκφραστεί στο πεδίο ορισμού του μετασχηματισμού Ζήτα ως:

$$D(z) = S(z)(1-A(z)) - Q(z)A(z)$$

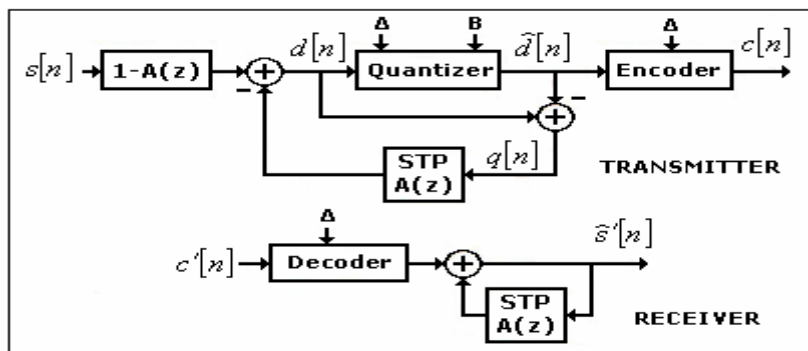
Το σχήμα 4.5 παρουσιάζει μια διαφορετική υλοποίηση του DPCM βασιζόμενη στην εξίσωση 4.9. Όπως φαίνεται στην εξίσωση 4.9, το $d[n]$ έχει δύο μέρη. Το πρώτο μέρος είναι μια εκδοχή του σήματος φωνής που έχει φιλτραριστεί από το βραχυπρόθεσμο φίλτρο, $1-A(z)$. Το δεύτερο μέρος είναι ο θόρυβος κβάντισης που έχει φιλτραριστεί από το φίλτρο πρόβλεψης, $A(z)$. Όπως φαίνεται στο σχήμα 4.5, το μέρος που περιλαμβάνει τη φωνή μπορεί να ληφθεί σε μια feed forward διαδρομή και το μέρος που περιλαμβάνει το θόρυβο μπορεί να ληφθεί από μια feedback διαδρομή. Αν υποθεθεί ότι αυτά τα δύο μέρη είναι ξεχωριστά, το feedback φίλτρο, $A(z)$, μπορεί να αντικατασταθεί από μια γενική feedback λειτουργία, $F(z)$. Σ' αυτά τα συστήματα, το σήμα φωνής εισόδου και ο θόρυβος κβάντισης φιλτράρονται με ξεχωριστά φίλτρα. Αυτό το σύστημα φαίνεται στο σχήμα 5.6. Από το σχήμα 4.6 φαίνεται ότι ο μετασχηματισμός Ζήτα του κωδικοποιημένου σήματος λάθους εξόδου

$$y[n] = s[n] - \hat{s}[n]$$

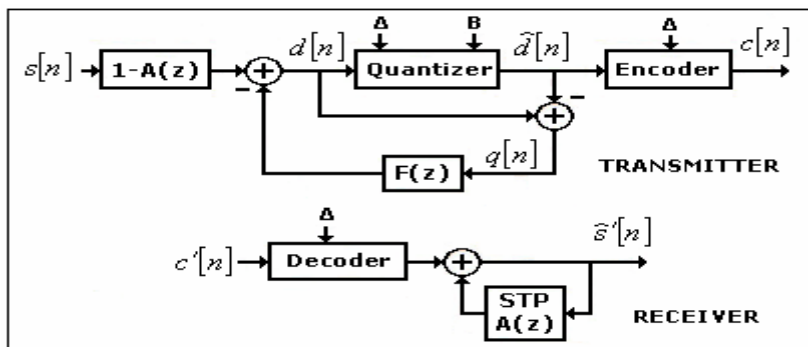
ορίζεται ως:

$$Y(z) = Q(z) \frac{1-F(z)}{1-A(z)} = \frac{Q(z)}{W(z)}$$

το οποίο είναι μια εκδοχή του θορύβου κβάντισης, $q[n]$, ο οποίος έχει φιλτραριστεί από το αντίστροφο του φίλτρου σχηματισμού θορύβου, $W(z)$. Έτσι αλλάζοντας το $F(z)$ και ως εκ τούτου το $W(z)$, το φάσμα του κωδικοποιημένου θορύβου μπορεί να σχηματιστεί με έναν αυθαίρετο τρόπο.



ΣΧΗΜΑ 4.5: Μια εναλλακτική παρουσίαση του ADPCM.



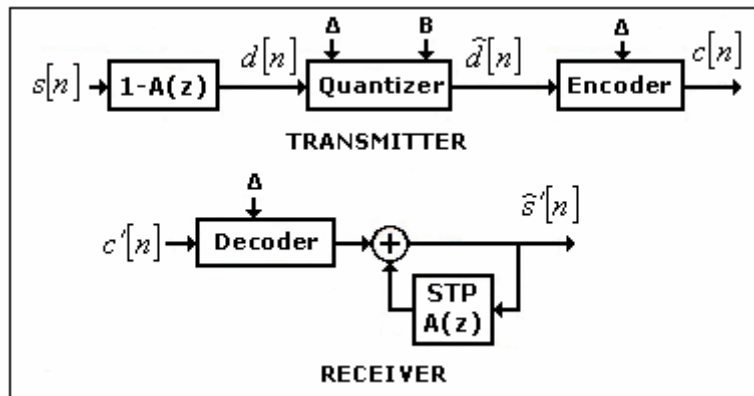
ΣΧΗΜΑ 4.6: Block διάγραμμα ενός θορύβου ανάδρασης κωδικοποιητή.

4.4.1. ΑΝΟΙΚΤΟΥ ΒΡΟΓΧΟΥ DPCM (D*PCM).

Το DPCM είναι μια ειδική περίπτωση ενός συστήματος του σχήματος 4.6 με $F(z)=A(z)$. Μια άλλη ειδική περίπτωση του συστήματος είναι όταν $F(z)=0$, το οποίο οδηγεί σε έναν βρόγχο χωρίς ανάδραση στον κωδικοποιητή. Το σύστημα αναφέρεται ως ένα ανοικτού βρόγχου DPCM ή D*PCM. Όπως φαίνεται στο σχήμα 4.7, το D*PCM αποτελείται από ένα προ-φίλτρο, έναν κβαντιστή και ένα "ταχυδρομικό" φίλτρο (post-filter). Εφόσον το post filter

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

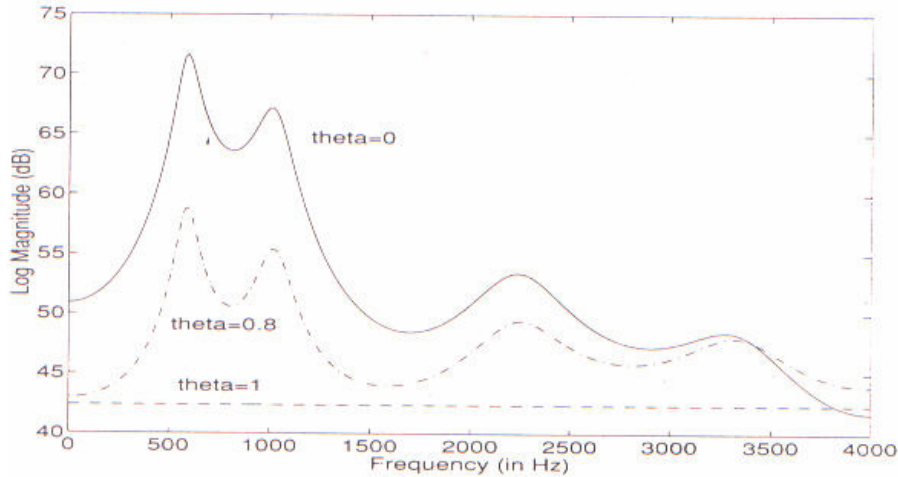
αντιπροσωπεύει το "φάκελο" του φάσματος φωνής, το φίλτρο αυτό σχηματίζει το φάσμα θορύβου κβάντισης να είναι πανομοιότυπο με το φάσμα φωνής εισόδου. Λόγω του ότι ο θόρυβος κβάντισης $q[n]$ φιλτράρεται μέσω του $H(z)$ (ένα IIR φίλτρο του οποίου το κέρδος είναι πάνω από 1), το D*PCM έχει πάντα ένα χαμηλότερο SNR από ένα DPCM. Παρόλ' αυτά υπάρχει ένα ρίσκο λάθους σωρού στο synthesizer το οποίο μπορεί να μειώσει την επίδοση του. Έτσι ένας συμβιβασμός μεταξύ DPCM και D*PCM προτιμάται συχνά.



ΣΧΗΜΑ 4.7: Block διάγραμμα ενός D*PCM.

4.4.2. DPCM ΜΕ ΑΝΑΔΡΑΣΗ ΘΟΡΥΒΟΥ.

Επειδή ο θόρυβος κβάντισης $q[n]$ φιλτράρεται αντίστροφα από το φίλτρο σχηματισμού λάθους, $W(z)$, και η φωνή περνάει μέσω του $H(z)$, το φάσμα της εξόδου του παράγοντα θορύβου μπορεί να σχηματιστεί αυθαίρετα χωρίς επηρεασμό της φωνής εξόδου. Μια συμβιβαστική επιλογή για το $F(z)$ μεταξύ των δύο άκρων ($F(z)=A(z)$ για το DPCM και $F(z)=0$ για το D*PCM) είναι $A(\theta z)$, όπου $0 < \theta < 1$. Οι οριακές τιμές του $\theta=1$ και $\theta=0$ αποδίδουν DPCM και D*PCM αντίστοιχα. Τυπικές τιμές του θ για σωστό feedback θόρυβο είναι στην περιοχή 0.7-0.95. Τέτοια επιλογή για το θ αυξάνει το εύρος πλάτους των μηδενικών σε $1-F(z)$ προκαλώντας το θόρυβο να φτάσει στην κορυφή σε πεδία formant όπου είναι λιγότερο ακουστός. Το σχήμα 5.8 δείχνει την απόκριση συχνότητας του αντίστροφου ζυγίσματος-λάθους φίλτρο (error-weighting filter), $(W(z))^{-1}$ για ένα frame της φωνομένης φωνής με $\theta=0$, $\theta=0.8$ και $\theta=1$. Το αντίστροφο φίλτρο δίνει έμφαση στο θόρυβο κωδικοποίησης όπου το σήμα έχει μεγάλη ενέργεια (το $W(z)$ είναι μικρό, συνεπώς κάνει το λόγο $Q(z)/W(z)$ μεγάλο) και χαμηλώνει το θόρυβο όταν το σήμα έχει μικρή ενέργεια. Σε πρακτικές εφαρμογές, ένας περιοριστής κορυφής (peak-limiter), με μια παράμετρο C , χρησιμοποιείται στο βρόγχο ανάδρασης για να εξασφαλίσει σταθερότητα, ειδικά με κοινή κβάντιση. Παρόλο που το NFC παράγει ένα SNR το οποίο είναι μεταξύ αυτού του DPCM και αυτού του D*PCM, η αντικειμενική του απόδοση είναι καλύτερη και από τα δύο αυτά, ειδικά σε χαμηλά bit rate.



ΣΧΗΜΑ 4.8

4.4.3. APC ΜΕ ΑΝΑΔΡΑΣΗ ΘΟΡΥΒΟΥ ΚΑΙ ΠΡΟΒΛΕΨΗ ΡΙΤΣΧ.

Η προσαρμοστική-προβλεπόμενη κωδικοποίηση μπορεί να συνδυαστεί με feedback θόρυβο για να βελτιώσει την απόδοση του APC. Ένα block διάγραμμα ενός APC-PP με feedback θόρυβο (APC-PPNF) φαίνεται στο σχήμα 4.9. Το APC-PPNF έχει ένα επιπλέον βρόγχο μακροχρόνιου predictor συγκρινόμενο με το NFC του σχήματος 4.6. Η πρόβλεψη προσαρμοστικού φάσματος, η πρόβλεψη προσαρμοστικού pitch και η ο προσαρμοστικός σχηματισμός θορύβου οδηγούν σε έναν ουσιαστικό κωδικοποιητή σε μεσαία bit rate (9.6-24kbps).

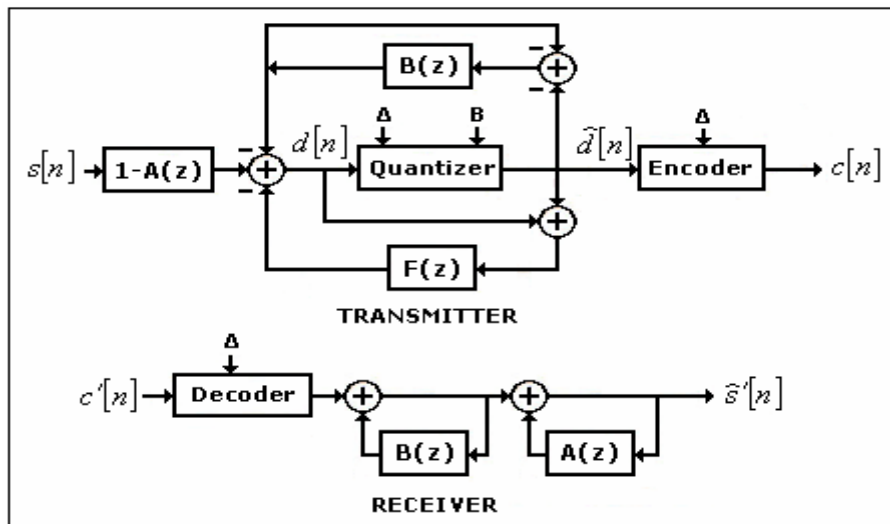
Ο πίνακας 4.4 δείχνει τις παραμέτρους ενός APC-PPNF. Οι παράμετροι αποτελούνται από τις παραμέτρους τους predictor, την παράμετρο του θορύβου feedback, την παράμετρο του περιοριστή κορυφής και τις παραμέτρους του κβαντιστή. Τα DPCM, D*PCM και NFC μπορούν να θεωρηθούν ως ειδικές περιπτώσεις των APC-NFPP. Τα DPCM και D*PCM μπορούν να υλοποιηθούν έχοντας $\theta=0$ και $\theta=1$ και ένα μακροχρόνιο συντελεστή πρόβλεψης $\beta=0$. Το NFC επιτυγχάνεται με $0<\theta<1$ και $\beta=0$ και τα APC-NFPP όταν το β είναι διάφορο του μηδενός και $0<\theta<1$. Ο πίνακας 4.5 ανακεφαλαιώνει αυτές τις περιπτώσεις.

parameters	name	range	typical value
predictor order	P	1 - 10	10
LPC frame size	I	80 - 360	160
LTP frame size	N	10 - 80	40
long-term memory size	L	128 - 512	256
error-weighting factor	θ	0 - 1	0.8
feedback-clipping limit	C	500 - 10000	3000
number of bits	B	1 - 14	3
quantizer parameters	--	--	--

ΠΙΝΑΚΑΣ 4.4: Παράμετροι ενός APC-PPNF.

coder type	θ	β
DPCM, ADPCM	1	0
D*PCM	0	0
NFC	0 - 1	non-zero
APC-PPNF	0 - 1	non-zero

ΠΙΝΑΚΑΣ 4.5: Οι συγκεκριμένες τιμές των θ και β για να ληφθεί DPCM, D*PCM και NFC.



ΣΧΗΜΑ 4.9: Block διάγραμμα ενός APC-NF.

ΚΕΦΑΛΑΙΟ ΠΕΜΠΤΟ

VOCODERS

5.1. ΙΣΤΟΡΙΑ ΤΩΝ VOCODERS

Ήταν το έτος 1928 όταν Homer W. Dudley των Bell Telephone Laboratories, του New Jersey άρχισε την αναζήτηση ενός τρόπου να μειωθεί το εύρος ζώνης που απαιτήθηκε για την τηλεφωνία, προκειμένου να ενισχυθεί η ικανότητα μετάδοσης. Η ιδέα του Homer ήταν να αναλύσει το σήμα φωνής, να το αποσυνθέσει και να το συνθέσει σε ένα φτωχό σήμα με μικρότερο εύρος ζώνης. Ονόμασε αυτήν την διαδικασία "parallel bandpass speech analysis and resynthesis" και το αντιλήφθηκε μέσω ενός πρωτοτύπου που ονομάστηκε *vocoder* (συντομία για τον "voice coder"). Ο *Vocoder* εξελίχθηκε σε ένα εμπορικότερο εφαρμόσιμο σχέδιο, μετονομάστηκε "σε *Voder*" και παρουσιάστηκε μπροστά σε ένα μεγάλο ακροατήριο στην παγκόσμια έκθεση του 1939. Οι πρώτοι σε κατασκευή *vocoders* ήταν τεράστιοι σωροί μηχανημάτων που συνδέθηκαν μαζί με μια αφθονία καλωδίων. Πιο πρόσφατα, στις αρχές της δεκαετίας του '70, ο Wendy Carlos και ο Rachel Elkind χρησιμοποίησαν σε πραγματικό χρόνο, αναλογικό *vocoder* που ο Bob Moog έκανε στις προδιαγραφές τους, για να αναπαραγάγουν τους φωνητικούς ήχους στην απόδοση της 9ης συμφωνίας του Beethoven για την ταινία του Stanley Kubrick, "κουρδιστό πορτοκάλι". Αυτό τελικά οδήγησε σε μια αναγέννηση ενδιαφέροντος για την κωδικοποίηση φωνής (*vocoding*) συμπεριλαμβανομένων πολλών βελτιωμένων *vocoding* αποτελεσμάτων και παραγώγων της ουσιαστικής ιδέας να παραχθούν τα ενδιαφέροντα ηχητικά και μουσικά αποτελέσματα.

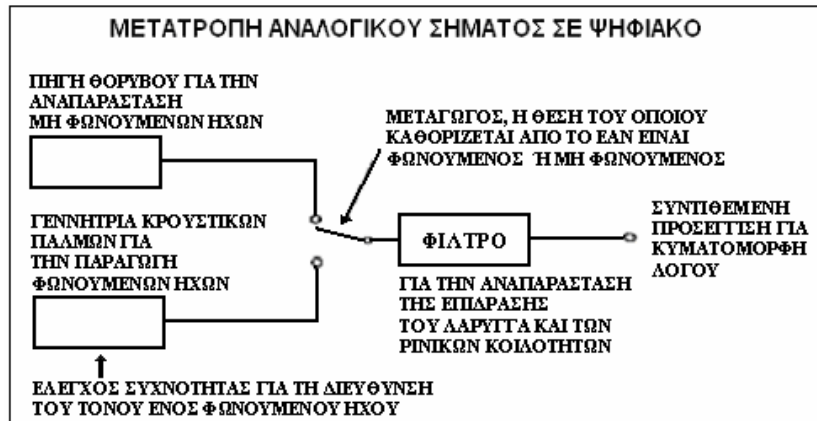
5.2. ΕΙΣΑΓΩΓΗ ΣΤΟΥΣ VOCODERS

Ακούμε με τη σειρά τον ένα μετά τον άλλο τους ανθρώπους να προφέρουν μια λέξη ή μία ακολουθία λέξεων. Ο ένας έχει υψηλό τόνο φωνής, ο άλλος χαμηλό. Ένας αρθρώνει καθαρά, άλλος μπερδεύει τα λόγια του και γενικά ο καθένας μιλά με τον τρόπο του. Ωστόσο καταλαβαίνουμε κάθε ομιλητή. Οι συλλογισμοί αυτοί καθιστούν φανερό ότι υπάρχει μία μεταβλητότητα (εντός ορίων) που επιτρέπεται στη κυματομορφή που πρέπει να φτάνει στο αυτί ενός ακροατή πριν εμφανιστεί απώλεια στην αναγνώριση της λέξης που ειπώθηκε. Αυτό που συμβαίνει σε εμάς λοιπόν είναι ότι για να μεταδώσουμε λόγο δεν χρειάζεται να μεταδώσουμε την ακριβή κυματομορφή που παρήγαγε ο ομιλητής. Μπορούμε κάλλιστα να μεταδώσουμε πληροφορία από την οποία να επανασκευαστεί μία κυματομορφή στο δέκτη η οποία, αντί να ταυτίζεται με αυτή που παρήγαγε ο ομιλητής, να είναι παρόμοια με αυτήν. Επιτρέποντάς μας αυτή την ευελιξία, προσδοκούμε να μπορούμε να λειτουργήσουμε ένα ψηφιακό σύστημα μετάδοσης με κωδικοποίηση πηγής με ένα χαμηλότερο bit rate. Πράγματι, η προσδοκία μας αυτή επιβεβαιώνεται. Οι κωδικοποιητές πηγής που χρησιμοποιούνται καλούνται *vocoders* και λειτουργούν σε ένα σημαντικό χαμηλότερο bit rate ακόμα και από τα ADM. Ένα αντιπροσωπευτικό διάστημα για bit rate ενός *vocoder* είναι από 1.2 έως 2.4kb/s. Ωστόσο, η προκύπτουσα φωνή που αναπαράχθηκε έχει συνθετικό ήχο και κάπως τεχνητή ποιότητα. Σαν αποτέλεσμα έχουμε, οι κωδικοποιητές φωνής να χρησιμοποιούνται για ειδικές εφαρμογές, όπου είναι αποδεκτή η ανταλλαγή ποιότητας λόγου για το πλεονέκτημα του χαμηλού bit rate.

5.2.1. ΜΟΝΤΕΛΟ ΦΩΝΗΣ

Φαίνεται ότι ο λόγος αποτελείται, ή τουλάχιστον μπορεί κάλλιστα να προσεγγιστεί, από μία ακολουθία φωνούμενων και μη φωνούμενων ήχων που διέρχονται μέσω ενός φίλτρου. Οι φωνούμενοι ήχοι παράγονται από τις ταλαντώσεις των φωνητικών χορδών. Οι μη φωνούμενοι ήχοι παράγονται όταν ένας ομιλητής προφέρει γράμματα όπως το "σ", "φ", "π", κ.λ.π. Σε αυτήν την τελευταία περίπτωση, οι ήχοι διαμορφώνονται με την εκπνοή του αέρα μέσω των χειλιών και των δοντιών. Στο Σχήμα 5.1 παρουσιάζεται μία γενικευμένη παράσταση ενός *vocoder*. Το φίλτρο παριστάνει την επίδραση των παραγόμενων ήχων από το στόμα, το λαιμό και τις ρινικές κοιλότητες του ομιλητή. Στον *vocoder*, οι φωνούμενοι ήχοι προσομοιώνονται από μία γεννήτρια κρουστικών παλμών της οποίας η συχνότητα είναι η θεμελιώδης συχνότητα ταλάντωσης των φωνητικών χορδών. Οι μη φωνούμενοι ήχοι προσομοιώνονται από μία πηγή θορύβου. Γενικά λοιπόν, όλοι οι *vocoders* χρησιμοποιούν τη διάταξη που παρουσιάζεται στο Σχήμα 5.1 για να παράγουν μία συντιθέμενη προσέγγιση μίας κυματομορφής ομιλίας. Διαφέρουν μόνο στις τεχνικές που χρησιμοποιούν για να

παράγουν τους φωνούμενους και μη φωνούμενους ήχους και στα χαρακτηριστικά και τη σχεδίαση του φίλτρου.



ΣΧΗΜΑ 5.1. Το μοντέλο ομιλίας που χρησιμοποιείται στους κωδικοποιητές φωνής.

5.2.2. ΧΡΗΣΕΙΣ VOCODER

- Συμπίεση λεκτικών στοιχείων στην τηλεφωνία (συμπεριλαμβανομένης της δορυφορικής τηλεφωνίας)
- Συμπίεση λεκτικών στοιχείων για την αποθήκευση
- Επεξεργασία του λόγου στα ακουστικά τεχνητά μέλη (ενισχύσεις ακρόασης και κοχλιωτά μοσχεύματα)
- Στις στρατιωτικές επικοινωνίες

5.3. ΚΑΤΗΓΟΡΙΕΣ VOCODER

Υπάρχουν τρεις βασικές κατηγορίες vocoder:

- Channel vocoders (κωδικοποιητές καναλιών)
- Formant vocoders
- LPC vocoder (Κωδικοποιητές γραμμικής πρόβλεψης)

5.3.1. CHANNEL VOCODERS (ΚΩΔΙΚΟΠΟΙΗΤΕΣ ΚΑΝΑΛΙΩΝ)

Ένα από τα πολλά συστήματα vocoders παρουσιάζονται στο Σχήμα 5.2. Σε αυτό το σύστημα κωδικοποίησης, το φάσμα της ομιλίας εισόδου διαιρείται σε 15 διαστήματα συχνοτήτων, καθένα εκ των οποίων έχει εύρος ζώνης 200Hz. Εάν η έξοδος από τα ζωνοπερατά φίλτρα ήταν ημιτονοειδής κυματομορφή καθορισμένου πλάτους, τότε ο ανορθωτής που είναι σε σειρά και το χαμηλοπερατό φίλτρο των 20Hz θα παρείχαν μία συνεχή τάση μεγέθους ανάλογου με το πλάτος αυτό. Αντί αυτής, στη πραγματική περίπτωση που παρουσιάζεται στο Σχήμα 5.2, κάθε χαμηλοπερατό φίλτρο των 20Hz θα παρέχει μία τάση που είναι ανάλογη με το πλάτος της εξόδου του συνδεδεμένου με αυτό ζωνοπερατό φίλτρο των 200Hz.

Επιπρόσθετα, η εισερχόμενη ομιλία εφαρμόζεται σε έναν διευκρινιστή συχνοτήτων που ακολουθείται από ένα χαμηλοπερατό φίλτρο των 20Hz. Όταν το σήμα είναι φωνούμενο, η έξοδος του φίλτρου παρέχει μία τάση που είναι ανάλογη με τη συχνότητα φωνής. Η συχνότητα αυτή είναι ο τόνος της φωνής. Η ιδιαιτερότητα του συνδυασμού διαχωριστή-φίλτρου έγκειται στο ότι όταν η ομιλία είναι φωνούμενη, η έξοδος του φίλτρου είναι μία τάση μικρότερη από αυτή που συναντάμε όταν η ομιλία είναι μη φωνούμενη. Σημειώνοντας την έξοδο του συνδυασμού διαχωριστή-φίλτρου, μπορούμε, χρησιμοποιώντας έναν ανιχνευτή, να καθορίσουμε εάν η ομιλία είναι φωνούμενη ή μη, και, εάν είναι φωνούμενη, τάση που ανιχνεύεται να καθοριστεί από τον τόνο.

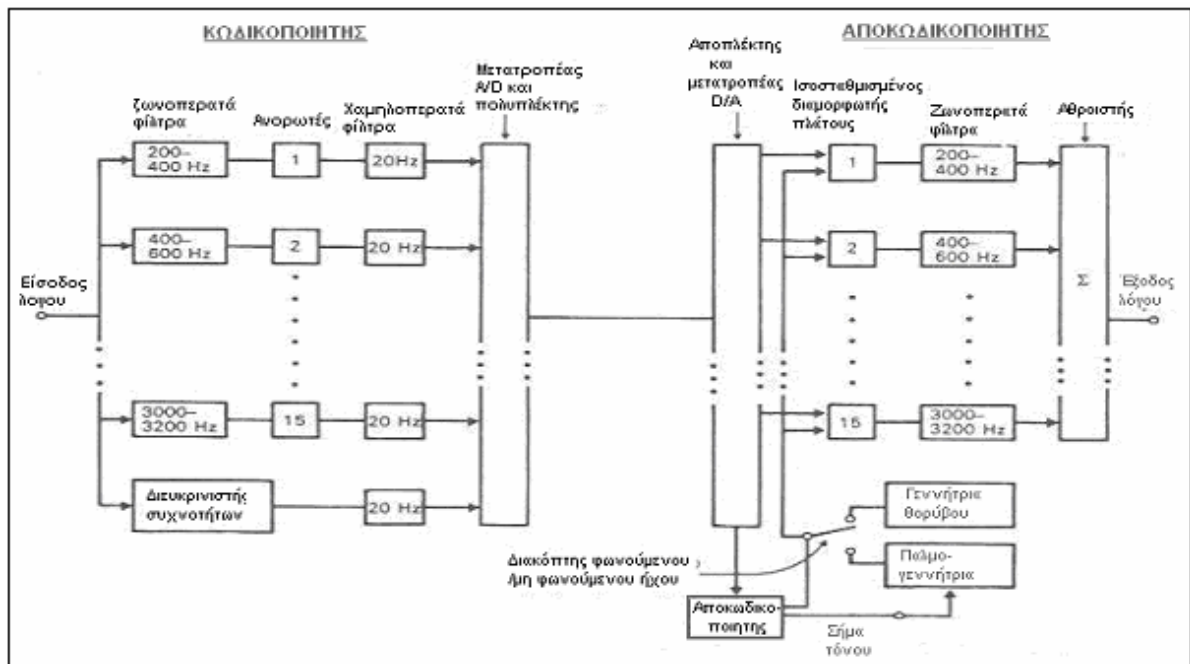
Οι έξοδοι των 16 χαμηλοπερατών φίλτρων 20Hz υφίστανται δειγματοληψία, πολυπλέκονται και μετατρέπονται από αναλογικές σε ψηφιακές. Εάν η δειγματοληψία γίνεται στο ρυθμό Nyquist των 40 δειγμάτων/s (που αντιστοιχεί σε σήματα με εύρος ζώνης 20Hz) και εάν χρησιμοποιούμε 3 bit/δείγμα για να παραστήσουμε κάθε δείγμα τάσης, τότε το bit rate είναι:

$$40 \frac{\text{δείγματα} / \text{sec}}{\text{φίλτρο}} \times 16 \text{ φίλτρα} \times 3 \text{ bits} / \text{δείγμα} = 1,9 \times 10^3 \text{ bits} / \text{sec} \quad (\text{Εξ. 1})$$

Σε οποιαδήποτε ειδική περίπτωση, το bit rate θα εξαρτάται από τον ρυθμό των χαμηλοπερατών φίλτρων, τις ζώνες διέλευσης τους και τον αριθμό των επιτρεπόμενων bit/δείγμα. Οι αντιπροσωπευτικοί bit rate ποικίλουν από $1,2 \times 1000$ έως $2,4 \times 1000$ b/s.

Όπως θα δούμε στην Εξ. 1, στο δέκτη του κωδικοποιητή φωνής, το σήμα αποπλέκεται και αποκωδικοποιείται, δηλαδή μετατρέπεται και πάλι στην αναλογική του μορφή. Σε αντιστοιχία με κάθε συνδυασμό φίλτρου-ανορθωτή στον κωδικοποιητή, παρέχεται στον αποκωδικοποιητή ένας ισοσταθμισμένος διαμορφωτής πλάτους και ένα ζωνοπερατό φίλτρο με την ίδια ζώνη διέλευσης. Θα θυμίσουμε ότι ένας ισοσταθμισμένος διαμορφωτής πλάτους είναι ένας διαμορφωτής που παρέχει μηδενική στάθμη φέροντος στην έξοδο για μηδενική διαμόρφωση στην είσοδο. Το φέρον εισόδου σε κάθε διαμορφωτή είναι ο θόρυβος ή η κυματομορφή της γεννήτριας παλμών. Η προς διαμόρφωση είσοδος είναι το πλάτος του σήματος (για καθένα από τους 15 διαμορφωτές-φίλτρα) που παρέχεται από τον κωδικοποιητή. Κάθε φορά που γίνεται δειγματοληψία, η πληροφορία για το πλάτος ενημερώνεται όπως συμβαίνει και με την πληροφορία για το αν η κυματομορφή του λόγου είναι φωνούμενη ή όχι, και, εάν είναι φωνούμενη, ο τόνος δίνεται από τον διαχωριστή.

Υποθέστε τώρα ότι σε κάποιο διάστημα δειγματοληψίας συμβαίνει ο λόγος να είναι φωνούμενος και η θεμελιώδης συχνότητα του παραγόμενου ήχου να είναι 450Hz. Ο ήχος αυτός θα παράγει ένα σήμα πλάτους στην έξοδο που συνδέεται με τον ανορθωτή 2. Ο ήχος αυτός θα έχει αρμονικές και έτσι μπορεί να υπάρχουν πλάτη στις εξόδους στις συχνότητες $2 \times 450 = 900\text{Hz}$, $3 \times 450 = 1350\text{Hz}$ κ.λ.π., που θα δούμε στις εξόδους των ανορθωτών 4, 6, κ.λ.π. Ο αποκωδικοποιητής θα λάβει τη πληροφορία ότι η ομιλία είναι φωνούμενη.



ΣΧΗΜΑ 5.2: Κωδικοποιητής φωνής καναλιού

Έτσι, ο διακόπτης του αποκωδικοποιητή θα συνδεθεί στη γεννήτρια παλμών και η συχνότητα της γεννήτριας παλμών θα τεθεί στον τόνο της φωνής. Τότε, η κυματομορφή της γεννήτριας παλμών αποτελεί ένα φέρον σαν είσοδο σε όλους τους διαμορφωτές. Ωστόσο, μόνο οι διαμορφωτές 2, 4, 6 κ.λ.π. θα παράγουν εξόδους. Οι εξόδοι αυτές είναι αναπαραγωγές της κυματομορφής της γεννήτριας παλμών διαμορφωμένες κατά πλάτος και συνεπώς περιλαμβάνουν μια θεμελιώδη συχνότητα και μία διαδοχή αρμονικών. Παρ' όλα αυτά, το φίλτρο που βρίσκεται μετά τον διαμορφωτή 2 θα καταστείλει όλες τις συχνότητες εκτός της αρμονικής, το φίλτρο που βρίσκεται μετά το διαμορφωτή 4 θα καταστείλει όλες τις συχνότητες εκτός της δεύτερης αρμονικής, κ.λ.π. Όλες αυτές οι κυματομορφές προστίθενται και η προκύπτουσα κυματομορφή εξόδου αποτελείται από ένα συνδυασμό μίας θεμελιώδους και των αρμονικών με τα ίδια σχετικά πλάτη που υπήρχαν στην ομιλία στην

είσοδο. Όταν η ομιλία στην είσοδο είναι φωνομένη, ο ήχος μοιάζει με θόρυβο και το φάσμα του εκτείνεται σε ολόκληρο το διάστημα των συχνοτήτων του λόγου, όχι απαραίτητα με ομοιόμορφο τρόπο. Στη περίπτωση αυτή μπορούμε να αναμένουμε ότι όλες οι οδεύσεις στο κωδικοποιητή θα παρέχουν μία έξοδο όπως και όλες οι οδεύσεις στον αποκωδικοποιητή.

Είναι επίσης ενδιαφέρον να σημειώσουμε ότι ένας κωδικοποιητής φωνής μπορεί να σχεδιασθεί ώστε να μεταδίδει τις εξόδους του ανιχνευτή της μεγαλύτερης περιβάλλουσας αγνοώντας εντελώς τις υπόλοιπες εξόδους. Για παράδειγμα εάν στέλλονται οι τρεις μεγαλύτερες εξόδους, 4bit/φίλτρο \times 3 φίλτρα = 12 πρόσθετα bit θα μεταδοθούν για να δηλώσουν στον δέκτη ποιες εξόδους φίλτρων στάλθηκαν. Στη περίπτωση αυτή, ο ρυθμός bit γίνεται :

$$40 \frac{\text{δείγματα} / \text{s}}{\text{φίλτρο}} \times 16 \text{φίλτρα} \times \frac{3\text{bits}}{\text{δείγμα}} + 40 \frac{\text{δείγματα} / \text{s}}{\text{δείγμα}} \times \frac{12\text{bits}}{\text{δείγμα}} = 840\text{bits} / \text{s} \quad (\text{Εξ. 2})$$

Δυστυχώς, η ποιότητα του σήματος υποβαθμίζεται σοβαρά. Ωστόσο, ο ρυθμός bit ελαττώνεται σημαντικά.

Οι vocoders καναλιών είναι γενικά πιο ικανοί στις διαδικασίες λιγότερης συμπίεσης σημάτων σε σχέση με τους LPC και formant vocoders. Ακόμα οι vocoders καναλιών τυπικά λειτουργούν μεταξύ 1 και 2 Kbits/s. Ακόμα κι αν αυτοί οι κωδικοποιητές είναι αποδοτικοί, παράγουν μια συνθετική ποιότητα και επομένως δεν χρησιμοποιούνται γενικά στα εμπορικά συστήματα.

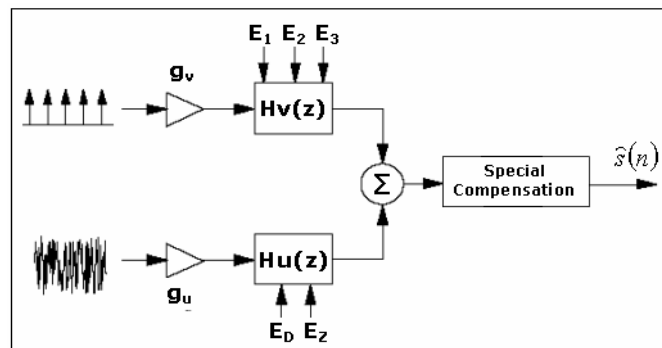
Εκτός από τους κλασσικούς vocoder καναλιών, το 1981 κατασκευάστηκαν οι πρώτοι ψηφιακοί channel vocoders από τον Gold. Οι ψηφιακοί vocoders καναλιών πρόσφεραν νέες βελτιωμένες ψηφιακές εφαρμογές. Τα πλεονεκτήματα των ψηφιακών channel vocoders είναι τα εξής:

- Αυξημένος αριθμός καναλιών (μεγαλύτερη ανάλυση)
- Φασματικές τεχνικές εξομάλυνσης στο σήμα διέγερσης με τη χρησιμοποίηση των τεχνικών DPCM
- Νέες εφαρμογές vocoders καναλιών, όπως, κανάλι vocoding των υπολοίπων LP και της καταστολής θορύβου.

5.3.2. FORMANT VOCODERS

Formant είναι οι vocoder στους οποίους η συχνότητα και το πλάτος των πρώτων τριών ή τεσσάρων formants (συχνοτήτων αντήχησης) εξάγονται και διαβιβάζονται. Η επανασύνθεση μπορεί να γίνει είτε σειριακά είτε παράλληλα, και οι διαφορετικές κατηγορίες λεκτικών ήχων αποκρίνονται καλύτερα σε διαφορετικές διαμορφώσεις. Οι formant vocoders είναι παρόμοιοι με τους channel vocoders (vocoders καναλιών) εκτός από το ότι τα χαρακτηριστικά συντονισμού του φίλτρου-bank στους formant vocoders προσαρμόζονται στις τροχιές των formants.

Αυτοί οι vocoders τυπικά λειτουργούν στο εύρος του 1 Kbit/s. Οι formant vocoders δεν είναι πολύ δημοφιλείς γιατί τα formants είναι δύσκολο να προβλεφθούν. Στο παρακάτω σχήμα φαίνεται ένας formant vocoder.

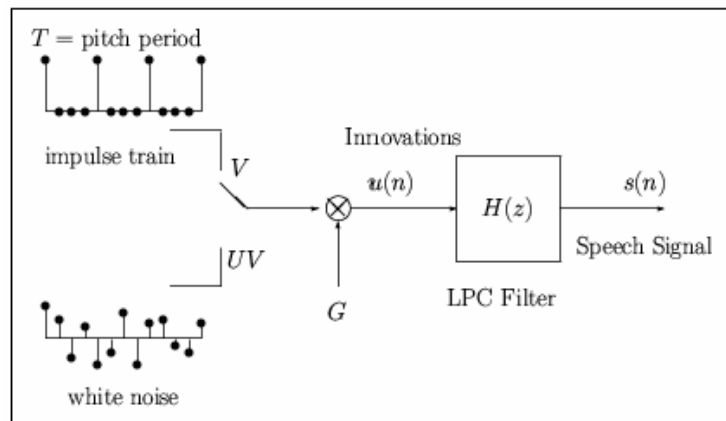


ΣΧΗΜΑ 5.3: Formant vocoder

5.3.3. ΚΩΔΙΚΟΠΟΙΗΤΕΣ ΓΡΑΜΜΙΚΗΣ ΠΡΟΒΛΕΨΗΣ (LPC)

Οι Linear predictive coders (LPC) πραγματοποιούν κωδικοποίηση στο πεδίο του χρόνου και αποτελούν υποκατηγορία των vocoders. Προσπαθούν να υπολογίσουν σημαντικά χαρακτηριστικά της φωνής από την κυματομορφή της ως προς το χρόνο. Αν και έχουν μεγάλες υπολογιστικές απαιτήσεις παρουσιάζουν πολύ καλές επιδόσεις σε μικρούς ρυθμούς δεδομένων (κάτω από τα 4,8 kbps).

Το παρακάτω πρότυπο (Σχ. 5.4) καλείται μαθηματικό πρότυπο LPC. Το πρότυπο λέει ότι το ψηφιακό λεκτικό σήμα είναι η έξοδος ενός ψηφιακού φίλτρου (αποκαλούμενου φίλτρο LPC) του οποίου η είσοδος είναι είτε μια σειρά ωθήσεων είτε μια ακολουθία θορύβου.



ΣΧΗΜΑ 5.4: Μαθηματικό πρότυπο LPC

Η ΣΧΕΣΗ ΜΕΤΑΞΥ ΤΩΝ ΦΥΣΙΚΩΝ ΚΑΙ ΜΑΘΗΜΑΤΙΚΩΝ ΠΡΟΤΥΠΩΝ		
Φωνητική περιοχή	↔	$H(z)$ (Φίλτρο LPC)
Αέρας	↔	$u(n)$ (innovation)
Φωνητική δόνηση χορδών	↔	V (φωνούμενος)
Φωνητική περίοδος δόνησης χορδών	↔	T (περίοδος pitch)
Fricatives και Plosives	↔	UV (μη φωνούμενος)
Ισχύς αέρα	↔	G (κέρδος)

Το φίλτρο LPC δίνεται από:

$$H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_{10} z^{-10}}$$

που ισοδύναμα μας λέει ότι η σχέση εισόδου-εξόδου του φίλτρου δίνεται από τη γραμμική εξίσωση διαφοράς:

$$s(n) + \sum_{i=1}^{10} a_i s(n-i) = u(n)$$

Το πρότυπο LPC μπορεί να αντιπροσωπευθεί με διανυσματική μορφή όπως:

$$\mathbf{A} = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, G, V/UV, T)$$

Το A αλλάζει περίπου κάθε 20 msec. Σε ένα ποσοστό δειγματοληψίας 8000 δειγμάτων/sec, τα 20 msec είναι ισοδύναμο με 160 δείγματα. Το ψηφιακό λεκτικό σήμα διαιρείται σε frames μεγέθους 20msec. Υπάρχουν 50 frames/second. Το πρότυπο λέει ότι:

$$\mathbf{A} = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, G, V/UV, T)$$

είναι ισοδύναμο με

$$\mathbf{S} = (s(0), s(1), \dots, s(159))$$

Κατά συνέπεια οι 160 τιμές S αντιπροσωπεύονται από τις 13 τιμές του A. Δεν υπάρχει σχεδόν καμία αντιληπτή διαφορά στο S εάν:

- Για τους φωνοϋμένους ήχους (V): η ακολουθία ώθησης μετατοπίζεται (ανεξάρτητα στην αλλαγή φάσης).
- Για τους μη φωνοϋμένους ήχους (UV): μια διαφορετική ακολουθία θορύβου χρησιμοποιείται.

5.3.3.1. ΑΝΑΛΥΣΗ LPC

Εξετάστε ένα frame του σήματος φωνής:

$$\mathbf{S} = (s(0), s(1), \dots, s(159))$$

Το σήμα $s(n)$ συσχετίζεται με την καινοτομία (innovation) $u(n)$ μέσω της γραμμικής εξίσωσης διαφοράς:

$$s(n) + \sum_{i=1}^{10} a_i s(n-i) = u(n)$$

Οι δέκα παράμετροι LPC (a_1, a_2, \dots, a_{10}) επιλέγονται για να ελαχιστοποιήσουν την ενέργεια της καινοτομίας (innovation) :

$$f = \sum_{n=0}^{159} u^2(n)$$

Χρησιμοποιώντας τον τυποποιημένο υπολογισμό, παίρνουμε τη παράγωγο f ως προς a_i και το θέτουμε ίσο με μηδέν:

$$\begin{aligned} df/da_1 &= 0 \\ df/da_2 &= 0 \\ &\dots \\ df/da_{10} &= 0 \end{aligned}$$

Έχουμε τώρα 10 γραμμικές εξισώσεις με 10 αγνώστους:

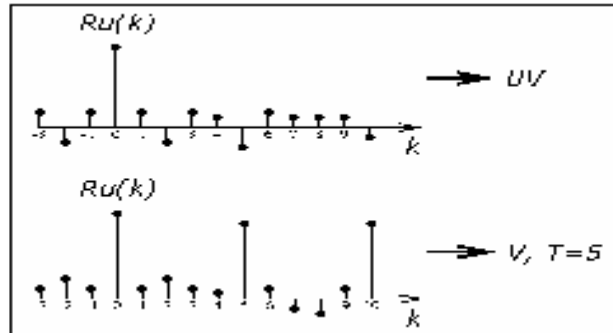
$$\begin{bmatrix} R(0) & R(1) & R(2) & R(3) & R(4) & R(5) & R(6) & R(7) & R(8) & R(9) \\ R(1) & R(0) & R(1) & R(2) & R(3) & R(4) & R(5) & R(6) & R(7) & R(8) \\ R(2) & R(1) & R(0) & R(1) & R(2) & R(3) & R(4) & R(5) & R(6) & R(7) \\ R(3) & R(2) & R(1) & R(0) & R(1) & R(2) & R(3) & R(4) & R(5) & R(6) \\ R(4) & R(3) & R(2) & R(1) & R(0) & R(1) & R(2) & R(3) & R(4) & R(5) \\ R(5) & R(4) & R(3) & R(2) & R(1) & R(0) & R(1) & R(2) & R(3) & R(4) \\ R(6) & R(5) & R(4) & R(3) & R(2) & R(1) & R(0) & R(1) & R(2) & R(3) \\ R(7) & R(6) & R(5) & R(4) & R(3) & R(2) & R(1) & R(0) & R(1) & R(2) \\ R(8) & R(7) & R(6) & R(5) & R(4) & R(3) & R(2) & R(1) & R(0) & R(1) \\ R(9) & R(8) & R(7) & R(6) & R(5) & R(4) & R(3) & R(2) & R(1) & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \\ a_{10} \end{bmatrix} = \begin{bmatrix} -R(1) \\ -R(2) \\ -R(3) \\ -R(4) \\ -R(5) \\ -R(6) \\ -R(7) \\ -R(8) \\ -R(9) \\ -R(10) \end{bmatrix}$$

όπου

$$\begin{aligned} R(k) &= \sum_{n=0}^{159-k} s(n)s(n+k) \\ &\equiv \text{autocorrelation of } s(n) \end{aligned}$$

Η παραπάνω εξίσωση πινάκων θα μπορούσε να λυθεί χρησιμοποιώντας:

- Την γκαουσιανή μέθοδος αποβολών.
- Οποιαδήποτε μέθοδος αντιστροφής πινάκων (MATLAB).
- Recursion Levinson-Durbin.

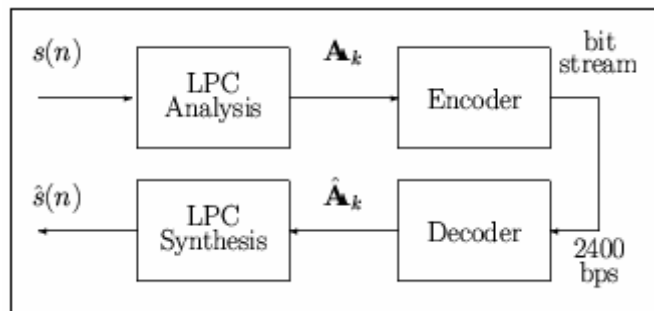


5.3.3.2. ΠΡΟΤΥΠΟ LPC-10

Αποκαλούμενος έτσι επειδή 10 συντελεστές LP χρησιμοποιούνται. Ποσοστό δειγμάτων 8 kHz, 180 δείγματα/frame, 50 frame/sec, 34 bit για το LSP (line spectrum pair), 7 bit που χρησιμοποιούνται για τον τόνο και τη σταθερότητα φωνής, 7 bit που χρησιμοποιούνται για το κέρδος. Δηλαδή συνολικά 48 bit ανά frame και ρυθμός μετάδοσης 2,4 Kbps όπως φαίνεται και από τον παρακάτω τύπο:

$$(34+7+7) \frac{\text{bits}}{\text{frames}} \cdot 50 \frac{\text{frames}}{\text{sec}} = 2.4 \text{ kbps}$$

Το παρακάτω είναι ένα block διάγραμμα vocoder 2,4 kbps LPC:



Οι συντελεστές LPC αντιπροσωπεύονται ως παράμετροι ζευγαριού φάσματος γραμμών (line spectrum pair - LSP). Το LSP είναι από μαθηματική άποψη ισοδύναμο (ένα προς ένα) με το LPC και πιο υποκειμενικό στην κβαντοποίηση. Το LSP υπολογίζεται ως εξής:

$$P(z) = 1 + (a_1 - a_{10})z^{-1} + (a_2 - a_9)z^{-2} + \dots + (a_{10} - a_1)z^{-10} - z^{-11}$$

$$Q(z) = 1 + (a_1 + a_{10})z^{-1} + (a_2 + a_9)z^{-2} + \dots + (a_{10} + a_1)z^{-10} + z^{-11}$$

Παραγοντοποιώντας τις παραπάνω εξισώσεις, παίρνουμε:

$$P(z) = (1 - z^{-1}) \prod_{k=2,4,\dots,10} (1 - 2 \cos \omega_k z^{-1} + z^{-2})$$

$$Q(z) = (1 + z^{-1}) \prod_{k=1,3,\dots,9} (1 - 2 \cos \omega_k z^{-1} + z^{-2})$$

όπου $\{\omega\}_{K=1}^{10}$ καλείται παράμετρος LSP.

LSP ταξινομούνται και οριοθετούνται:

$$0 < \omega_1 < \omega_2 < \dots < \omega_{10} < \pi$$

LSP συσχετίζεται από ένα frame στο επόμενο από το LPC. Το μέγεθος των frame είναι 20msec. Υπάρχουν 50frames/sec 2400bps είναι ισοδύναμο με 48bits/frame. Αυτά τα κομμάτια διατίθενται ως εξής:

Parameter Name	Parameter Notation	Rate (bits/frame)
LPC (LSP)	$\{a_k\}_{k=1}^{10}$ ($\{\omega_k\}_{k=1}^{10}$)	34
Gain	G	7
Voiced/Unvoiced & Period	$V/UV, T$	7
Total		48

Τα 34 bit για το LSP διατίθενται ως εξής:

LSP	No. of Bits
ω_1	3
ω_2	4
ω_3	4
ω_4	4
ω_5	4
ω_6	3
ω_7	3
ω_8	3
ω_9	3
ω_{10}	3
Total	34

Το κέρδος G , κωδικοποιείται χρησιμοποιώντας ένα 7bit ανομοιόμορφο κλιμακωτό κβαντιστή (ένας 1-διαστατικός διάνυσμα κβαντοποίησης). Για την εκφρασμένη ομιλία, οι τιμές T των σειρών από 20 έως 146, $V/UV, T$ κωδικοποιούνται από κοινού ως εξής:

V/UV	T	Encoded Value
UV	—	0
V	20	1
V	21	2
V	22	3
V	23	4
:	:	:
:	:	:
V	146	127

Μειονεκτήματα:

- Μπορεί να ακούγεται παρασιτικός θόρυβος (buzzy)
- Η φτωχή LP μοντελοποίηση οδηγεί σε μεγάλα εύρη ζώνης και στη γρήγορη αλλοίωση της διέγερσης παλμού
- Τα σφάλματα φωνής παράγουν σημαντικές παραμορφώσεις
- Δεν είναι προσαρμοσμένο στο μοντέλο οξύ μουσικού τόνου (nasals)
- Χρησιμοποιείται μόνο στα μοντέλα φωνής - δεν λειτουργεί εάν υπάρχει παρασιτικός θόρυβος [background noise] (δηλ. μη προσαρμοσμένο στις κινητές τηλεφωνικές εφαρμογές χωρίς περαιτέρω εργασία).

Στοιχεία που πρέπει να εξεταστούν:

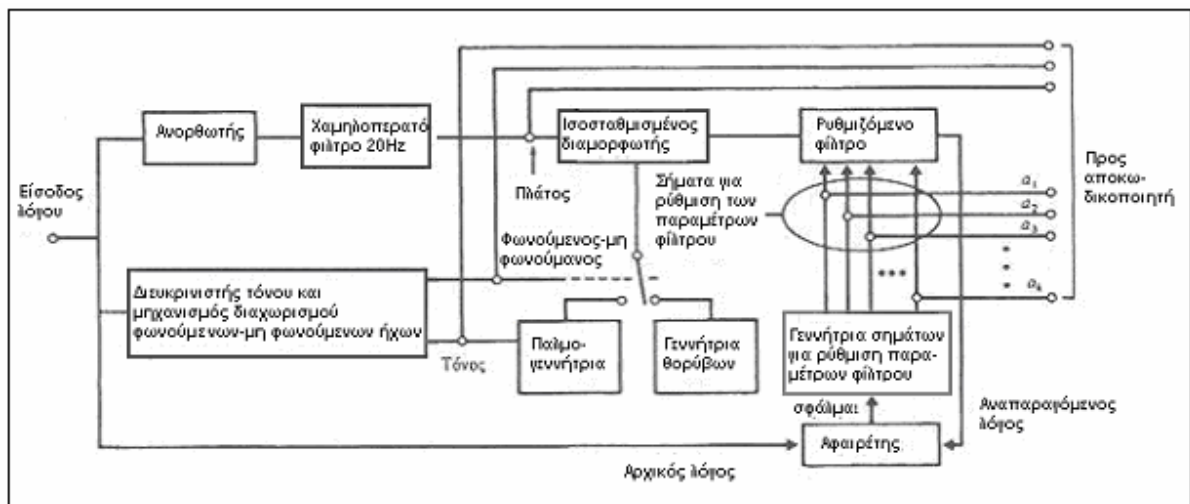
- Γλωττική παλμοδιαμόρφωση (Glottal pulse shaping)
- Ενημέρωση σύγχρονων παραμέτρων τόνου φωνής
- Λεπτός συντονισμός της φωνητικής σταθερότητας
- Διαχωρισμός της ομιλίας και του θορύβου
- Εκμετάλλευση των χρονικών συσχετισμών των ακουστικών διανυσμάτων

5.3.3.3. ΛΕΙΤΟΥΡΓΙΑ LPC

Αναφερόμενοι στο Σχήμα 5.2, παρατηρούμε ότι ο μηχανισμός με τον οποίο αναπαράγουμε την ομιλία στον αποκωδικοποιητή είναι ο ακόλουθος: Διαθέτουμε στον αποκωδικοποιητή μια γεννήτρια θορύβου και μια γεννήτρια παλμών. Παρέχουμε

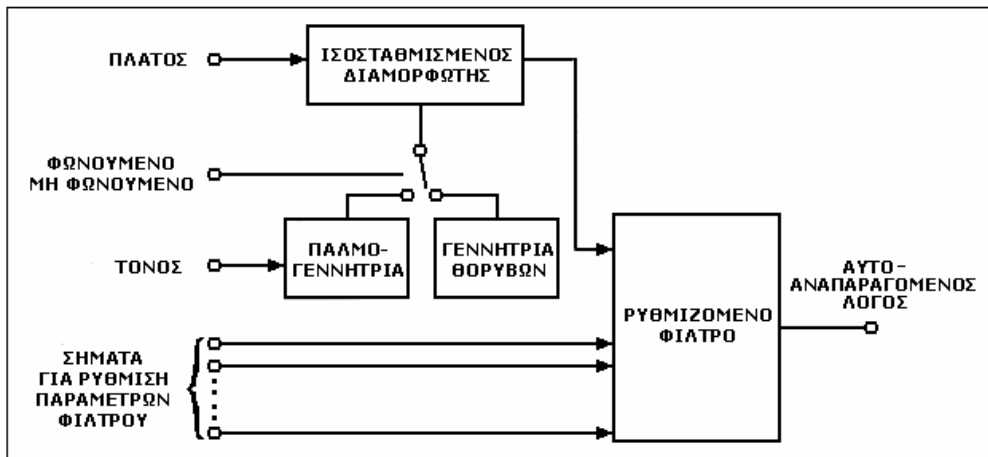
πληροφορία στο αποκωδικοποιητή για το εάν η ομιλία είναι φωνούμενη ή όχι και, εάν είναι αν είναι φωνούμενη, τον τόνο. Οι έξοδοι αυτών των γεννητριών ήχου, δηλαδή ο θόρυβος και η γεννήτρια παλμών διέρχονται στη συνέχεια από ένα φίλτρο με ρυθμιζόμενες παραμέτρους. Το φίλτρο αποτελείται από διαμορφωτές πλάτους και ζωνοπερατά φίλτρα. Τα ρυθμιζόμενα σήματα είναι οι διαμορφώνουσες είσοδοι στους διαμορφωτές. Τα σήματα αυτά καθορίζουν τις σχετικές αναλογίες θεμελιώδους και αρμονικών οι οποίες συνδυάζονται στην έξοδο του αποκωδικοποιητή.

Συμβαίνει να μπορούμε να βελτιώσουμε την απόδοση ενός συστήματος κωδικοποίησης φωνής με τους ακόλουθους τρόπους: Αρχικά, ας χρησιμοποιήσουμε ένα φίλτρο το οποίο έχει μεγαλύτερη ευελιξία στη δυνατότητα ρύθμισης του. Στη συνέχεια, ας διαθέσουμε το ίδιο φίλτρο στον κωδικοποιητή. Κάλιστα θα είμαστε τότε ικανοί να αναπαράγουμε την ομιλία όχι μόνο στον αποκωδικοποιητή αλλά και στον κωδικοποιητή. Έτσι ο κωδικοποιητής θα είναι ικανός να "ακούσει" πως ακούγεται ο ήχος όταν αναπαράγεται στον αποκωδικοποιητή. Τελικά ας συγκρίνουμε στον κωδικοποιητή, την αρχική ομιλία και την αναπαραχθείσα της εκδοχή και, όπως σε κάθε σύστημα ελέγχου με ανάδραση, ας χρησιμοποιήσουμε τις διαφορές, δηλαδή το σήμα σφάλματος που εμφανίζεται, για να ρυθμίσουμε τις παραμέτρους του φίλτρου για ελαχιστοποίηση του σφάλματος. Με αυτό τον τρόπο λειτουργεί ο κωδικοποιητής γραμμικής πρόβλεψης (LPC) του Σχήματος. 7.5.



ΣΧΗΜΑ 5.5: Ένας απλοποιημένος κωδικοποιητής γραμμικής πρόβλεψης

Όπως και στον απλούστερο κωδικοποιητή, παρέχεται ένας διαχωριστής συχνοτήτων και άλλο υλικό (hardware) για να κάνει διαθέσιμο ένα τόνο και ένα σήμα φωνούμενο - μη φωνούμενο. Χρησιμοποιείται ένας απλός ανορθωτής και ένα φίλτρο για να παράγει μόνο ένα πλάτος σήματος. Δίνουμε επίσης το αναγκαίο υλικό (hardware) για την αναπαραγωγή της ομιλίας, δηλαδή μια γεννήτρια παλμών, μια γεννήτρια θορύβου, ένα διακόπτη και έναν ισοσταθμισμένο διαμορφωτή. Το σήμα που χρησιμοποιείται για να ρυθμίσει τις παραμέτρους του ρυθμιζόμενου φίλτρου είναι το σήμα διαφοράς (σφάλμα) μεταξύ της αρχικής ομιλίας και της αναπαραχθείσας μορφής όπως δημιουργήθηκε στην έξοδο του φίλτρου. Αυτό το σήμα σφάλματος εφαρμόζεται σε μια διάταξη που υπολογίζει και παράγει τα σήματα που χρειάζεται να εφαρμοστούν στο ρυθμιζόμενο φίλτρο για να ελαχιστοποιηθεί το σήμα σφάλματος. Ο αποκωδικοποιητής παρουσιάζεται στο σχήμα 5.6.



ΣΧΗΜΑ 5.6: Απλοποιημένος αποκωδικοποιητής γραμμικής πρόβλεψης

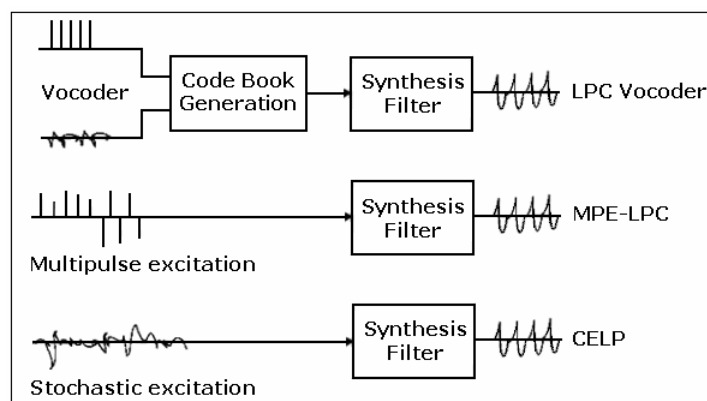
Λαμβάνει το πλάτος, σήματα φωνούμενα - μη φωνούμενα και τόνους. Αυτά χρησιμοποιούνται σε σύνδεση με το διαμορφωτή και τις γεννήτριες παλμού και θορύβου, όπως στον κωδικοποιητή, για να παρέχουν μία είσοδο στο ρυθμιζόμενο φίλτρο. Τα σήματα ρύθμισης των παραμέτρων φίλτρου επίσης λαμβάνονται και χρησιμοποιούνται, όπως στον κωδικοποιητή, για να ρυθμίσουν τα χαρακτηριστικά του φίλτρου για βέλτιστη αναπαραγωγή φωνής.

Δεν παρουσιάζεται καθαρά στα σχήματα 5.5 και 5.6, αλλά μολαταύτα θα πρέπει να γίνει καταληπτό ότι, όπως και στον απλούστερο κωδικοποιητή φωνής του σχήματος 5.2, το μεταδιδόμενο σήμα θα πρέπει να είναι το πολυπλεγμένο με διαίρεση χρόνου των ξεχωριστών σημάτων που πρόκειται να μεταδοθούν. Τυπικά χρησιμοποιούνται 18 σήματα αϊ για τη ρύθμιση του φίλτρου. Εάν κάνουμε δειγματοληψία όπως προηγουμένως με ρυθμό 40δείγματα/sec και κωδικοποιούμε κάθε τιμή δείγματος με 3bit, τότε το bit rate R είναι:

$$R = \frac{(18+3)\text{σήματα} \times 40\text{δείγματα ανά sec}}{\text{σήμα} \times 3\text{bits ανά δείγμα}} = 2,52\text{kb/s}$$

5.3.3.4. ΕΝΑΛΛ. ΜΕΘΟΔΟΙ ΔΗΜΙΟΥΡΓΙΑΣ ΤΟΥ ΣΗΜΑΤΟΣ ΔΙΕΓΕΡΣΗΣ ΣΤΟΝ ΑΠΟΚΩΔ. LPC

Υπάρχουν διάφοροι τρόποι LPC που διαφέρουν στον τρόπο με τον οποίο το error σήμα αναδημιουργείται στο δέκτη. Τρεις εναλλακτικές μέθοδοι δημιουργίας του σήματος διέγερσης στον αποκωδικοποιητή LPC φαίνονται στο παρακάτω σχήμα.

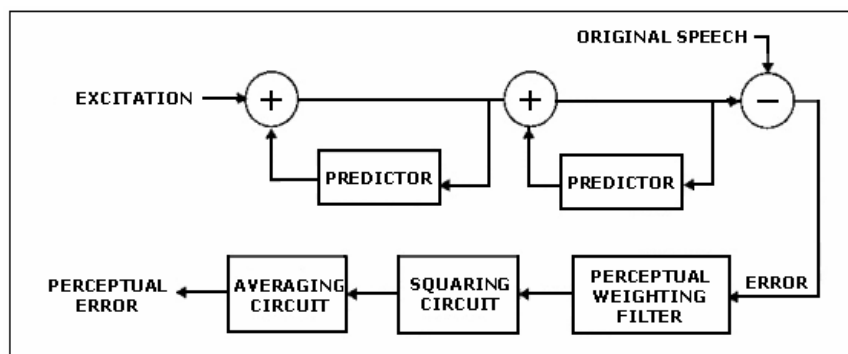


α) LPC Vocoder: Η πρώτη μέθοδος η οποία είναι αρκετά δημοφιλής για τη δημιουργία της διέγερσης χρησιμοποιεί δύο πηγές στο δέκτη. Η μία παράγει λευκό θόρυβο και η άλλη μια σειρά παλμών με ρυθμό αυτό της τρέχουσας θεμελιώδης συχνότητας. Η μέθοδος αυτή εμφανίζει δύο μεγάλα προβλήματα. Η θεμελιώδη συχνότητα η οποία αποστέλλεται από τον πομπό είναι πολλές φορές δύσκολο να εξαχθεί. Επιπλέον, σε αυτή τη μέθοδο πραγματοποιείται αλληλεπίδραση μεταξύ των φάσεων των συνιστωσών του παλμού

διέγερσης με αποτέλεσμα να παράγεται ένας οξύς τόνος στη φωνή που συνθέεται. Τα παραπάνω προβλήματα μετριάζονται στις δύο άλλες προσεγγίσεις.

b) Multi-pulse Excited LPC (MPE-LPC): Ο Atal έδειξε ότι όσο καλά και αν τοποθετηθεί ο παλμός, η διέγερση ενός και μόνο παλμού ανά περίοδο της θεμελιώδης συχνότητας παράγει αρκετά μεγάλη παραμόρφωση στο σήμα. Γι' αυτό το λόγο πρότεινε τη χρησιμοποίηση περισσοτέρων παλμών (συγκεκριμένα 8 ανά περίοδο) και τη διαδοχική ρύθμιση των θέσεων των παλμών καθώς και των πλατών ώστε να ελαχιστοποιηθεί το μέσο φασματικό τετραγωνικό σφάλμα. Αυτή η τεχνική ονομάζεται multi-pulse excited LPC και έχει ως αποτέλεσμα καλύτερη ποιότητας φωνή. Η τεχνική αυτή έχει το πλεονέκτημα ότι η πρόβλεψη υπολογίζεται καλύτερα χρησιμοποιώντας περισσότερους παλμούς ανά περίοδο της θεμελιώδης συχνότητας και επίσης ότι δεν είναι απαραίτητος ο υπολογισμός της θεμελιώδης συχνότητας.

c) Code-Excited LPC (CELP): Σε αυτή τη μέθοδο ο κωδικοποιητής και ο αποκωδικοποιητής έχουν προκαθορισμένο codebook (βιβλίο κωδικών) στοχαστικών σημάτων διέγερσης. Για κάθε σήμα φωνής ο πομπός ψάχνει στο codebook των στοχαστικών σημάτων για την εύρεση της βέλτιστης διέγερσης η οποία θα χρησιμοποιηθεί στο LPC φίλτρο. Στη συνέχεια μεταδίδεται ο δείκτης του codebook που αντιστοιχεί στο καλύτερο ταίριασμα. Στο δέκτη, ο δείκτης αυτός χρησιμοποιείται ώστε να επιλεγεί το σωστό σήμα διέγερσης για το φίλτρο σύνθεσης. Οι κωδικοποιητές τύπου CELP είναι αρκετά πολύπλοκοι και απαιτούν μεγάλη υπολογιστική ισχύ για την λειτουργία τους. Παρέχουν, όμως, υψηλής ποιότητας φωνή, με το πλεονέκτημα ότι η διέγερση κωδικοποιείται μόνο με 0.25 bits ανά δείγμα. Οι κωδικοποιητές CELP επιτυγχάνουν ρυθμό μετάδοσης έως 4.8kbrps. Στο παρακάτω σχήμα φαίνεται η διαδικασία για την επιλογή του βέλτιστου σήματος διέγερσης.



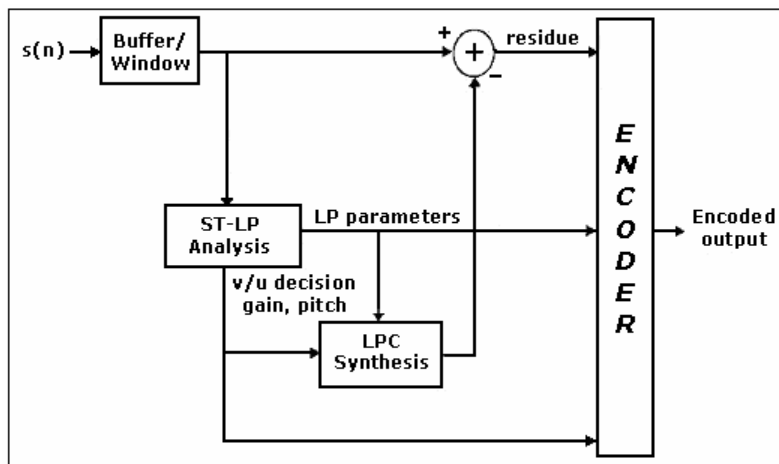
Για καλύτερη κατανόηση θα περιγράψουμε τη διαδικασία με ένα παράδειγμα. Έστω ότι θέλουμε να κωδικοποιήσουμε ένα σύντομο block σήματος φωνής διάρκειας 5ms. Χρησιμοποιώντας συχνότητα δειγματοληψίας 8 kHz κάθε block αποτελείται από 40 δείγματα φωνής. Θεωρώντας ρυθμό δεδομένων 1/4 bit ανά δείγμα υπολογίζουμε ότι απαιτούνται 10 bit ανά block. Κάθε μέλος του codebook παρέχει 40 δείγματα του σήματος διέγερσης μαζί με ένα παράγοντα διαβάθμισης ο οποίος αλλάζει κάθε 5ms block. Τα διαβαθμισμένα δείγματα περνούν διαδοχικά από δύο αναδρομικά φίλτρα, τα οποία εισάγουν περιοδικότητα στο σήμα και ρυθμίζουν το φάσμα του. Τα παραγόμενα δείγματα φωνής του δεύτερου φίλτρου συγκρίνονται με τα δείγματα του πραγματικού σήματος φωνής και σχηματίζουν το error σήμα. Το σήμα αυτό υφίσταται επιπλέον επεξεργασία μέσω ενός γραμμικού φίλτρου (Perceptual Weighting Filter) το οποίο ενισχύει τις συχνότητες που είναι σημαντικές για την καλύτερη αντίληψη της φωνής και εξασθενεί τις λιγότερες σημαντικές. Το codebook, δηλ. η γεννήτρια διέγερσης (generator excitation) είναι γνωστή και στον κωδικοποιητή και στον αποκωδικοποιητή. Αφού έχει βρει τη βέλτιστη ακολουθία $u(k)$, μόνο ο δείκτης της επιλεγμένης εισόδου πρέπει να διαβιβαστεί. Είναι αρκετά προφανές ότι μια εξαντλητική αναζήτηση όλων των πιθανών διανυσμάτων διέγερσης του codebook απαιτεί μια πολύ υψηλή πολυπλοκότητα του κωδικοποιητή-αποκωδικοποιητή που δεν μπορεί συχνά να παρασχεθεί ακόμη και από τους σύγχρονους επεξεργαστές σημάτων. Επομένως, πολλοί από τους δομημένους codebooks έχουν ερευνηθεί στο παρελθόν, για να ξεπεράσουν αυτό το πρόβλημα πολυπλοκότητας. Οι παραλλαγές των ανάλυση-από-σύνθεσης κωδικοποιητών (CELP) είναι δομές όπως ACELP (αλγεβρικό CELP), RPE (κανονική διέγερση παλμού), MPE (πολυπαλμικός CELP), VSELP (Vector-Sum Excited Linear Prediction).

Οι LPC Vocoders μπορούν να πετύχουν πολύ χαμηλούς ρυθμούς μετάδοσης αλλά χαρακτηρίζονται από κακή ποιότητα:

- Η περιοδική διέγερση ακούγεται αφύσικη

- Έχουμε πολλά σφάλματα στην απόφαση για voicing
- Δεν μοντελοποιούνται σωστά οι ένρινοι ήχοι.
- Δεν μπορούν να κωδικοποιήσουν background.

Τα προβλήματα αυτά μπορούν να αντιμετωπιστούν με την κωδικοποίηση και μετάδοση του σφάλματος πρόβλεψης όπως στους RELP vocoders. Το σύστημα κωδικοποίησης RELP βασίζεται ουσιαστικά στην ίδια λογική με αυτή του ADPCM συστήματος. Αφού υπολογιστούν οι παράμετροι του μοντέλου (παράγοντες του LPC) και οι παράγοντες διέγερσης (απόφαση αν είναι εύφωνο/άφωνο το φώνημα, θεμελιώδη συχνότητα, κέρδος) από το frame της φωνής, το σήμα φωνής συνθέτεται στον πομπό και αφαιρείται από το πραγματικό σήμα. Έτσι σχηματίζεται το σήμα σφάλματος (residual), το οποίο κβαντίζεται, κωδικοποιείται και μεταδίδεται στο δέκτη μαζί με τις υπόλοιπες παραμέτρους του LPC μοντέλου. Στο δέκτη συντίθεται μια προσέγγιση του αρχικού σήματος φωνής προσθέτοντας το residual σήμα στο σήμα που παράγεται, χρησιμοποιώντας παράλληλα και τις παραμέτρους του μοντέλου. Η ποιότητα του σήματος φωνής που συντίθεται βελτιώνεται επειδή υπάρχει η επιπρόσθετη πληροφορία του residual error σήματος. Στο παρακάτω σχήμα φαίνεται το διάγραμμα βαθμίδων του συστήματος κωδικοποίησης RELP:



Το πλεονέκτημα του RELP κωδικοποιητή είναι ότι παρέχει καλή ποιότητα φωνής έχοντας χαμηλή πολυπλοκότητα. Η ποιότητα φωνής, όμως, περιορίζεται από το συνολικό θόρυβο που εισάγει η διεργασία της αναπαραγωγής υψηλών συχνοτήτων καθώς και από τα λάθη που πραγματοποιούνται κατά τη μετάδοση.

ΕΠΙΛΟΓΟΣ - ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1. Η ΣΥΝΕΧΗΣ ΑΝΑΓΚΗ ΓΙΑ ΤΗ ΣΥΜΠΙΕΣΗ ΦΩΝΗΣ

Η συμπίεση φωνής είναι κεντρική στις τεχνολογίες της υπεραστικής επικοινωνίας, της υψηλής ποιότητας αποθήκευσης φωνής, και της αποκρυπτογράφησης μηνυμάτων. Η συμπίεση συνεχίζει να είναι μια βασική τεχνολογία στις επικοινωνίες παρά την υπόσχεση των οπτικών μέσων μετάδοσης σχετικά απεριόριστου εύρους ζώνης. Αυτό είναι λόγω της συνεχούς και, στην πραγματικότητα, αυξανόμενης ανάγκης μας να χρησιμοποιήσουμε τα περιορισμένης ζώνης μέσα όπως ραδιοφωνικές και δορυφορικές συνδέσεις, και περιορισμένης αναλογίας bit (bit rate) μέσα απομνημόνευσης όπως είναι τα CD ROM και οι μνήμες πυριτίου. Η αποθήκευση και αρχειοθέτηση των μεγάλων όγκων των προφορικών πληροφοριών καθιστούν τη συμπίεση φωνής ουσιαστική ακόμη και στα πλαίσια των σημαντικών αυξήσεων της χωρητικότητας των οπτικών και στερεάς κατάστασης (solid-state) μνημών.

Η τεχνολογία χαμηλού bit rate είναι ένας βασικός παράγοντας στην ικανοποίηση της αυξανόμενης ζήτησης για τις νέες ψηφιακές ασύρματες υπηρεσίες επικοινωνιών. Εντυπωσιακή πρόοδος έχει σημειωθεί κατά τη διάρκεια των τελευταίων χρόνων στην κωδικοποίηση φωνής, με υψηλή ποιότητα στα χαμηλά bit rate και με χαμηλότερο κόστος. Μόλις δέκα χρόνια πριν, η υψηλής ποιότητας φωνή δεν θα μπορούσε να παραχθεί σε bit rate κάτω από 24 kbps. Σήμερα, μπορούμε να προσφέρουμε υψηλή ποιότητα στα 8 kbps. Χρησιμοποιώντας νέες τεχνικές για την κωδικοποίηση και ισοστάθμιση (equalization) καναλιών, είναι δυνατό να μεταδοθεί η φωνή των 8 kbps με πιο σταθερό τρόπο διαμέσου του κινητού ραδιοφωνικού καναλιού, παρόλο το θόρυβο, την εξασθένιση του σήματος και τις παρεμβολές. Η έρευνα για την κατάλληλα συντονισμένη κωδικοποίηση πηγής και καναλιών απαιτείται για να βρεθεί μια καλή λύση σε αυτό το πρόβλημα.

Τα ασύρματα κανάλια επικοινωνίας πάσχουν από την πολλαπλών διαδρομών παρεμβολή, παράγοντας ποσοστό λάθους πάνω από 10%. Η πρόκληση για τη έρευνα φωνής είναι να παραχθεί ψηφιακή φωνή που μπορεί να διαβιβαστεί με υψηλή ποιότητα διαμέσου των δικτύων επικοινωνίας με την παρουσία καναλικών λαθών μέχρι 10%. Ένας κωδικοποιητής φωνής που λειτουργεί στα 2 kbps θα παρείχε αρκετά bit για τη διόρθωση τέτοιων καναλικών λαθών, αν υποθεθεί ότι ο συνολικός ρυθμός μετάδοσης είναι από 4 έως 8 kbps.

Το bit rate των 2 kbps έχει μια ελκυστική συνέπεια και στην αποθήκευση φωνής. Σε αυτό το bit rate, περισσότερο από 2 ώρες συνεχούς ομιλίας μπορούν να αποθηκευτούν σε ένα chip μνήμης 16 Mbit, επιτρέποντας υπηρεσίες μηνύματος φωνής στα διάφορα προσωπικά τερματικά επικοινωνίας, επεκτείνοντας σημαντικά τις ικανότητες των ψηφιακών αυτόματων τηλεφωνητών. Οι θεμελιώδεις πρόοδοι στην κατανόηση για την παραγωγή και αντίληψη της ομιλίας απαιτούνται για να επιτευχθεί υψηλής ποιότητας ομιλία στα 2 kbps.

Οι εφαρμογές της ευρείας ζώνης κωδικοποίησης φωνής περιλαμβάνουν υψηλής ποιότητας ακουστική διάσκεψη (audio conferencing) με εύρος ζώνης φωνής 7 KHz και bit rate από 16 έως 32 kbps, και υψηλής ποιότητας στερεοφωνική διάσκεψη (stereo conferencing) και dual-language προγραμματισμό πάνω σε μια βασική σύνδεση ISDN. Τέλος, η συμπίεση ενός εύρους ζώνης 20 KHz σε ρυθμούς που φτάνουν τα 64 kbps θα δημιουργήσει νέες ευκαιρίες στην ακουστική μετάδοση και δικτύωση, στις ηλεκτρονικές εκδόσεις, στην τηλεδιδασκαλία, στα διαδικτυακά παιχνίδια, στα υπομνήματα πολυμέσων και στην αποθήκευση βάσεων δεδομένων.

6.2. ΟΙ ΣΤΟΧΟΙ ΤΗΣ ΤΕΧΝΟΛΟΓΙΑΣ

Δεδομένου ότι δεν υπάρχει καμία αυστηρή μαθηματική φόρμα για την εντροπία φωνής, ένας φυσικός στόχος στη κωδικοποίηση φωνής είναι η επίτευξη υψηλής ποιότητας σε bit rates που είναι τουλάχιστον κατά ένα παράγοντα δύο χαμηλότερα από τους αριθμούς που παρέχουν στις μέρες μας υψηλή ποιότητα, δηλαδή 4 kbps για την τηλεφωνική ομιλία, 8 kbps για την ομιλία ευρείας ζώνης και 24 kbps για την ομιλία που φτάνει την ποιότητα

των CD. Αυτοί οι αριθμοί αντιπροσωπεύουν ένα bit rate περίπου 0,5 bit ανά δείγμα σε κάθε περίπτωση.

Μια άλλη πρόκληση είναι η πραγματοποίηση αλγορίθμων όσον αφορά τις πρακτικές ατέλειες όπως ο θόρυβος εισόδου και τα λάθη κατά την μετάδοση.

Τέλος, ένα κύριο σύνολο προκλήσεων έχει να κάνει με την πραγματοποίηση των ανωτέρω στόχων σε χαμηλά επίπεδα πολυπλοκότητας εφαρμογής.

Σε όλες αυτές τις αναζητήσεις, περιοριζόμαστε από τη γνώση μας σε διάφορες μεμονωμένες αρχές, και με τον τρόπο που αυτές αλληλεπιδρούν. Οι πρόοδοι απαιτούνται στην κατανόησή μας για κωδικοποίηση, επικοινωνίες και δίκτυα, παραγωγής φωνής και ακρόασης, και ψηφιακή επεξεργασία σήματος.

Συζητώντας τις κατευθύνσεις της έρευνας και την πρόβλεψη ποιες απ' αυτές μπορεί να είναι επιτυχείς, δεν αναμένουμε απαραίτητως να είμαστε ακριβείς. Εντούτοις, μπορεί να χρειαστεί να καθοριστούν μερικές γενικές ερευνητικές κατευθύνσεις, που θα καλύπτουν τα προφανή καθώς επίσης και τα θεωρητικά σημεία της έρευνας.

6.3. ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ

6.3.1. ΚΩΔΙΚΟΠΟΙΗΣΗ, ΕΠΙΚΟΙΝΩΝΙΑ ΚΑΙ ΔΙΚΤΥΩΣΗ

Τα τελευταία χρόνια, υπήρξε σημαντική πρόοδος στους θεμελιώδεις τομείς της κωδικοποίησης πηγής: εύκαμπτες μέθοδοι ανάλυσης χρόνου-συχνότητας, προσαρμοστική διανυσματική κβάντιση και αθόρυβη κωδικοποίηση. Οι αναγκαίες εφαρμογές αυτών των τεχνικών για την κωδικοποίηση φωνής είναι σχετικά "ανώριμες". Οι συμπληρωματικές πρόοδοι στην κωδικοποίηση καναλιών και στη δικτύωση περιλαμβάνουν την κωδικοποιημένη διαμόρφωση για τα ασύρματα κανάλια και τα ενσωματωμένα πρωτόκολλα μετάδοσης για τη δικτύωση. Τα κοινά σχέδια κωδικοποίησης πηγής, κωδικοποίησης καναλιών και δικτύωσης θα είναι η βάση για την ασύρματη επικοινωνία, ειδικά στα πλαίσια των εφαρμογών πολυμέσων.

6.3.2. ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ

Στην τεχνολογία της εποχής μας, ένας επεξεργαστής σημάτων γενικού σκοπού είναι ικανός να κάνει σχεδόν 100 εκατομμύρια αριθμητικές πράξεις ανά δευτερόλεπτο, και ένα τετραγωνικό εκατοστόμετρο της μνήμης πυριτίου μπορεί να αποθηκεύσει περίπου 25 Mbit των πληροφοριών. Η μνήμη και επεξεργαστική ισχύς που είναι διαθέσιμη σε ένα chip, αναμένονται να συνεχίσουν να εξελίσσονται σημαντικά κατά τα επόμενα έτη.

Η αποδοτικότητα του επεξεργαστή όπως μετρείται σε MIPS ανά δευτερόλεπτο της κατανάλωσης ισχύος αναμένεται επίσης να βελτιωθεί κατά μια τάξη μεγέθους. Εντούτοις, για να προσαρμόσουμε τους αλγόριθμους κωδικοποίησης πολύ υψηλότερης πολυπλοκότητας σε αυτές τις συσκευές, θα χρειαστούμε συνεχείς προόδους στον τρόπο με τον οποίο ταιριάζουμε αρχιτεκτονικές επεξεργασίας με σύνθετους αλγορίθμους. Τα ζητήματα της κατανάλωσης ισχύος και της ζωής μπαταριών είναι ιδιαίτερα κρίσιμα για τις προσωπικές υπηρεσίες επικοινωνίας και τα φορητά τερματικά πληροφοριών.

ΠΑΡΑΡΤΗΜΑ

Name	standardized by	description	bit rate (kb/s)	sampling rate (kHz)	frame size (ms)	remarks
<u>(ADPCM)</u> <u>DVI</u>	Intel, <u>IMA</u>	ADPCM	32	8	sample	
<u>G.711</u>	<u>ITU-T</u>	Pulse code modulation (PCM)	64	8	sample	mu-law (US, Japan) and A-law (Europe) companding
<u>G.721</u>	<u>ITU-T</u>	Adaptive differential pulse code modulation (ADPCM)	32	8	sample	Now described in G.726; obsolete.

<u>G.722</u>	<u>ITU-T</u>	7 kHz audio-coding within 64 kbit/s	64	16	sample	Subband-codec that divides 16 kHz band into two subbands, each coded using ADPCM
<u>G.722.1</u>	<u>ITU-T</u>	Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss	24/32	16	20	See also
<u>G.723</u>	<u>ITU-T</u>	Extensions of Recommendation G.721 adaptive differential pulse code modulation to 24 and 40 kbit/s for digital circuit multiplication equipment application	24/40	8	sample	Superseded by G.726; obsolete. This is a completely different codec than G.723.1.
<u>G.723.1</u>	<u>ITU-T</u>	Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s	5.6/6.3	8	30	Part of H.324 video conferencing. DSP Group . It encodes speech or other audio signals in frames using linear predictive analysis-by-synthesis coding. The excitation signal for the high rate coder is Multipulse Maximum Likelihood Quantization (MP-MLQ) and for the low rate coder is Algebraic-Code-Excited Linear-Prediction (ACELP).
<u>G.726</u>	<u>ITU-T</u>	40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)	16/24/32/40	8	sample	ADPCM; replaces G.721 and G.723.
<u>G.727</u>	<u>ITU-T</u>	5-, 4-, 3- and 2-bit/sample embedded adaptive differential pulse code modulation (ADPCM)	var.	?	sample	ADPCM. Related to G.726.
<u>G.728</u>	<u>ITU-T</u>	Coding of speech at 16 kbit/s using low-delay code excited linear prediction	16	8		CELP. Annex J offers variable-bit rate operation for DCME.
<u>G.729</u>	<u>ITU-T</u>	Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)	8	8	10	Low delay (15 ms)
<u>GSM 06.10</u>	ETSI	RegularPulse Excitation LongTerm Predictor (RPE-LTP)	13	8	22.5	Used for GSM cellular telephony.
<u>LPC10e (FIPS 1015)</u>	US Govt.	Linear-predictive codec	2.4	8	22.5	10 coefficients.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] *Speech coding: A Computer Laboratory Textbook*
Thomas P. Barnwell III, Kambiz Nayebi, Craig H. Richardson,
Georgia Tech Digital Signal Processing Laboratory Series
- [2] *Ψηφιακή επεξεργασία σήματος*
Monson H. Hayes, Εκδόσεις Τζιόλα
- [3] *Τηλεπικοινωνιακά συστήματα*
Taub/Schilling, Εκδόσεις Τζιόλα

ΠΗΓΕΣ ΔΙΑΔΥΚΤΙΟΥ

- [04] http://en.wikipedia.org/wiki/Human_Voice
- [05] http://en.wikipedia.org/wiki/Vocal_tract
- [06] http://en.wikipedia.org/wiki/Speech_communication
- [07] http://en.wikipedia.org/wiki/Speech_production
- [08] <http://en.wikipedia.org/wiki/Phonetics>
- [09] http://en.wikipedia.org/wiki/A-law_algorithm
- [10] http://en.wikipedia.org/wiki/%CE%9C-law_algorithm
- [11] http://en.wikipedia.org/wiki/Delta_modulation
- [12] http://en.wikipedia.org/wiki/Signal-to-noise_ratio
- [13] <http://www.ind.rwth-aachen.de>
- [14] <http://mi.eng.cam.ac.uk/~ajr/SA95/node79.html>
- [15] <http://cslu.cse.ogi.edu/HLTsurvey/ch10node3.html>
- [16] <http://cslu.cse.ogi.edu/HLTsurvey/ch10node4.html>
- [17] http://scl.ece.ucsb.edu/pubs/pubs_E
- [18] <http://www.utdallas.edu>
- [19] <http://www.cs.columbia.edu>
- [20] <http://www.xilinx.com>
- [21] <http://www.mobile.ecs.soton.ac.uk>