

ΤΜΗΜΑ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ
ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΙΚΕΣ
ΣΥΜΠΙΕΣΗΣ ΦΩΝΗΣ
ΚΑΙ ΟΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥΣ

Σπουδαστές : Σμπρίνης Ανέστης – Χρήστος Στεφανόπουλος
Επιβλέπων Καθηγητής: Λιαπέρδος Ιωάννης

- 2011 -

Περιεχόμενα

Κεφάλαιο 1Εισαγωγή

- 1.2 Ιστορική αναδρομή γενικά για επικοινωνίες στη χώρα μας

Κεφάλαιο 2Ήχος

- 2.1 Ήχος και Εφαρμογές
 - 2.1.1 Ψηφιακός ήχος
 - 2.1.2 MIDI ήχος
 - 2.1.3 Εξέλιξη της κωδικοποίησης ψηφιακού ήχου
 - 2.1.4 Τα πρότυπα MPEG

Κεφάλαιο 3Κωδικοποίηση φωνής

- 3.1 Γενικά για την κωδικοποίηση φωνής
 - 3.1.2 Χρησιμότητα της κωδικοποίησης φωνής
 - 3.1.3 Πρότυπα κωδικοποίησης φωνής
- 3.2 Κατηγοριοποίηση κωδικοποιητών φωνής
 - 3.2.1 Κωδικοποίηση με γραμμική πρόβλεψη (Linear Predictive Coding - LPC)
 - 3.2.2 Ορόσημα στην κωδικοποίηση φωνής
 - 3.2.3 Αξιολόγηση κωδικοποιητών φωνής
- 3.3 Το πρότυπο LPC-10e (FS-1015)
 - 3.3.1 Γενικά
 - 3.3.2. Μοντέλο
- 3.4 Ψηφιακή επεξεργασία σήματος
 - 3.4.1 Τι είναι η επεξεργασία σήματος
 - 3.4.2 Από το αναλογικό στο ψηφιακό σήμα

Κεφάλαιο 4.....Συμπίεση

- 4.1 Τεχνικές συμπίεσης ψηφιακού ήχου**
 - 4.1.1 Mu-law και A-law PCM**
 - 4.1.2 ADPCM**
 - 4.1.3 LPC και CELP**
- 4.2 Φωνή συμπίεσης σε VoIP**
- 4.3 Δεδομένα συμπίεσης ήχου**
- 4.4 Χρήσιμοι μετασχηματισμοί για αλγόριθμους συμπίεσης**

Κεφάλαιο 5.....Επίλογος – Συμπεράσματα

- 5.1 Η συνεχής ανάγκη για συμπίεση φωνής**
- 5.2 Οι στόχοι της τεχνολογίας**
- 5.3 Μελλοντικές κατευθύνσεις**
 - 5.3.1 Κωδικοποίηση, επικοινωνία και δικτύωση**
 - 5.3.2 Ψηφιακή επεξεργασία σήματος**

ΚΕΦΑΛΑΙΟ 1

1.1 Εισαγωγή

Το βασικό ζητούμενο κατά τα πρώτα στάδια της ψηφιακής επεξεργασίας ήχου ήταν η κωδικοποίηση της ανθρώπινης φωνής. Καθώς το ήδη εκτεταμένο αναλογικό τηλεφωνικό δίκτυο (PSTN) άρχισε να ψηφιοποιείται κατά τη δεκαετία του 1970, αναζητήθηκαν τρόποι μείωσης του ρυθμού μετάδοσης της φωνής ώστε να εξοικονομηθεί εύρος ζώνης στα τηλεφωνικά κανάλια. Η πρώτη τεχνική ψηφιακής κωδικοποίησης φωνής που χρησιμοποιήθηκε στο τηλεφωνικό δίκτυο ήταν το PCM, με συχνότητα δειγματοληψίας 8 kHz και μη ομοιόμορφο κβαντιστή 256 επιπέδων (8 bits ανά δείγμα), οπότε απαιτείται ρυθμός μετάδοσης 64 kbps. Έκτοτε έγιναν προσπάθειες μείωσης αυτού του ρυθμού μετάδοσης με παράλληλη διατήρηση της ποιότητας του ήχου. Όμως, παρότι οι προσπάθειες αυτές απέδωσαν καλύτερες τεχνικές κωδικοποίησης, το PCM τελικά εδραιώθηκε στο τηλεφωνικό δίκτυο κυρίως λόγω της σημαντικής υποδομής PCM που είχε ήδη εγκατασταθεί.

Καταλυτικό ρόλο στην έρευνα σχετικά με την κωδικοποίηση φωνής έπαιξε η αύξηση των ασύρματων δικτύων και υπηρεσιών κάθε μορφής. Το τηλεπικοινωνιακό φάσμα συχνοτήτων έγινε σιγά-σιγά «σπάνιος πόρος» (rare resource) και το διαθέσιμο εύρος ζώνης για οποιαδήποτε μετάδοση περιορίστηκε. Έτσι, η ανάγκη για συμπίεση φωνής έγινε έκδηλη, ιδιαίτερα στην αναπτυσσόμενη βιομηχανία της κινητής τηλεφωνίας. Σήμερα πλέον, υπάρχει πληθώρα προτύπων και εξεζητημένων τεχνικών για την κωδικοποίηση φωνής ενώ περαιτέρω έρευνα στοχεύει στην βελτίωση της ποιότητας του ήχου για τους πολύ χαμηλούς ρυθμούς μετάδοσης.

Η συμπίεση των σημάτων ομιλίας έχει πολλές πρακτικές εφαρμογές. Ένα παράδειγμα είναι η ψηφιακή κυψελοειδής τεχνολογία, όπου πολλοί χρήστες μοιράζονται το ίδιο εύρος ζώνης συχνοτήτων. Η συμπίεση επιτρέπει σε περισσότερους χρήστες να μοιράζονται το σύστημα από ό, τι είναι συνήθως δυνατόν.

Ένα άλλο παράδειγμα είναι στην ψηφιακή αποθήκευση φωνής (π.χ. αυτόματους τηλεφωνητές). Για ένα δεδομένο μέγεθος μνήμης, η συμπίεση επιτρέπει την αποθήκευση μεγαλύτερου αριθμού μηνυμάτων.

1.2 Ιστορική αναδρομή γενικά για επικοινωνίες στη χώρα μας

Κάθε φορά που εκφράζουμε μια ιδέα ή αναμεταδίδουμε μια πληροφορία, επικοινωνούμε. Μπορεί να θέλουμε να ευχηθούμε, να στείλουμε ένα μήνυμα, να προειδοποιήσουμε για έναν κίνδυνο, να ευχαριστήσουμε, να διασκεδάσουμε ή να χρησιμοποιήσουμε κάποια άλλη από τις αναρίθμητες μορφές επικοινωνίας. Για να επικοινωνήσουμε όμως, χρειαζόμαστε μια γλώσσα ή έναν κώδικα καθώς και ένα μέσο για να μεταφερθεί το μήνυμα.

Αν βρισκόμαστε στον ίδιο χώρο με κάποιον μπορούμε να επικοινωνήσουμε με λέξεις. Συχνά όμως πρέπει να επικοινωνήσουμε με κάποιον που βρίσκεται μακριά ή θέλουμε να καταγράψουμε μόνιμα κάποιες σκέψεις ή πληροφορίες.

Αυτό επιτυγχάνεται χρησιμοποιώντας το γραπτό αντί του προφορικού λόγου το τηλέφωνο, το φαξ, το ραδιόφωνο και η τηλεόραση είναι τρόποι άμεσης αποστολής λέξεων και εικόνων μεταξύ μεγάλων αποστάσεων. Αποκαλούνται τηλεπικοινωνίες και οι εφευρέσεις αυτές επιτρέπουν στους ανθρώπους όλου του κόσμου να επικοινωνούν μεταξύ τους.

Η Ελλάδα από αρχαιοτάτων χρόνων λόγω του πολιτισμού της και της γεωγραφικής της θέσης είχε δεσπόζουσα θέση ανάμεσα στους λαούς και των τριών Ηπείρων με τις οποίες συνορεύει. Η επιτυχία αυτή οφειλόταν κατά κύριο λόγο στη συνεργασία της με τους λαούς αυτούς, χάρη στα συστήματα επικοινωνιών που επινόησε και κατά καιρούς χρησιμοποίησε σε όλες τις ιστορικές περιόδους. Αν μελετήσει κάποιος τις τεχνικές, τα μέσα και τους τρόπους επικοινωνίας κατά την αρχαιότητα, απορεί και θαυμάζει για τη φαντασία που ανέπτυξαν αρκετοί λαοί, και ειδικότερα οι Έλληνες, στο θέμα της από μακριά επικοινωνίας, δηλαδή της τηλε-επικοινωνίας. Θα εκπλαγεί μάλιστα όταν διαπιστώσει ότι αρκετές μέθοδοι επικοινωνίας που χρησιμοποιούνταν έως χθες ή χρησιμοποιούνται ακόμη και σήμερα, δεν είναι τίποτε άλλα παρά βελτιώσεις παλαιότερων μεθόδων τηλεπικοινωνίας με την προσθήκη κάποιων μηχανικών μέσων.

Οι αρχαίες τηλεπικοινωνίες αποτελούν μια πολύ γλαφυρή πτυχή της αρχαίας ελληνικής κοινωνίας, που πολλά έχει να μας διδάξει εάν μάλιστα σκεφτούμε πόσο σοβαρή και επιτακτική ήταν η ανάγκη για γρήγορη, μυστική και πλήρη ενημέρωση αντιλαμβανόμεστε την τεράστια σημασία που είχαν οι αρχαίες τηλεπικοινωνίες κάθε μορφής. Ας δούμε λοιπόν επιγραμματικά μερικές τεχνικές και τεχνάσματα επικοινωνίας που χρησιμοποιήθηκαν για να καλύψουν αυτή τη μεγάλη ανάγκη. Αγγελιοφόροι, πεζοί και έφιπποι, τα ταχυδρομικά περιστέρια, οπτικά σήματα με φωτιά, Φρυκτωρίες (Πύργοι αναμετάδοσης οπτικών σημάτων) κλπ.

ΚΕΦΑΛΑΙΟ 2

Ήχος

2.1 Ήχος και Εφαρμογές

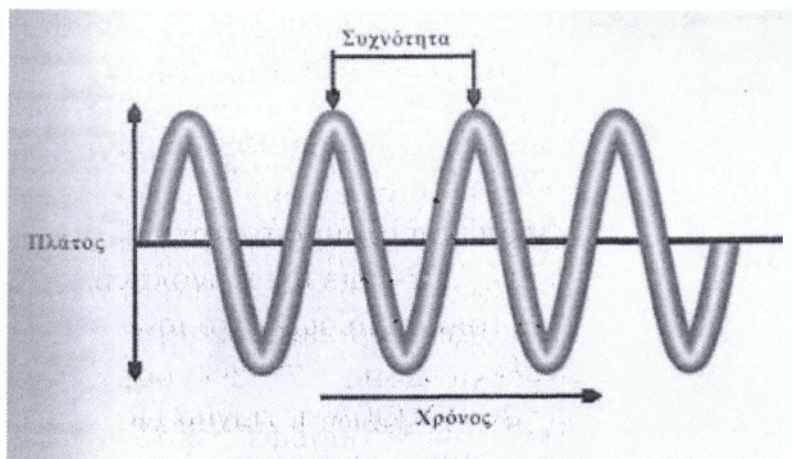
Μέχρι σήμερα, η βιομηχανία των υπολογιστών δεν έχει επιδείξει ιδιαίτερα μεγάλο ενδιαφέρον για τον ήχο στις εφαρμογές πολυμέσων. Αυτό οφείλεται κυρίως στην εστίαση των προσπαθειών στην εισαγωγή του video αλλά και στη δυσκολία καθορισμού της χρησιμότητας του ήχου σε επαγγελματικές εφαρμογές (business applications). Σαν αποτέλεσμα, ο ήχος στα πολυμέσα περιορίζεται στις ψυχαγωγικές και εκπαιδευτικές εφαρμογές.

Παρ' όλα αυτά, είναι φανερό ότι ο ήχος, κατάλληλα συνδυασμένος με τα άλλα είδη πληροφορίας, μπορεί να κάνει μια εφαρμογή πολυμέσων πιο αποτελεσματική. Ιδιαίτερα σε εκπαιδευτικές εφαρμογές και περίπτερα πληροφοριών (information kiosks) η αφήγηση και ο σχολιασμός των όσων παρουσιάζονται στην οθόνη βοηθά στην μετάδοση του μηνύματος ενώ η κατάλληλη ηχητικά υπόκρουση καθιστά την παρακολούθηση της εφαρμογής πιο ευχάριστη. Το μοναδικό χαρακτηριστικό του ήχου να γίνεται αντιληπτός χωρίς να έχουμε την προσοχή μας εστιασμένη, καθιστά τα ηχητικά σήματα αναντικατάστατα στην απόσπαση της προσοχής του χρήστη.

Υπάρχουν ορισμένες κατηγορίες εφαρμογών όπου ο ήχος αποτελεί την καρδιά του συστήματος. Εφαρμογές που έχουν ως αντικείμενο την μουσική ή ακόμα εφαρμογές που προορίζονται για ανθρώπους με προβλήματα όρασης κάνουν εκτενή και αποτελεσματική χρήση του ήχου. Με την πρόοδο της τεχνολογίας, το ενδιαφέρον για την εφαρμογή της αναγνώρισης και σύνθεσης ομιλίας σε επαγγελματικές εφαρμογές μεγαλώνει. Ήδη έχουν εμφανιστεί τα πρώτα δείγματα συστημάτων χειρισμού ενός υπολογιστή με προφορικές εντολές και υπαγόρευσης κειμένου στον υπολογιστή. Ο ήχος παράγεται από μια πηγή και μεταδίδεται με περιοδικά κύματα τα οποία χαρακτηρίζονται από το πλάτος και τη συχνότητα (πλήθος των διακυμάνσεων στη μονάδα του χρόνου). Στον υπολογιστή έχουμε δύο είδη ήχου. Τον ψηφιοποιημένο και τον ήχο MIDI.

2.1.1 Ψηφιακός ήχος

Τα ηχητικά σήματα που προέρχονται από εξωτερικές πηγές είναι αναλογικά και επομένως πρέπει να μετατραπούν σε ψηφιακά για να είναι αναγνωρίσιμα από τον υπολογιστή. Η ψηφιοποίηση γίνεται από τον μετατροπέα αναλογικού σήματος σε ψηφιακό (Analog to Digital Converter- ADC). Ο ADC παίρνει δείγματα του αναλογικού σήματος με κάποιο συγκεκριμένο ρυθμό. Για κάθε δείγμα υπολογίζεται το πλάτος του σήματος και στη συνέχεια στρογγυλοποιείται προς τη πλησιέστερη ακέραια τιμή. Είναι προφανές ότι όσο περισσότερα δείγματα παίρνουμε στη μονάδα του χρόνου, τόσο πιο υψηλή είναι η ποιότητα του ψηφιοποιημένου ήχου, καθώς προσεγγίζει κατά πολύ το αρχικό αναλογικό σήμα.

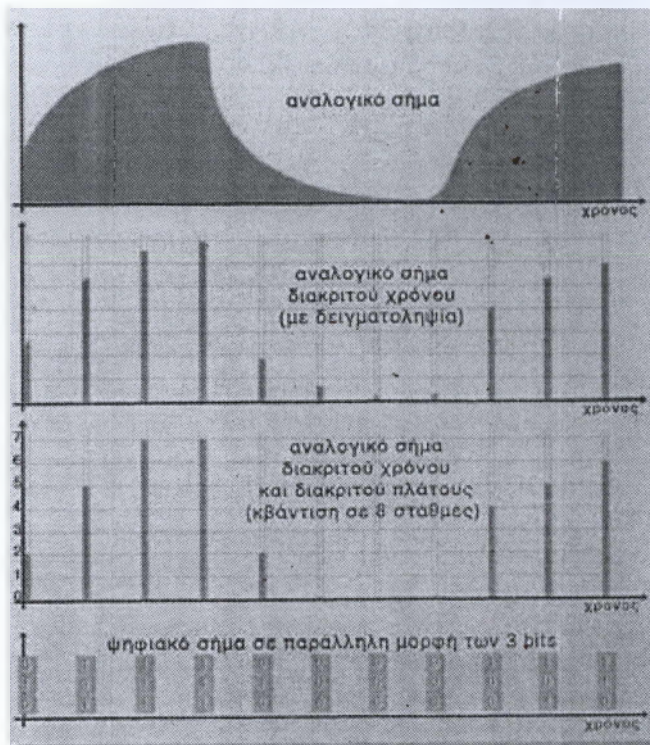


Η δειγματοληψία (sampling) καθορίζεται από δύο παραμέτρους:

- Το μέγεθος του δείγματος (sampling size) που είναι ο αριθμός των δυαδικών ψηφίων που χρησιμοποιούνται για τη καταγραφή του πλάτους του δείγματος, καθορίζοντας έτσι το πλήθος των διακριτών σταθμών τάσης στις οποίες χωρίζεται το αρχικό ηχητικό σήμα. Τα καθιερωμένα πρότυπα είναι 8, 12, 16 και 32bits.
- Ο ρυθμός δειγματοληψίας (sampling rate) που καθορίζει τον αριθμό των δειγμάτων που παίρνει η κάρτα ήχου σε 1 sec και μετριέται σεKHz. Οι καθιερωμένες συχνότητες είναι 8 KHz, 11.025

KHz(τηλεφωνική ποιότητα), 22.05 KHz (ραδιοφωνική ποιότητα) και 44.1KHz (ποιότητα μουσικού CD). Θα πρέπει να σημειωθεί ότι οι συχνότητες δειγματοληψίας δεν ορίζονται αυθαίρετα αλλά σχετίζονται άμεσα με τη συχνότητα του αναλογικού σήματος.

Σύμφωνα με το θεώρημα του Nyquist για να έχουμε πιστότητα στην αναπαραγωγή ενός αναλογικού σήματος θα πρέπει η συχνότητα δειγματοληψίας να είναι διπλάσια από την αντίστοιχη του αναλογικού σήματος. Με βάση τη πρόταση αυτή έχει καθοριστεί ο ρυθμός δειγματοληψίας για στερεοφωνική μουσική στα 44.1 KHz, συχνότητα η οποία υπερκαλύπτει το διπλάσιο του άνω ορίου των ακουστικών συχνοτήτων από τον άνθρωπο (20 KHz). Αντίστοιχα για ψηφιοποίηση ανθρώπινης φωνής (το όριο της οξύτερης ανθρώπινης φωνής είναι 10 KHz) έχουμε αρκετά ικανοποιητικό αποτέλεσμα με ρυθμό δειγματοληψίας 22.05 KHz. Για τη ψηφιακή μουσική έχει καθιερωθεί το πρότυπο του μουσικού CD, όπου ο ήχος εγγράφεται στερεοφωνικά με δείγμα 16 bit και ρυθμό δειγματοληψίας 44.1 KHz, αλλά απαιτεί διπλάσιο αποθηκευτικό χώρο. Η απαιτούμενη χωρητικότητα για την αποθήκευση ψηφιοποιημένου ήχου δίνεται από τη σχέση χωρητικότητα(bits)=K x R(Hz) x S(bits) x t(sec), όπου K ο αριθμός των καναλιών ηχογράφησης, R ο ρυθμός δειγματοληψίας, S το μέγεθος του δείγματος και t η χρονική διάρκεια. Κατά την αναπαραγωγή του ψηφιοποιημένου ήχου είναι απαραίτητη η μετατροπή των ψηφιακών δεδομένων σε αναλογικό σήμα, ώστε να είναι δυνατή η απόδοση τους από τις ηχητικές εξόδους του συστήματος (π.χ. τα ηχεία). Η κάρτα ήχου διεκπεραιώνει τη διαδικασία αυτή μέσω του μετατροπέα ψηφιακού σήματος σε αναλογικό (Digital to Analog Converter-DAC). Ο DAC δέχεται τα αριθμητικά δεδομένα που αντιστοιχούν στον ψηφιοποιημένο ήχο, με συχνότητα ίση με το ρυθμό δειγματοληψίας του ADC και παράγει στη έξοδο του την 11 αντίστοιχη τάση που δημιουργεί το αναλογικό ηχητικό σήμα. Τα αρχεία ήχου κατατάσσονται σε δύο κατηγορίες με βάση τη διαδικασία παραγωγής τους: α)τα αρχεία ψηφιακού ήχου (digit audio files) που περιλαμβάνουν φωνή, μουσική ή ήχους από εξωτερικές πηγές με κατάληξη .wav, και β)τα αρχεία midi που παράγονται από μουσικά όργανα μέσω κατάλληλου εξοπλισμού MIDI με κατάληξη mid.



2.1.2 MIDI ΉΧΟΣ

Το πρότυπο αυτό αναπτύχθηκε στη αρχή της δεκαετίας του 80 και περιέχει πρότυπα για την επικοινωνία μουσικών οργάνων με υπολογιστή. Το MIDI καθορίζει πως κωδικοποιούνται τα διάφορα στοιχεία μιας μουσικής παρτιτούρας καθώς και τα όργανα που συμμετέχουν.

Υπάρχει η δυνατότητα χρησιμοποίησης 127 οργάνων και ηχητικών εφέ. Ένας υπολογιστής με MIDI interface μπορεί να χειριστεί συσκευές που ακολουθούν αυτό το πρότυπο όπως ηλεκτρονικά synthesizers. Στα αρχεία MIDI δεν αποθηκεύεται ψηφιοποιημένο σήμα αλλά μια σειρά ειδικών πληροφοριών που απαιτούνται για την αναπαραγωγή του ήχου. Συγκεκριμένα καταγράφονται τέσσερις αριθμοί που περιγράφουν ένα μουσικό γεγονός όπως το πάτημα ενός πλήκτρου στο πιάνο. Ο πρώτος καταγράφει τη νότα, ο δεύτερος το όργανο που τη παρήγαγε, ο τρίτος την ένταση με την οποία πατήθηκε η νότα και ο τέταρτος το χρόνο που έμεινε πατημένο το πλήκτρο.

Επειδή τα αρχεία MIDI περιέχουν κώδικα και όχι ψηφιακές πληροφορίες κυματομορφών έχουν σημαντικά μικρότερο μέγεθος σε σχέση με αρχεία ψηφιοποιημένου ήχου (WAV). Ένα λεπτό μουσικής σε αρχείο WAV χρειάζεται όπως είδαμε 10,1 MB, ενώ το αντίστοιχο σε MID περίπου 6KB. Ο ήχος MIDI μπορεί να χωριστεί σε 16 διαφορετικά κανάλια . Κάθε κανάλι αντιστοιχεί με ένα όργανο και με κατάλληλο συνδυασμό όλων των καναλιών μπορούμε να φτιάξουμε μια ολοκληρωμένη σύνθεση.

2.1.3 Εξέλιξη της κωδικοποίησης ψηφιακού ήχου

Η κωδικοποίηση ψηφιακού ήχου παρουσιάστηκε ως ιδέα στις αρχές του 20^{ου} αιώνα. Η πρώτη τεχνική κωδικοποίησης ήταν η απλή δειγματοληψία και κβαντισμός ενός αναλογικού σήματος, που ονομάστηκε παλμοκωδική διαμόρφωση (Pulse Code Modulation - PCM). Ωστόσο, μέχρι το 1950 οι εφαρμογές ψηφιακού ήχου ήταν σχεδόν ανύπαρκτες καθώς η απαραίτητη ηλεκτρονική τεχνολογία βρισκόταν ακόμα σε βρεφικό στάδιο. Τις δεκαετίες του 1950 και 1960 έχουμε την εμφάνιση μερικών πιο εξεζητημένων τεχνικών (διαμόρφωση δέλτα, DPCM) καθώς και περιορισμένες υλοποιήσεις ψηφιακής επεξεργασίας ήχου στους πρώτους υπολογιστές. Οι σημαντικότερες όμως εξελίξεις σημειώθηκαν από το 1970 και έπειτα, υποβοηθούμενες τόσο από την τεχνολογική έκρηξη των ψηφιακών συστημάτων (συμπεριλαμβανομένων των ηλεκτρονικών υπολογιστών) όσο και από τις προόδους στον ευρύτερο τομέα της ψηφιακής επεξεργασίας σήματος (ανακάλυψη ψηφιακών φίλτρων, FFT κλπ).

Έκτοτε το θεωρητικό υπόβαθρο της ψηφιακής επεξεργασίας ήχου έχει τελειοποιηθεί, ενώ πληθώρα μεθόδων κωδικοποίησης έχουν βρει εφαρμογή στους τομείς των ενσύρματων και ασύρματων επικοινωνιών αλλά και στη μουσική βιομηχανία. Ένας από τους κυριότερους στόχους της έρευνας στην περιοχή της ψηφιακής επεξεργασίας ήχου είναι η συμπίεση ενός σήματος ήχου, χωρίς να μεταβληθεί αισθητά η ποιότητα. Για τον σκοπό αυτό αναπτύχθηκαν πολλές τεχνικές, οι οποίες χρησιμοποιούν διάφορες προσεγγίσεις κωδικοποίησης. Οι πιο επιτυχημένες μέθοδοι συμπίεσης έχουν προτυποποιηθεί από διεθνείς οργανισμούς και τυγχάνουν ευρείας αποδοχής παγκοσμίως, ενώ η έρευνα για τη βελτίωση των

υπαρχόντων και την εισαγωγή νέων προτύπων συνεχίζεται μέχρι σήμερα.

Γενικά, τα διάφορα πρότυπα ομαδοποιούνται σε δύο μεγάλες κατηγορίες:

1. Πρότυπα που αποσκοπούν στην κωδικοποίηση σημάτων φωνής

Τα πρότυπα αυτά κωδικοποιούν ένα σήμα με την παραδοχή ότι πρόκειται για σήμα φωνής. Είναι γνωστό ότι ένα σήμα φωνής έχει συγκεκριμένη μορφή και ιδιότητες. Εκμεταλλευόμενοι αυτές τις ιδιότητες της φωνής έχουν αναπτυχθεί αλγόριθμοι κωδικοποίησης που επιτυγχάνουν μεγάλη συμπίεση. Οι πιο γνωστοί και ευρέως αποδεκτοί φορείς προτυποποίησης στην κωδικοποίηση φωνής είναι η ITU (International Telecommunication Union) και ο ευρωπαϊκός φορέας ETSI (European Telecommunications Standards Institute).

2. Πρότυπα που αποσκοπούν στην κωδικοποίηση οποιασδήποτε κυματομορφής ήχου

Τα πρότυπα αυτά είναι γενικότερα των προηγούμενων και βρίσκουν εφαρμογή κυρίως στην κωδικοποίηση σημάτων μουσικής. Μία μεγάλη κατηγορία τέτοιων τεχνικών κωδικοποίησης εντάσσονται στο είδος της αντιληπτικής κωδικοποίησης (perceptual coding). Κατά την αντιληπτική κωδικοποίηση λαμβάνονται υπόψιν τα χαρακτηριστικά της ανθρώπινης ακοής, ώστε τα σφάλματα κβαντισμού που προκύπτουν από τον μεγάλο βαθμό συμπίεσης να μην γίνονται αντιληπτά από το ανθρώπινο ακουστικό αισθητήριο.

Το πιο αντιπροσωπευτικό πρότυπο αντιληπτικής κωδικοποίησης είναι το MPEG-1 Audio Layer 3 ή MP3. Κύριος φορέας προτυποποίησης αυτών των προτύπων είναι ο ISO και η IEC.

2.1.4 Τα πρότυπα MPEG

Η ομάδα Moving Pictures Experts Group (MPEG) του ISO/IEC αναπτύσσει πρότυπα για video και ήχο από το 1988. Το πρότυπο που προέκυψε από την πρώτη φάση είναι το MPEG-1 ενώ η έρευνα συνεχίζεται στο MPEG-2. Για κάποιο χρονικό διάστημα υπήρχε σε εξέλιξη το MPEG-3 αλλά εγκαταλείφθηκε και ένα τμήμα του ενσωματώθηκε στο MPEG-2.

Κάθε ένα από αυτά στοχεύει σε εφαρμογές με διαφορετικές απαιτήσεις ποιότητας και εύρους ζώνης και εκτός από τεχνικές συμπίεσης κινούμενης εικόνας περιέχει και ένα κομμάτι που αναφέρεται στην συμπίεση του ήχου. Οι τεχνικές συμπίεσης ήχου δεν περιορίζονται στη φωνή αλλά είναι σχεδιασμένες για τη γενικότερη περίπτωση της μουσικής.

Με δυο λόγια θα μπορούσαμε να περιγράψουμε το MPEG-1 ως εξής:

Το ηχητικό κομμάτι του MPEG-1 δεν αποτελεί έναν αλγόριθμο συμπίεσης, αλλά μια οικογένεια τριών διαφορετικών τεχνικών κωδικοποίησης και συμπίεσης ήχου. Αυτές οι οικογένειες ονομάζονται MPEG-Audio Layer-1, Layer-2, Layer-3.

Και οι τρεις αυτές τεχνικές στηρίζονται στην ίδια αρχή: η συμπίεση γίνεται συνδυάζοντας ένα είδος κωδικοποίησης μετασχηματισμού και sub-band division. Οι διαφορές αυτών των τριών στρωμάτων εντοπίζονται στο τελικό στάδιο της κβαντοποίησης.

Τα βασικά βήματα που ακολουθούνται είναι:

1. Χωρισμός του φάσματος σε 32 τμήματα (sub-bands)
2. Εφαρμόζεται στο σήμα ένας ταχύς μετασχηματισμός Fourier (Fast Fourier Transform)
3. Ένα ψύχο-ακουστικό μοντέλο εφαρμόζεται στο μετασχηματισμένο σήμα για να υπολογιστεί το ελάχιστο επίπεδο θορύβου που γίνεται αντιληπτό από το μέσο ακροατή.

Το MPEG-1 προβλέπει δύο ηχητικά κανάλια. Αυτά μπορούν να απλά (μονοφωνικά), διπλά (δύο μονοφωνικά κανάλια), απλά στερεοφωνικά (ένα κανάλι μεταφέρει το αριστερό ηχητικό σήμα και το άλλο το δεξί) ή από κοινού στερεοφωνικά (joint stereo, το ένα κανάλι μεταφέρει το άθροισμα και το άλλο τη διαφορά των σημάτων). Το πρότυπο χρησιμοποιεί 16bits για την κωδικοποίηση των δειγμάτων ενώ η συχνότητα δειγματοληψίας είναι 44.1kHz, 48kHz ή 32kHz.

Οι επιδόσεις κάθε στρώματος είναι:

- MPEG-Audio Layer-1: επιτρέπει την κατασκευή πολύ απλών κωδικοποιητών και αποκωδικοποιητών θυσιάζοντας βέβαια ένα μέρος της ποιότητας. Ως αποτέλεσμα, η ηχητική ποιότητα είναι μέτρια ενώ το bandwidth που απαιτείται είναι αυξημένο: 192 ή 256Kbps ανά κανάλι.
- MPEG-Audio Layer-2: ο αλγόριθμος αυτή της κατηγορίας έχει βελτιστοποιηθεί για ένα εύρος ζώνης 96 ή 128Kbps ανά μονοφωνικό κανάλι. Η ποιότητα είναι εφάμιλλη του CD.
- MPEG-Audio Layer-3: έχει την καλύτερη επίδοση από τα τρία στρώματα. Η ποιότητα του είναι υποδεέστερη αλλά πολύ κοντά σε αυτή του CD. Το βασικό του πλεονέκτημα είναι ότι απαιτεί μόνο 64Kbps. Σε σύγκριση με MPEG-Audio Layer-2 στα 64Kbps λειτουργεί πολύ καλύτερα.

Να σημειωθεί ότι αυτά τα στρώμα έχουν προς τα πίσω συμβατότητα, δηλαδή μια συσκευή MPEG-Audio Layer-3 αποκωδικοποιεί και δεδομένα κωδικοποιημένα κατά MPEG-Audio Layer-1 ή 2.

ΚΕΦΑΛΑΙΟ 3

ΚΩΔΙΚΟΠΟΙΗΣΗ ΦΩΝΗΣ

3.1 Γενικά για την κωδικοποίηση φωνής

Η φωνή είναι ένα αναλογικό σήμα και για τη μετάδοσή της μέσα από ένα ψηφιακό σύστημα, όπως ένα IP δίκτυο, είναι απαραίτητη η μετατροπή της σε ψηφιακή μορφή. Η μετατροπή αυτή λαμβάνει χώρα στους codecs (coder/decoder). Η πιο διαδεδομένη τεχνική μετατροπής αναλογικού σήματος σε ψηφιακή μορφή είναι η παλμοκωδική διαμόρφωση (Pulse Code Modulation - PCM).

Η PCM μετατρέπει τη φωνή σε ψηφιακή μορφή, με δειγματοληψία 8000 δειγμάτων το δευτερόλεπτο (125 msec μεταξύ δύο διαδοχικών δειγμάτων) και μετατροπή κάθε δείγματος σε δυαδικό κώδικα, η τιμή του οποίου αναπαριστά την τιμή του αναλογικού σήματος τη στιγμή της δειγματοληψίας. Στην περίπτωση της standard PCM, χρησιμοποιείται δυαδικός κώδικας των 8 bits και έτσι απαιτείται κανάλι χωρητικότητας 64kbps. Τα δείγματα αυτά χρησιμοποιούνται για την αναπαραγωγή της φωνής στο δέκτη, με ανασύνθεση κατά προσέγγιση της αρχικής αναλογικής κυματομορφής.

Για τη μείωση του απαιτούμενου bandwidth που απαιτείται για τη μετατροπή της φωνής σε ψηφιακή μορφή από τα 64kbps που απαιτεί η standard PCM, έχουν αναπτυχθεί από την ITU μια σειρά από πρότυπα για τη συμπίεση φωνής. Τα πρότυπα αυτά μπορεί να θεωρηθεί ότι ανήκουν σε δύο οικογένειες. Η πρώτη οικογένεια αποτελείται από πρότυπα τα οποία στηρίζονται στην PCM και χρησιμοποιούν ρυθμό δειγματοληψίας 8000 δειγμάτων ανά δευτερόλεπτο. Σκοπός των τεχνικών αυτών είναι η μείωση της περίσσειας της πληροφορίας μέσω της αποσυσχέτισης των δειγμάτων. Για τη μείωση του απαιτούμενου bandwidth, χρησιμοποιούνται για την κωδικοποίηση των δειγμάτων κώδικες των 4 ή 2 bits και η τιμή του δείγματος στην περίπτωση αυτή αντιπροσωπεύει τη διαφορά του δείγματος από το αμέσως προηγούμενο. Η μείωση του μήκους των χρησιμοποιούμενων κωδικών έχει σαν αποτέλεσμα τη μείωση του bandwidth στα 32 ή στα 16 kbps (Adaptive Differential PCM - ADPCM), η οποία όμως συνοδεύεται από μία ανάλογη μείωση στην ποιότητα της φωνής.

Η δεύτερη οικογένεια προτύπων επιτυγχάνει ακόμα μεγαλύτερη μείωση του απαιτούμενου bandwidth, ενώ ταυτόχρονα διατηρεί ικανοποιητική την ποιότητα της φωνής. Τα πρότυπα αυτά χρησιμοποιούν για την κωδικοποίηση της φωνής έναν ειδικό αλγόριθμο, τον LPC (Linear Predictive Code), ο οποίος πραγματοποιεί μια προσομοίωση - μοντελοποίηση της ανθρώπινης ομιλίας με αποτέλεσμα την αυξημένη απόδοση και τη μείωση του απαιτούμενου bandwidth .

Για την αξιολόγηση των διαφόρων αλγορίθμων συμπίεσης ως προς την ποιότητα της φωνής που αναπαράγουν, χρησιμοποιείται το *Mean Opinion Score (MOS)* με κλίμακα από 0 έως 5. MOS μεταξύ 4 και 5 θεωρείται απόλυτη ποιότητα, μεταξύ 3 και 4 θεωρείται τηλεπικοινωνιακή ποιότητα, ενώ MOS λιγότερο του 3 θεωρείται συνθετική ποιότητα. Στον παρακάτω πίνακα παρουσιάζονται οι διάφοροι αλγόριθμοι συμπίεσης και το αντίστοιχο MOS.

Μέθοδος Συμπίεσης	MOS score	Καθυστέρηση αλγορίθμου (msec)
64kbps PCM (G.711)	4.4	0.75
32kbps ADPCM (G.726)	4.2	1
16kbps LD-CELP (G.728)	4.2	3.6
8kbps CS-ACELP (G.729)	4.2	10
8kbps CS-ACELP (G.729a)	4.2	10
6.3kbps MPMLG (G.723.1)	3.98	30
5.3kbps ACELP (G.723.1)	3.5	30

3.1.2 Χρησιμότητα της κωδικοποίησης φωνής

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. _εδομένου ότι η πολυπλοκότητα αυξάνεται, η καθυστέρηση φωνής αυξάνεται εξαιτίας του γεγονότος ότι παίρνει περισσότερο για να κωδικοποιηθεί και να αποκωδικοποιηθεί ένα τμήμα της ομιλίας. Οι καθυστερήσεις που είναι μεγαλύτερες από το 100msec μπορούν να γίνουν ενοχλητικές στο χρήστη.

3.1.3 Πρότυπα κωδικοποίησης φωνής

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 φωνητικών και μη σημάτων. Πολλά από τα πρώτα πρότυπα κωδικοποίησης φωνής δημιουργήθηκαν από το υπουργείο άμυνας των ΗΠΑ (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 κωδικοποιητής γραμμικής πρόβλεψης ο οποίος λειτουργεί στα 2.4 kbps. Το δεύτερο και πιο πρόσφατο DoD πρότυπο είναι μιας μορφής code-excited κωδικοποιητής γραμμικής πρόβλεψης (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.

3.2 Κατηγοριοποίηση κωδικοποιητών φωνής

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)

a) **Waveform coders**: 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 διανεύουμε περισσότερα bit στις σημαντικές υποζώνες έτσι ώστε ο θόρυβος σε αυτές τις περιοχές συχνότητων 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 παλμοδική διαμόρφωση (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 συχνοτήτων 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 παλμοδικής διαμόρφωσης (DPCM) και διαμόρφωσης δέλτα (Delta Modulation - DM) κάνουν απευθείας χρήση ενός μακροχρόνιου σταθερού στατιστικού μοντέλου για την παραγωγή φωνής. Η προσαρμοστική διαφορική παλμοδική διαμόρφωση (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): 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 εμπεριέχει 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,

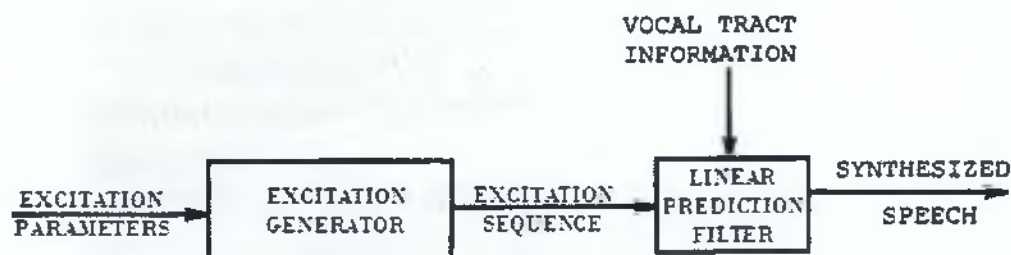
οι παράμετροι ενός φίλτρου σύνθεσης LPC κβαντίζονται και μεταδίδονται ως δευτερεύουσες πληροφορίες (κανάλι παραμέτρων). Επιπλέον, τα περιοδικά (φωνητικά) μέρη του λεκτικού σήματος μοντελοποιούνται από ένα δεύτερο φίλτρο, το οποίο καλείται μακροπρόθεσμο φίλτρο πρόβλεψης (Long Term Predictive - LTP). Η 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

3.2.1 Κωδικοποίηση με γραμμική πρόβλεψη (Linear Predictive Coding - LPC)

Στην κωδικοποίηση με γραμμική πρόβλεψη, η φωνητική οδός μοντελοποιείται με ένα γραμμικό ψηφιακό φίλτρο, το οποίο διεγείρεται από μία κυματομορφή που προσομοιώνει τη λειτουργία των φωνητικών χορδών.



γενική μορφή μοντέλου γραμμικής πρόβλεψης LPC

Σε έναν κωδικοποιητή LPC, τα δείγματα της φωνής χωρίζονται σε ομάδες στο πεδίο του χρόνου, που ονομάζονται πλαίσια (frames). Κάθε frame περιέχει έναν αριθμό γειτονικών δειγμάτων. Στη συνέχεια το κάθε frame αναλύεται ώστε να εξαχθούν οι παράμετροι της κυματομορφής διέγερσης και του ψηφιακού

φίλτρου γραμμικής πρόβλεψης. Αυτές οι παράμετροι, χαρακτηριστικές του frame, υφίστανται κβαντισμό και μεταδίδονται στο δέκτη.

Το γραμμικό ψηφιακό φίλτρο που προσομοιώνει τη φωνητική οδό έχει συνήθως τη μορφή:

$$y[n] = \sum_{i=1}^M a_i y[n-i] + Gx[n] \quad , \quad \text{όπου} \quad x[n]: \text{είσοδος φίλτρου}$$

$y[n]$: έξοδος φίλτρου

M : τάξη προβλέπτη

G : κέρδος φίλτρου

Δηλαδή το κάθε δείγμα ήχου προσεγγίζεται από έναν γραμμικό συνδυασμό των προηγούμενων δειγμάτων, ενώ ταυτόχρονα λαμβάνεται υπόψιν και η συνεισφορά της τρέχουσας διέγερσης της φωνητικής οδού (είσοδος $x[n]$). Οι συντελεστές πρόβλεψης αι υπολογίζονται έτσι ώστε να ελαχιστοποιείται το μέσο τετραγωνικό σφάλμα μεταξύ της αυθεντικής και της συνθετικής κυματομορφής φωνής. Τέλος, η μορφή και οι παράμετροι της κυματομορφής διέγερσης παρουσιάζουν μεγάλη ποικιλία και στην ουσία είναι αυτές που διαφοροποιούν τα πολυάριθμα πρότυπα μεταξύ τους.

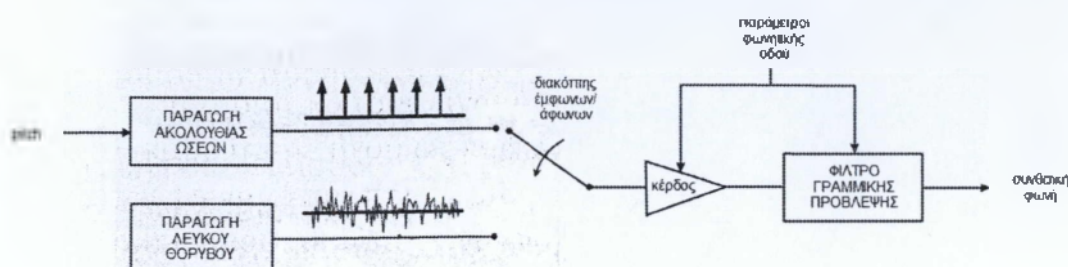
Ο αποκωδικοποιητής LPC λαμβάνει τις κβαντισμένες παραμέτρους του μοντέλου και, μετά τον αποκβαντισμό, τις χρησιμοποιεί για να παράγει συνθετικό ήχο σύμφωνα με το σχήμα

Διακρίνονται οι εξής δύο κατηγορίες προτύπων κωδικοποίησης LPC:

1. *Analysis/synthesis (Open-loop vocoders)*

Οι vocoders ανοιχτού βρόχου είναι οι πρώτοι και πιο απλοί κωδικοποιητές LPC φωνής που κατασκευάστηκαν. Κύριο χαρακτηριστικό τους είναι ότι προσπαθούν να χαρακτηρίσουν το κάθε frame ήχου ως έμφωνο ή άφωνο, ανάλογα με το είδος των φωνημάτων που περιέχονται σε αυτό. Σε περίπτωση που το frame χαρακτηριστεί έμφωνο, υπολογίζεται επίσης μια εκτίμηση της περιόδου pitch. Ο χαρακτηρισμός του frame και πιθανόν η τιμή του pitch αποστέλλονται στο δέκτη μαζί με τις παραμέτρους του φίλτρου πρόβλεψης. Εκεί η κυματομορφή διέγερσης παράγεται με τη βοήθεια ενός διακόπτη, ο οποίος επιλέγει μεταξύ μιας περιοδικής ακολουθίας ώσεων με περίοδο

pitch (για έμφωνο frame) και μιας κυματομορφής λευκού θορύβου (για άφωνο frame), κατά μίμηση της λειτουργίας του ανθρώπινου φωνητικού συστήματος.



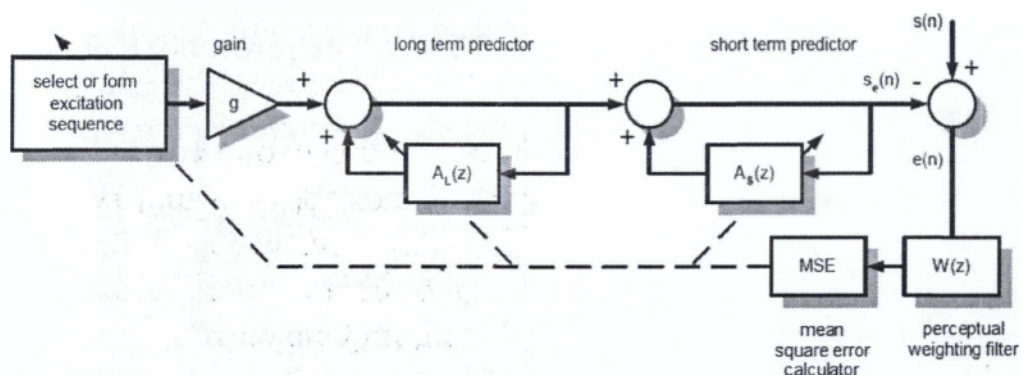
αποκωδικοποιητής ενός open-loop vocoder

2. Analysis by Synthesis (Closed-loop coders ή hybrid coders)

Η ποιότητα του παραγόμενου ήχου από έναν vocoder ανοιχτού βρόχου είναι γενικά χαμηλή, διότι το μοντέλο του φωνητικού συστήματος που χρησιμοποιείται δεν ανταποκρίνεται επακριβώς στην πραγματικότητα. Ειδικότερα, η προσέγγιση της λειτουργίας των φωνητικών χορδών με μία ακολουθία ώσεων που ισαπέχουν μεταξύ τους και έχουν ίδιο πλάτος προκαλεί ένα αφύσικο βουητό σε μερικά έμφωνα τμήματα του συνθετικού ήχου. Ακόμα σε περιβάλλοντα θορύβου η αναπαραγόμενη φωνή μπορεί να αλλιωθεί τόσο ώστε να είναι δύσκολα κατανοητή. Για να αντιμετωπιστούν αυτά τα προβλήματα αναπτύχθηκαν οι κωδικοποιητές κλειστού βρόχου, οι οποίοι επιτυγχάνουν σημαντικά καλύτερη ποιότητα ήχου και ευρωστία στο θόρυβο. Αυτοί κωδικοποιούν εκτός από τις παραμέτρους της φωνητικής οδού και τα χαρακτηριστικά της κυματομορφής διέγερσης, επιλέγοντας από ένα σύνολο τύπων κυματομορφών. Οι κυματομορφές αυτές μπορεί είτε να είναι εκ των προτέρων γνωστές σε πομπό και δέκτη είτε να σχηματίζονται δυναμικά κατά την κωδικοποίηση. Η κατάλληλη κυματομορφή διέγερσης επιλέγεται στον κωδικοποιητή έτσι ώστε να ελαχιστοποιείται το σφάλμα μεταξύ της αυθεντικής και της συνθετικής κυματομορφής ήχου. Για να γίνει αυτό, ο κωδικοποιητής προσομοιώνει τη λειτουργία του αποκωδικοποιητή, δηλαδή, για κάθε ομάδα παραμέτρων που εξετάζεται, συντίθεται το αποκωδικοποιημένο σήμα ήχου και συγκρίνεται με το αυθεντικό. Για αυτό το λόγο οι τεχνικές αυτές ονομάζονται και analysis-by-synthesis.

Κατά τον υπολογισμό του σφάλματος λαμβάνονται υπόψιν και τα χαρακτηριστικά της ακοής (αντιληπτική κωδικοποίηση), διότι το ανθρώπινο αυτί δεν αντιλαμβάνεται ως διαφορετικούς δύο ήχους που είναι αρκετά όμοιοι μεταξύ τους.

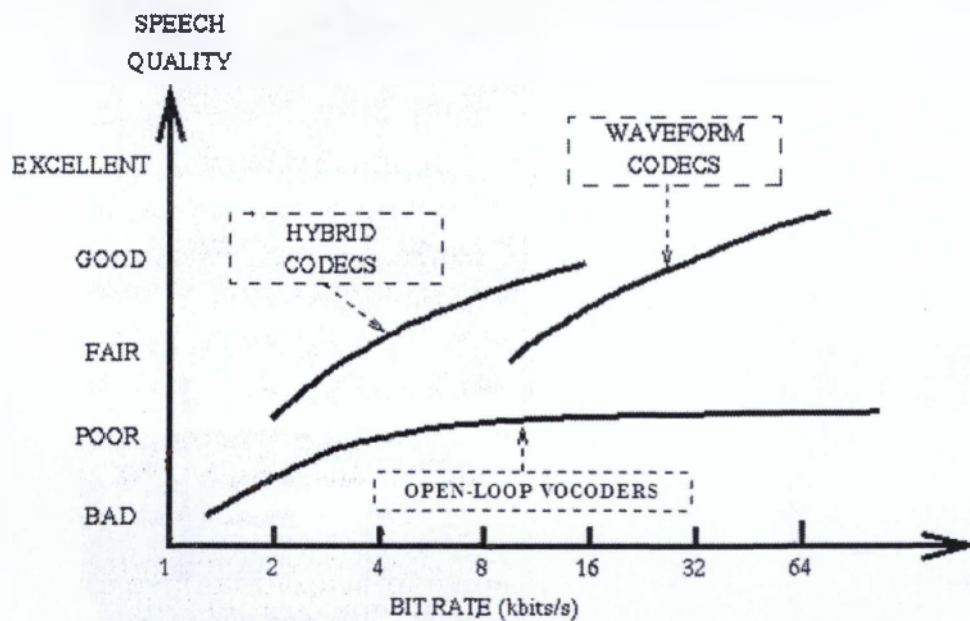
Το σχηματικό διάγραμμα ενός κωδικοποιητή κλειστού βρόχου φαίνεται στο σχήμα. Το γραμμικό φίλτρο πρόβλεψης χωρίζεται σε έναν μακροπρόθεσμο προβλέπτη (long term predictor), ο οποίος μοντελοποιεί την περιοδικότητα της φωνής που οφείλεται στο pitch, και έναν βραχυπρόθεσμο προβλέπτη (short term predictor), που μοντελοποιεί την φωνητική οδό όπως και στους κωδικοποιητές ανοιχτού βρόχου. Το φίλτρο $W(z)$ στο σχήμα καταστέλλει τις συστατικές του σφάλματος που δεν γίνονται αντιληπτές από το ανθρώπινο ακουστικό αισθητήριο.



διάγραμμα κωδικοποιητή κλειστού βρόχου

Οι κωδικοποιητές κλειστού βρόχου θεωρούνται από μερικούς ως ένα ενδιάμεσο είδος μεταξύ των κωδικοποιητών κυματομορφής και των κωδικοποιητών πηγής. Από τη μία πλευρά χρησιμοποιούν το μοντέλο του ανθρώπινου φωνητικού συστήματος αλλά ταυτόχρονα εξετάζουν την αρχική και την παραγόμενη κυματομορφή επιχειρώντας να ελαχιστοποιήσουν το σφάλμα που προκύπτει, όπως ακριβώς συμβαίνει με τους κωδικοποιητές κυματομορφής (πχ ADPCM). Για αυτό ονομάζονται και υβριδικοί κωδικοποιητές (hybrid coders). Πάντως, η βασική φιλοσοφία πίσω από τη σχεδίασή τους όσο και η προέλευσή τους, τους εντάσσει αναμφισβήτητα στην οικογένεια των vocoders.

Στο σχήμα συγκρίνεται η ποιότητα του αναπαραγόμενου ήχου και ο ρυθμός μετάδοσης μεταξύ των τριών κατηγοριών κωδικοποιητών φωνής (waveform coders, open-loop vocoders, closed-loop coders).



Ποιότητα ήχου και bit rate των διαφόρων κατηγοριών κωδικοποιητών φωνής.

3.2.2 Ορόσημα στην κωδικοποίηση φωνής

Στον παρακάτω πίνακα παρουσιάζεται μια ιστορική αναδρομή της κωδικοποίησης φωνής από τις απαρχές της μέχρι και σήμερα.

Έτος	Γεγονός
1804	Ο Wolfgang von Kempelen κατασκευάζει τον πρώτο μηχανικό vocoder. Επρόκειτο για μια μηχανή, κατασκευασμένη από εξαρτήματα όπως ένα φουσερό, ένα γλωσσίδι μουσικού οργάνου, μοχλούς και σωλήνες από σόμπα. Η μηχανή αυτή λέγεται ότι μπορούσε να αρθρώσει ολόκληρες φράσεις στην αγγλική, γαλλική, ιταλική και γερμανική γλώσσα.
1926	Ο Paul M. Rainey συλλαμβάνει για πρώτη φορά την ιδέα της ψηφιοποίησης ενός αναλογικού σήματος (PCM) και κατοχυρώνει την εφεύρεσή του με πατέντα.
1939	Ο Homer Dudley επιδεικνύει στη Νέα Υόρκη τον VODER (Voice Operating Demonstrator), τον πρώτο ηλεκτρικό αναλογικό vocoder που κατασκευάστηκε στα εργαστήρια της Bell. Επρόκειτο για ένα αρμόνιο συνδεδεμένο με μία τράπεζα ήχων. Ο χειριστής πιέζοντας πλήκτρα καθόριζε τη διαμόρφωση της φωνητικής οδού, ενώ με μία χειρολαβή επέλεγε μεταξύ έμφωνων και άφωνων φωνημάτων.
1943	Το SIGSALY είναι ο πρώτος ψηφιακός vocoder και το πρώτο σύστημα που χρησιμοποίησε PCM. Κατασκευάζεται στην Αμερική κατά τη διάρκεια του Β' Παγκοσμίου Πολέμου και χρησιμοποιείται για την επικοινωνία των ηγετών των συμμάχων σε ανώτατο επίπεδο. Το σύστημα συνδύαζε ψηφιοποίηση με αρχές κρυπτογραφίας και κωδικοποίησης φωνής.

- 1952
Εισάγεται η τροποποίηση δέλτα (delta modulation), η οποία είναι το πρώτο είδος κωδικοποίησης DPCM (Differential Pulse Code Modulation).
- 1957
Προτείνεται το μ -law PCM, το οποίο χρησιμοποιεί μη ομοιόμορφο κβαντισμό.
- 1962
Εισαγωγή του PCM στο τηλεφωνικό δίκτυο PSTN των ΗΠΑ, το PSTN αρχίζει να γίνεται ψηφιακό.
- 1972
Το πρότυπο G.711 της ITU προτυποποιεί την κωδικοποίηση PCM στα 64 kbps και ορίζει τις παραλλαγές μ -law και A-law PCM.
- 1974
Αναπτύσσεται το ADPCM (Adaptive Differential Pulse Code Modulation) στα εργαστήρια της Bell, μια παραλλαγή του DPCM που χρησιμοποιεί προσαρμοζόμενο κβαντιστή.
- 1982
Οι B. Atal και J. Remde εισάγουν τον πρώτο κωδικοποιητή κλειστού βρόχου analysis-by-synthesis. Η κωδικοποίηση MPE (Multi-Pulse Excitation) χρησιμοποιεί ως κυματομορφή διέγερσης μια απεριοδική ακολουθία ώσεων με μεταβλητά πλάτη, που δημιουργείται δυναμικά για κάθε frame ήχου.
- 1984
Το πρότυπο G.721 της ITU προτυποποιεί την κωδικοποίηση ADPCM στα 32 kbps.
Το αμερικανικό υπουργείο άμυνας εκδίδει το πρότυπο FS-1015, γνωστό και ως LPC-10, το οποίο χρησιμοποιεί γραμμική πρόβλεψη και επιτυγχάνει ρυθμό μετάδοσης 2.4 kbps.

1985

Οι M. Schroeder και B. Atal εισάγουν τον κωδικοποιητή κλειστού βρόχου CELP (Code Excited Linear Prediction), ο οποίος, μαζί με τις πολλές παραλλαγές του, είναι ο δημοφιλέστερος αντιπρόσωπος των τεχνικών analysis-by-synthesis. Κύριο χαρακτηριστικό του είναι ότι η κυματομορφή διέγερσης επιλέγεται από ένα σύνολο εκ των προτέρων γνωστών κυματομορφών.

1986

Οι P. Kroon και R. Sluyter αναπτύσσουν την κωδικοποίηση κλειστού βρόχου RPE (Regular Pulse Excitation), όπου η κυματομορφή διέγερσης είναι μια περιοδική ακολουθία ώσεων με μεταβλητά πλάτη.

1987

Εισάγεται η κωδικοποίηση ACELP (Algebraic CELP).

1990

Εισάγεται η κωδικοποίηση VSELP (Vector Sum Excited Linear Prediction), παραλλαγή της CELP, από τους I. Gerson και M. Jasiuk.

1991

Το αμερικανικό υπουργείο άμυνας εκδίδει το πρότυπο FS-1016, το οποίο χρησιμοποιεί κωδικοποίηση CELP και λειτουργεί στα 4.8 kbps.

1992

Η ITU εκδίδει το πρότυπο G.728, το οποίο χρησιμοποιεί μια παραλλαγή του CELP, το LD-CELP (Low Delay CELP), για να μειώσει την καθυστέρηση στην κωδικοποίηση που προκαλείται από το μεγάλο μέγεθος των frames. Λειτουργεί στα 16 kbps. Ο ευρωπαϊκός φορέας ETSI δημοσιεύει την πρώτη έκδοση του προτύπου GSM 6.10 για χρήση στην κινητή τηλεφωνία. Χρησιμοποιείται κωδικοποίηση RPE στα 12.2 kbps (GSM full rate channel).

1994

Εισάγεται η παραλλαγή RCELP (Relaxation Code Excited Linear Prediction) από τους W. B. Kleijn και D. Nahumi.

1995

Έκδοση του ETSI GSM 06.20, με κωδικοποίηση VSELP στα 5.6 kbps (GSM half rate channel)

1996

Η ITU εκδίδει το πρότυπο G.729, το οποίο χρησιμοποιεί την παραλλαγή CSACELP (Conjugate Structure Algebraic Code Excited Linear Prediction) για κωδικοποίηση στα 6.4, 8 και 11.8 kbps.

1997

Έκδοση του ETSI GSM 06.60, με κωδικοποίηση CS-ACELP στα 12.2 kbps (GSM enhanced full rate). Το αμερικανικό υπουργείο άμυνας εκδίδει το πρότυπο MELP (Mixed Excitation Linear Prediction), το οποίο λειτουργεί στα 2.4 kbps και αντικαθιστά το προϋπάρχον FS-1015. Αργότερα εισάγονται προσθήκες στο MELP που επιτρέπουν μετάδοση και στα 1.2 kbps.

1999

Έκδοση του ETSI GSM 06.90 ή GSM Adaptive Multi-Rate, το οποίο υποστηρίζει διαφορετικούς ρυθμούς μετάδοσης μεταξύ 4.75 kbps και 12.2 kbps σε κωδικοποίηση CS-ACELP.

3.2.3 Αξιολόγηση κωδικοποιητών φωνής

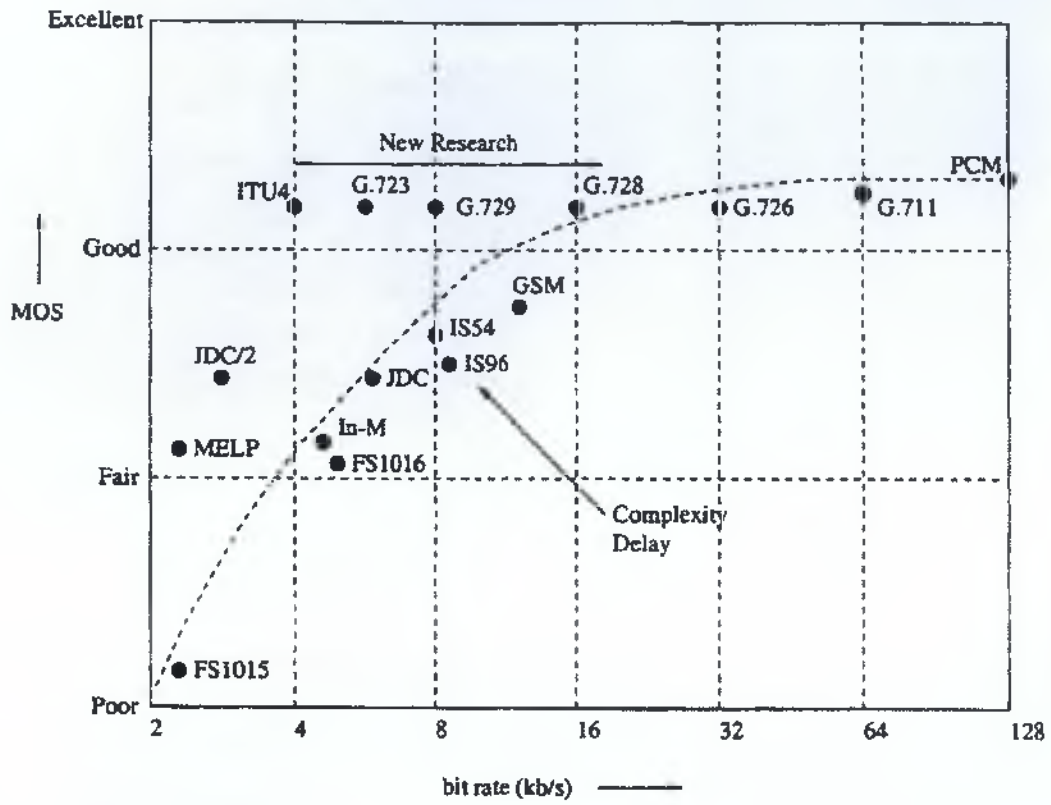
Για την αξιολόγηση της ποιότητας των διαφόρων κωδικοποιητών φωνής χρησιμοποιούνται διάφορα κριτήρια, τα οποία είναι κυρίως υποκειμενικά κριτήρια. Τα υποκειμενικά κριτήρια βασίζονται στην κρίση μιας ομάδας ανθρώπων για να χαρακτηρίσουν την ποιότητα ενός προτύπου, σε αντιδιαστολή με τα αντικειμενικά κριτήρια, τα οποία είναι μετρήσεις ορισμένων μεγεθών του αρχικού και ανακατασκευασμένου σήματος, που σχετίζονται με την ποιότητα της προσέγγισης. Για παράδειγμα το SNR (Signal to Noise Ratio) είναι ένα αντικειμενικό κριτήριο. Το πιο ευρέως χρησιμοποιούμενο υποκειμενικό κριτήριο είναι το MOS (Mean Opinion Score). Σύμφωνα με αυτό, ζητείται από μία ομάδα 20 έως 60 ανεκπαίδευστων ατόμων να βαθμολογήσουν την ποιότητα της φωνής που παράγεται από ένα συγκεκριμένο πρότυπο. Στη συνέχεια, η μέση τιμή της βαθμολογίας λαμβάνεται ως βαθμός ποιότητας του προτύπου.

Η κλίμακα βαθμολόγησης είναι η παρακάτω:

<u>Βαθμός MOS</u>	<u>Ποιότητα Φωνής</u>	<u>Χαρακτηριστικά Φωνής</u>
5	εξαιρετική (excellent)	απόλυτη διαύγεια φωνής
4	καλή (good)	καθαρή φωνή
3	ικανοποιητική (fair)	ποιότητα επικοινωνιών
2	χαμηλή (poor)	συνθετική φωνή
1	κακή (bad)	ακατανόητη φωνή

Coding Standard	Year	Bit Rate (kbps)	MOS	MIPS	Frame Size (msec)
Linear PCM	1948	128	4.5	0	0.125
G.711 μ -law PCM	1972	64	4.3	0.01	0.125
G.721 ADPCM	1984	32	4.1	2	0.125
G.722 ADPCM	1984	48/56/64	4.1	5	0.125
G.726/G.727 ADPCM	1990	16/24/32/40	4.1	2	0.125
G.728 LD-CELP	1992	16	4.0	30	0.625
G.729 CS-ACELP	1996	8	4.0	20	10
G.723.1 MPC-MLQ	1995	6.3/5.3	4.0/3.7	11	10
GSM FR RPLPC/LTP	1987	13	3.7	6	22.5
GSM HR VSELP	1994	5.6	3.5	14	22.5
IS-54 VSELP	1989	8	3.6	14	22.5
IS-96 QCELP	1993	1.2/2.4/4.8/9.6	3.5	15	22.5
FS-1015 LPC10(e)	1984	2.4	2.3	7	22.5
FS-1016 CELP	1991	4.8	3.0	16	30/7.5
NSA MELP	1996	2.4	3.2	40	22.5

Χαρακτηριστικά προτύπων κωδικοποίησης φωνής



Ποιότητα φωνής συναρτήσει του ρυθμού μετάδοσης για διάφορα πρότυπα.

3.3 Το πρότυπο LPC-10e (FS-1015)

3.3.1 ΓΕΝΙΚΑ

Το Υπουργείο Άμυνας των ΗΠΑ προσπαθούσε από τις αρχές της δεκαετίας του 1970 να αναπτύξει ένα πρότυπο για κωδικοποίηση φωνής σε πολύ χαμηλούς ρυθμούς μετάδοσης, χρησιμοποιώντας την κωδικοποίηση γραμμικής πρόβλεψης (Linear Predictive Coding). Το πρότυπο αυτό προοριζόταν για χρήση από τις αμερικανικές ένοπλες δυνάμεις και έπρεπε να έχει τα εξής χαρακτηριστικά:

1) Ψηφιακή επεξεργασία ήχου: σήμερα βέβαια δεν νοείται πρότυπο επεξεργασίας ήχου που να χρησιμοποιεί αναλογικά σήματα, αλλά εκείνη την εποχή η ψηφιακή επεξεργασία ήχου ήταν μια σχετικά πρωτόγνωρη πρακτική.

2) Πολύ χαμηλό ρυθμό μετάδοσης: Το πρότυπο επρόκειτο να χρησιμοποιηθεί σε στρατιωτικά περιβάλλοντα, όπου το διαθέσιμο εύρος ζώνης θα ήταν περιορισμένο (λόγω επίδρασης των καιρικών συνθηκών, έλλειψης τηλεπικοινωνιακής υποδομής, εχθρικών παρεμβολών κλπ). Έτσι, για να ολοκληρωθεί επιτυχώς μια εκπομπή, ήταν αναγκαίο να μειωθεί το bitrate όσο το δυνατόν περισσότερο.

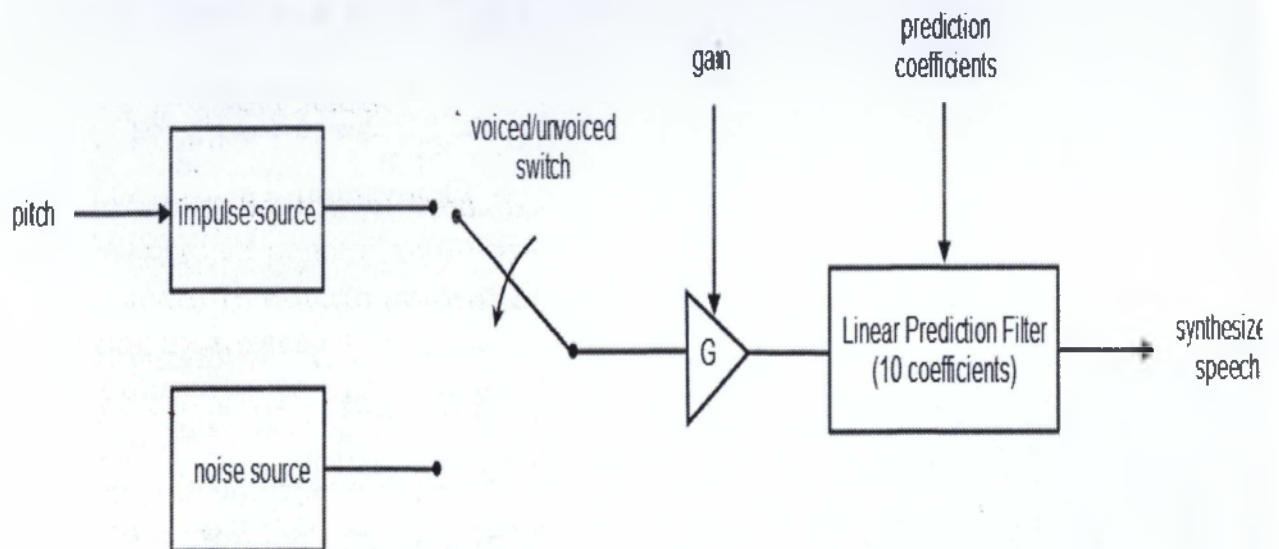
3) Βέλτιστη δυνατή ποιότητα ήχου: Στους χαμηλούς ρυθμούς μετάδοσης υπάρχει έντονο το πρόβλημα της κακής ποιότητας ήχου. Η φωνή που θα αναπαραγόταν στον δέκτη θα έπρεπε να είναι κατανοητή από τον ακροατή. Περισσότερη βαρύτητα όμως δινόταν στην μείωση του ρυθμού μετάδοσης παρά στην ποιότητα.

4) Ασφάλεια: Σε περίπτωση υποκλοπής της επικοινωνίας, η κωδικοποιημένη φωνή έπρεπε να μην μπορεί να αποκωδικοποιηθεί εύκολα. Για τον σκοπό αυτό πάντως χρησιμοποιήθηκαν περισσότερο μέθοδοι κρυπτογραφίας παρά η ίδια η τεχνική κωδικοποίησης.

Αποτέλεσμα αυτής της προσπάθειας ήταν η έκδοση του προτύπου **FS-1015 (Federal Standard 1015)** το 1984. Το πρότυπο έγινε ευρύτερα γνωστό ως **LPC-10 (Linear Predictive Coding - 10 coefficients)**, διότι χρησιμοποιούσε γραμμική πρόβλεψη με φίλτρο πρόβλεψης 10 συντελεστών. Αργότερα, μερικές προσθήκες στην αρχική μορφή του προτύπου είχαν ως αποτέλεσμα την δημοσίευση της βελτιωμένης έκδοσης **LPC-10e (LPC-10 enhanced)** κατά το έτος 1987.

3.3.2. Μοντέλο

Ως ένας κλασικός vocoder ανοιχτού βρόχου, το πρότυπο LPC-10e χρησιμοποιεί το μοντέλο γραμμικής πρόβλεψης με κατηγοριοποίηση έμφωνων/άφωνων τμημάτων (voiced-unvoiced speech model), όπως φαίνεται στο σχήμα:



Σχήμα – Μοντελοποίηση φωνής στο πρότυπο LPC-10e

3.4 ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ

3.4.1 Τι είναι η επεξεργασία σήματος

Ως Επεξεργασία Σήματος ορίζεται ο κλάδος της επιστήμης που ασχολείται με τις μεθόδους διαχείρισης ηλεκτρικών (ως επί το πλείστον) σημάτων. Η έννοια της διαχείρισης σήματος μπορεί να περιλαμβάνει τη σύλληψη, επεξεργασία, μετάδοση, αποθήκευση ή ανάκτηση ενός σήματος. Γενικότερα μπορεί να ειπωθεί ότι ο στόχος της επεξεργασίας σήματος είναι η μετατροπή ενός αρχικού σήματος σε ένα νέο σήμα, το οποίο έχει ορισμένα επιθυμητά χαρακτηριστικά. Μέχρι τα μέσα περίπου του 20ού αιώνα η επεξεργασία σήματος αφορούσε εξ' ολοκλήρου αναλογικά σήματα, δηλαδή σήματα που αναπαρίστανται με συνεχή τρόπο τόσο ως προς το πεδίο ορισμού όσο και ως προς το πεδίο τιμών τους. Ένα αναλογικό σήμα είναι πρακτικά η τιμή κάποιου ηλεκτρικού μεγέθους ενός κυκλώματος, πχ η τάση ενός κόμβου, η ένταση του ρεύματος ενός κλάδου, το ηλεκτρικό φορτίο ενός πυκνωτή κλπ. Η επεξεργασία αυτών των σημάτων γίνεται από ηλεκτρικά κυκλώματα, τα οποία αποτελούνται με τη σειρά τους από ηλεκτρονικά εξαρτήματα (αντιστάσεις, πηνία, πυκνωτές, τρανζίστορς, διόδους κλπ).

Διάφορα είδη κυκλωμάτων μπορούν να χρησιμοποιηθούν για διαφορετικούς σκοπούς (ενισχυτές, φίλτρα κ.α.).

3.4.2 Από το αναλογικό στο ψηφιακό σήμα

Η ραγδαία ανάπτυξη των ψηφιακών συστημάτων κατά το δεύτερο μισό του 20ου αιώνα επέφερε σαρωτικές αλλαγές σε όλα τα πεδία δραστηριοτήτων του ηλεκτρολόγου μηχανικού. Ο κλάδος της επεξεργασίας σήματος δεν θα μπορούσε να μείνει ανεπηρέαστος. Σύντομα φάνηκε ότι η ψηφιοποίηση των αναλογικών σημάτων και η ψηφιακή επεξεργασία τους έχει πολλά πλεονεκτήματα έναντι της προγενέστερης αναλογικής επεξεργασίας. Έτσι η ψηφιακή επεξεργασία σήματος έγινε το επίκεντρο της επιστημονικής έρευνας μέχρι σήμερα, παρότι και η αναλογική προσέγγιση δεν έχει παραμεριστεί εντελώς και ακόμα χρησιμοποιείται σε διάφορες εφαρμογές.

Η μετάβαση από το αναλογικό στο ψηφιακό σήμα γίνεται σε δύο στάδια: πρώτα με δειγματοληψία και στη συνέχεια με κβαντισμό.

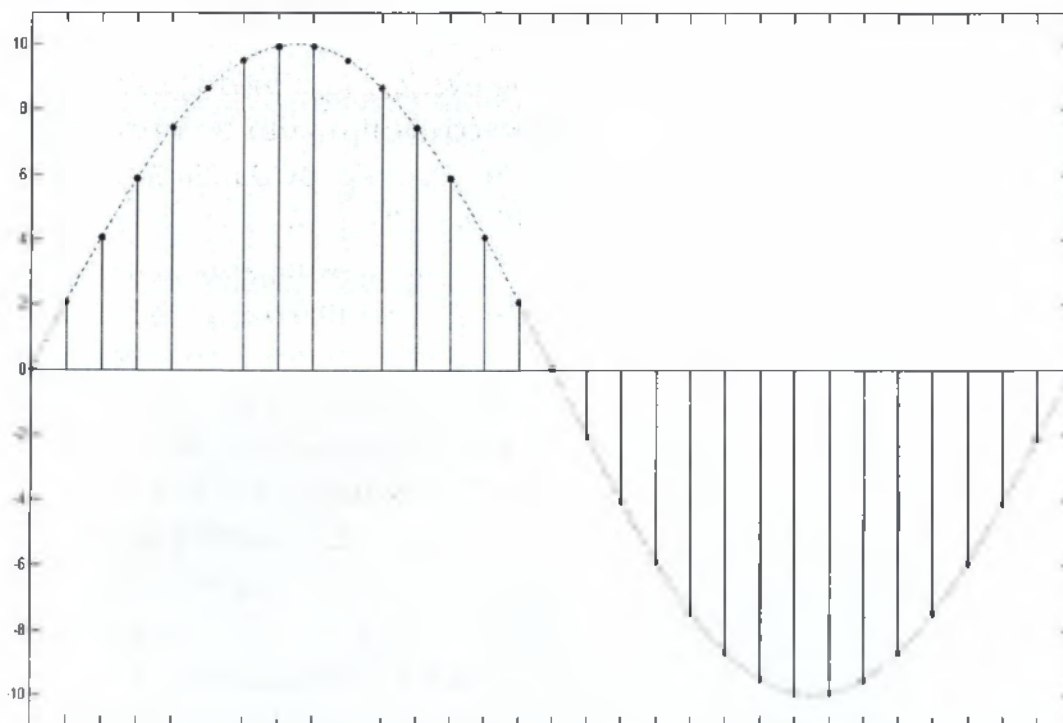
Θεωρώντας ότι το εξεταζόμενο σήμα είναι μονοδιάστατο με ανεξάρτητη μεταβλητή τον χρόνο μπορούμε να ορίσουμε:

Δειγματοληψία είναι η καταγραφή της τιμής του σήματος ανά τακτά χρονικά διαστήματα T . Το διάστημα T λέγεται περίοδος δειγματοληψίας. Έτσι προκύπτει ένα νέο διακριτό σήμα: $x[n] = x(nT)$
Μέσω της δειγματοληψίας δηλαδή επιτυγχάνεται η διακριτοποίηση του πεδίου ορισμού του σήματος.

Κβαντισμός είναι ο χωρισμός του πεδίου τιμών του σήματος σε ένα πεπερασμένο αριθμό συνεκτικών διαστημάτων και η αντιστοίχιση κάθε τέτοιου διαστήματος σε ένα και μοναδικό σύμβολο. Μέσω του κβαντισμού δηλαδή επιτυγχάνεται η διακριτοποίηση του πεδίου τιμών του σήματος.

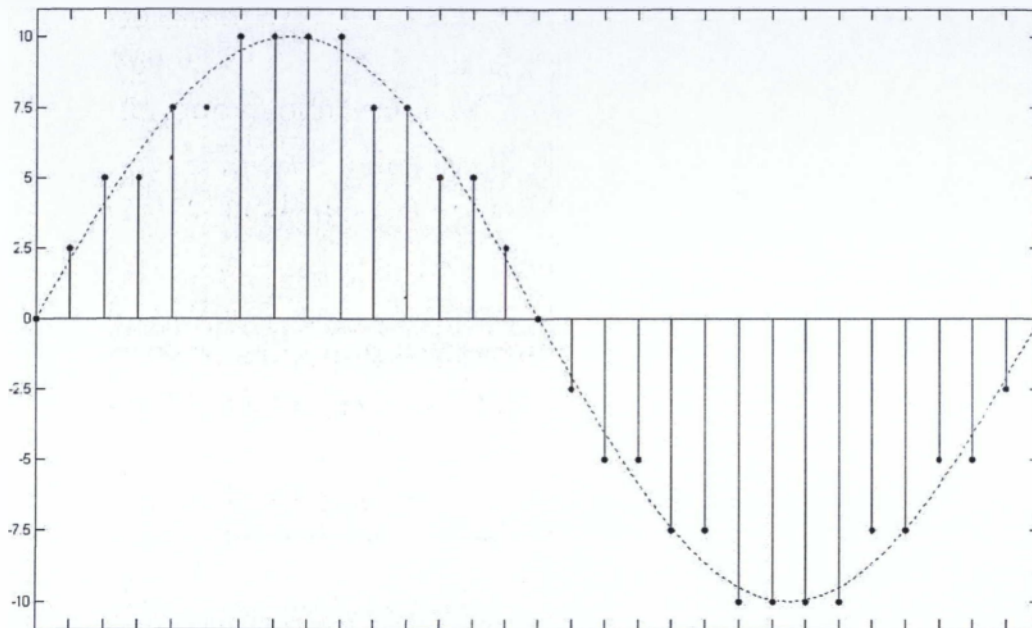
Η δειγματοληψία ενός αναλογικού σήματος ακολουθούμενη από κβαντισμό ονομάζεται ψηφιοποίηση του σήματος.

(α)



Δειγματοληψία (α) και κβαντισμός (β) ενός σήματος ημιτόνου

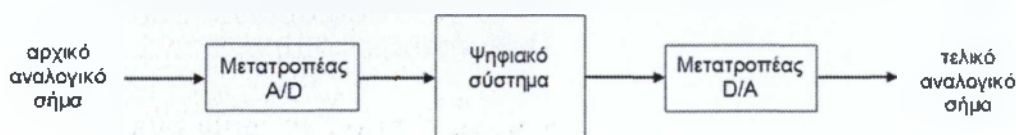
(β)



Σε ένα ψηφιακό σύστημα επεξεργασίας σήματος η ψηφιοποίηση των αναλογικών σημάτων επιτυγχάνεται μέσω ενός μετατροπέα A/D (Analog to Digital Converter). Ένας μετατροπέας A/D δέχεται στην είσοδό του ένα αναλογικό σήμα και παράγει δυαδικές λέξεις, οι οποίες αναπαριστούν το ψηφιοποιημένο σήμα.

Έτσι ένα ψηφιακό σύστημα μπορεί πλέον να επεξεργαστεί το σήμα. Η έξοδος του ψηφιακού συστήματος πρέπει συνήθως να μετατραπεί ξανά σε αναλογική μορφή. Αυτό επιτυγχάνεται μέσω ενός μετατροπέα D/A (Digital to Analog Converter), ο οποίος εκτελεί τον αποκβαντισμό και την ανακατασκευή του αναλογικού σήματος από το αντίστοιχο ψηφιακό.

Το σχηματικό διάγραμμα ενός συστήματος ψηφιακής επεξεργασίας



Θεώρημα Nyquist

Η τέλεια ανακατασκευή ενός αρχικού σήματος από το σύνολο των δειγμάτων του είναι δυνατή αν και μόνο αν η συχνότητα δειγματοληψίας είναι τουλάχιστον διπλάσια της μέγιστης συχνότητας που εμφανίζεται στο αρχικό σήμα. Δηλαδή πρέπει να ισχύει: f_s μεγαλύτερο ή ίσο $2F_{\max}$,
όπου f_s : συχνότητα δειγματοληψίας,
 $\max F$: μέγιστη συχνότητα αρχικού σήματος.

Σε αντίθεση με τη δειγματοληψία, ο κβαντισμός δεν είναι αναστρέψιμη διαδικασία. Έτσι πάντα χάνεται πληροφορία όταν ένα σήμα ψηφιοποιείται και στη συνέχεια από-ψηφιοποιείται. Η ψηφιακή επεξεργασία σήματος βρίσκει σήμερα πάρα πολλές εφαρμογές σε ποικίλα επιστημονικά πεδία.

Οι κυριότερες από αυτές εμπίπτουν στην κατηγορία της επεξεργασίας και συμπίεσης πολυμέσων (ήχου, βίντεο εικόνας). Από τα υπόλοιπα πεδία εφαρμογής ενδεικτικά αναφέρουμε την βιοϊατρική τεχνολογία, τα συστήματα radar, συστήματα συλλογής σεισμικών δεδομένων, πρόβλεψη καιρού, οικονομικά μοντέλα κλπ.

Κεφάλαιο 4

Συμπίεση

Ο ήχος είναι γενικά δύσκολο να συμπιεστεί γιατί η ακοή είναι πιο ευαίσθητη στις αλλοιώσεις του ήχου σε σχέση με την όραση. Εξαιτίας του ενδιαφέροντος για συμπίεση του ήχου στην τηλεφωνία, έχουν αναπτυχθεί πολλές τεχνικές αποτελεσματικής κωδικοποίησης της ομιλίας. Στην μουσική τα πράγματα είναι πιο δύσκολα γιατί και οι απαιτήσεις ποιότητας είναι μεγαλύτερες και το εύρος ζώνης της είναι σημαντικά μεγαλύτερο από αυτό της ανθρώπινης φωνής.

Στον παρακάτω πίνακα παρουσιάζονται οι συχνότητες δειγματοληψίας και το μήκος της λέξης κατά την κβαντοποίηση για διάφορα πρότυπα συμπίεσης ήχου. Εκτός από την κατάλληλη επιλογή αυτών των παραμέτρων ανάλογα με τις απαιτήσεις της εφαρμογής, περαιτέρω συμπίεση μπορεί να επιτευχθεί με αφαίρεση των σιωπηλών τμημάτων και με καλύτερες μεθόδους κωδικοποίησης όπως: μη γραμμική PCM όπως λογαριθμική ή μ -law, διαφορική PCM και προσαρμοστική διαφορική PCM.

Συχνότητα Δειγματοληψίας (kHz)	Κβαντοποίηση (bits)	Τεχνική Κωδικοποίησης	Ποιότητα
44.1	16	PCM	Hi-fi
37.8	8	ADPCM	Hi-fi
37.8	8	ADPCM	FM μετάδοση (μουσική)
18.9	8	ADPCM	AM μετάδοση (ομιλία)
8	8	PCM	Τηλεφωνική

Ηχητική ποιότητα και μέθοδος ψηφιοποίησης

4.1 Τεχνικές συμπίεσης ψηφιακού ήχου

Ο ασυμπίεστος ψηφιακός ήχος απαιτεί μεγάλο εύρος ζώνης για να μεταδοθεί. Υπάρχουν πολλές τεχνικές που χρησιμοποιούνται για να συμπίεσουν τον ψηφιακό ήχο. Μερικές από αυτές χρησιμοποιούνται στα συστήματα videoconference και αναλύονται παρακάτω. Με δύο λόγια, μπορούμε να πούμε ότι αυτές είναι τεχνικές οι οποίες μπορούν να κατορθώσουν πραγματικού χρόνου συμπίεσεις και αποσυμπίεσεις μέσω κώδικα (software) ή μέσω υλικού (hardware). Μερικές τεχνικές εφαρμόζονται σε ηχητικά σήματα γενικώς, ενώ άλλες είναι σχεδιασμένες ειδικά για σήματα ομιλίας - λόγου.

4.1.1 Mu-law και A-law PCM

Με την PCM κωδικοποίηση, κάθε δείγμα αντιπροσωπεύεται από μία κωδική λέξη. Η uniform PCM χρησιμοποιεί μία ομοιόμορφη ποσότητα βήματος των διαστημάτων. Πραγματοποιώντας τον μετασχηματισμό, η κλίμακα βήματος του διαστήματος μπορεί να αλλαχθεί σε λογαριθμική, επιτρέποντας έτσι σε ένα μεγαλύτερο εύρος να καλυφθεί με τον ίδιο αριθμό των bits. Υπάρχουν δύο κυρίως χρησιμοποιούμενοι μετασχηματισμοί: ο mu-law και ο A-law. Αυτοί οι μετασχηματισμοί επιτρέπουν 8 bits/sample να αντιπροσωπεύουν το ίδιο εύρος που θα επιτυγχάνονταν με 14 bits/sample της uniform PCM. Αυτό μεταφράζεται σε σχέση συμπίεσης 1:1,75 (αρχική ποσότητα πληροφοριών:συμπιεσμένη ποσότητα πληροφοριών). Εξαιτίας αυτής της λογαριθμικής φύσης της μετατροπής, δείγματα μικρού πλάτους είναι κωδικοποιημένα με μεγαλύτερη ακρίβεια σε σχέση με δείγματα μεγάλου πλάτους.

Οι mu-law και A-law μέθοδοι κωδικοποίησης είναι τυποποιημένοι στην Διεθνή Ένωση Τηλεπικοινωνιών (International Telecommunication Union - ITU) με το στάνταρ G.711 "Pulse Code Modulation των συχνοτήτων ομιλίας". Το mu-law PCM format κωδικοποίησης χρησιμοποιείται κυρίως στα ISDN δίκτυα της Βόρειας Αμερικής και της Ιαπωνίας, ενώ το A-law PCM format κωδικοποίησης

στις άλλες χώρες. Το G.711 είναι ένα στάνταρ ήχου καθορισμένο στο H.320 στάνταρ και στα 8kHz με 8 bits/sample και ένα κανάλι ήχου, η mu-law ή η A-law PCM απαιτεί bandwidth των 64kbps.

4.1.2 ADPCM

Η PCM μέθοδος κωδικοποίησης κωδικοποιεί κάθε ηχητικό σήμα ανεξάρτητα από τα συνεχόμενα δείγματα. Παρόλα αυτά, συνήθως συνεχόμενα δείγματα είναι ίδια μεταξύ τους και το πλάτος του δείγματος μπορεί να προβλεφθεί με ακρίβεια χρησιμοποιώντας την τιμή των συνεχόμενων δειγμάτων. Για παράδειγμα, μία απλή μέθοδος πρόβλεψης είναι να υποθέσουμε ότι το επόμενο δείγμα θα είναι ίδιο με το τρέχον δείγμα. Η ADPCM (Adaptive Differential Pulse Code Modulation) μέθοδος κωδικοποίησης υπολογίζει την διαφορά μεταξύ κάθε δείγματος και του υποθετικού επόμενου και κωδικοποιεί τη διαφορά (ο όρος διαφορά εννοεί διαφορικό). Λίγα bits, για την ακρίβεια τέσσερα χρειάζονται για να κωδικοποιηθεί η διαφορά από ένα ολόκληρο εύρος δείγματος. Αυτή η συμπίεση δίνει ταχύτητα εξόδου 32kbit/sec. Οι κωδικοποιητές μπορούν να προσαρμόσουν τα χαρακτηριστικά του σήματος αλλάζοντας τις παραμέτρους της υποθετικής ποσότητας. Με αυτόν τον τρόπο η ADPCM κωδικοποίηση επιτυγχάνει σχέση συμπίεσης 1:2 σε σύγκριση με την mu-law ή την A-law PCM. Διαφορετικές μορφές της ADPCM κωδικοποίησης περιλαμβάνουν τον τρόπο με τον οποίο η υποθετική τιμή υπολογίζεται και πώς η υποθετική ποσότητα προσαρμόζει τα χαρακτηριστικά του σήματος.

Πολλά συστήματα videoconference χρησιμοποιούν ADPCM μεθόδους κωδικοποίησης. Ο ITU έχει ορίσει ορισμένα πρότυπα περιγράφοντας διαφορετικές ADPCM μεθόδους όπως: G.721, G.722, G.723, G.726, G.727. Μία από τις μεθόδους κωδικοποίησης ήχου που καθορίζεται από το H.320 είναι το G.722 "7kHz κωδικοποίηση ήχου με 64kbps" που χρησιμοποιεί SB-ADPCM (Sub-Band ADPCM) κωδικοποίηση. Ο G.722 κωδικοποιητής δειγματοληπτεί στα 16kHz με 14 bits ακρίβειας. Με την SB-ADPCM μέθοδο, η συχνότητα σπάει σε δύο κομμάτια (υψηλή και χαμηλή) και τα σήματα σε κάθε κομμάτι κωδικοποιούνται με ADPCM. Το G.722 έχει τρεις τρόπους λειτουργίας: στα 64, 56 και 48kbps. Με τα 56 ή 48kbps τα επιπρόσθετα 8 ή 16 kbps bandwidth μπορούν να χρησιμοποιηθούν για άλλα δεδομένα. Το G.726 αντικαθιστά το G.721 και G.723, για να επιτρέψει μετατροπή μεταξύ 64kbit/s PCM κωδικοποίησης και καναλιών 40, 32, 24 ή

16kbit/s. Το G.727 είναι μία επέκταση του G.726 και εκδόθηκε για την ενσωμάτωση του ADPCM σε κανάλια 40, 32, 24 ή 16kbit/s, με συγκεκριμένο σκοπό να χρησιμοποιηθεί σε συστήματα που μεταφέρουν ομιλία σε πακέτα χρησιμοποιώντας το πρωτόκολλο PVP (Packetized Voice Protocol), το οποίο ορίζεται από το G.764.

4.1.3 LPC και CELP

Υπάρχουν μερικές μέθοδοι κωδικοποίησης σχεδιασμένες ειδικά για ομιλία. Χρησιμοποιώντας χαρακτηριστικά μοντέλα ήχου, αυτές μπορούν να παρουσιάσουν καλά αποτελέσματα για δεδομένα λόγου. Παρόλα αυτά, αυτές οι μέθοδοι συνήθως δεν δουλεύουν καλά για τα άλλα ηχητικά σήματα. Δύο μέθοδοι κωδικοποίησης που έχουν σχεδιαστεί για σήματα ομιλίας είναι οι LPC και CELP.

Η τεχνική LPC (Linear Predictive Coding, κωδικοποίηση γραμμικής πρόβλεψης) χρησιμοποιείται για να συμπιέζει ήχο στα 16kbit/s ή και πιο χαμηλά. Σε αυτή τη μέθοδο ο κωδικοποιητής ταιριάζει την ομιλία σε ένα απλό αναλυτικό μοντέλο φωνητικής δέσμης. Μόνο οι παράμετροι οι οποίες περιγράφουν το καλύτερο τρόπο με τον οποίο ταιριάζει η ομιλία στο μοντέλο αυτό στέλνονται στον αποκωδικοποιητή. Ένας αποκωδικοποιητής LPC χρησιμοποιεί τις παραμέτρους που του δίνονται για να δημιουργήσει συνθετική ομιλία η οποία ακούγεται αρκετά όμοια ως προς την πραγματική ομιλία. Το αποτέλεσμα είναι κατανοητό αλλά η ομιλία ακούγεται σαν να προέρχεται από μηχανή.

Η GSM (Groupe Speciale Mobile) κωδικοποίηση χρησιμοποιεί μία παραλλαγή του LPC που ονομάζεται RPE-LPC (Regular Pulse Excited - Linear Predictive Code). Το GSM ξεκίνησε σαν ένα Ευρωπαϊκό πρότυπο τηλεφωνικής ομιλίας και συμπιέζει 160 δείγματα των 13bit (2080bits) σε 260bits με σχέση συμπίεσης 1:8. Για 8kHz δειγματοληψία, η GSM κωδικοποίηση απαιτεί bandwidth 13kbps. Ο CELP (Code Excited Linear Prediction) κωδικοποιητής κάνει την ίδια φωνητική μοντελοποίηση στην περιοχή των συχνοτήτων της ομιλίας. Επιπλέον υπολογίζει τα λάθη μεταξύ των εισαγόμενων δεδομένων ομιλίας και του μοντέλου και μεταδίδει τις παραμέτρους του μοντέλου και την παρουσίαση των λαθών. Τα λάθη εμφανίζονται σαν ευρετήριο σε ένα κύριο βιβλίο κωδικών μοιρασμένο μεταξύ κωδικοποιητή και αποκωδικοποιητή. Από αυτό προέρχεται το όνομα "Code Excited". Τα παραπάνω δεδομένα και οι υπολογισμοί

παράγουν μία υψηλής ποιότητας κωδικοποίηση από τους απλούς συγκριτικά LPC κωδικοποιητές. Το G.728 πρότυπο της ITU το οποίο είναι ένα από τα ηχητικά format κωδικοποίησης που περιγράφεται από το H.320 χρησιμοποιεί μία παραλλαγή του CELP το LD-CELP (Low Delay CELP). Το G.728 απαιτεί bandwidth 16kbps και αρκετά περίπλοκους υπολογισμούς που συνήθως εκτελούνται σε ειδικό hardware.

4.2 Συμπίεση Φωνής στο VoIP

Υπάρχουν τόσοι πολλοί παράγοντες που επηρεάζουν την ποιότητα φωνής: η σύνδεση ευρείας ζώνης, το εύρος ζώνης, το υλικό, το λογισμικό και η ίδια η τεχνολογία. Οι παράγοντες του εύρους ζώνης, του υλικού και του λογισμικού είναι υπό τον έλεγχο μας- μπορούμε να τους αλλάξουμε και να τους βελτιώσουμε, Έτσι ώστε όταν μιλάμε για ποιότητα φωνής σε VoIP, συνήθως εννοούμε την υποκείμενη τεχνολογία την ίδια, κατι το οποίο είναι περα απο τον ελεγχο μας ως χρηστες . Ένα σημαντικό στοιχείο της τεχνολογίας VoIP είναι η συμπίεση δεδομένων.

Συμπίεση δεδομένων είναι μια διαδικασία κατά την οποία τα δεδομένα φωνής συμπιέζονται για να είναι λιγοτερο ογκώδη για μεταφορα. Η συμπίεση του λογισμικού (που ονομάζεται codec) κωδικοποιεί τα σήματα φωνής σε ψηφιακά δεδομένα που τα συμπιέζουν σε ελαφρύτερα πακέτα που στη συνέχεια μεταφέρονται μέσω του Διαδικτύου. Στον τοπο προορισμού, τα πακέτα αποσυμπιέζονται, τους δίνεται το αρχικό τους μέγεθος (αν και όχι πάντα) και μετατρέπονται πάλι σε αναλογική φωνή, έτσι ώστε να μπορεί ο χρήστης να ακούσει.

Τα Codecs δεν χρησιμοποιούνται μόνο για συμπίεση, αλλά και για την κωδικοποίηση, η οποία είναι η μετάφραση της αναλογικής φωνής σε ψηφιακά δεδομένα, τα οποία μπορούν να μεταδοθούν μέσω δικτύων IP.

Η ποιότητα και η αποτελεσματικότητα του λογισμικού συμπίεσης συνεπώς, έχει μεγάλη επίδραση στην ποιότητα φωνής σε VoIP συνομιλίες. Υπάρχουν καλές τεχνολογίες συμπίεσης και λιγότερο

καλές. Για να το εξηγήσουμε καλύτερα, κάθε τεχνολογία συμπίεσης έχει σχεδιαστεί για συγκεκριμένη χρήση κάτω από συγκεκριμένες συνθήκες. Μετά τη συμπίεση, ορισμένες τεχνολογίες συμπίεσης υφίστανται ορισμένες απώλειες όσον αφορά τα δεδομένα bits, ακόμα και τα πακέτα. Αυτό έχει ως αποτέλεσμα την κακή ποιότητα φωνής. Το VoIP κωδικοποιεί και συμπιέζει τα δεδομένα φωνής με τέτοιο τρόπο, ώστε ορισμένα από τα στοιχεία του audio stream χανονται. Αυτό ονομάζεται συμπίεση με απώλειες. Η απώλεια δεν είναι σκληρό πλήγμα για την ποιότητα της φωνής, αφού μεγάλο μέρος της γίνεται επίτηδες. Για παράδειγμα, οι ήχοι που δεν μπορούν να ακουστούν από το ανθρώπινο αυτί (συχνότητας κάτω ή πάνω από αυτή του φάσματος ακοής) απορρίπτονται επειδή θα είναι άχρηστοι. Το ίδιο συμβαίνει και με τη σιωπή. Επίσης παρόλο που χανονται και λεπτά κλάσματα του ακουστικού ήχου, τα μικροσκοπικά κομμάτια αυτά δεν μας εμποδίζουν από το να αντιληφθούμε τι λέγεται.

Σήμερα, οι τεχνολογίες συμπίεσης είναι τόσο προχωρημένες ώστε η έξοδος φωνής να είναι σχεδόν τέλεια. Αλλά ένα πρόβλημα έγκειται στην επιλογή του λογισμικού συμπίεσης: διαφορετικό λογισμικό συμπίεσης προσαρμόζεται σε διαφορετικές ανάγκες. Για παράδειγμα, υπάρχουν ορισμένα για φωνή, ορισμένα για δεδομένα και κάποια για φαξ. Αν δοκιμάσετε να στείλετε fax χρησιμοποιώντας το λογισμικό φωνητικής συμπίεσης, η ποιότητα θα υποφέρει.

Συμπίεση δεδομένων: όταν αναπτυχθεί και χρησιμοποιηθεί αποτελεσματικά, μπορεί να είναι το σημαντικότερο στοιχείο που ωθεί το VoIP, από άποψη ποιότητας φωνής, πιο πάνω και από τη σταθερή τηλεφωνία. Αυτό μπορεί να είναι εφικτό εφ'όσον τα λοιπά στοιχεία (εύρος ζώνης, υλικό κ.λπ.), είναι ευνοϊκά. Από τη στιγμή που η συμπίεση ελαφραίνει το φορτίο των δεδομένων προς μεταφορά σε ένα ορισμένο χρονικό διάστημα, καλύτερα αποτελέσματα μπορούν να επιτευχθούν.

4.3 Δεδομένα συμπίεσης ήχου

Συμπίεσης ήχου είναι μια μορφή συμπίεσης δεδομένων με στόχο τη μείωση της απαίτησης εύρους ζώνης μετάδοσης του ψηφιακού ήχου και το μέγεθος αποθήκευσης των αρχείων ήχου . Έχου συμπίεση αλγόριθμοι εφαρμόζονται σε λογισμικό ηλεκτρονικών υπολογιστών , όπως κωδικοποιητές ήχου . Generic αλγόριθμοι συμπίεσης δεδομένων έχουν κακές επιδόσεις με τα δεδομένα ήχου, σπάνια μείωση του μεγέθους των δεδομένων πολύ κάτω από το 87% από το αρχικό, και δεν έχουν σχεδιαστεί για χρήση σε εφαρμογές σε πραγματικό χρόνο. Ως εκ τούτου, έχουν δημιουργηθεί αλγόριθμοι ειδικά βελτιστοποιημένα ήχου χωρίς απώλειες και με απώλειες. Με απώλεια αλγόριθμοι προσφέρει μεγαλύτερες ταχύτητες συμπίεσης και χρησιμοποιούνται γενικά στις συσκευές ήχου των καταναλωτών.

Σε αμφότερες τις απώλειες και χωρίς απώλειες συμπίεση, πλεονασμού πληροφοριών μειώνεται, χρησιμοποιώντας μεθόδους όπως η κωδικοποίηση , αναγνώρισης προτύπων και γραμμική πρόβλεψη να μειωθεί η ποσότητα των πληροφοριών που χρησιμοποιούνται για να αντιπροσωπεύουν τις μη συμπιεσμένων δεδομένων.

Το trade-off μεταξύ ελαφρώς μειωμένη ποιότητα ήχου και μετάδοσης ή το μέγεθος αποθήκευσης αντισταθμίζεται από το τελευταίο για τις περισσότερες πρακτικές εφαρμογές ήχου με τον οποίο οι χρήστες δεν μπορούν να αντιληφθούν την απώλεια της ποιότητας απόδοση της αναπαραγωγής. Για παράδειγμα, ένα Compact Disc κατέχει περίπου μία ώρα ασυμπίεστου μουσική υψηλής πιστότητας, λιγότερο από 2 ώρες μουσικής με συμπίεση, ή 7 ώρες μουσικής με συμπίεση, στο MP3 format σε μέτρια ποσοστά.

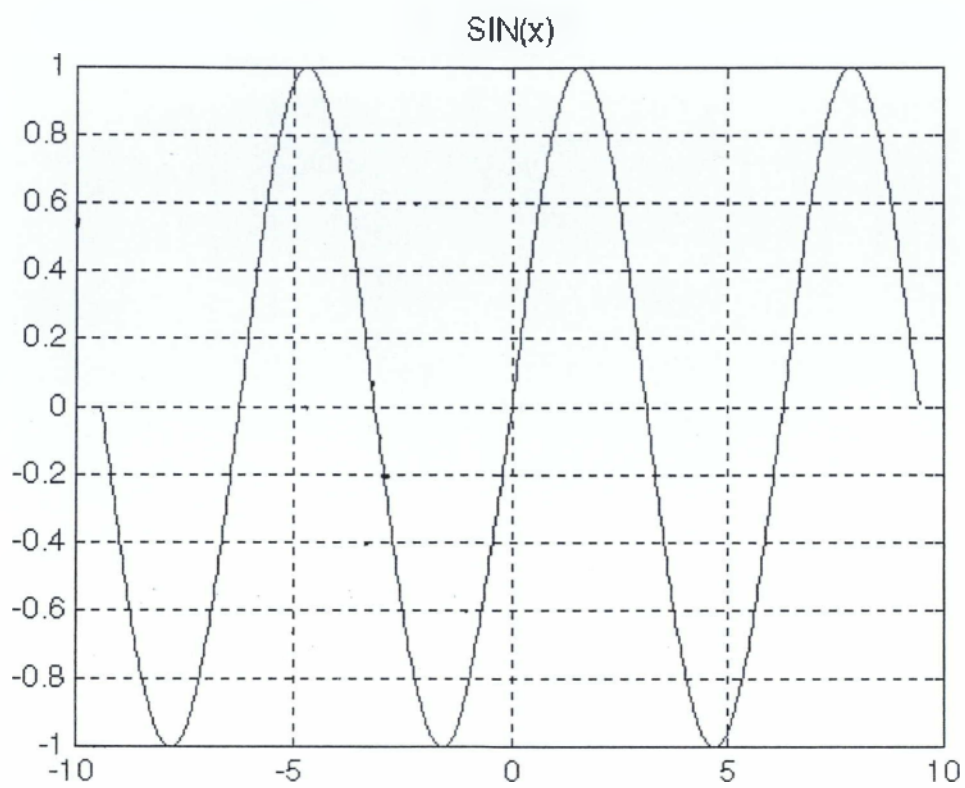
4.4 Χρήσιμοι μετασχηματισμοί για αλγόριθμους συμπίεσης

Τα δεδομένα εισόδου μπορούν να μετασχηματιστούν με αρκετούς τρόπους προκειμένου να γίνει πιο εύκολη η εφαρμογή των τεχνικών συμπίεσης. Ο πιο συνηθισμένος μετασχηματισμός στις σημερινές τεχνικές είναι ο Διακριτός Συνημιτονικός Μετασχηματισμός (Discrete Cosine Transform). Αυτός είναι μια παραλλαγή του διακριτού μετασχηματισμού Fourier.

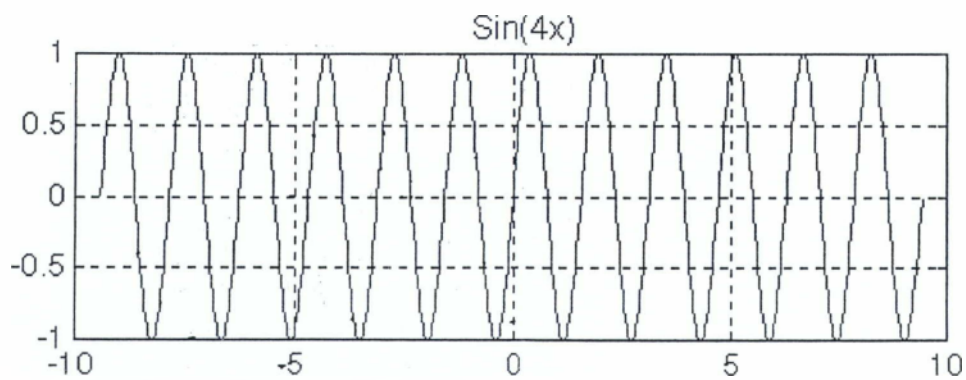
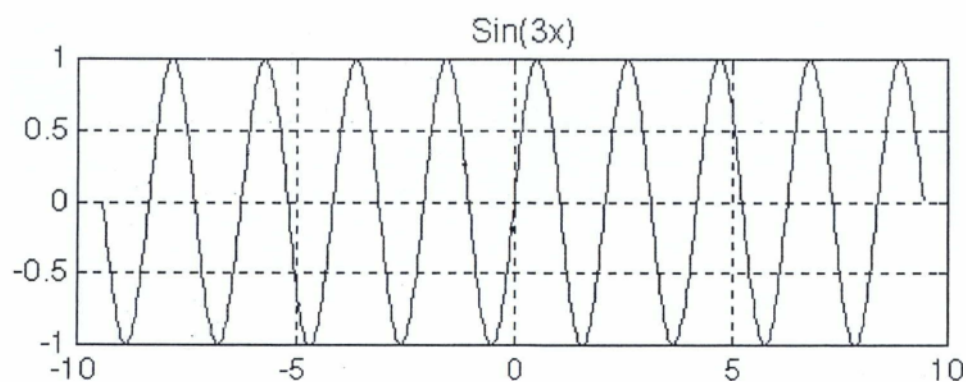
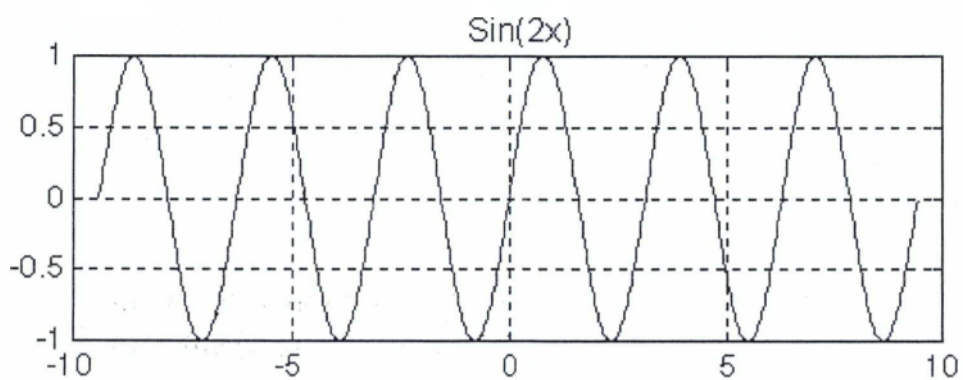
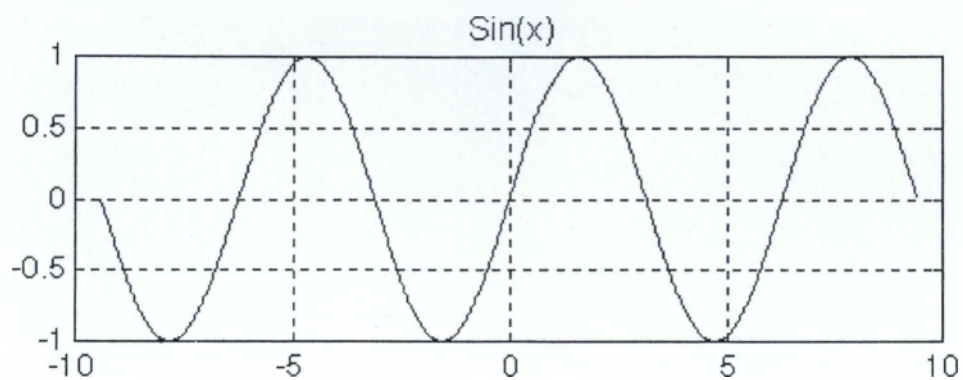
Κάθε σήμα είτε είναι ήχος είτε είναι βίντεο μπορεί να θεωρηθεί μια περιοδική κυματομορφή. Αν σκεφτούμε ότι μια ακολουθία ήχων είναι διαμόρφωση ενός ηχητικού κύματος, αντίστοιχα η σάρωση μιας εικόνας από την κάμερα αποδίδει ένα κύμα το οποίο έχει περιοδική μορφή στο πεδίο του χρόνου. Είναι δυνατόν να μετασχηματίσουμε το αρχικό σήμα από το πεδίο του χρόνου σε μία συνάρτηση Fourier, η οποία θα είναι το άθροισμα ενός αριθμού περιόδων, και η κάθε μία περίοδος θα είναι μια συγκεκριμένη συχνότητα ή μήκος κύματος. Μπορούμε να θεωρήσουμε αυτούς τους συντελεστές σαν την συμβολή σε έναν αριθμό καθαρών ημιτονοειδών συχνοτήτων (γνωστό και ως πυκνότητα φάσματος), που συνδυαζόμενοι αποδίδουν το πραγματικό σήμα.

$$g(\omega) = K_1 \int_{-\infty}^{+\infty} f(t) \exp(-j\omega t) dt$$
$$f(t) = K_2 \int_{-\infty}^{+\infty} g(\omega) \exp(-j\omega t) dt$$

Η ημιτονική κυματομορφή φαίνεται στο παρακάτω σχήμα.

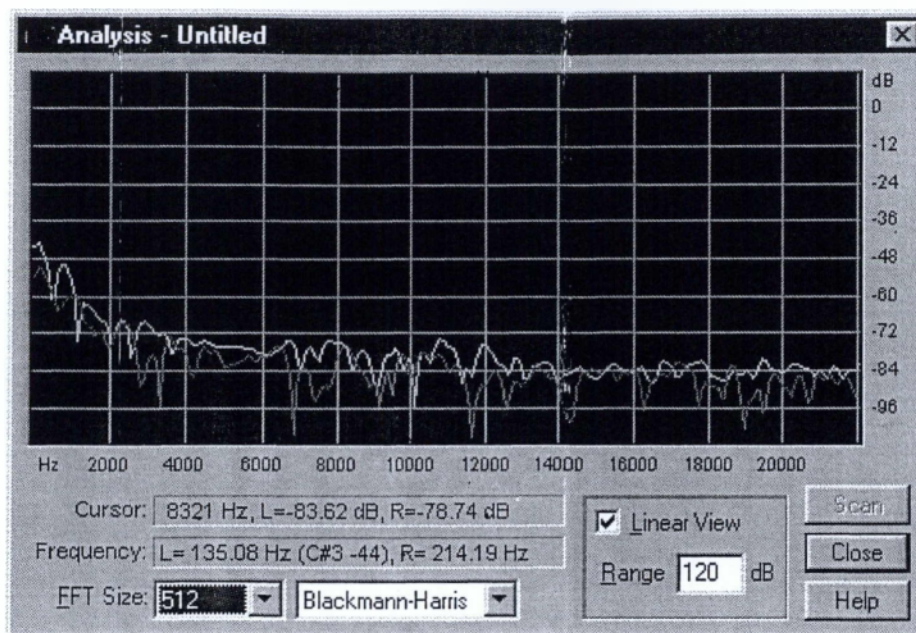


Στην επόμενη γραφική παράσταση παρουσιάζεται ένα σει ημιτονικών κυματομορφών ($x, 2x, 3x$ κ.ο.κ.)



Μπορεί να φανταστεί κανείς τις αρμονικές να σαρώνουν ένα τυπικό σήμα ήχου όπως αυτό που φαίνεται στην εικόνα που ακολουθεί και να εξάγουν ένα φάσμα όπως αυτό που φαίνεται στο παρακάτω σχήμα ή ένα σετ συντελεστών που απεικονίζουν την συμβολή της κάθε συχνότητας στο συγκεκριμένο τμήμα του σήματος.





Οι τεχνικές συμπίεσης φωνής εκμεταλλεύονται τις ιδιότητες των σημάτων φωνής. Στο πεδίο του χρόνου μπορούμε να δούμε ότι υπάρχει μεγάλη ομοιότητα μεταξύ δειγμάτων συνεχόμενης ομιλίας. Αυτό σημαίνει ότι ένα σύστημα το οποίο μόνο στέλνει διαφορές μεταξύ δειγματοληπτικών τιμών θα πετύχει μια μερική συμπίεση (η τεχνική DPCM κατορθώνει και γλυτώνει 1bit σε σχέση με το καθαρό PCM πρωτόκολλο).

Μπορούμε να δούμε ότι υπάρχουν πολύ περισσότερες τιμές σε δείγματα χαμηλής έντασης από ότι στα δείγματα με υψηλή ένταση. Αυτό σημαίνει ότι θα μπορούσαμε να χρησιμοποιήσουμε περισσότερα bit για να αναπαραστήσουμε τις χαμηλές τιμές από ότι για να αναπαραστήσουμε τις υψηλές. Μπορούμε να χρησιμοποιήσουμε την τεχνική ADPCM (Adaptive Differential Pulse Code Modulation) η οποία επιτυγχάνει μέχρι και 50% την συμπίεση σε σχέση με την απλή PCM χωρίς καμία προφανή απώλεια στην ποιότητα του ήχου, και είναι και σχετικά φτηνή στην εφαρμογή της.

Μια πιο έξυπνη τεχνική συμπίεσης στηρίζεται σε δύο παράγοντες: μία εκτίμηση του “μοντέλου” της ομιλίας και του “μοντέλου” του ακροατή. Τέτοιες τεχνικές συνήθως αναγνωρίζουν την ομιλία και παράγουν ένα σετ φίλτρων τα οποία εκπέμπονται στον δέκτη (ακροατή) και χρησιμοποιούνται για να επανακατασκευάσουν τον ήχο στην πλευρά του δέκτη χρησιμοποιώντας πηγή μίας μόνο συχνότητας και ένα offset “λευκού” θορύβου. Παραδείγματα CODEC

που βασίζονται σε αυτή την τεχνική είναι τα LPC και τα CELP. Αν στα παραπάνω προσθέσουμε και ένα μοντέλο το οποίο μας δείχνει το πως αντιλαμβάνονται οι άνθρωποι τον ήχο μας οδηγεί σε μία πιο ακριβή αλλά σαφώς πολύ αποδοτική μέθοδο συμπίεσης σαν αυτή που χρησιμοποιείται στα CODEC που βασίζονται στην MPEG τεχνική.

Στο πεδίο των διακριτών συνημίτονων μπορούμε να αντικαταστήσουμε το $\sin(\omega t) + j\cos(\omega t)$ απλά με το $\cos(t)$, και τα ολοκληρώματα γίνονται αθροίσματα. Ο χρόνος t απλά παραπέμπει σε ένα σημείο στο συνολικό ψηφιακό δείγμα (για μια εικόνα). Φυσικά μία εικόνα είναι δύο διαστάσεων πράγμα που σημαίνει ότι πρέπει να κάνουμε τις μετατροπές για δύο χωρικές συντεταγμένες, αντί για μία χρονική συντεταγμένη.

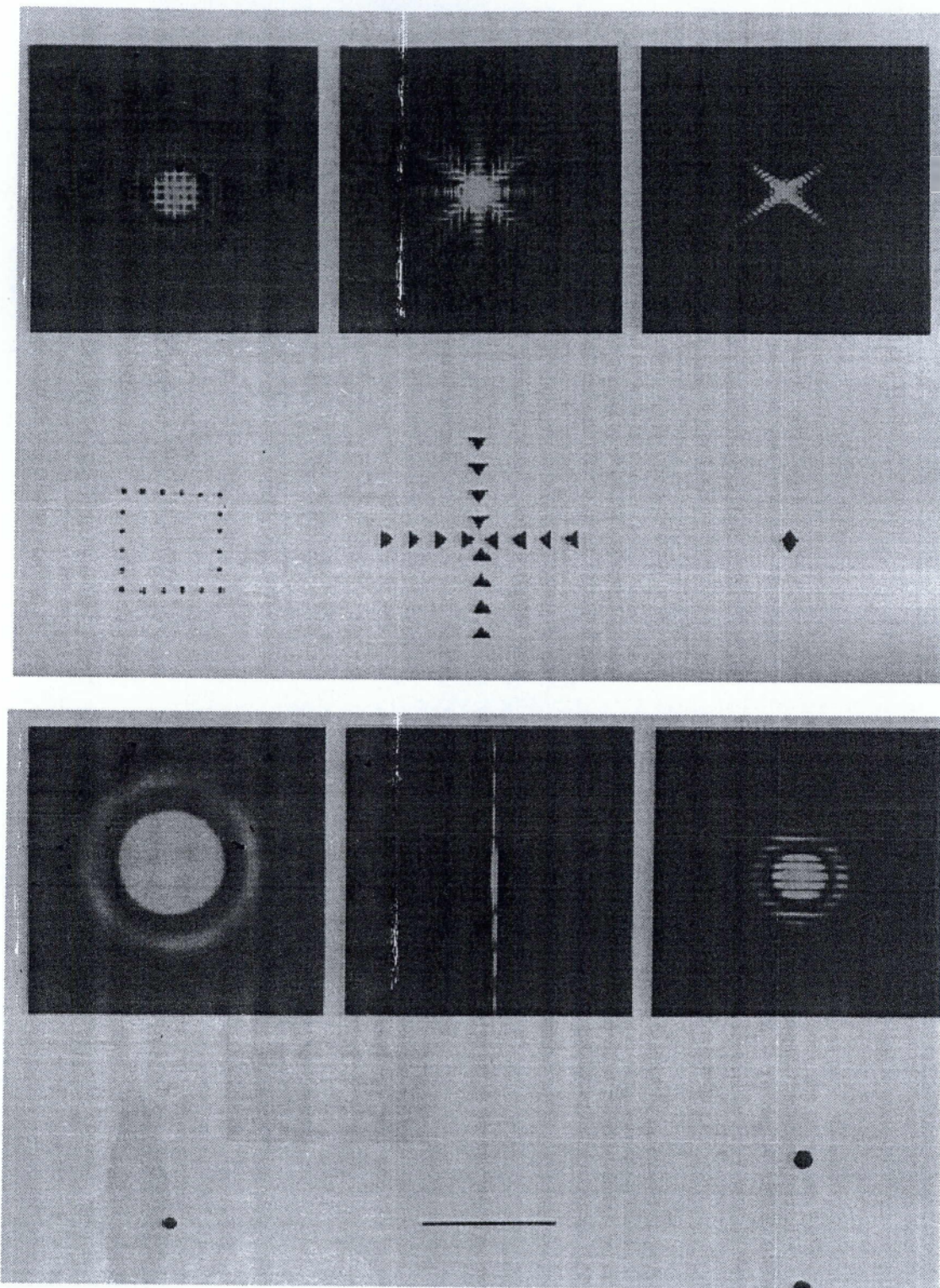
Έτσι λοιπόν προκειμένου να μετατρέψουμε ένα μπλοκ δεδομένων εικόνας της τάξης 8×8 pixels στο πεδίο της συχνότητας εφαρμόζουμε το εξής:

$$DCT[i, j] = \sum_{x=0}^7 \sum_{y=0}^7 \cos(2\pi i(x_i + y_j)) f(x, y)$$

Ας δούμε όμως πως δουλεύει αυτή η σχέση στην πράξη. Στο παρακάτω σχήμα μπορούμε να δούμε ορισμένες τεχνητές εικόνες και τους μετασχηματισμούς τους.

Στον μετασχηματισμό, μπορούμε να δούμε πως φαίνεται η εικόνα που δώσαμε ως είσοδο. Για παράδειγμα η κάτω αριστερή γωνία του DCT αντιπροσωπεύει την DC συνισταμένη, ενώ οι γωνίες αντιπροσωπεύουν πληροφορία για τις οριζόντιες, κάθετες και λοιπές συχνότητες.

Τέλος σκόπιμο είναι να επισημάνουμε ότι υπάρχει ένας αριθμός γνωστών τεχνικών για την βελτιστοποίηση των συνημιτονικών μετασχηματισμών και των μετασχηματισμών Fourier, που βασίζονται στον κερματισμό στο πεδίο του χρόνου. Αυτές οι τεχνικές ουσιαστικά προέρχονται από το γεγονός ότι υπάρχει μεγάλη συμμετρία στον μετασχηματισμό και η σχέση $x=2, i=1$, μας δίνει την ίδια τιμή συνημίτονου με την σχέση $x=1, i=2$.



Κεφάλαιο 5

Επιλογος Συμπεράσματα

5.1 Συνεχής ανάγκη για τη συμπίεση φωνής

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 (solidstate) 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.

5.2 Οι στόχοι της τεχνολογίας

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.

5.3 Μελλοντικές κατευθύνσεις

5.3.1 Κωδικοποίηση, επικοινωνία και δικτύωση

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.

5.3.2 Ψηφιακή επεξεργασία σήματος

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.

Βιβλιογραφία

- [1] Ψηφιακή επεξεργασία σήματος
Monson H. Hayes, Εκδόσεις Τζιόλα
- [2] Τηλεπικοινωνιακά συστήματα
Taub/Schilling, Εκδόσεις Τζιόλα

Πηγές διαδικτύου

[03] http://en.wikipedia.org/wiki/Human_Voice

[04] http://en.wikipedia.org/wiki/Vocal_tract

[05] http://en.wikipedia.org/wiki/Speech_communication

[06] http://en.wikipedia.org/wiki/Speech_production

[07] <http://web.teipir.gr/WWWbeldia/greek/Diploma/kefalaio3/346.html>

[08] <http://www.it.uom.gr/project/MultimediaTechnologyNotes/chap2b-3.htm>

[09] <http://vivliothmmy.ee.auth.gr/>