ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

ΠΕΡΙΓΡΑΦΗ ΑΝΘΡΩΠΙΝΗΣ ΦΩΝΗΣΗΣ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΗΧΟΥ (CODEC) ΣΤΑ ΠΡΟΤΥΠΑ ΚΙΝΗΤΩΝ ΕΠΙΚΟΙΝΩΝΙΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μιχαήλ Τσαντίλας

Επιβλέπων : Αθανάσιος Δ. Παναγόπουλος
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2024

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

ΠΕΡΙΓΡΑΦΗ ΑΝΘΡΩΠΙΝΗΣ ΦΩΝΗΣΗΣ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΗΧΟΥ (CODEC) ΣΤΑ ΠΡΟΤΥΠΑ ΚΙΝΗΤΩΝ ΕΠΙΚΟΙΝΩΝΙΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μιχαήλ Τσαντίλας

Επιβλέπων : Αθανάσιος Δ. Παναγόπουλος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 7η Ιουνίου 2024.

Αθανάσιος Δ. Παναγόπουλος	Γεώργιος Φικιώρης	Γεώργιος Ματσόπουλος
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2024

.....
Μιχαήλ Τσαντίλας

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Στη Βάσω και στον Δημήτρη, που με έσπρωξαν να μπω.

Στη Σάντη, που με έσπρωξε να βγω.

Περίληψη

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η ανάλυση του πώς λειτουργούν οι αλγόριθμοι κωδικοποίησης ομιλίας και μουσικής, στο πλαίσιο των δικτύων κινητών επικοινωνιών, από μία σκοπιά υψηλού επιπέδου.

Στα πρώτα δύο κεφάλαια γίνεται μία σύντομη ιστορική αναδρομή στους σταθμούς της εξέλιξης των τηλεπικοινωνιών, και στις διαδοχικές γενιές δικτύων κινητών επικοινωνιών.

Στα επόμενα τέσσερα κεφάλαια, που αποτελούν τον κύριο κορμό της εργασίας, εξηγούνται οι βασικές αρχές του ήχου, το πώς λειτουργεί η ανθρώπινη ομιλία και ακοή, και ποια μοντέλα αυτών αξιοποιούνται από τους σχεδιαστές των codec ήχου, προκειμένου η ηχητική πληροφορία να μεταδοθεί ασύρματα, και με αποτελεσματικότητα ως προς το κόστος και την ποιότητα.

Στο τελευταίο κεφάλαιο γίνεται σύντομη επισκόπηση τεσσάρων δημοφιλών codec, τα οποία χρησιμοποιούνται ευρύτατα σήμερα, σε εφαρμογές των δικτύων κινητών επικοινωνιών.

Λέξεις κλειδιά

αλγόριθμος κωδικοποίησης, codec, κινητές επικοινωνίες, ήχος, ομιλία, μουσική

Abstract

The subject of this diploma thesis is the analysis of how coding algorithms for speech and music work, in the field of mobile communication networks, from a high level standpoint.

In the first two chapters, we give a brief historical overview of significant turning points in the evolution of telecommunications, and of the consecutive generations of mobile communication networks.

Afterwards, and across four chapters that constitute the main body of this work, we analyse the basic principles of sound, how human speech and hearing works, and which models of them are used by sound codecs designers, in order to wirelessly transmit audio information, in a cost-effective and quality-mindful way.

In the last chapter, we give an overview of four popular audio codecs that are widely used today, in mobile communication network applications.

Key words

coding algorithm, codec, mobile communications, sound, audio, speech, music

Ευχαριστίες

Ολοκληρώνοντας έναν κύκλο σπουδών που κράτησε πολύ παραπάνω από όσο αναμενόταν, θα ήθελα πριν από όλους να ευχαριστήσω τον Καθηγητή του Εθνικού Μετσόβιου Πολυτεχνείου κ. Αθανάσιο Παναγόπουλο, που δέχτηκε να ασχοληθεί με την περίπτωσή μου, και που ενδιαφέρθηκε να βρει και να μου αναθέσει ένα τόσο ενδιαφέρον θέμα. Κύριε Παναγόπουλε, σας είμαι ευγνώμων.

Θα ήθελα, επίσης, να ευχαριστήσω όλους τους καθηγητές και τις καθηγήτριες, και το προσωπικό της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Ε.Μ.Π., για τις γνώσεις, τις εμπειρίες, και τις υπηρεσίες, όλα αυτά τα χρόνια.

Ευχαριστίες πρέπουν και στους/στις φίλους/ες, συναδέλφους/ισες, συμφοιτητές/τριες, για τις στιγμές που μοιραστήκαμε κατά καιρούς: Γιάννης Πεχλιβανίδης, Ισαάκ Σολωμονίδης, Κώστας Μαρμαρινός, Σίμος Ματτές, Σωτήρης Ρήγας, Σπύρος Χόρτης, Δημήτρης Αθανασόπουλος, Άγγελος Μοσχούδης, Επιστήμη Τσέτσου, Γεωργία Χολέβα, Μάνος Πουλής.

Τέλος, ευχαριστώ την ευρύτερη οικογένειά μου, για όλη τη στήριξη στη μέχρι τώρα πορεία μου.

Μ.Τ.

Απρίλιος 2024

Περιεχόμενα

Κατάλογος σχημάτων

Κεφάλαιο 1

Εισαγωγή: Ορολογία και (λίγη) Ιστορία

1.1 Επικοινωνία

Επικοινωνία (επί + κοινωνία) ονομάζουμε τη διαδικασία ανταλλαγής (μετάδοσης - λήψης) μηνυμάτων (πληροφοριών, σκέψεων, ιδεών, συναισθημάτων κλπ.), ανάμεσα σε έναν πομπό και έναν δέκτη. Η επικοινωνία πραγματοποιείται με τη χρήση ενός κώδικα, δηλαδή ενός συστήματος συμβόλων (προφορικού ή γραπτού λόγου, κινήσεων, σημάτων ήχων κλπ.) [1].

Παρότι νοείται επικοινωνία και μεταξύ οντοτήτων του φυσικού περιβάλλοντος (ζώων, φυτών κλπ.), σε ό,τι αφορά την ανθρωπότητα η ικανότητα της επικοινωνίας αναπτύχθηκε και εξελίχθηκε μέσα στους αιώνες, και αποτέλεσε το βασικό όπλο για την ανάδειξη του είδους σε κυρίαρχο του πλανήτη. Η εξέλιξη αυτή βοηθήθηκε -επιταχύνθηκε- από την εφεύρεση τρόπων καταγραφής της πληροφορίας και της γνώσης, και μεταβίβασής τους σε κάθε επόμενη γενιά· από την εφεύρεση, δηλαδή, νέων τρόπων επικοινωνίας.

Σήμερα, ο όρος απαντάται σχεδόν σε όλους τους τομείς της ανθρώπινης δραστηριότητας· κοινωνικής, οικονομικής, πολιτιστικής κλπ. Ο τρόπος που επιτυγχάνεται -ή όχι- η επικοινωνία διαμορφώνει δυναμικές και αποφασίζει το πώς εξελίσσονται οι ανθρώπινες κοινωνίες.

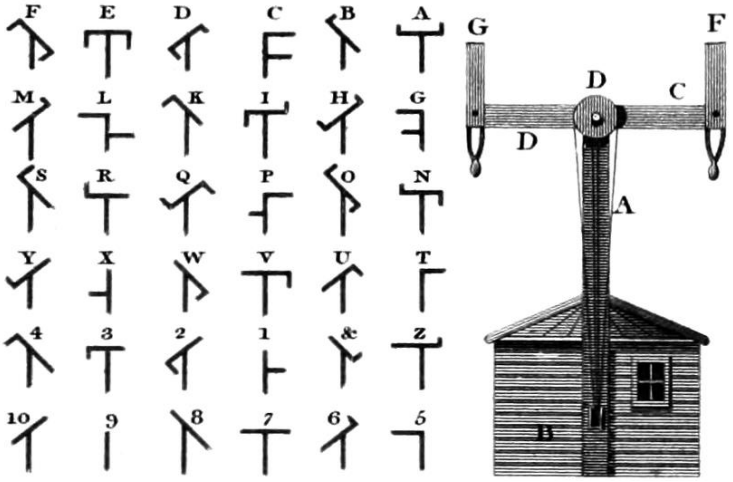
1.2 Τηλεπικοινωνίες

Με τον όρο *τηλεπικοινωνίες* (τηλέ + επικοινωνία) αναφερόμαστε στις

τεχνολογίες που εξαλείφουν τις αποστάσεις μεταξύ προσώπων, χωρών, ηπείρων -που παρακάμπτουν την ανάγκη για φυσική (διά ζώσης) επικοινωνία.

Ο όρος δημιουργήθηκε από τον Γάλλο λογοτέχνη και μηχανικό Édouard Estaunié (1862-1942), στο βιβλίο του *Traité Pratique de Télécommunication Électrique (Télégraphic-Téléphonie)* (1904), και αναγνωρίστηκε επίσημα το 1932 από την International Telecommunication Union (ITU). Στην ελληνική γλώσσα, ολοένα και περισσότερο πλέον χρησιμοποιείται στη θέση του ο απλουστευμένος όρος *επικοινωνίες*.

Παρά την τόσο πρόσφατη επινόηση του όρου, πάντως, η ιστορία των τηλεπικοινωνιών απλώνεται χιλιετίες στο παρελθόν. Από τα πρώτα μέσα που χρησιμοποιήθηκαν ήταν ο ήχος των τυμπάνων (Αφρική) και τα σήματα καπνού (Βόρεια Αμερική, Κίνα). Είναι, άλλωστε, γνωστές και οι αρχαιοελληνικές φρυκτωρίες, ένα σύστημα οπτικής επικοινωνίας με τη χρήση πυρσών.



Σχήμα 1.1 Ο οπτικός τηλέγραφος, με τον κώδικα γραμμάτων και συμβόλων του [3].

Υπήρχαν επίσης τα ταχυδρομικά περιστέρια, μεταφορά επιστολών με άλογα και άμαξες, καθώς και υδραυλικά συστήματα οπτικής επικοινωνίας. Ώσπου, στα 1790, ο Γάλλος μηχανικός Claude Chappe ξεκίνησε να πειραματίζεται με ένα σύστημα οπτικής επικοινωνίας, το οποίο εξελίχθηκε στον Οπτικό Τηλέγραφο.

Το ξεκίνημα για εκμετάλλευση του ηλεκτρισμού στον τομέα των τηλεπικοινωνιών έγινε στα 1726, από επιστήμονες όπως οι Pierre-Simon Laplace, André-Marie Ampère και Carl Friedrich Gauss. Χρειάστηκαν πάντως σχεδόν 100 χρόνια για να φτάσουμε στον πρώτο ηλεκτρικό τηλέγραφο: τον κατασκεύασε ο Francis Ronalds, το 1816. Χρειάστηκαν πολλές ακόμα προσπάθειες, από τον Charles Wheatstone και τον Samuel Morse μεταξύ άλλων, ώσπου η νέα τεχνολογία να επικρατήσει, και να φτάσει τελικά να συνδέσει και τις δύο κραταιές ηπείρους του Δυτικού Κόσμου, την Ευρώπη και τη Βόρεια Αμερική (1866).

Εξέλιξη της τηλεγραφίας αποτέλεσε η τηλεφωνία, η οποία δεν άργησε να περάσει σε εμπορική εκμετάλλευση εκατέρωθεν του Ατλαντικού: στα 1878-’79 οι κάτοικοι του Νιού Χέιβεν και του Λονδίνου μπορούσαν να κάνουν χρήση -και είχαν να ευχαριστήσουν γι’ αυτό κυρίως τους Alexander Graham Bell και Gardiner Greene Habbard. Στις επόμενες δεκαετίες, η χρήση του τηλεφώνου έμελλε να ενώσει όλη την υφήλιο.

Ήδη, πάντως, από τα τέλη του 19ου αιώνα προετοιμαζόταν μία άλλη επανάσταση: ο Ιταλός Guglielmo Marconi, μια... αλλοπαρμένη μεγαλοφυΐα που... αυθαδίασε απέναντι σε επιστήμονες εγνωσμένου κύρους της εποχής, πατώντας στα χνάρια του Γερμανού Heinrich Hertz, δάμαζε τα ραδιοκύματα με μια σειρά πειραμάτων που θα άλλαζαν πολλά. Η επιτυχής προσπάθειά του, το 1901, να στήσει ραδιοεπικοινωνιακή γέφυρα ανάμεσα στον Καναδά και τη Βρετανία, κατέρριψε την επικρατούσα άποψη της εποχής -που υποστηριζόταν και από τον σπουδαίο κατά τα άλλα Henri Poincaré- ότι κάτι τέτοιο ήταν αδύνατο· και του χάρισε το βραβείο Νομπέλ Φυσικής [4].

Ο αντίκτυπος της συγκεκριμένης εξέλιξης υπήρξε, όπως είναι γνωστό, τεράστιος. Από το ραδιόφωνο και την τηλεόραση, έως τις δορυφορικές επικοινωνίες και τα δίκτυα κινητών επικοινωνιών, οι εφαρμογές υπήρξαν πολλές, και καταλυτικές για την εξέλιξη των ανθρώπινων κοινωνιών.

Κεφάλαιο 2

Λίγη περισσότερη Ιστορία: Αναδρομή στην εξέλιξη των Δικτύων Κινητών Επικοινωνιών

Παρότι υπάρχουν επιμέρους διαφωνίες για το ποια ακριβώς είναι τα όριά τους, είναι κοινώς αποδεκτό ότι μέχρι σήμερα έχουν υπάρξει πέντε διαφορετικές γενιές δικτύων κινητών επικοινωνιών -και αντίστοιχες γενιές προτύπων που τις διέπουν. Κάθε γενιά συνήθως αναφέρεται με τη συντομογραφία XG, όπου X ο αύξων αριθμός της γενιάς. Στις ασύρματες επικοινωνίες που προηγήθηκαν της πρώτης γενιάς αποδίδεται συχνά στη βιβλιογραφία η συντομογραφία 0G.

Καθώς κάθε επόμενη γενιά, ανάλογα με την περιοχή, συνυπήρχε σε κάποιον βαθμό με τις προηγούμενες, σε γενικές γραμμές τα νεότερα πρωτόκολλα και οι νεότερες συσκευές διέθεταν την απαραίτητη προς-τα-πίσω συμβατότητα.

2.1 Η... μηδενική γενιά (0G)

Τα πρώτα συστήματα κινητής τηλεφωνίας που είχαν εμπορική χρήση εμφανίστηκαν, διόλου τυχαία, μετά τη λήξη του Β΄ Παγκοσμίου Πολέμου. Ονομάστηκαν εκ των υστέρων *0G*, δηλαδή μηδενικής γενιάς, ακολουθώντας, προς τα πίσω, τον τρόπο ονοματοθεσίας των μετέπειτα γενεών συστημάτων.

Κινητά, ασύρματα συστήματα συζητούνταν ως ενδεχόμενο και ως σχεδιασμός ήδη από δεκαετίες, όταν το 1946 εμφανίστηκε το πρώτο τέτοιο εμπορικό προϊόν. Συνέβη στις Η.Π.Α., στο Σεντ Λούις της Πολιτείας Μιζούρι, ενώ το ξεκίνημα για το Ηνωμένο Βασίλειο έγινε αρκετά αργότερα, το 1958, με το “System1”. Επρόκειτο για εφαρμογές VHF ραδιοεπικοινωνίας, αναλογικές φυσικά, που δεν είχαν καμία σχέση με την κυψελωτή ψηφιακή τεχνολογία των μελλοντικών συστημάτων.

Τα εν λόγω συστήματα πρόσφεραν αποκλειστικά υπηρεσίες φωνής. Τα διαθέσιμα κανάλια ήταν εξαιρετικά περιορισμένα, και η ποιότητα φτωχή. Υπήρχαν επίσης πολλά ζητήματα ιδιωτικότητας και παρεμβολών. Αρχικά, υπήρχε η δυνατότητα μονόπλευρης ομιλίας (Push to Talk, PTT), όμως αργότερα, σε πρότυπα όπως τα Mobile Telephone Service (MTS) και Improved MTS (IMTAS) της Bell Systems στην Αμερική, και Advanced MTS (AMTS) στην Ιαπωνία, ενσωματώθηκε η δυνατότητα για full-duplex επικοινωνία.

Σχήμα 2.1 Ο Reginald Blevins, Γενικός Διευθυντής της βρετανικής ταχυδρομικής υπηρεσίας, εγκαινιάζει το πρώτο ραδιοτηλέφωνο για οχήματα (1959) [29].

Η φορητότητα των συσκευών των εν λόγω συστημάτων ήταν ιδιαίτερα περιορισμένη, λόγω του όγκου και του βάρους τους, αλλά και λόγω της φτωχής απόδοσης των μπαταριών. Είναι χαρακτηριστικό το γεγονός ότι αρχικά σχεδιάζονταν για εγκατάσταση αποκλειστικά σε οχήματα, και μόνο αργότερα κυκλοφόρησαν και σε μορφή φορητής βαλίτσας.

2.2 Τα πρώτα κυψελωτά συστήματα (1G)

Τα δίκτυα πρώτης γενιάς ήταν αναλογικά. Η πρώτη εμπορική εφαρμογή τους έγινε το 1979, στην Ιαπωνία, ενώ στη δεκαετία του ’80 επεκτάθηκαν στις

Η.Π.Α. Οι υπηρεσίες που προσφέρονταν ήταν αποκλειστικά φωνητικών κλήσεων. Άλλοι περιορισμοί τους είχαν να κάνουν με τη χρήση ραδιοσυχνοτήτων, οπότε υπήρχαν ζητήματα ασφάλειας/ιδιωτικότητας, θορύβου, και παρεμβολής από άλλα σήματα.

Το 1983, στις Η.Π.Α., τέθηκε σε εφαρμογή ένα πρότυπο συστήματος αναλογικής κινητής τηλεφωνίας, ανεπτυγμένο από την Bell Labs σε συνεργασία με τη Motorola, που ονομάστηκε Advanced Mobile Phone System (AMPS). Η χρήση του επεκτάθηκε και σε άλλες χώρες της Αμερικής, της Ασίας και της Ωκεανίας, ενώ παρέμεινε κυρίαρχο στη Βόρεια Αμερική για περισσότερα από είκοσι χρόνια.

Η εν λόγω τεχνολογία χρησιμοποιούσε ξεχωριστή συχνότητα (κανάλι) για κάθε κλήση, κάτι που απαιτούσε μεγάλο εύρος ζώνης συχνοτήτων. Στην ουσία, το AMPS έμοιαζε πολύ με τα συστήματα της 0G γενιάς, αλλά είχε το πλεονέκτημα της επαναχρησιμοποίησης των συχνοτήτων, στην περίπτωση που υπήρχε ικανή απόσταση ανάμεσα στις θέσεις. Εδώ χρησιμοποιήθηκε για πρώτη φορά η έννοια της «κυψέλης», καθώς η προς κάλυψη περιοχή χωριζόταν νοητά σε εξαγωνικά τμήματα.

Ένα άλλο πρότυπο, ευρωπαϊκής προέλευσης, είναι το Nordic Mobile Telephony (NMT). Εφαρμόστηκε το 1981 στη Σουηδία και στη Νορβηγία, και γρήγορα επεκτάθηκε σε άλλες χώρες της Σκανδιναβικής Χερσονήσου, της Ευρώπης και της Ασίας.

Άλλα πρότυπα της γενιάς αυτής ήταν τα TACS (Total Access Communications System) στο Ηνωμένο Βασίλειο, C-450 στη Δυτική Γερμανία, Radiocom 2000 στη Γαλλία, RTMI στην Ιταλία, και TZ-801, TZ-802, TZ-803 και JTACS (Japan Total Access Communications System) στην Ιαπωνία.

2.3 Το πέρασμα στον ψηφιακό κόσμο (2G)

Η δεύτερη γενιά προτύπων για τις κινητές επικοινωνίες εισήχθη σε εμπορική χρήση το 1991, στη Φινλανδία· ήταν τότε που αποδόθηκε η ονομασία 1G στα συστήματα που είχαν προηγηθεί. Πλέον τα ραδιοσήματα ήταν ψηφιακά, ενώ στα 1G πρότυπα ήταν αναλογικά, και μόνο η επικοινωνία μεταξύ των σταθμών βάσης και των χρηστών γινόταν με ψηφιακό τρόπο.

Οι βελτιώσεις της 2G γενιάς αφορούσαν στην ασφάλεια, με χρήση ψηφιακής κρυπτογράφησης (κυρίως στην επαφή χρήση και σταθμού βάσης), στην πιο αποδοτική χρήση του διαθέσιμου συχνοτικού φάσματος, που επέτρεπε περισσότερους χρήστες ανά ζώνη συχνοτήτων, και στη δυνατότητα ανταλλαγής μηνυμάτων, γραπτών (Short Message/Messaging Service, SMS) αλλά και πολυμεσικών (Multimedia Messaging Service, MMS).
Το πιο γνωστό πρότυπο της 2G οικογένειας είναι το Global System for Mobile Communications ή GSM, το οποίο χρησιμοποιούνταν στο μεγαλύτερο μέρος της υφηλίου, πλην της Ιαπωνίας. Στη Βόρεια Αμερική περισσότερο από το GSM χρησιμοποιούνταν τα Digital AMPS και cdmaOne, ενώ στην Ιαπωνία επικρατούσαν πρωτίστως το Personal Digital Cellular (PDC), και δευτερευόντως το Personal Handy-phone System (PHS).

Εξέλιξη της 2G γενιάς αποτέλεσε η 2.5G, που αφορούσε στο πρότυπο General Packet Radio Service (GPRS). Εδώ χρησιμοποιούνταν μεταγωγή πακέτου, παράλληλα με τη μεταγωγή κυκλώματος, με αποτέλεσμα την αύξηση της ταχύτητας σε αρκετές περιπτώσεις. Τα δίκτυα GPRS εξελίχθηκαν σε δίκτυα 2.75G ή EDGE (Enhanced Data rates for GSM Evolution), με αύξηση του ρυθμού μετάδοσης δεδομένων, και στη συνέχεια σε 2.85G ή EDGE Evolution. Πρόκειται για τεχνολογίες με προς τα πίσω συμβατότητα, που βελτίωσαν διάφορες πλευρές του GSM.

Τα δίκτυα GSM άρχισαν να καταργούνται σε διάφορα σημεία του κόσμου, κατά τα τέλη της δεκαετίας του 2010. Παρ’ όλα αυτά, εξακολουθούν να χρησιμοποιούνται σε κάποιες περιοχές, ενώ η ευρεία και εκτεταμένη χρήση

του εν λόγω προτύπου οδήγησε στο να χρησιμοποιείται ακόμα η ονομασία του προκειμένου να αναφερθεί κανείς στην οικογένεια προτύπων που προέκυψαν ως εξελίξεις αυτού.

2.4 Διαδίκτυο εν κινήσει (3G)

Τα συστήματα τρίτης γενιάς υλοποιήθηκαν για πρώτη φορά το 1998, στην Ιαπωνία, από την NTT, αλλά η πρώτη εμπορική χρήση τους έγινε τρία χρόνια αργότερα, στη Νότια Κορέα, από την SK Telecom.

Τα πρότυπα της συγκεκριμένης γενιάς δικτύων χρησιμοποιούσαν τη μέθοδο CDMA (Code-Division Multiple Access) και μεταγωγή πακέτου, στις συχνότητες 850, 1900 και 2100 MHz.
Το πλέον κυρίαρχο από τα πρότυπα της εν λόγω γενιάς ήταν το WCDMA (Wideband CDMA), το οποίο ήταν βασισμένο στο GSM. Πρόκειται για παραλλαγή του UMTS (Universal Mobile Telecommunications System), το οποίο αναπτύχθηκε από την 3GPP (3rd Generation Partnership Project).

Τα εν λόγω πρότυπα πρόσφεραν μέγιστες ταχύτητες της τάξης των 2 Mbps, που όμως ίσχυαν μόνο για ακίνητους χρήστες· στην πράξη η μέγιστη ταχύτητα ήταν 384 Kbps, η οποία και πάλι ήταν αισθητά βελτιωμένη σε σχέση με τα πρότυπα δεύτερης γενιάς. Πλέον υπήρχε η δυνατότητα να υποστηριχθούν βιντεοκλήσεις, mobile internet και streaming περιεχομένου, ενώ η χρέωση των χρηστών μπορούσε πλέον να γίνεται με βάση τον όγκο δεδομένων και όχι με βάση τον χρόνο χρήσης.

Ένα εξελιγμένο πρότυπο υπήρξε το HSPA (High Speed Packet Access) που σηματοδότησε την 3.5G γενιά, με θεωρητικές μέγιστες ταχύτητες 5.76 Mbps στο uplink και 14.4 Mbps στο downlink, και πραγματικές 2000 Kbps και 500 Kbps αντίστοιχα. Ακολούθησε το HSPA+ (3.75G), το οποίο πρόσφερε τη δυνατότητα MIMO (Multiple Inputs Multiple Outputs), δηλαδή χρήση πολλαπλών κεραιών για εκπομπή και λήψη, με αποτέλεσμα οι θεωρητικές

μέγιστες ταχύτητες μετάδοσης να διαμορφωθούν σε 22 Mbps στο uplink και 168 Mbps στο downlink.

2.5 Πιο γρήγορα, πιο ψηλά, πιο δυνατά (4G)

Τα δίκτυα τέταρτης γενιάς προτυποποιήθηκαν το 2008, από την ITU (International Telecommunication Union), με την έκδοση της IMT Advanced (International Mobile Communications-Advanced). Σε επικαιροποίηση, το 2010, της εν λόγω προτυποποίησης συμπεριλήφθηκαν πρότυπα όπως το LTE (Long Term Evolution) και το WiMAX (Worldwide Interoperability for Microwave Access).

Η πρώτη εμπορικά διαθέσιμη υλοποίηση LTE δικτύου έγινε το 2009, σε Νορβηγία και Σουηδία, στο Όσλο και στη Στοκχόλμη αντίστοιχα, για να εξαπλωθεί στη συνέχεια στο μεγαλύτερο μέρος της υφηλίου.

Οι κυριότερες τεχνολογίες που χρησιμοποιούνται από όλα τα πρότυπα τέταρτης γενιάς είναι η MIMO (Multiple Inputs Multiple Outputs), η OFDM (Orthogonal Frequency-Division Multiplexing) και η OFDMA (Orthogonal Frequency-Division Multiple Access). Χρησιμοποιούνται επίσης πολλαπλές ζώνες συχνοτήτων: 700, 800, 900, 1700, 1800, 1900, 2100, και 2600 MHz.

Πλεονεκτήματα των προτύπων 4G γενιάς είναι οι ιδιαίτερα υψηλές ταχύτητες και το χαμηλό latency. Για το LTE έχουμε 150 Mbps στο downlink και 50 Mbps στο uplink, και για το LTEA (LTE Advanced) 1 Gbps και 500 Mbps, αντίστοιχα.

Τα πρότυπα της εν λόγω γενιάς μπορούν να προσφέρουν πληθώρα εφαρμογών, όπως τηλεφωνία VoIP (Voice over Internet Protocol), τηλεόραση 3D, βιντεοκλήσεις με πολλούς συμμετέχοντες, υπηρεσίες νέφους, mobile web και mobile gaming.

Παρότι ήδη σε κάποια σημεία του κόσμου τα δίκτυα 4G έχουν αρχίσει να αντικαθίστανται από δίκτυα της επόμενης γενιάς, παραμένουν κραταιά, και είναι βέβαιο ότι θα παραμείνουν μέχρι το τέλος της τρέχουσας δεκαετίας.

2.6 Η επανάσταση του τώρα (5G)

Η 3GPP κατηγοριοποιεί ως δίκτυο πέμπτης γενιάς οποιοδήποτε σύστημα χρησιμοποιεί την τεχνολογία-πρότυπο 5G New Radio (5G NR). Πρόκειται για μία τεχνολογία που σχεδιάστηκε ως το παγκοσμίως συμφωνημένο πρότυπο, και η οποία βασίστηκε στην OFDMA -αποτελεί δηλαδή εξέλιξη του LTE. Οι πρώτες εργασίες ξεκίνησαν το 2015, ενώ η πρώτη εμπορική εφαρμογή πραγματοποιήθηκε στα τέλη του 2018.

Η τεχνολογία πέμπτης γενιάς χρησιμοποιεί κυμαινόμενο εύρος ζωνών συχνοτήτων, αλλά και κυμαινόμενο μέγεθος κυψέλης, ανάλογα με την εφαρμογή. Έτσι, υπάρχουν και διαφορετικά πρότυπα για την κάθε περίπτωση: enhanced Mobile Broadband (eMBB), massive Machine Type Communication (mMTC), Ultra Reliable & Low Latency Communication (URLLC).

Χρησιμοποιούνται επίσης τεχνολογίες όπως, μεταξύ άλλων, massive MIMO (μέχρι και 64x64 κεραίες), Edge computing (αρχιτεκτονική κατανεμημένων υπολογιστικών συστημάτων, όπου χρησιμοποιούνται servers που βρίσκονται πιο κοντά στην πηγή των προς επεξεργασία δεδομένων), και Beamforming (όπου τα ψηφιακά δεδομένα στέλνονται σε πολλαπλές ροές, ενώ η κατευθυντικότητα συστοιχιών κεραιών προσαρμόζεται για βέλτιστη εκπομπή και λήψη).

Αποτέλεσμα όλων των παραπάνω, αλλά και των πολλών ακόμα βελτιώσεων της 5G γενιάς σε σχέση με τις προηγούμενες, είναι η παροχή στους χρήστες ιδιαίτερα αυξημένων ταχυτήτων, σε συνδυασμό με πολύ χαμηλή χρονική υστέρηση. Αλλά και οι εφαρμογές πληθαίνουν, με εξελίξεις που αφορούν

στους τομείς του Internet of Things (IoT), και των επικοινωνιών μηχανής-με-μηχανή.

Κεφάλαιο 3

Ο ήχος ως κύμα

3.1 Ο ήχος και οι διαφορετικές πλευρές του

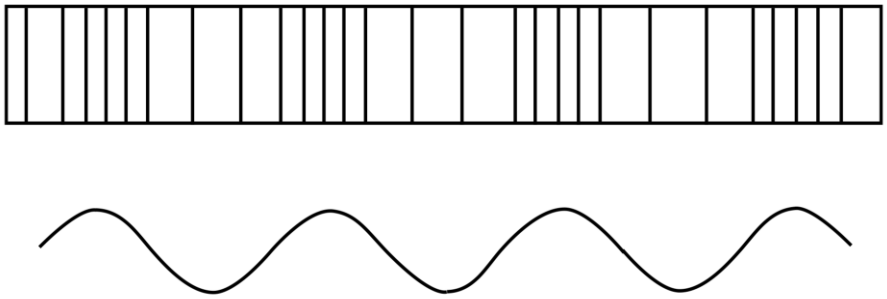
Σε ό,τι αφορά στην καθημερινότητα, ο όρος *ήχος* αναφέρεται στην ερμηνεία που δίνει ο ανθρώπινος εγκέφαλος στο φυσικό ερέθισμα που φτάνει στα αφτιά του. Από αυτήν την άποψη, η αντίληψη του ήχου εμπεριέχει μεγάλο βαθμό υποκειμενικότητας. Όσα σχετίζονται με αυτήν την πλευρά του φαινομένου μελετώνται από την επιστημονικό κλάδο της Ψυχοακουστικής.

Όμως, υπάρχει και μια άλλη πλευρά του ήχου, που μετράται αντικειμενικά, και μελετάται από τις επιστήμες της Ακουστικής και της Ηλεκτροακουστικής, οι οποίες ουσιαστικά αποτελούν παρακλάδια της Φυσικής.

Ως φυσικό, λοιπόν, φαινόμενο, ο ήχος περιγράφεται ως ακουστική ενέργεια, η οποία διαδίδεται με τη μορφή διαμήκους μηχανικού κύματος σε ένα ελαστικό μέσο. Για να παραχθεί ήχος θα πρέπει να υπάρχει ένα δονούμενο σώμα, το οποίο θα οδηγήσει σε ταλάντωση τα γύρω του μόρια του αέρα, και αυτά με τη σειρά τους θα μεταδώσουν την ταλάντωση στα διπλανά τους κ.ο.κ. Το αποτέλεσμα είναι η δημιουργία στον χώρο διαδοχικών πυκνώσεων και αραιώσεων της ύλης. Μιλάμε τότε για *ηχητικό κύμα*, το οποίο προκαλείται από μία *ηχητική πηγή*. Το πιο συνηθισμένο *μέσο διάδοσης* είναι ο αέρας, όμως ένα ηχητικό κύμα μπορεί να ταξιδέψει μέσα σε οποιοδήποτε μέσο· αέριο, υγρό ή στερεό. Στην ουσία, όλος ο υλικός χώρος μπορούμε να πούμε ότι χορεύει αέναα, δονούμενος από χίλια μύρια ηχητικά κύματα.

Όταν το ηχητικό κύμα συναντήσει έναν δέκτη, π.χ. το ανθρώπινο αφτί, οι μεταβολές της πίεσης του αέρα ανιχνεύονται από τους βιολογικούς

μηχανισμούς της ακοής, και μεταφέρονται στον εγκέφαλο ως σήματα του



Σχήμα 3.1 Ημιτονικό κύμα (κάτω), και οι αντίστοιχες πυκνώσεις και αραιώσεις στο υλικό μέσο (πάνω) [9].

νευρικού συστήματος. Εκεί ερμηνεύονται -ως ομιλία, ως μουσική, ως θόρυβος κλπ.

3.2 Βασικά χαρακτηριστικά του ήχου

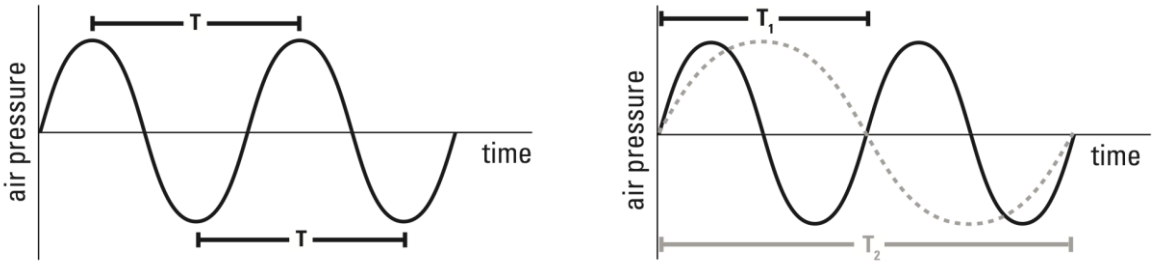
Για να προσδιορίσουμε πλήρως ένα ηχητικό κύμα, χρειάζεται να γνωρίζουμε κάποια βασικά χαρακτηριστικά του: συχνότητα, συχνοτικό φάσμα, ένταση, διάρκεια.

3.2.1 Συχνότητα και συχνοτικό φάσμα

Η *συχνότητα* είναι το μέγεθος που μετρά τον αριθμό των πλήρων ταλαντώσεων των σωματιδίων του υλικού μέσου, ανά δευτερόλεπτο. Χρησιμοποιώντας την καμπύλη του ημιτόνου ως απεικόνιση της απλούστερης ταλάντωσης που μπορεί να πραγματοποιήσει ένα μόριο, και χρησιμοποιώντας το σύμβολο f για τη συχνότητα, ισχύει η σχέση:

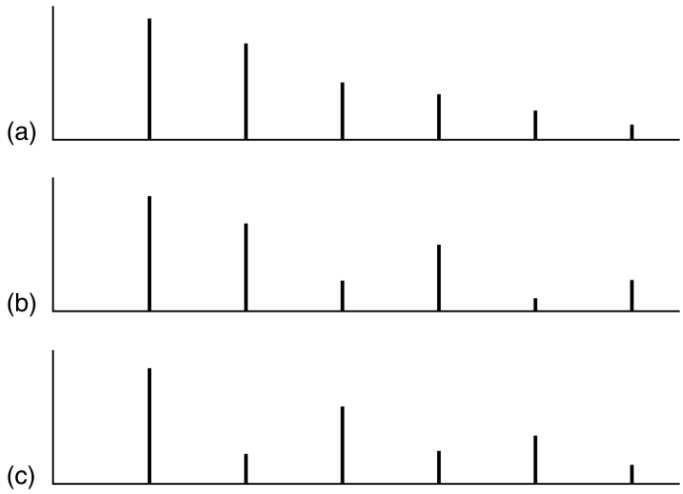
$$f = \frac{1}{T} \quad [3.1]$$

όπου T η περίοδος, δηλαδή ο χρόνος που χρειάζεται το μόριο για να επιστρέψει στη θέση ισορροπίας του έπειτα από μία πλήρη ταλάντωση.



Σχήμα 3.2 Αριστερά, ημιτονική κυματομορφή και η περίοδός της, T. Δεξιά, δύο ημιτονικές κυματομορφές, ίδιου πλάτους, αλλά διαφορετικής περιόδου ($T_2 = 2T_1$) [11].

Παρότι υπάρχουν ήχοι που προσεγγίζονται από το μοντέλο της ημιτονοειδούς κυματομορφής (π.χ. εκείνος του διαπασών ή οι υψηλές νότες ενός φλάουτου), οι περισσότεροι ήχοι, φυσικοί και τεχνητοί, δεν είναι απλοί αλλά σύνθετοι· αποτελούνται, δηλαδή, από πολλές συχνότητες. Από αυτές, εκείνη που κυριαρχεί ονομάζεται *βασική ή θεμελιώδης* (fundamental), και οι υπόλοιπες αποκαλούνται *μερικές* ή *παράγωγες* (partials). Η θεμελιώδης



Σχήμα 3.3 Φασματικό περιεχόμενο τριών κυμάτων, τα οποία έχουν ίδια ισχύ θεμελιώδους συχνότητας, αλλά διαφορετικής ισχύος παράγωγες [12].

συχνότητα είναι εκείνη που καθορίζει το *τονικό ύψος* του ήχου, ενώ το πλήθος και η διάταξη των παραγώγων αποτελούν το *φάσμα συχνοτήτων* του,

και καθορίζουν τη *χροιά* ή *ηχόχρωμα* αυτού. Με άλλα λόγια, θα διακρίνουμε δύο ήχους που έχουν την ίδια θεμελιώδη συχνότητα από το διαφορετικό συχνοτικό φάσμα τους.

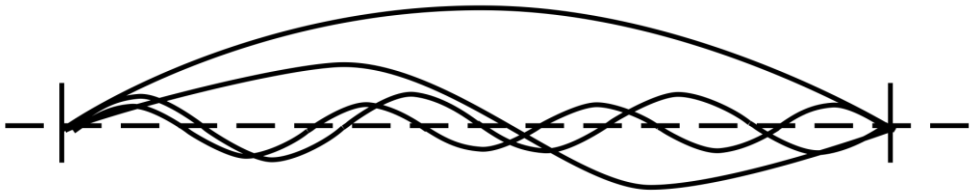
Στο φάσμα συχνοτήτων ενός συγκεκριμένου ήχου, η θεμελιώδης συχνότητα είναι η μικρότερη, και έχει συνήθως τη μεγαλύτερη ένταση από όλες τις συχνότητες του φάσματος -αν και υπάρχουν ήχοι των οποίων η θεμελιώδης υπολείπεται σε ένταση. Η σχετική ισχύς των αρμονικών, καθώς και η συνολική ισχύς του ήχου, η οποία επηρεάζεται από το πόσο έντονα δονείται η ηχητική πηγή, είναι παράγοντες που επηρεάζουν την ποιότητά του και το συχνοτικό του φάσμα.

Το φαινόμενο της ύπαρξης παράγωγων συχνοτήτων μπορεί να εξηγηθεί από τη μελέτη των ταλαντώσεων μίας τεντωμένης χορδής. Όταν αυτή διεγερθεί και αρχίσει να πάλλεται, η ταλάντωσή της θα γίνει ταυτόχρονα και ως προς το συνολικό της μήκος, που αντιστοιχεί στη θεμελιώδη συχνότητα, αλλά και ως προς το 1/2 αυτού, το 1/3 κ.ο.κ. Τα ακέραια πολλαπλάσια της θεμελιώδους ονομάζονται *αρμονικές συχνότητες* (harmonics). Συνήθως, η θεμελιώδης ονομάζεται 1η αρμονική, και τα ακέραια πολλαπλάσιά της αποκαλούνται, κατ' αναλογία, 2η αρμονική (διπλάσια συχνότητα της θεμελιώδους), 3η αρμονική (τριπλάσια) κ.ο.κ. Συχνότητες στο φάσμα ενός ήχου, οι οποίες δεν αποτελούν ακέραιο πολλαπλάσιο της θεμελιώδους, ονομάζονται *μη αρμονικές* (inharmonics). Υπάρχουν, τέλος, και οι αποκαλούμενες *υποαρμονικές συχνότητες* (subharmonics), οι οποίες είναι ακέραια υποπολλαπλάσια της θεμελιώδους, και έχουν να κάνουν κυρίως με τους ήχους των μουσικών οργάνων, και όχι με τους φυσικούς ήχους.

Αντί του όρου «φάσμα συχνοτήτων», λόγω και όσων αναφέρθηκαν παραπάνω, χρησιμοποιούνται ισοδύναμα οι όροι «αρμονικό περιεχόμενο», «αρμονική δομή» και «αρμονικό φάσμα».

Η διαδικασία της ανάλυσης ενός ήχου σε αρμονικές συχνότητες, και η σχηματική αναπαράσταση αυτής (όπως στο Σχήμα 3.3), ονομάζεται *αρμονική*

ανάλυση στο πεδίο της συχνότητας ή *ανάλυση Fourier*[1]. Σημειώνεται ότι η



Σχήμα 3.4 Η θεμελιώδης και οι αρμονικές συχνότητες ταλάντωσης μίας χορδής [9].

απεικόνιση του Σχήματος 3.3 είναι απλουστευμένη: στην πραγματικότητα κάθε κατακόρυφη γραμμή αντιπροσωπεύει μία ευρύτερη περιοχή συχνοτήτων. Η ευρύτητα αυτής σχετίζεται με την «καθαρότητα» της αρμονικής συχνότητας: όσο πιο στενή είναι η περιοχή, τόσο πιο καθαρή, και τείνουσα προς μία μοναδική συχνότητα, είναι η αρμονική.

Το ανθρώπινο αφτί μπορεί να ανιχνεύσει συχνότητες περίπου από 16 Hz, και μέχρι 20 kHz. Άτομα νεαρής ηλικίας ενδέχεται να έχουν ακόμα μεγαλύτερο εύρος, ξεκινώντας από τα 12 Hz και φτάνοντας έως τα 25 kHz. Με την πάροδο του χρόνου, οι άνθρωποι χάνουν σημαντικό μέρος της ακουστικής ευαισθησίας τους: το άνω όριο πέφτει έως τα 12 ή και τα 8 kHz. Ονομάζουμε τους ήχους που βρίσκονται έξω από το φάσμα της ανθρώπινης ακοής *υπόηχους* (<16 Hz) και *υπέρηχους* (>20 kHz).

Η ανθρώπινη ακοή κατανέμει τους ήχους με βάση τη συχνότητά τους κατά τρόπο σχετικιστικό, κατατάσσοντάς τους ως πιο πρίμους (ή οξείς), σε σχέση με άλλους, που αποκαλούνται μπάσοι. Έτσι, οι διάφοροι ήχοι μπορούν να διαταχθούν σε μία κλίμακα συχνοτήτων, όπως οι νότες ενός πιάνου διατάσσονται στο κλαβιέ, με τους πιο χαμηλούς σε συχνότητα να βρίσκονται στο αριστερό άκρο του, και τους πιο υψίσυχνους στο δεξί.

[1] Προς τιμή του Γάλλου μαθηματικού Jean-Baptiste-Joseph Fourier (1768-1830).

3.2.2 Ένταση

Όταν μία ηχητική πηγή δονείται, προκαλώντας τη δημιουργία ηχητικού κύματος, η ταλάντωσή της εξασθενεί με την πάροδο του χρόνου, ώσπου τελικά σβήνει. Μέρος της ενέργειας που προσφέρεται στην πηγή, ώστε να την κάνει να ταλαντωθεί, μετατρέπεται σε ακουστική ενέργεια, η οποία μεταφέρεται μέσω του ηχητικού κύματος στον περιβάλλοντα την πηγή χώρο. Η ακουστική ενέργεια που μεταφέρεται ανά δευτερόλεπτο ονομάζεται *ακουστική ισχύς*, και μετράται σε Watt (W).

Ονομάζουμε *ένταση του ήχου*, σε κάποιο σημείο του χώρου και για δεδομένη κατεύθυνση διάδοσής του, την ακουστική ισχύ που διαπερνά μία επιφάνεια ενός τετραγωνικού μέτρου, η οποία είναι κάθετη στην κατεύθυνση διάδοσης. Μονάδα μέτρησης της έντασης είναι τα Watts ανά m^2 (W/m^2). Η ένταση μειώνεται καθώς αυξάνεται η απόσταση του σημείου όπου πραγματοποιούμε τη μέτρηση από την πηγή, με ρυθμό ανάλογο του τετραγώνου της απόστασης.

Καθώς στην πράξη είναι δύσκολο να μετρηθεί η ένταση του ήχου, έχει οριστεί η *ακουστική πίεση*, ένα μέγεθος το οποίο περιγράφει τις διακυμάνσεις της πίεσης του αέρα γύρω από ένα σημείο ηρεμίας. Εδώ η μονάδα μέτρησης είναι τα Newton ανά m^2 (N/m^2) ή τα dyne ανά cm^2. Τα δύο μεγέθη, η ένταση I και η rms τιμή της ακουστικής πίεσης P συνδέονται μέσω της σχέσης:

$$I = \frac{P^2}{\rho \cdot \upsilon} \quad [3.2]$$

όπου ρ η πυκνότητα του μέσου στο οποίο διαδίδεται το ηχητικό κύμα, και υ η ταχύτητα του κύματος στο συγκεκριμένο μέσο. Η εν λόγω σχέση δηλώνει ότι η ακουστική πίεση του κύματος μειώνεται όσο αυτό απομακρύνεται από την πηγή, με ρυθμό ανάλογο της απόστασης.

Λόγω του ότι η μέτρηση της έντασης και της ακουστικής πίεσης, κατά τον απόλυτο τρόπο που προηγουμένως περιγράφηκε, δεν είναι εύκολη στην πράξη, έχει αναπτυχθεί μία διαφορετική οπτική και μέθοδος. Συγκεκριμένα, και έπειτα από πειράματα, διαπιστώθηκε ότι ένας απλός τόνος συχνότητας 1,000 Hz γίνεται μόλις αντιληπτός όταν η έντασή του κυμαίνεται γύρω από την τιμή 2.5×10^{-12} W/m^2. Έτσι, επιλέχθηκε η τιμή των 10^{-12} W/m^2 ως τιμή αναφοράς, και ονομάστηκε *ένταση κατωφλίου*. Ορίστηκε, στη συνέχεια, η *στάθμη έντασης* (intensity level) *L*, για την οποία ισχύει η σχέση:

$$L = 10 \cdot \log\left(\frac{I}{I_0}\right) \quad [3.3]$$

όπου *I* η προς μέτρηση ένταση ήχου, και I_0 η ένταση κατωφλίου.

Η χρήση της λογαριθμικής συνάρτησης στην παραπάνω σχέση δηλώνει ότι πια περνάμε σε μετρήσεις που γίνονται με σχετικό τρόπο, συγκρίνοντας με την τιμή αναφοράς. Η μονάδα μέτρησης είναι πλέον το Decibel (dB), και τα 0 dB αντιστοιχούν στην κατώτατη τιμή έντασης, δηλαδή στα 10^{-12} W/m^2. Μία στάθμη 10 dB δηλώνει δεκαπλάσια ένταση, τα 100 dB σημαίνουν εκατονταπλάσια ένταση κ.ο.κ. Έχει διαπιστωθεί ότι το όριο ανεκτικότητας για την ανθρώπινη ακοή βρίσκεται στα 120 dB.

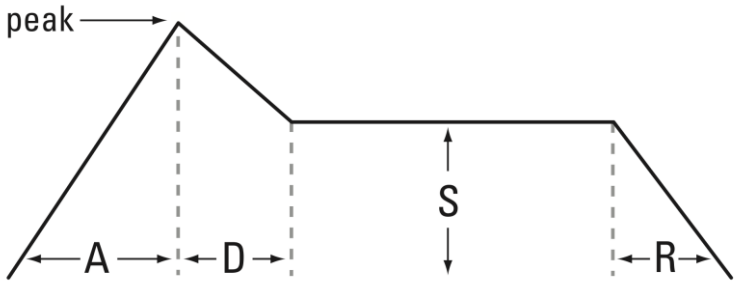
Αντίστοιχα, ορίζεται και η *στάθμη ακουστικής πίεσης* (sound pressure level, *SPL*), για την οποία ισχύει:

$$SPL = 20 \cdot \log\left(\frac{P}{P_0}\right) \quad [3.4]$$

όπου *P* η προς μέτρηση rms τιμή της ακουστικής πίεσης, και P_0 η rms τιμή της ακουστικής πίεσης που αντιστοιχεί στην ένταση κατωφλίου, και ισούται με 2×10^{-5} N/m^2 ή 2×10^{-4} dyne/cm^2.

3.2.3 Διάρκεια

Ο κάθε ήχος δεν είναι, ως γνωστόν, ένα στιγμιαίο φαινόμενο: η έντασή του μεταβάλλεται κατά το χρονικό διάστημα που αυτός διαρκεί. Ονομάζουμε *χρονική περιβάλλουσα* τη νοητή καμπύλη που απεικονίζει τις μεταβολές του πλάτους της κυματομορφής του εκάστοτε ήχου. Κατ' ουσία, η περιβάλλουσα αποτελεί μία αναπαράσταση της δυναμικής ανάπτυξης της έντασης σε συνάρτηση με τον χρόνο.



Σχήμα 3.5 Χρονική περιβάλλουσα (ADSR), χαρακτηριστική των έγχορδων και των πνευστών μουσικών οργάνων. Χωρίζεται σε τέσσερα τμήματα: μέτωπο ή ατάκα (Attack), εξασθένηση ή πτώση (Decay), διάρκεια (Sustain), αποδέσμευση (Release) [11].

Κατά τη διάρκεια της εξέλιξης ενός ήχου, το αρμονικό περιεχόμενό του δεν μεταβάλλεται ομοιόμορφα. Οι διάφορες αρμονικές ακολουθούν η καθεμιά τη δική της πορεία μεταβολής· τη δική της περιβάλλουσα. Το γεγονός αυτό αποτελεί ένα από τα πλέον ουσιώδη χαρακτηριστικά ενός ήχου, συμπεριλαμβανομένης της ανθρώπινης φωνής.

3.3 Η διάδοση του ήχου

3.3.1 Η ταχύτητα του ήχου

Η *ταχύτητα* διάδοσης των ηχητικών κυμάτων, δηλαδή το πόσο γρήγορα διαδίδονται οι ταλαντώσεις στο μέσο διάδοσης, εξαρτάται από τη θερμοκρασία και από το υλικό (αέριο, υγρό, στερεό). Ισχύει η σχέση:

$$\upsilon = \sqrt{\gamma \cdot P_0 / \rho} \quad [3.5]$$

όπου υ η ταχύτητα, γ ένας συντελεστής που εξαρτάται από τη θερμοκρασία και το υλικό μέσο, P_0 η στατική πίεση του αέρα, και ρ η πυκνότητά του.

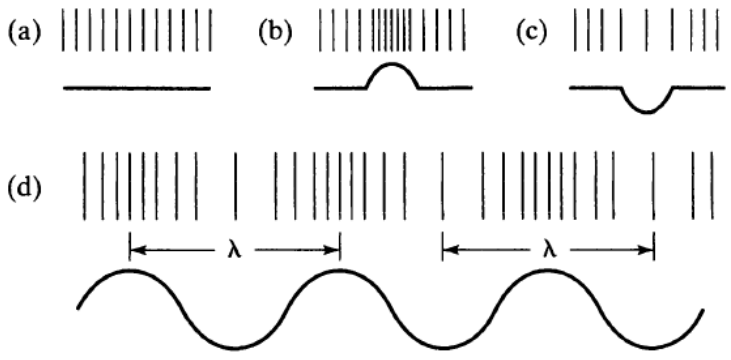
Υπό κανονικές συνθήκες (θερμοκρασία 22°C, πίεση 1 atm), η ταχύτητα του ήχου στον αέρα είναι 344 m/sec. Στα υγρά η ταχύτητά του είναι πολύ μεγαλύτερη (1,430 m/sec στο νερό), και στα στερεά ακόμα περισσότερο - αναμενόμενο, αφού σε αυτές τις καταστάσεις η ύλη είναι πολύ πυκνότερη.

Ενώ εξαρτάται, όπως προαναφέρθηκε, από τη θερμοκρασία και το υλικό, η ταχύτητα του ήχου δεν επηρεάζεται από τη συχνότητά του, ούτε από την πίεση, αφού, όταν αλλάζει η τελευταία, αλλάζει και η πυκνότητα, με αποτέλεσμα ο όρος P_0/ρ στην εξίσωση [3.5] να παραμένει σταθερός.

Ισχύει επίσης η σχέση:

$$\upsilon = \lambda \cdot f \quad [3.6]$$

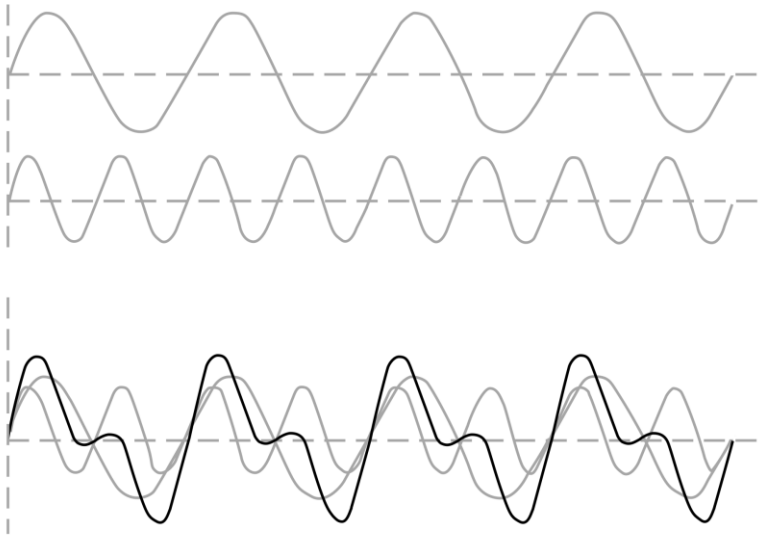
η οποία συνδέει τη συχνότητα f, με την ταχύτητα υ και το μήκος κύματος λ. Από αυτήν προκύπτει ότι μήκος κύματος και συχνότητα είναι αντιστρόφως ανάλογα μεγέθη.



Σχήμα 3.6 Εγκάρσια αναπαράσταση (a) χορδής σε ισορροπία, (b) πύκνωσης, (c) αραίωσης, και (d) σειράς πυκνώσεων και αραιώσεων [10].

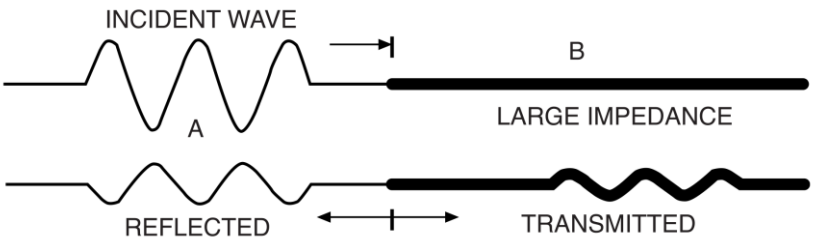
3.3.2 Φαινόμενα κατά τη διάδοση του ήχου

Κατά τη διάδοσή του, ένα ηχητικό κύμα μπορεί να συμβάλλει με άλλα κύματα που διαδίδονται στην ίδια περιοχή του μέσου. Η συμβολή δύο κυμάτων μπορεί να είναι είτε ενισχυτική (όταν η διαφορά φάσης τους είναι άρτιο πολλαπλάσιο του π) είτε αποσβεστική (όταν είναι περιττό πολλαπλάσιο του π). Επιπλέον, το κύμα συναντά αντικείμενα, τα οποία λειτουργούν ως εμπόδια. Από το μέγεθος και την υφή αυτών των αντικειμένων προκαλούνται μία σειρά φαινομένων, τα οποία αλλοιώνουν την αρχική μορφή του κύματος.



Σχήμα 3.7 Συμβολή δύο κυμάτων [9].

Υπάρχουν, κατ' αρχάς, η *ανάκλαση*, η *απορρόφηση* και η *διέλευση* του ήχου. Αυτά παρατηρούνται όταν το ηχητικό κύμα συναντήσει αντικείμενο/εμπόδιο με διαστάσεις μεγάλες σε σύγκριση με το μήκος κύματός του. Τότε, ένα μέρος του κύματος ανακλάται, δηλαδή επιστρέφει προς την κατεύθυνση από την οποία ερχόταν, ένα άλλο μέρος απορροφάται από το αντικείμενο (ανάλογα με το υλικό από το οποίο αυτό αποτελείται), και ένα τρίτο μέρος διαπερνά το αντικείμενο και συνεχίζει να διαδίδεται στην αρχική κατεύθυνση.



Σχήμα 3.8 Κύμα που διαδίδεται σε τεντωμένη χορδή, και συναντά τμήμα της με μεγαλύτερη αντίσταση διάδοσης, με αποτέλεσμα μέρος του να ανακλαστεί [9].

Παράγωγο της ανάκλασης είναι η *αντήχηση* ή *μετήχηση*. Εμφανίζεται κατά τη διάδοση σε κλειστό χώρο, όπου ο ήχος εξακολουθεί να υπάρχει μετά τη σίγαση της ηχητικής πηγής, μέσω των πολλαπλών ανακλάσεων στις επιφάνειες του χώρου. Αυτή η παράταση ύπαρξης του ήχου γίνεται αντιληπτή ως «βάθος».

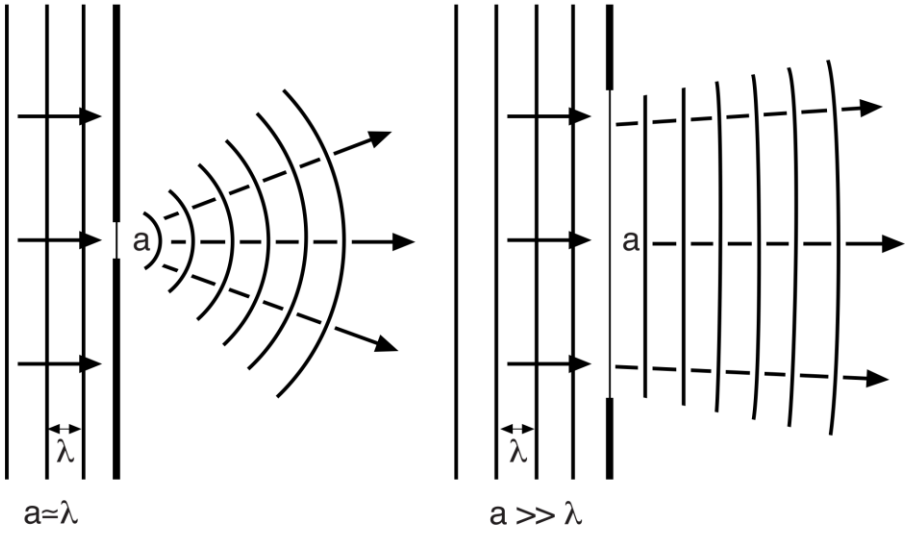
Η *ηχώ* σχετίζεται επίσης με την ανάκλαση, αλλά αυτή αφορά στην επανάληψη ενός ήχου, τη δημιουργία δηλαδή ακριβών αντιγράφων του αρχικού, αλλά μειωμένης έντασης σε σχέση με αυτόν. Διαφέρει από την αντήχηση, στο ότι δίνει την αίσθηση ενός ανεξάρτητου ήχου σε σχέση με τον αρχικό, και όχι της επιμήκυνσης αυτού.

Έπειτα, υπάρχει το φαινόμενο της *περίθλασης*. Αυτό παρατηρείται όταν οι διαστάσεις του εμποδίου είναι συγκρίσιμες ή μικρότερες από το μήκος κύματος του ήχου. Σε αυτήν την περίπτωση το κύμα περιθλάται, δηλαδή περνά γύρω ή πάνω από το αντικείμενο. Με αυτόν τον τρόπο, ο ήχος μπορεί να φτάσει σε σημεία που δεν είναι «ορατά» από τη θέση της πηγής του.

Όταν το ηχητικό κύμα περνά από ένα υλικό μέσο σε ένα άλλο ή όταν αλλάζουν οι συνθήκες που επικρατούν στο μέσο διάδοσης, τότε παρατηρείται το φαινόμενο της *διάθλασης*, της εκτροπής δηλαδή της τροχιάς του κύματος. Κάτι τέτοιο παρατηρείται όταν π.χ. ένα ηχητικό κύμα περάσει από ένα στρώμα θερμού αέρα σε στρώμα με χαμηλότερη θερμοκρασία.

Τα προαναφερθέντα φαινόμενα (ανάκλαση, απορρόφηση, διάθλαση, περίθλαση) εξαρτώνται από τη συχνότητα του κύματος, από τη γωνία

πρόσκρουσης στο εμπόδιο ή εισόδου στο μέσο διάδοσης, από την υφή και τη



Σχήμα 3.9 Περίθλαση κύματος που συναντά εμπόδιο με οπή, το μέγεθος της οποίας συγκρίνεται με το μήκος κύματός του (αριστερά), ή είναι πολύ μεγαλύτερο αυτού (δεξιά) [9].

δομή του εμποδίου ή του μέσου διάδοσης· αλλά όχι από την ένταση του ήχου.

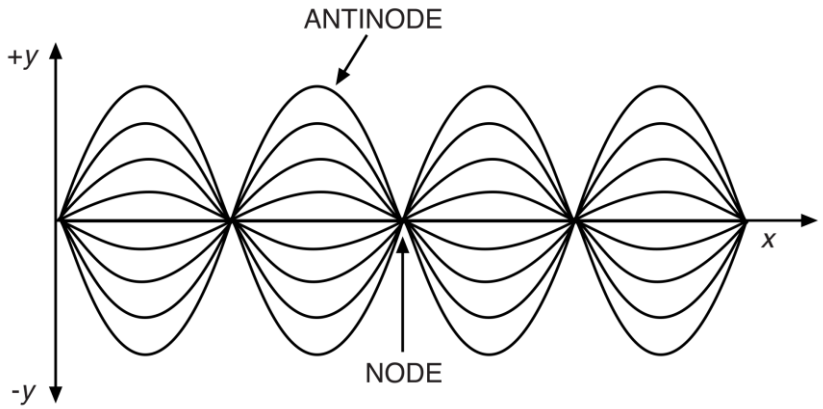
Πέρα από τα παραπάνω φαινόμενα, υπάρχει ένα ακόμα, το οποίο παρατηρείται όταν η πηγή και ο δέκτης/ακροατής κινούνται, το ένα σε σχέση με το άλλο, ή και τα δύο ταυτόχρονα. Πρόκειται για το *φαινόμενο Doppler*[2], και αφορά στη μεταβολή της συχνότητας που αντιλαμβάνεται ο ακροατής: μικρότερη της πραγματικής όταν η απόστασή του από την πηγή βαίνει αυξανόμενη, και μεγαλύτερη όταν η απόσταση μικραίνει.

Η *ακουστική διασπορά*, ένα ακόμα φαινόμενο, αφορά στον διαχωρισμό ενός ήχου στις επιμέρους/παράγωγες συχνότητές του. Κάτι τέτοιο παρατηρείται όταν οι παράγωγες συχνότητες έχουν διαφορετικές ταχύτητες διάδοσης εντός του μέσου, και διαθλώνται η καθεμιά με διαφορετικό τρόπο.
Υπάρχουν επίσης τα *στάσιμα κύματα*, τα οποία προκύπτουν όταν μία ηχητική πηγή εκπέμπει εντός παράλληλων ανακλαστικών επιφανειών.

[2] Προς τιμή του Αυστριακού μαθηματικού και φυσικού Christian Andreas Doppler (1803-1853).

Δημιουργούνται τότε περιοχές εντός του ενδιάμεσου χώρου όπου η ακουστική πίεση είναι πολύ μικρή ή μηδενική (αποκαλούνται *κόμβοι*), και άλλες στις οποίες η πίεση είναι μεγάλη (ονομάζονται *κοιλίες*). Οι θέσεις αυτών εξαρτώνται από τη συχνότητα και την απόσταση ανάμεσα στις επιφάνειες.

Άλλα σχετικό φαινόμενο είναι ο *συντονισμός*. Προκαλείται όταν ένα ηχητικό κύμα προσπίπτει σε ένα αντικείμενο του οποίου η ιδιοσυχνότητα ταλάντωσης είναι ίδια με εκείνη του κύματος. Τότε το αντικείμενο δονείται, δημιουργώντας νέα ηχητικά κύματα.



Σχήμα 3.10 Στάσιμα κύματα [9].

Όλα τα προαναφερθέντα φαινόμενα μπορεί να είναι επιθυμητά ή όχι, ανάλογα με την περίσταση και την εφαρμογή.

Κεφάλαιο 4

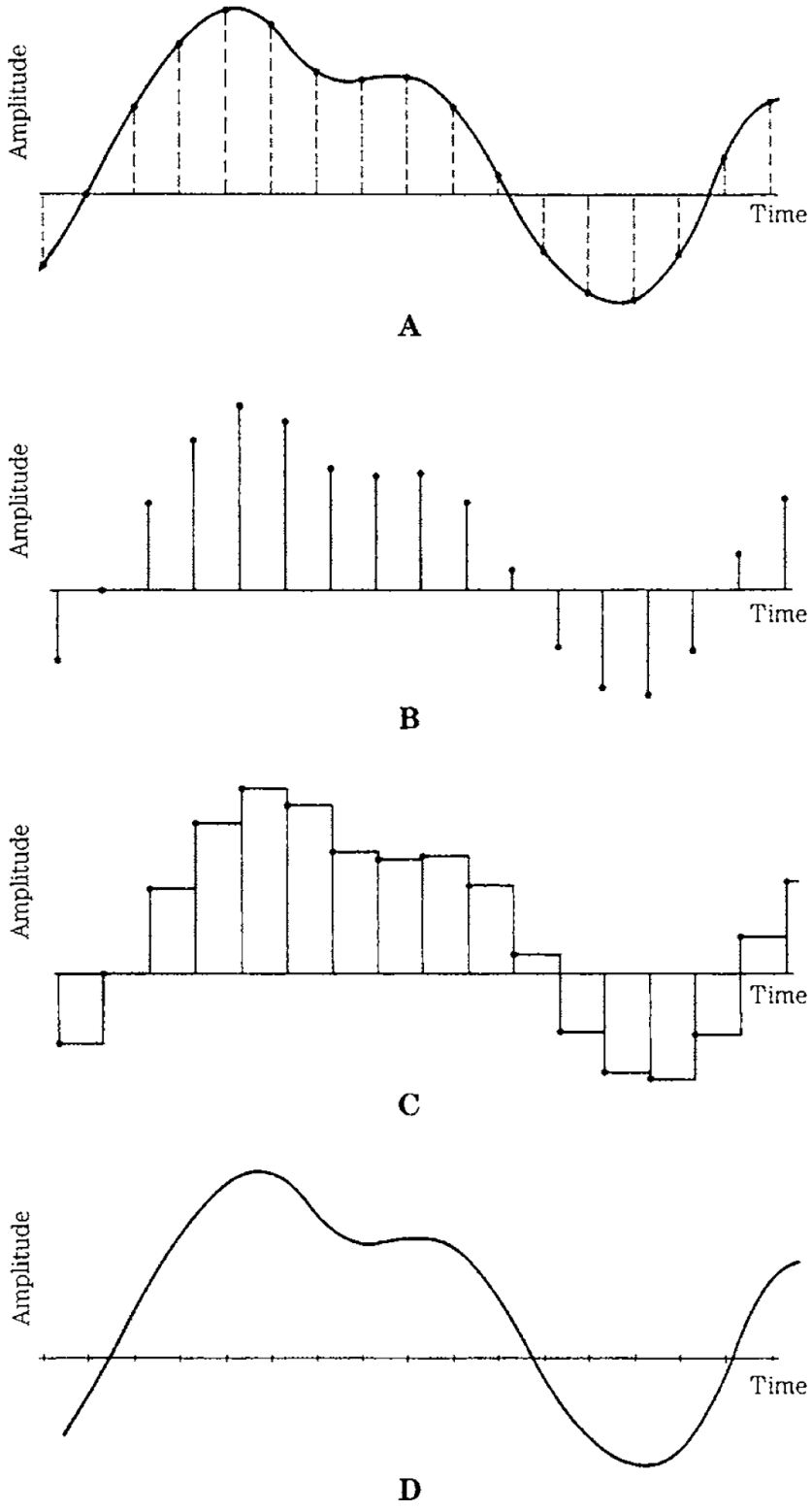
Ο ήχος ως σήμα: Codec, και η περίπτωση της φωνής

4.1 Ο ήχος ως επεξεργάσιμο ηλεκτρικό σήμα

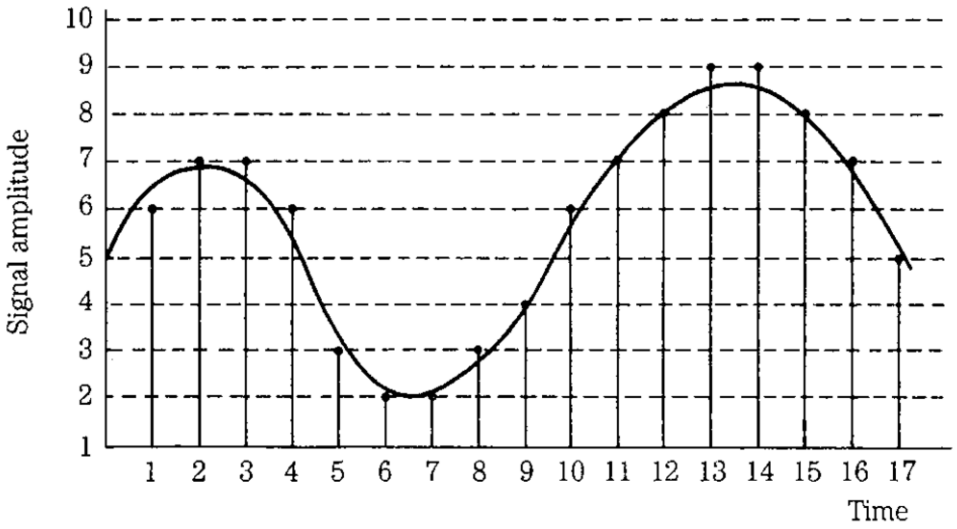
Η τεχνολογία μάς δίνει τη δυνατότητα να «συλλάβουμε» τα ηχητικά κύματα, να τα αποθηκεύσουμε, να τα επεξεργαστούμε, και να τα μεταδώσουμε σε μεγάλες αποστάσεις. Σήμερα, κάθε συσκευή κινητής επικοινωνίας (τηλέφωνο, tablet, laptop κλπ.) διαθέτει ενσωματωμένο μικρόφωνο, καθώς και κατάλληλο υλικό (hardware) και λογισμικό (software), προκειμένου να επιτελεί τον ρόλο αυτόν ανά πάσα στιγμή.

Από τη στιγμή που μπαίνει στην εξίσωση η προαναφερθείσα διαδικασία, κάνουμε πια λόγο όχι για ηχητικό κύμα, αλλά για *ακουστικό σήμα* (audio signal). Σε ό,τι αφορά στο πώς θα το επεξεργαστούμε, και στο πώς -αλλά και σε τι μορφή- θα το μεταδώσουμε, μεγάλο ρόλο παίζει το είδος και το πλαίσιο της εφαρμογής.

Κατ’ αρχάς, όταν μιλάμε για δίκτυα κινητών επικοινωνιών, εδώ και δεκαετίες ισχύει ότι έχουμε να κάνουμε με ψηφιακή τεχνολογία. Πράγμα που σημαίνει ότι το αναλογικό σήμα που μάς παρέχεται στην έξοδο του μικροφώνου θα πρέπει να μετατραπεί σε ψηφιακό. Αφού μεταδοθεί, στον δέκτη ακολουθείται η αντίστροφη διαδικασία, της μετατροπής του ψηφιακού σήματος σε αναλογικό. Συνοπτικά, οι διαδικασίες αυτές απεικονίζονται στα Σχήματα 4.1 και 4.2.



Σχήμα 4.1 Μετατροπή αναλογικού ζωνοπερατού σήματος σε ψηφιακό, και αντίστροφα. **A**. Δειγματοληψία. **B.** Αριθμητικές τιμές των δειγμάτων, προς αποθήκευση ή μετάδοση. **C.** Συγκράτηση των τιμών των δειγμάτων και δημιουργία κλιμακωτής αναπαράστασης του σήματος. **D.** Βαθυπερατό φίλτρο ανακατασκευής της κυματομορφής εισόδου, με χρήση μεθόδου παρεμβολής [15].



Σχήμα 4.2 Στάδιο κβάντισης κατά την ψηφιοποίηση αναλογικού σήματος. Η κάθε τιμή πλάτους αντιστοιχίζεται στην πλησιέστερη από μία σειρά από πεπερασμένες, προκαθορισμένες στάθμες. Κατά τη διαδικασία εισάγεται το *σφάλμα κβάντισης* [15].

4.2 Κωδικοποίηση ομιλίας και ήχου

Η ανάγκη για επικοινωνία είναι βασικό χαρακτηριστικό της ανθρώπινης φύσης, και ο πιο κοινός τρόπος για ανταλλαγή πληροφοριών παραμένει το δίπολο ομιλία/ακρόαση. Παρότι στα δίκτυα κινητών επικοινωνιών η επικοινωνία μπορεί να επιτευχθεί με διάφορους επιπλέον τρόπους, ο βασικότερος εξακολουθεί να είναι ο απευθείας διάλογος μεταξύ δύο (ή περισσότερων) ανθρώπων. Πράγμα που σημαίνει ότι είναι οι φωνές που πρέπει να μεταδοθούν σε μεγάλες (ή μεγαλύτερες) αποστάσεις, με τρόπο αποτελεσματικό (σε ταχύτητα αλλά και ακρίβεια), ώστε να επιτευχθεί η απρόσκοπτη επικοινωνία.

Η *κωδικοποίηση ομιλίας και ήχου* είναι η διαδικασία μέσω της οποίας επιδιώκεται η αναπαράσταση ενός ψηφιοποιημένου σήματος μέσω μίας σειράς bit, με όσο το δυνατό λιγότερα bit ανά ηχητικό δείγμα, διατηρώντας παράλληλα ένα επιθυμητό επίπεδο ποιότητας. Χρησιμοποιείται επίσης, αν και όχι με την ίδια συχνότητα, ο όρος *συμπίεση* (compression). Σε κάθε περίπτωση, απώτερος στόχος της διαδικασίας είναι το κωδικοποιημένο σήμα να μεταδοθεί μέσω ενός διαύλου επικοινωνίας, ή να αποθηκευτεί· και να

μπορεί στη συνέχεια να αποκωδικοποιηθεί -να ανακτηθεί, δηλαδή, από τη ληφθείσα σειρά bit ένα αναλογικό ηχητικό σήμα, κατάλληλο για ακρόαση.

Η κωδικοποίηση ήχου είναι ένα υποσύνολο του αντικειμένου της επεξεργασίας σήματος, και είναι ένα αχανές πεδίο, το οποίο, κατά τις τελευταίες δεκαετίες, έχει εξελιχθεί σε εξαιρετικό βαθμό, έχει δε αξιοποιηθεί στα πλαίσια πληθώρας εφαρμογών. Οι εξελίξεις στη μικροηλεκτρονική έχουν παίξει φυσικά τον ρόλο τους σε αυτό, όπως και οι προσπάθειες για προτυποποίηση (που κάνουν τις διάφορες μεθόδους ευρύτερα αποδεκτές), καθώς και οι επιχειρηματικές βλέψεις για ικανοποίηση των αναγκών, αλλά και για εκμετάλλευση των ευκαιριών για προσφορά καινοτόμων προϊόντων και υπηρεσιών.

Σε κάθε περίπτωση, είναι χρήσιμο να τονίζεται πάντα ότι το όλο πεδίο της κωδικοποίησης ήχου σχετίζεται με την ανθρώπινη πρόσληψη του ήχου, και έτσι πάντοτε υπεισέρχεται σε αυτό ένας βαθμός απροσδιοριστίας: δεν υπάρχει εδώ η έννοια του απολύτως ορθού ή του απολύτως εσφαλμένου για όλες τις περιπτώσεις, και εξ αυτού δεν είναι πάντοτε δυνατό να λάβουμε μαθηματικές αποδείξεις για όλες τις αποφάσεις που καλούμαστε να λάβουμε κατά τη ροή της διαδικασίας. Πολλές από τις λύσεις που δίνονται, δηλαδή, ανακαλύπτονται ή/και αιτιολογούνται με λιγότερο ή περισσότερο διαισθητικές μεθόδους.

Όποια μέθοδος και αν ακολουθηθεί τελικά κατά την κωδικοποίηση και την αποκωδικοποίηση, υπάρχει μία σειρά ζητημάτων που πρέπει πάντοτε να λαμβάνονται υπόψη. Κατ' αρχάς, χρειάζεται ακρίβεια κατά την ποσοτικοποίηση των βασικών παραμέτρων του ηχητικού σήματος, στα πεδία του χρόνου και της συχνότητας. Χρειάζεται, επίσης, προσεκτική εκτίμηση για το ποιες είναι εκείνες οι πλευρές του σήματος οι οποίες ελέγχονται εκούσια από τον ομιλητή, και οι οποίες είναι απαραίτητες στον ακροατή, προκειμένου να επιτευχθεί η επικοινωνία. Ακόμη, απαιτείται μέριμνα ώστε να καταστεί δυνατή η αντιμετώπιση της τεράστιας ποικιλίας που ενυπάρχει στην προφορική επικοινωνία, τόσο εξαιτίας της διαφορετικότητας των ομιλητών, όσο και λόγω των περιβαλλοντικών συνθηκών εντός των οποίων

ηχεί -και «συλλαμβάνεται» από τις συσκευές- η κάθε φωνή. Τέλος, είναι απαραίτητη η γνώση των δυνατοτήτων -και των ορίων- των ηλεκτρονικών υπολογιστών, ώστε αυτά να χρησιμοποιηθούν όσο το δυνατό πιο αποτελεσματικά για την επίτευξη του τελικού στόχου.

4.3 Τι είναι codec

Ο όρος *codec* αποτελεί ακρωνύμιο των όρων coder/decoder (κωδικοποιητής/αποκωδικοποιητής), και αναφέρεται σε μία συσκευή ή/και σε ένα λογισμικό που κωδικοποιεί/αποκωδικοποιεί μία ροή δεδομένων.

Στο πεδίο της κωδικοποίησης ήχου, ο όρος codec αναφέρεται σε κάποιες προτυποποιημένες διαδικασίες (αλγορίθμους) που ακολουθούνται κατά την κωδικοποίηση (και, αντεστραμμένες, κατά την αποκωδικοποίηση). Υπάρχει μεγάλη ποικιλία διαθέσιμων codec, διαφόρων κατηγοριών, ανάλογα και με το πρότυπο κινητών επικοινωνιών στα πλαίσια του οποίου εντάσσεται το κάθε ένα.

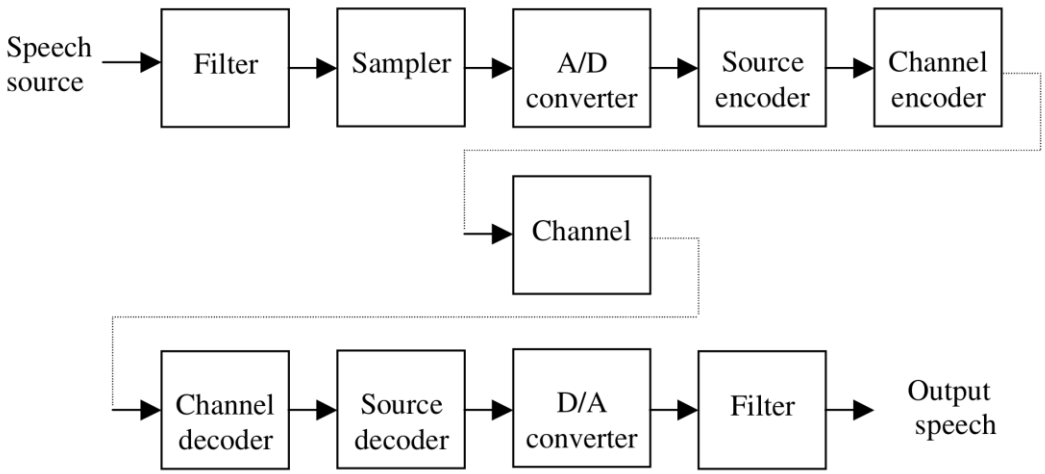
Ως *αλγόριθμος* ορίζεται μία οποιαδήποτε πλήρως ορισμένη υπολογιστική διαδικασία, αποτελούμενη από διακριτά βήματα, η οποία λαμβάνει ως είσοδο μία τιμή ή ένα σετ τιμών, και παράγει ως έξοδο μία άλλη τιμή ή ένα άλλο σετ τιμών. Πολλά προβλήματα επεξεργασίας σήματος, ανάμεσά τους και το ζήτημα της κωδικοποίησης ήχου, μπορούν να μετατραπούν σε υπολογιστικά προβλήματα. Η σειρά των βημάτων, οι οδηγίες δηλαδή, μπορούν είτε να δοθούν σε έναν επεξεργαστή εν είδει λογισμικού, είτε να «μεταφραστούν» σε επίπεδο hardware, παίρνοντας τη μορφή ψηφιακού κυκλώματος.

4.4 Η δομή ενός συστήματος κωδικοποίησης ομιλίας

Όπως προαναφέρθηκε, η βασική πληροφορία την οποία τα δίκτυα κινητών

επικοινωνιών καλούνται να μεταδώσουν αφορά στην ανθρώπινη φωνή. Έτσι, στο παρόν κεφάλαιο θα εστιάσουμε σε αντίστοιχα codec, τα οποία, όμως, πολύ συχνά, καλύπτουν και τις ανάγκες μετάδοσης οποιουδήποτε ηχητικού σήματος συλλάβει το μικρόφωνο, έστω κι αν τα αποτελέσματα παρουσιάζουν ποιοτικές διακυμάνσεις.

Στο Σχήμα 4.3 δίνεται το μπλοκ διάγραμμα ενός συστήματος κωδικοποίησης ομιλίας. Το αναλογικό σήμα συνεχούς χρόνου ψηφιοποιείται μέσω μίας αλυσίδας σταδίων, η οποία περιλαμβάνει φίλτρο (το οποίο απαλείφει την επικάλυψη συχνοτήτων), δειγματολήπτη (ο οποίος μετατρέπει το σήμα σε διακριτού χρόνου), και μετατροπέα αναλογικού σε ψηφιακό (ο οποίος υποθέτουμε ότι εφαρμόζει ομοιόμορφη κβάντιση). Η έξοδος αυτών των βαθμίδων αναφέρεται ως *ψηφιακή ομιλία*.



Σχήμα 4.3 Μπλοκ διάγραμμα συστήματος κωδικοποίησης και αποκωδικοποίησης ομιλίας [13].

Τα περισσότερα συστήματα κωδικοποίησης φωνής σχεδιάστηκαν για την υποστήριξη τηλεπικοινωνιακών αναγκών, και θέτουν ως προαπαιτούμενο τον περιορισμό του συχνοτικού περιεχομένου των σημάτων στην περιοχή μεταξύ 300 και 3,400 Hz. Σύμφωνα με το θεώρημα Nyquist[3], η συχνότητα δειγματοληψίας θα πρέπει να είναι τουλάχιστον διπλάσια του εύρους ζώνης

[3] Προς τιμή του Σουηδοαμερικανού φυσικού και ηλεκτρονικού μηχανικού Harry Nyquist (1889-1976).

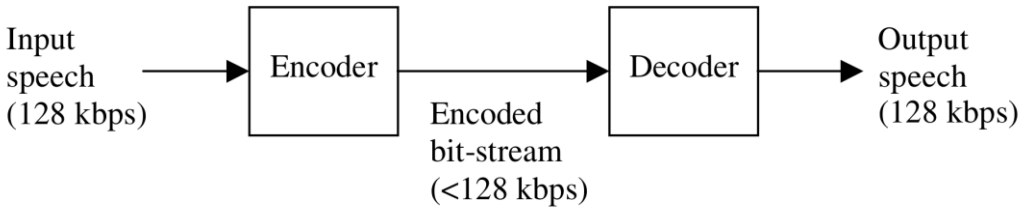
του σήματος συνεχούς χρόνου, έτσι ώστε να αποφευχθεί επικάλυψη συχνοτήτων. Συνήθως επιλέγεται η τιμή των 8 kHz. Προκειμένου να μετατραπούν τα αναλογικά δείγματα σε ψηφιακά, με χρήση ομοιόμορφης κβάντισης, και με ποιότητα αντίστοιχη της ενσύρματης τηλεφωνίας (toll quality), ώστε η ψηφιακή ομιλία να μη μπορεί πρακτικά να διακριθεί σε σύγκριση με το ζωνοπερατό σήμα εισόδου, είναι απαραίτητο να χρησιμοποιηθούν περισσότερα από 8 bit ανά δείγμα. Αν, π.χ., επιλέξουμε 16 bit ανά δείγμα (δηλαδή υψηλή ποιότητα), και συχνότητα δειγματοληψίας 8 kHz, ο *ρυθμός δυαδικών ψηφίων* (bit rate) θα είναι 16 x 8 = 128 kbps.

Αυτός ο ρυθμός δυαδικών ψηφίων (ή *ρυθμός ψηφιακών δεδομένων* ή απλώς *ρυθμός δεδομένων*), που αποκαλείται και *ρυθμός δυαδικών ψηφίων εισόδου*, είναι εκείνος τον οποίο ο κωδικοποιητής πηγής επιχειρεί να μειώσει. Στην έξοδό αυτού λαμβάνουμε την κωδικοποιημένη ψηφιακή φωνή, η οποία έχει σημαντικά χαμηλότερο bit rate. Ρυθμοί δεδομένων που χρησιμοποιούνται είναι 64 kbps για πολλά από τα ενσύρματα δίκτυα, και 10 έως 13 kbps για την κινητή τηλεφωνία. Συνήθως γίνεται διαχωρισμός ανάμεσα στη *μετάδοση στενής ζώνης* (narrow band, NB) -π.χ. τηλεφωνία, με συχνότητα δειγματοληψίας 8 kHz-, και την *ευρυζωνική μετάδοση* (wideband, WB) -π.χ. διαδικτυακές εφαρμογές, με συχνότητα δειγματοληψίας 16 kHz.

Στη συνέχεια, η κωδικοποιημένη ψηφιακή φωνή περνά από τον κωδικοποιητή καναλιού, ο οποίος παρέχει προστασία από σφάλματα, προκειμένου η ροή bit να μεταδοθεί μέσω του καναλιού επικοινωνίας, το οποίο μπορεί να εισάγει θόρυβο και παρεμβολή. Σε κάποιες περιπτώσεις, ο κωδικοποιητής πηγής και ο κωδικοποιητής καναλιού μπορεί να εφαρμοστούν στο ίδιο βήμα.

Στην πλευρά του δέκτη, ο αποκωδικοποιητής καναλιού επεξεργάζεται τα προστατευμένα από σφάλματα δεδομένα, προκειμένου να ανακτήσει το κωδικοποιημένο περιεχόμενο. Στη συνέχεια, ο αποκωδικοποιητής πηγής μάς δίνει το σήμα ψηφιακής φωνής, το οποίο μετατρέπεται σε αναλογικό συνεχούς χρόνου, μέσω του μετατροπέα ψηφιακού σήματος σε αναλογικό, και του φιλτραρίσματος.

Στο Σχήμα 4.4 απεικονίζεται η πλέον απλοποιημένη εκδοχή ενός codec ομιλίας, το οποίο αποτελείται από το ζεύγος κωδικοποιητή και αποκωδικοποιητή πηγής. Σε αυτό το σύστημα εστιάζουμε στην παρόν κεφάλαιο.



Σχήμα 4.4 Απλοποιημένο μπλοκ διάγραμμα codec ομιλίας [13].

4.5 Επιθυμητά χαρακτηριστικά των codec ομιλίας

Ο κύριος στόχος κάθε κωδικοποίησης φωνής είναι είτε να μεγιστοποιήσει την αντιληπτή ποιότητα για έναν δεδομένο ρυθμό δεδομένων, είτε να ελαχιστοποιήσει τον ρυθμό δεδομένων για μία δεδομένη ποιότητα. Ο ρυθμός δεδομένων με τον οποίο η φωνή θα πρέπει να μεταδοθεί (ή να αποθηκευτεί) εξαρτάται από το κόστος της μετάδοσης (ή της αποθήκευσης), από το κόστος της κωδικοποίησης του σήματος ψηφιακής φωνής, και από τις απαιτήσεις για ποιότητα. Σε όλους τους κωδικοποιητές φωνής, το ανακτηθέν σήμα διαφέρει σε σχέση με το αρχικό, καθώς επιχειρείται να μειωθεί ο ρυθμός δεδομένων, γεγονός που αφαιρεί από την ακρίβεια, αλλά και εξαλείφει πλεονασμούς που ενυπάρχουν στο αρχικό σήμα. Αποτέλεσμα αυτών των επεμβάσεων είναι να έχουμε ένα σχήμα κωδικοποίησης με απώλειες (lossy).

Επιγραμματικά, κάποια επιθυμητά χαρακτηριστικά για ένα codec φωνής/ήχου είναι τα παρακάτω:

- Χαμηλός ρυθμός δεδομένων. Έτσι εξασφαλίζεται μικρό εύρος ζώνης για μετάδοση ή/και μικρό αποτύπωμα αποθήκευσης.

- Υψηλή ποιότητα ομιλίας. Υπάρχουν πολλές διαστάσεις στην αντίληψη της ποιότητας, όπως διακριτότητα, φυσικότητα, ευχαριστότητα, και αναγνώριση του ομιλητή.

- Προσαρμοστικότητα στους διαφορετικούς ομιλητές. Το codec θα πρέπει να σχεδιαστεί με τέτοιον τρόπο ώστε να μπορεί να εξυπηρετεί ομιλητές όλων των ηλικιών και φύλων, και χρήστες όσο το δυνατόν περισσότερων γλωσσών. Πρόκειται για διόλου ευκαταφρόνητης δυσκολίας εγχείρημα.

- Ανθεκτικότητα σε σφάλματα του καναλιού.

- Καλή απόδοση σε διαφορετικού τύπου σήματα. Στις τηλεπικοινωνιακές ζεύξεις μπορεί να είναι παρόντα και άλλα σήματα, π.χ. μουσική ή τόνοι σηματοδοσίας. Δεν είναι όλα τα codec ικανά να αναπαραστήσουν με πιστότητα κάθε σήμα, αλλά σε κάθε περίπτωση θα πρέπει να προβλέπεται η αποφυγή ενοχλητικών θορύβων και παραμορφώσεων.

- Μικρό μέγεθος μνήμης και μικρή υπολογιστική πολυπλοκότητα. Πρόκειται για χαρακτηριστικά που μειώνουν το κόστος εφαρμογής του codec.

- Μικρή χρονική υστέρηση κωδικοποίησης. Στην όλη διαδικασία υπεισέρχεται αναπότρεπτα χρονική υστέρηση, η οποία υπολογίζεται ως ο χρόνος που περνά από τη στιγμή που ένα δείγμα ομιλίας φτάνει στην είσοδο του κωδικοποιητή, μέχρι τη στιγμή που το ίδιο δείγμα εμφανίζεται στην έξοδο του αποκωδικοποιητή. Υπέρβαση κάποιων σχετικών ορίων δημιουργεί προβλήματα σε συνομιλίες πραγματικού χρόνου.

Πολλά από τα προαναφερθέντα χαρακτηριστικά έρχονται σε σύγκρουση μεταξύ τους. Για παράδειγμα, ο χαμηλός ρυθμός bit έρχεται σχεδόν πάντα σε σύγκρουση με την ανάγκη για ποιότητα ομιλίας, ενώ η χρονική υστέρηση και το bit rate είναι ένα ακόμα ζευγάρι - διελκυστίνδα. Σε κάθε σχεδιαστική απόπειρα απαιτείται η επίτευξη συμβιβασμών.

4.6 Κατηγοριοποίηση των codec

Υπάρχουν διάφορα κριτήρια με βάση τα οποία μπορεί να γίνει η κατηγοριοποίηση των διαφόρων codec, και στο συγκεκριμένο ζήτημα επικρατεί ενίοτε κάποια σύγχυση. Παρακάτω αναφέρονται κάποια από αυτά τα κριτήρια, και οι αντίστοιχες κατηγοριοποιήσεις.

4.6.1 Κατηγοριοποίηση με βάση τον ρυθμό δυαδικών ψηφίων

Όλα τα codec ομιλίας σχεδιάζονται με στόχο τη μείωση του bit rate αναφοράς των 128 kbps. Οι διαφορετικές τεχνικές κωδικοποίησης οδηγούν αναπόφευκτα σε διαφορετικά bit rate, και το ελάχιστο bit rate που μπορεί να επιτευχθεί σε κάθε δεδομένη περίπτωση εξαρτάται από το περιεχόμενο της ομιλίας, αλλά και από άλλους παράγοντες. Από γλωσσολογική άποψη, μία λογική εκτίμηση για την ελάχιστη αυτή τιμή θα ήταν τα 100 bps.

Οι βασικές κατηγορίες codec είναι οι υψηλού bit rate (ανώτερου των 15 kbps), μέσου bit rate (5 έως 15 kbps), χαμηλού bit rate (2 έως 5 kbps), και πολύ χαμηλού bit rate (μικρότερου των 2 kbps).

4.6.2 Κατηγοριοποίηση με βάση τη μέθοδο κωδικοποίησης

Εξετάζοντας τα διάφορα codec με βάση την τεχνική κωδικοποίησης που χρησιμοποιεί το καθένα, έχουμε κατά βάση δύο μεγάλες κατηγορίες: τα *codec κυματομορφής* και τα *παραμετρικά codec* ή *vocoder*. Κάθε άλλη κατηγορία αποτελεί παρακλάδι μία εκ των δύο, ή συνδυασμό τους.

Τα codec κυματομορφής επιδιώκουν να διατηρήσουν όσο το δυνατό πιο πιστά την αρχική κυματομορφή του σήματος, και για αυτό μπορούν γενικά να εφαρμοστούν σε οποιοδήποτε σήμα. Έχουν καλύτερα αποτελέσματα σε υψηλά bit rate -στην πράξη, από 32 kbps και πάνω. Η ποιότητά τους μπορεί

να εκτιμηθεί με βάση τον λόγο σήματος προς θόρυβο (SNR), μεταξύ της εισόδου του κωδικοποιητή και της εξόδου του αποκωδικοποιητή. Η *Παλμοκωδική Διαμόρφωση* (Pulse Code Modulation, PCM), που είναι το απλούστερο ως προς τη λειτουργία του audio codec, κατατάσσεται σε αυτή την κατηγορία, όπως και οι διάφορες παραλλαγές του.

Τα παραμετρικά codec υποθέτουν ότι το σήμα ομιλίας παράγεται από ένα *μοντέλο*, το οποίο ελέγχεται από κάποιες *παραμέτρους*. Κατά την κωδικοποίηση, οι παράμετροι αυτές εκτιμώνται με βάση το σήμα εισόδου, και είναι αυτές που μεταδίδονται ως κωδικοποιημένη ροή bit, αποτελούμενη από *πλαίσια*, διάρκειας 10 έως 30 msec. Καθώς δεν γίνεται κάποια απόπειρα διατήρησης της αρχικής κυματομορφής, κριτήρια όπως ο SNR δεν έχουν χρησιμότητα. Η αντιληπτή ποιότητα του αποκωδικοποιημένου σήματος σχετίζεται άμεσα με την ακρίβεια της μοντελοποίησης. Λόγω αυτού του περιορισμού, τα codec αυτής της κατηγορίας έχουν κακή απόδοση σε άλλα σήματα, πλην της ομιλίας. Το πιο επιτυχημένο, από τα πολλά, μοντέλο είναι εκείνο της *Γραμμικής Πρόβλεψης*, όπου ο μηχανισμός της φώνησης αντιπροσωπεύεται από ένα χρονομεταβλητό φίλτρο. Η εν λόγω κατηγορία λειτουργεί καλά για χαμηλά bit rate, συνήθως μεταξύ 2 και 5 kbps. Ένα σχετικό παράδειγμα codec είναι η *Γραμμική Προβλεπτική Κωδικοποίηση*.

Τα *υβριδικά codec* συνδυάζουν χαρακτηριστικά των δύο προηγούμενων κατηγοριών: χρησιμοποιείται μοντέλο παραγωγής ομιλίας, αλλά επιχειρείται και η παραγωγή μίας κυματομορφής που θα είναι όσο το δυνατό πιο κοντά στην αρχική. Ο βαθμός επιτυχίας συχνά μετράται μέσω ενός σήματος σφάλματος. Η εν λόγω κατηγορία codec επικρατεί σε μεσαία bit rate, και η πιο γνωστή περίπτωση είναι η *Γραμμική Πρόβλεψη Με Διέγερση Κώδικα*, με τις διάφορες παραλλαγές της.

Από τεχνικής άποψης, η βασική διαφορά ανάμεσα σε ένα υβριδικό codec και σε ένα παραμετρικό είναι ότι το πρώτο επιχειρεί να αναπαραστήσει το σήμα διέγερσης του μοντέλου παραγωγής ομιλίας, το οποίο μεταδίδεται σαν μέρος της κωδικοποιημένης ροής bit, ενώ το δεύτερο απορρίπτει όλη την πληροφορία που αφορά σε λεπτομέρειες του σήματος διέγερσης, και κρατά

μόνο απολύτως βασικές παραμέτρους. Ένα υβριδικό codec τείνει να συμπεριφέρεται σαν codec κυματομορφής στα υψηλά bit rate, και σαν παραμετρικό codec στα χαμηλά, ενώ έχει μέτρια έως καλή απόδοση στα μεσαία bit rate.

Τα *codec νευρωνικών δικτύων* αποτελούν στην ουσία υποκατηγορία, ή παρακλάδι, των codec κυματομορφής. Εκμεταλλεύονται τις ραγδαίες εξελίξεις των τελευταίων δύο δεκαετιών στον χώρο των τεχνητών νευρωνικών δικτύων, και εφαρμόζουν μη γραμμικές αλγοριθμικές μεθόδους. Είναι ιδιαίτερα κοστοβόρα ως προς την υπολογιστική πολυπλοκότητα, και αυτός είναι ο λόγος για τον οποίο προς το παρόν δεν μπορούν να αμφισβητήσουν την πρωτοκαθεδρία της παλιάς φρουράς. Είναι βέβαιο, πάντως, ότι στα επόμενα χρόνια οι υβριδικές τεχνικές που θα τα ενσωματώσουν θα κερδίζουν ολοένα και περισσότερο έδαφος.

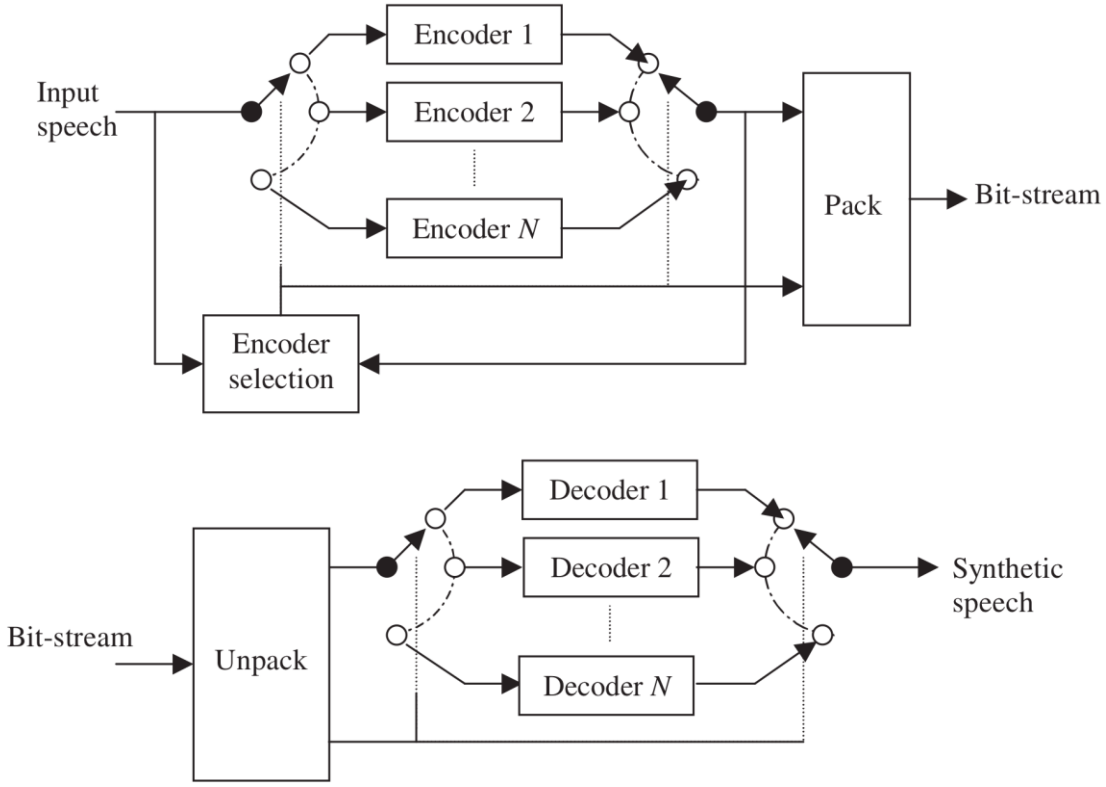
4.6.3 Κατηγοριοποίηση με βάση τη σταθερότητα του αλγορίθμου κωδικοποίησης

Πέρα από τα codec που χρησιμοποιούν σε κάθε περίπτωση έναν συγκεκριμένο μηχανισμό κωδικοποίησης, οπότε και έχουν ένα σταθερό bit rate, υπάρχουν και άλλα, τα οποία εφευρέθηκαν προκειμένου να εκμεταλλεύονται τη δυναμική φύση της ομιλίας, και να προσαρμόζονται σε συνθήκες που μεταβάλλονται με τον χρόνο.

Ένα τέτοιο codec διαθέτει διάφορους διακριτούς τρόπους κωδικοποίησης, από τους οποίους επιλέγεται ένας κάθε φορά. Η επιλογή μπορεί να γίνεται είτε με έλεγχο από την πηγή (με την επιλογή να βασίζεται σε στατιστικά του σήματος εισόδου), είτε με έλεγχο από το δίκτυο (με την επιλογή να γίνεται μέσω κάποιας εξωτερικής εντολής που αφορά στις ανάγκες του δικτύου ή στις συνθήκες που επικρατούν στο κανάλι).

Τα περισσότερα πολυτροπικά codec έχουν μεταβλητό bit rate, με κάθε

ξεχωριστό τρόπο να έχει τη δική του, σταθερή τιμή. Η μεταβλητότητα του bit rate επιτρέπει ευελιξία, που σημαίνει βελτιωμένη αποτελεσματικότητα και σημαντική μείωση στο μέσο bit rate.



Σχήμα 4.5 Μπλοκ διαγράμματα κωδικοποιητή (επάνω) και αποκωδικοποιητή (κάτω) ενός codec με πολλαπλή επιλογή αλγορίθμου, η οποία ελέγχεται από την πηγή [13].

4.7 Παραγωγή και μοντελοποίηση ομιλίας

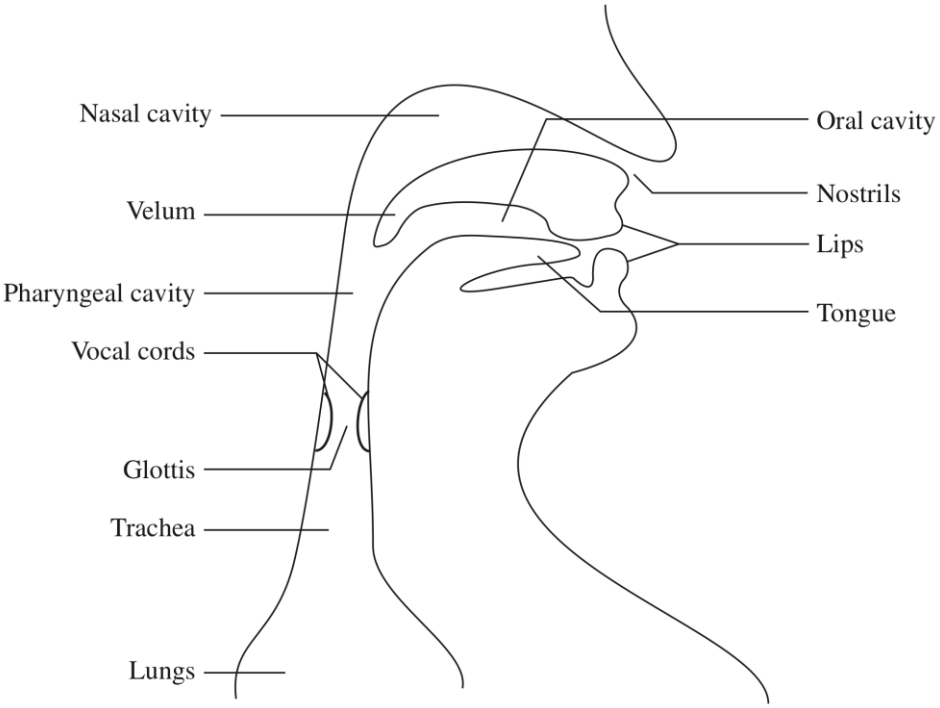
4.7.1 Φώνηση και σήματα ομιλίας

Το ανθρώπινο σύστημα παραγωγής φωνής και ομιλίας αποτελείται από μία σειρά ανατομικών δομών, μία απλουστευμένη απεικόνιση των οποίων φαίνεται στο Σχήμα 4.6.

Η διαδικασία παραγωγής ομιλίας ονομάζεται *φώνηση* (phonation). Η ομιλία, εν ολίγοις, παράγεται ως ένα ηχητικό κύμα το οποίο εξέρχεται από τα

ρουθούνια και το στόμα, όταν αέρας εξωθείται από τους πνεύμονες, και η ροή του διαταράσσεται από διάφορα εμπόδια μέσα στο σώμα.

Η όλη δομή μπορεί να ερμηνευθεί ως ένα φίλτρο, το οποίο αποτελείται από τις τρεις κύριες κοιλότητες: ρινική, στοματική, φαρυγγική. Το φίλτρο διεγείρεται από τον αέρα που έρχεται από τους πνεύμονες, και έχει ως φορτίο στην κύρια έξοδό του την εμπέδηση που σχετίζεται με τα χείλη.



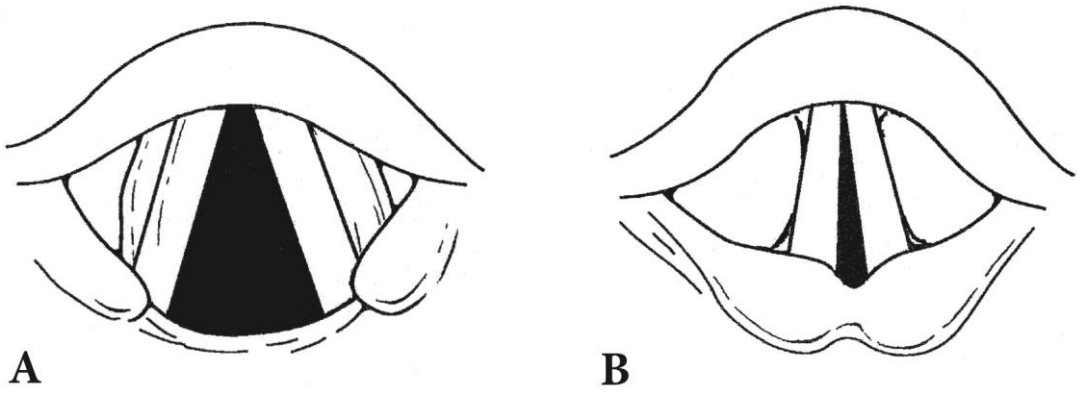
Σχήμα 4.6 Απλουστευμένη αναπαράσταση της ανθρώπινης φωνητικής οδού [15].

Χρησιμοποιούμε τον όρο *φωνητική οδός* κυρίως για να αναφερθούμε στο σύνολο που αποτελείται από δύο κοιλότητες, τη φαρυγγική και τη στοματική. Η ρινική οδός ξεκινά στον ουρανίσκο και τελειώνει στα ρουθούνια της μύτης. Όταν ο ουρανίσκος χαμηλώνει, η ρινική οδός ενώνεται ακουστικά με τη φωνητική οδό, και παράγει ρινικούς ήχους της ομιλίας.

Η μορφή και το σχήμα της φωνητικής και της ρινικής οδού αλλάζουν συνεχώς με τον χρόνο, δημιουργώντας ένα ακουστικό φίλτρο με χρονομεταβλητή απόκριση συχνότητας. Καθώς ο αέρας από τους πνεύμονες

περνά μέσα από τις οδούς, το συχνοτικό φάσμα διαμορφώνεται από τη συχνοτική επιλεκτικότητα αυτών. Κάθε συχνότητα συντονισμού του σωλήνα της φωνητικής οδού αποκαλείται *διαμορφωτής* (formant), και εξαρτάται από το σχήμα και τις διαστάσεις της φωνητικής οδού. Οι διαμορφωτές είναι κατ' ουσία οι συχνότητες που ενισχύονται, και εντοπίζονται ως κορυφές της περιβάλλουσας του συχνοτικού φάσματος. Αντιληπτικά πειράματα έχουν δείξει ότι οι πιο κρίσιμοι διαμορφωτές είναι ο πρώτος (*F1*) και ο δεύτερος (*F2*), και δευτερευόντως ο τρίτος (*F3*).

Μέσα στον λάρυγγα βρίσκεται ένα από τα πλέον σημαντικά μέρη του συστήματος φώνησης, που δεν είναι άλλο από τις *φωνητικές χορδές* ή *φωνητικές πτυχές*. Η θέση τους εντοπίζεται στο ύψος του «μήλου του Αδάμ» -της προεξοχής στο μπροστινό μέρος του λαιμού που έχουν οι περισσότεροι άρρενες. Πρόκειται για ένα ζεύγος ελαστικών λωρίδων, αποτελούμενων από μύες και βλενώδη μεμβράνη, που ανοίγουν και κλείνουν γρήγορα κατά την ομιλία. Η ταχύτητα με την οποία ανοιγοκλείνουν είναι μοναδική για κάθε άνθρωπο, και καθορίζει το γνώρισμα και την προσωπικότητα της φωνής του.



Σχήμα 4.7 Η σύγκλιση των φωνητικών χορδών **A.** κατά την εισπνοή, **B.** κατά την έναρξη της φώνησης [30].

Για να υπάρξει φώνηση, πρέπει να επικρατήσουν κάποιες συγκεκριμένες συνθήκες στο σύστημα που την υποστηρίζει. Αρχικά, θα πρέπει οι φωνητικές χορδές να έχουν έρθει στη θέση φώνησης -δηλαδή να έχουν προσεγγίσει η μία την άλλη- ή πλησίον αυτής (π.χ. να είναι ακόμα και κλειστές). Επίσης,

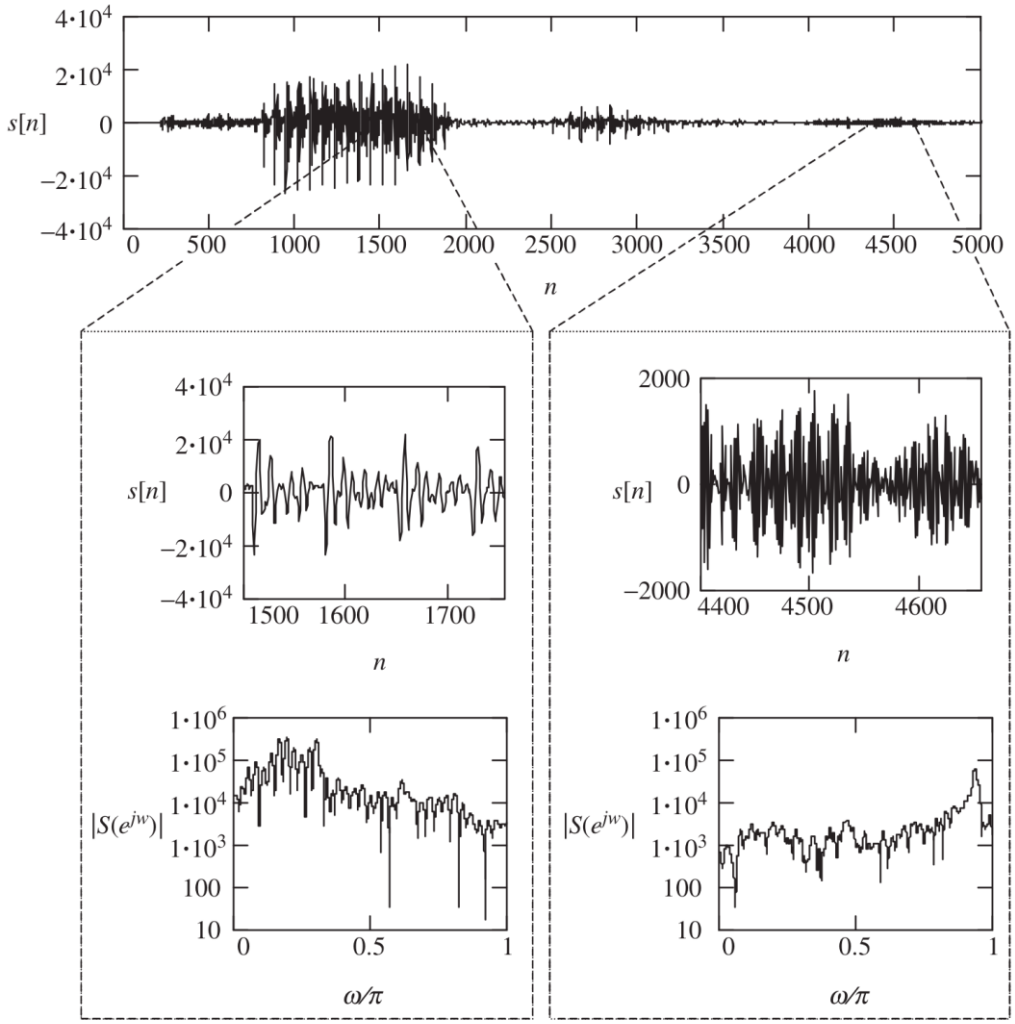
θα πρέπει αυτές να επιμηκυνθούν κατάλληλα· το πόσο και πώς αυτής της επιμήνυνσης θα καθορίσει τη θεμελιώδη συχνότητα δόνησής τους. Τέλος, θα πρέπει να υπάρχει ροή αέρα από τους πνεύμαονες, πράγμα που συνεπάγεται εισπνοή.

4.7.2 Κατηγοριοποίηση των σημάτων ομιλίας

Οι ήχοι που παράγει κατά την ομιλία του ο άνθρωπος -οποιοσδήποτε άνθρωπος, ανεξάρτητα από τη γλώσσα που χρησιμοποιεί- ονομάζονται *φθόγγοι*, και μελετώνται από τη Φωνητική, έναν κλάδο της Γλωσσολογίας. Ανάλογα αν μελετάται η άρθρωση των φθόγγων από τα φωνητικά όργανα του ομιλητή, οι ακουστικές ιδιότητες των παραγόμενων φθόγγων, ή η πρόσληψη αυτών των φθόγγων από τον ακροατή, έχουμε τις υποκατηγορίες της Αρθρωτικής, της Ακουστικής, και της Ακροατικής ή Προσληπτικής Φωνητικής, αντίστοιχα. Οι φθόγγοι κατηγοριοποιούνται βασικά σε *σύμφωνα* και *φωνήεντα*, ενώ υπάρχει και μία ενδιάμεση κατηγορία, τα *ημίφωνα*.

Τα σήματα ομιλίας μπορεί, χονδρικά, να κατηγοριοποιηθούν ως *ηχηρά* ή *άηχα*. Στα ηχηρά κατατάσσονται τα φωνήεντα και κάποια από τα σύμφωνα, ενώ στα άηχα ανήκουν τα υπόλοιπα σύμφωνα. Οι ηχηροί φθόγγοι παράγονται όταν οι φωνητικές χορδές πάλλονται με τέτοιον τρόπο, ώστε η ροή του αέρα από τους πνεύμονες να διακόπτεται περιοδικά, δημιουργώντας μία σειρά από παλμούς που διεγείρουν τη φωνητική οδό. Για την παραγωγή άηχων φθόγγων, αντίθετα, οι φωνητικές χορδές παραμένουν στατικές.

Στο πεδίο του χρόνου, οι ηχηροί φθόγγοι έχουν το χαρακτηριστικό της ισχυρής περιοδικότητας, με τη θεμελιώδη συχνότητα να ονομάζεται *συχνότητα τονικού ύψους* ή απλά *τονικό ύψος*, και να συμβολίζεται με *F0*. Για τους άνδρες, το τονικό ύψος κυμαίνεται μεταξύ 50 και 250 Hz, ενώ για τις γυναίκες συνήθως μεταξύ 120 και 500 Hz. Οι άηχοι φθόγγοι, από την άλλη, δεν παρουσιάζουν κανενός είδους περιοδικότητα και η φύση τους είναι ουσιαστικά τυχαία.



Σχήμα 4.8 Κυματομορφή ομιλίας άρρενα. Σε μεγέθυνση ένα ηχηρό (αριστερά) και ένα άηχο πλαίσιο, καθένα μεγέθους 256 δειγμάτων, μαζί με τις αντίστοιχες καμπύλες των μετασχηματισμών Fourier [13].

Είναι απαραίτητο να σημειωθεί ότι η κατηγοριοποίηση που συζητούμε εδώ ενδέχεται να μην είναι απόλυτα ξεκάθαρη για κάθε πλαίσιο, καθώς κατά τις μεταβάσεις από ηχηρό σε άηχο φθόγγο, ή το αντίστροφο, υπάρχει τυχαιότητα και οιονεί περιοδικότητα, οι οποίες καθιστούν δύσκολη την κατάταξη του πλαισίου σε μία από τις δύο κατηγορίες.

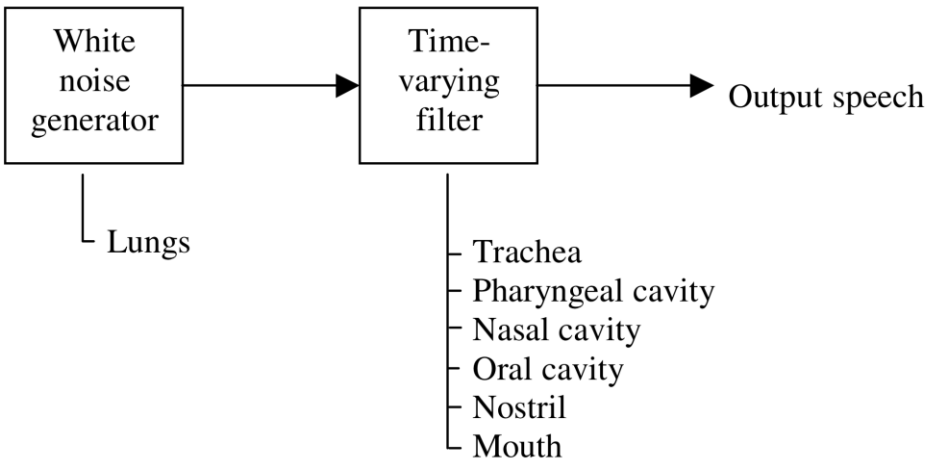
Για τα περισσότερα codec, η επεξεργασία του κάθε σήματος γίνεται σε βάση πλαισίου, με καθένα από αυτά να αποτελείται από έναν πεπερασμένο αριθμό δειγμάτων. Το μήκος του πλαισίου επιλέγεται με τέτοιον τρόπο ώστε τα στατιστικά του σήματος εντός του να παραμένουν σχεδόν σταθερά. Αυτό το

μήκος είναι συνήθως μεταξύ 20 και 30 msec, ή 160 και 240 δείγματα, για συχνότητα δειγματοληψίας 8 kHz.

4.7.3 Μοντελοποίηση του συστήματος φώνησης

Το σύστημα φώνησης του ανθρώπου μπορεί να μοντελοποιηθεί με χρήση μιας αρκετά απλής δομής. Οι πνεύμονες, οι οποίοι παρέχουν τον αέρα ή την ενέργεια που θα διεγείρει τη φωνητική οδό, αντιπροσωπεύονται από μία πηγή λευκού θορύβου, ενώ το ακουστικό μονοπάτι μέσα στο σώμα, με όλα του τα μέρη, αναπαρίσταται από ένα χρονομεταβλητό φίλτρο. Ή όλη ιδέα απεικονίζεται στο Σχήμα 4.9.

Το εν λόγω απλό μοντέλο βρίσκεται στον πυρήνα πολλών αλγορίθμων κωδικοποίησης ομιλίας. Χρησιμοποιώντας διάφορες τεχνικές που θα αναλυθούν στη συνέχεια, είναι δυνατό να εκτιμηθούν οι παράμετροι του χρονομεταβλητού φίλτρου.



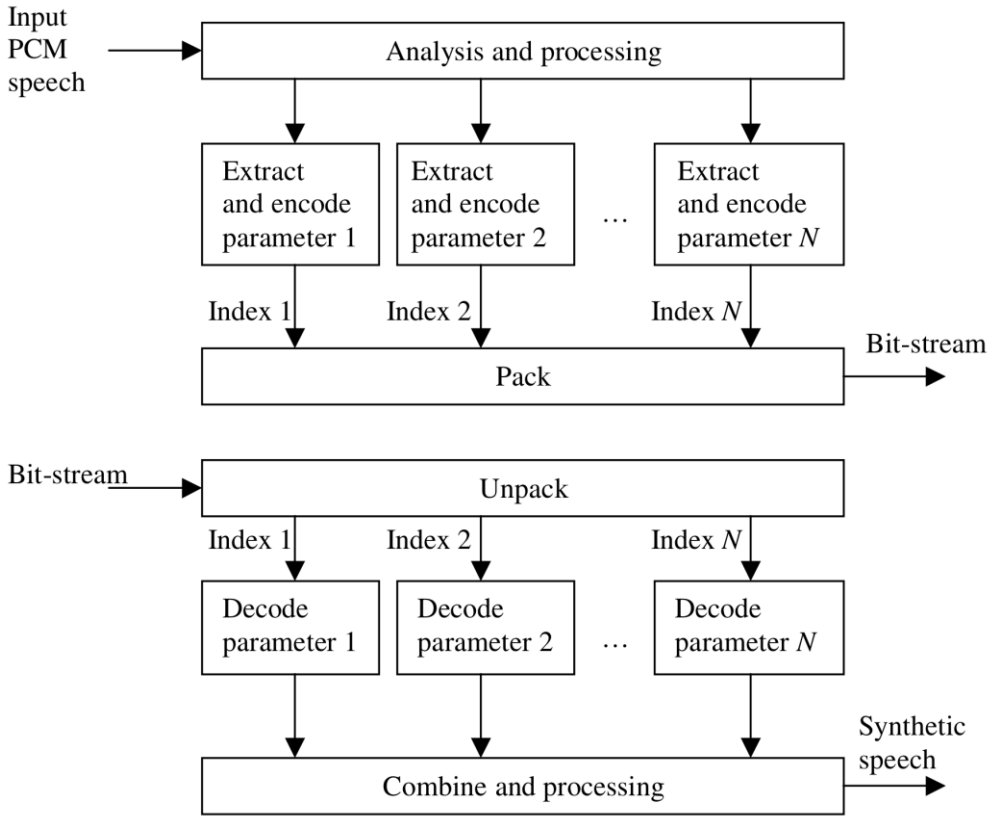
Σχήμα 4.9 Μοντέλο του ανθρώπινου συστήματος φώνησης [13].

Η παραδοχή που κάνουμε στα πλαίσια του μοντέλου που συζητάμε εδώ είναι ότι η κατανομή της ενέργειας του σήματος ομιλίας στο πεδίο της συχνότητας οφείλεται συνολικά στο χρονομεταβλητό φίλτρο, με τους πνεύμονες να

παράγουν ένα σήμα διέγερσης το οποίο είναι ουσιαστικά λευκός θόρυβος επίπεδου φάσματος. Το μοντέλο είναι ιδιαίτερα αποτελεσματικό και έχουν αναπτυχθεί αρκετά αναλυτικά εργαλεία γύρω από αυτό.

4.7.4 Γενική δομή ενός codec ομιλίας

Στο Σχήμα 4.11 απεικονίζεται ένα γενικό μπλοκ διάγραμμα ενός κωδικοποιητή και ενός αποκωδικοποιητή ομιλίας.



Σχήμα 4.10 Γενική δομή του κωδικοποιητή και του αποκωδικοποιητή ενός codec ομιλίας [13].

Στον κωδικοποιητή γίνεται επεξεργασία και ανάλυση της ομιλίας εισόδου, ώστε να εξαχθούν οι παράμετροι που αναπαριστούν το υπό εξέταση πλαίσιο δειγμάτων. Αυτές οι παράμετροι κωδικοποιούνται ή κβαντίζονται, και οι δυαδικοί δείκτες μεταδίδονται ως συμπιεσμένη ροή bit, αφού μπουν σε

πακέτα -αφού τοποθετηθούν δηλαδή με μια συγκεκριμένη, προκαθορισμένη σειρά.

Ο αποκωδικοποιητής, τώρα, αντλεί από τα πακέτα τη ροή των bit, και οι ανακτηθέντες δυαδικοί δείκτες οδηγούνται στον αντίστοιχο αποκωδικοποιητή παραμέτρων, έτσι ώστε να ληφθούν οι κβαντισμένες παράμετροι. Αυτές οι τελευταίες συνδυάζονται και γίνεται επεξεργασία τους, ώστε να παραχθεί η συνθετική ομιλία.

Κάθε codec έχει το δικό του, παρόμοιο μπλοκ διάγραμμα. Ο σχεδιαστής του εκάστοτε αλγορίθμου είναι εκείνος που καλείται να αποφασίσει τη λειτουργία και τα χαρακτηριστικά των διαφόρων μπλοκ επεξεργασίας, ανάλυσης και κβάντισης. Οι επιμέρους αποφάσεις του είναι εκείνες που θα καθορίσουν την απόδοση και τα χαρακτηριστικά του codec.

4.8 Το ανθρώπινο σύστημα ακοής

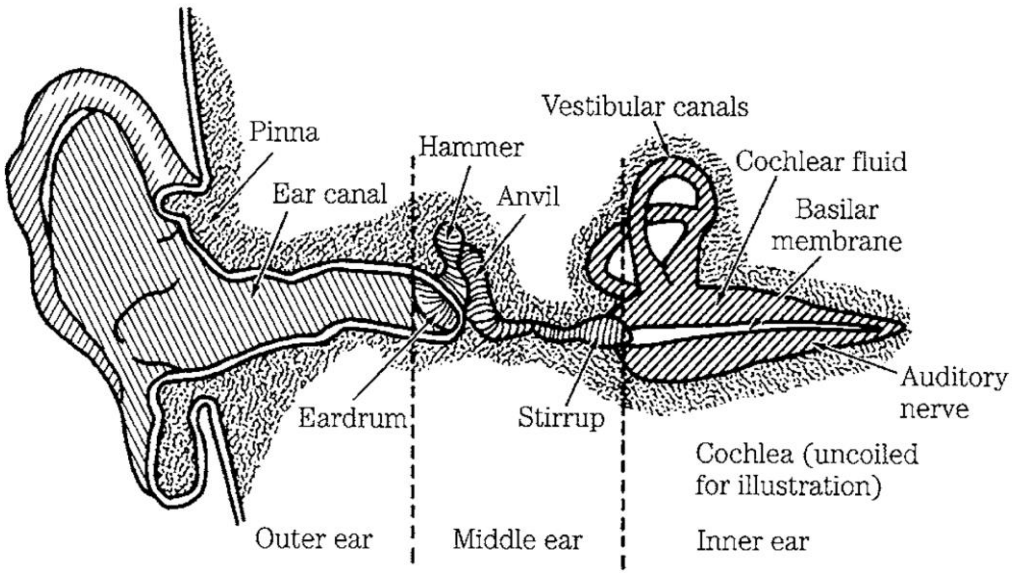
Η γνώση του τρόπου λειτουργίας του ανθρώπινου συστήματος ακοής παίζει σημαντικό ρόλο στη σχεδίαση των codec ομιλίας και ήχου. Η κατανόηση του πώς εκλαμβάνονται οι ήχοι οδηγεί στην πιο αποτελεσματική κατανομή των διατιθέμενων πόρων, και τελικά στη βελτίωση της αποτελεσματικότητας ως προς τη μείωση του κόστους. Σήμερα, τα περισσότερα codec σχεδιάζονται με στόχο την προσαρμογή στις ιδιότητες του συστήματος ακοής.

4.8.1 Η δομή του συστήματος ακοής

Το ανθρώπινο αφτί χωρίζεται σε τρία βασικά μέρη: το *εξωτερικό*, το *μέσο*, και το *εσωτερικό αφτί*.

Το εξωτερικό αφτί περιλαμβάνει το *πτερύγιο*, το *ακουστικό κανάλι*, και το *τύμπανο*. Το πτερύγιο χρησιμεύει για να κατευθύνει τα ηχητικά κύματα προς

το κανάλι, αλλά και για συνδρομή στον εντοπισμό της πηγής του ήχου -αφού εμποδίζει ήχους που προέρχονται από το πίσω μέρος. Το τύμπανο είναι μία μεμβράνη, η οποία μπορεί να δονείται σε ένα μεγάλο εύρος συχνοτήτων.



Σχήμα 4.11 Απλοποιημένη αναπαράσταση της φυσιολογίας του ανθρώπινου αφτιού [15].

Το μέσο αφτί είναι μία μικρή κοιλότητα, μέσα στην οποία εντοπίζονται τρία οστάρια: η *σφύρα*, ο *άκμωνας*, και ο *αναβολέας*. Η σφύρα είναι προσκολλημένη στο τύμπανο, και μεταδίδει, μέσω του άκμωνα και του αναβολέα, τις δονήσεις του τυμπάνου στην *ωοειδή θυρίδα*, στην οποία είναι προσκολλημένος ο αναβολέας. Ουσιαστικά, τα τρία οστάρια αποτελούν ένα σύστημα μοχλών, το οποίο ανιχνεύει τη δόνηση του τυμπάνου, την ενισχύει, και τη μεταφέρει στη μεμβράνη της ωοειδούς θυρίδας. Επίσης, σε περιπτώσεις πολύ ισχυρών ήχων, περιορίζει την πίεση και προστατεύει το εσωτερικό αφτί.

Η κοιλότητα του μέσου αφτιού περιέχει αέρα, και επικοινωνεί με τον φάρυγγα μέσω της *ευσταχιανής σάλπιγγας*, η οποία είναι συνήθως κλειστή. Στην κατάποση και στο χασμουρητό, όμως, ανοίγει, ώστε να εξισορροπηθούν οι πιέσεις μέσα και έξω από το τύμπανο. Κάτω από την ωοδειδή θυρίδα υπάρχει η *στρογγυλή θυρίδα*, η οποία επίσης φράσσεται με μεμβράνη.

Το εσωτερικό αφτί περιλαμβάνει τον *κοχλία* και το *όργανο του Corti*. Ο κοχλίας αποτελείται από τρία παράλληλα σωληνοειδή κανάλια, τα οποία είναι γεμάτα με υγρό, και τα οποία χωρίζονται μεταξύ τους από τη *μεμβράνη του Reissner* και τη *βασική μεμβράνη*. Το πάνω κανάλι ονομάζεται *αιθουσαία κλίμακα*, το μεσαίο *κοχλιακός πόρος*, και το κάτω *τυμπανική κλίμακα*. Το όργανο του Corti βρίσκεται στη μία πλευρά της βασικής μεμβράνης, και περιέχει τις απολήξεις των ακουστικών νεύρων, οι οποίες έχουν τη μορφή μικρών τριχών και εκτείνονται σε όλο το μήκος του κοχλία. Οι ίνες των νεύρων ξεκινούν από τα τριχοειδή κύτταρα και καταλήγουν στον πυρήνα του κοχλία, όπου σχηματίζουν ένα «μονό καλώδιο», το οποίο καταλήγει στον εγκέφαλο.

Όταν ένας ήχος εισέρχεται στο κανάλι και φτάνει στο τύμπανο, το θέτει σε ταλάντωση ανάλογη της συχνότητάς του. Η κίνηση του τυμπάνου μεταφέρεται στην ωοειδή θυρίδα, της οποίας η μεμβράνη αρχίζει επίσης να δονείται. Οι δονήσεις της μεταφέρονται στο υγρό του κοχλία, και έτσι τίθενται σε κίνηση τα τριχοφόρα κύτταρα, τα οποία με τη σειρά τους διεγείρουν τα ακουστικά νεύρα.

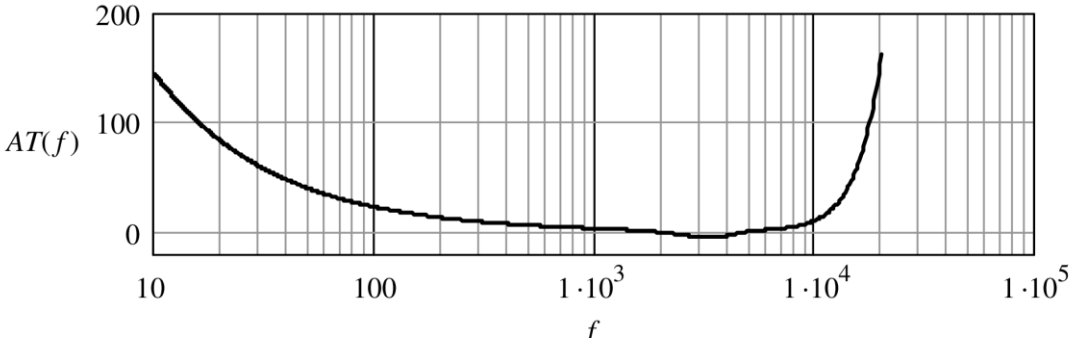
Διαφορετικά σημεία της βασικής μεμβράνης αντιδρούν διαφορετικά, ανάλογα με τη συχνότητα του εισερχόμενου ηχητικού κύματος. Έτσι, τριχοειδή κύτταρα που βρίσκονται σε διαφορετικές θέσεις κατά μήκος της μεμβράνης διεγείρονται από ήχους διαφορετικών συχνοτήτων. Οι αντίστοιχοι νευρώνες που μεταφέρουν τη διέγερση στα ανώτερα ακουστικά κέντρα διατηρούν την εν λόγω συχνοτική εξειδίκευση. Κάπως έτσι, το όλο σύστημα λειτουργεί σαν ένας αναλυτής συχνοτικού φάσματος· και ο χαρακτηρισμός του συστήματος είναι ευκολότερος αν γίνει στο πεδίο της συχνότητας.

4.8.2 Κατώφλι ακουστότητας

Το *κατώφλι ακουστότητας* ενός ήχου είναι η ελάχιστη ανιχνεύσιμη στάθμη αυτού, όταν απουσιάζει οποιοσδήποτε άλλος ήχος. Το μέγεθος χαρακτηρίζει

το ποσό ενέργειας που χρειάζεται ένας καθαρός τόνος ώστε να ανιχνευθεί από τον ακροατή σε ένα περιβάλλον χωρίς θορύβους.

Στο Σχήμα 4.12 απεικονίζεται μία τυπική καμπύλη κατωφλίου ακουστότητας, όπου ο οριζόντιος άξονας αναπαριστά τη συχνότητα (σε Hz), ενώ ο κάθετος το κατώφλι ακουστότητας (σε dB), για μία συχνότητα αναφοράς της τάξης των 10^{-12} W/m^2. Η συγκεκριμένη καμπύλη απεικονίζει τη μέση συμπεριφορά, ενώ είναι διαφορετική για κάθε ξεχωριστό άτομο.



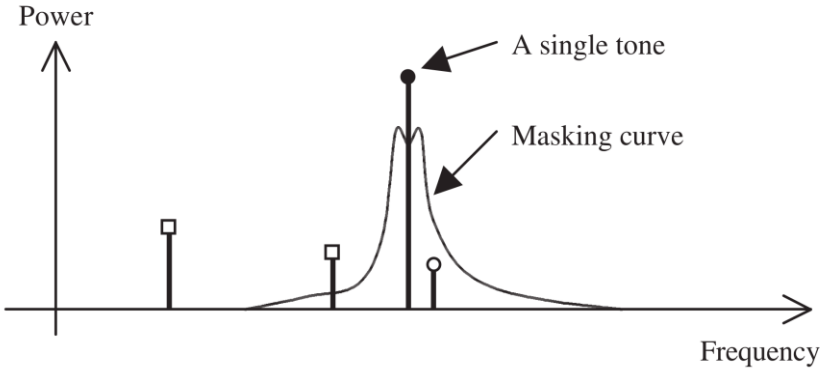
Σχήμα 4.12 Μία τυπική καμπύλη κατωφλίου ακουστότητας [13].

Όπως διαπιστώνεται από τη μελέτη της καμπύλης, το ανθρώπινο σύστημα ακοής κατά μέσο όρο είναι περισσότερο ευαίσθητο στη ζώνη συχνοτήτων από 1 έως 5 kHz, ενώ το κατώφλι ανεβαίνει απότομα στις πολύ χαμηλές και στις πολύ υψηλές συχνότητες.

Ο σχεδιαστής ενός codec μπορεί κάλλιστα να αξιοποιήσει το συγκεκριμένο χαρακτηριστικό, καθώς κάθε σήμα έντασης μικρότερης του κατωφλίου ακουστότητας δεν χρειάζεται να ληφθεί υπόψη, ενώ η ζώνη μεταξύ 1 και 5 kHz θα πρέπει να αναπαρασταθεί με απόδοση περισσότερων πόρων, αφού κάθε παραμόρφωση εντός αυτής της περιοχής θα γίνει πιο εύκολα αντιληπτή.

4.8.3 Φαινόμενο μάσκας ή κάλυψης

Το *φαινόμενο μάσκας* ή *απόκρυψης* ή *κάλυψης* παρατηρείται όταν ένας ήχος καθίσταται μη αντιληπτός, εξαιτίας της παρουσίας ενός άλλου ήχου. Για παράδειγμα, ένας καθαρός τόνος μπορεί να καλύψει γειτνιάζοντα σήματα· η ικανότητα κάλυψης είναι αντιστρόφως ανάλογη της απόλυτης διαφοράς συχνοτήτων. Μία σχετική περίπτωση απεικονίζεται στο Σχήμα 4.13, όπου



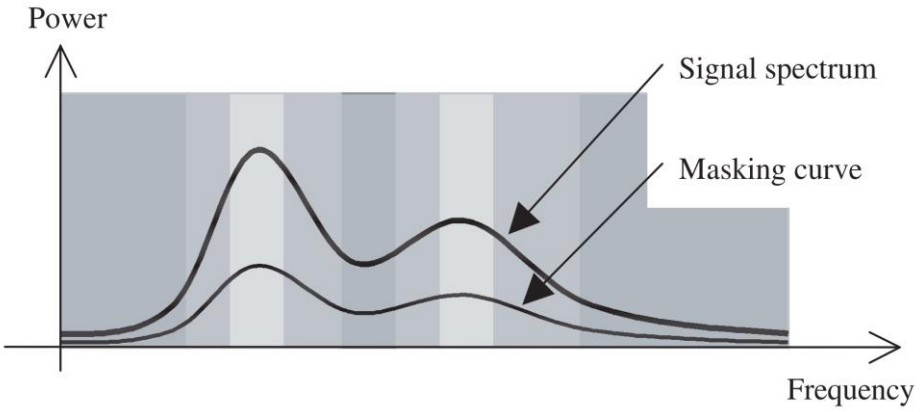
Σχήμα 4.13 Καμπύλη κάλυψης που σχετίζεται με έναν απλό τόνο. Τόνοι που σημειώνονται με τετράγωνο γίνονται αντιληπτοί, ενώ ο τόνος που σημειώνεται με κύκλο και του οποίου η ισχύς βρίσκεται κάτω από την καμπύλη δεν γίνεται αντιληπτός [13].

ένας απλός τόνος παράγει μία καμπύλη κάλυψης, η οποία καθιστά ανεπαίσθητο κάθε σήμα με ισχύ μικρότερη αυτής. Όσο αυξάνει η ένταση του σήματος αναφοράς (του καθαρού τόνου στη συγκεκριμένη περίπτωση), τόσο αυξάνει και η ικανότητα κάλυψης.

Τα χαρακτηριστικά της καμπύλης κάλυψης διαφέρουν για το κάθε άτομο, και για τον προσδιορισμό της απαιτείται μία διαδικασία μετρήσεων στο εργαστήριο.

Το εν λόγω φαινόμενο μπορεί να αξιοποιηθεί και για τους σκοπούς της κωδικοποίησης. Μία ανάλυση των συχνοτικού περιεχομένου ενός σήματος, λ.χ., θα μπορούσε να οδηγήσει στον εντοπισμό των περιοχών στις οποίες το σήμα είναι περισσότερο επιρρεπές σε παραμόρφωση. Στο Σχήμα 4.14 απεικονίζεται ένα τυπικό φάσμα, το οποίο αποτελείται από μία σειρά από

περιοχές άλλοτε υψηλής (κορυφές), και άλλοτε χαμηλής (κοιλάδες) ισχύος. Όπως γίνεται σαφές, η καμπύλη κάλυψης ακολουθεί τη μορφή της καμπύλης του αρχικού φάσματος, και έτσι σήματα με ισχύ κάτω από την καμπύλη κάλυψης είναι μη αντιληπτά. Άρα οι κορυφές μπορούν να αντέξουν περισσότερο την παραμόρφωση, σε σχέση με τις κοιλάδες· οπότε ένα καλά σχεδιασμένο codec θα πρέπει να εξασφαλίζει ότι οι κοιλάδες θα διατηρηθούν.



Σχήμα 4.14 Παράδειγμα συχνοτικού φάσματος, και της σχετικής καμπύλης κάλυψης. Οι περιοχές με σκούρο φόντο έχουν μικρή ανοχή σε παραμόρφωση, σε αντίθεση με εκείνες που έχουν πιο ανοιχτού χρώματος φόντο [13].

4.8.4 Αντιληπτότητα φάσης

Στις περισσότερες περιπτώσεις σχεδίασης codec ήχου, η εστίαση του μεγαλύτερου ποσοστού των προσπαθειών εντοπίζεται στην πληροφορία πλάτους του εκάστοτε σήματος. Αυτό συμβαίνει επειδή κυριαρχεί η άποψη ότι η ανθρώπινη ακοή δεν αντιλαμβάνεται δεδομένα που αφορούν στη φάση του σήματος.

Υπάρχουν αρκετά επιχειρήματα και στοιχεία που στηρίζουν την τελευταία πρόταση. Για παράδειγμα, ένας καθαρός τόνος και η χρονικά μετατοπισμένη εκδοχή του παράγουν κατ' ουσία την ίδια ακουστική αίσθηση. Επίσης, η αντιληπτότητα του θορύβου σχετίζεται κυρίως με το φάσμα πλάτους. Παρ' όλο τον μικρό ρόλο που παίζει η φάση στην αντιληπτότητα του ήχου, είναι

επιθυμητή μια κάποια φροντίδα για διατήρηση της εν λόγω πληροφορίας κατά την κωδικοποίηση, καθώς μέσω αυτής επιτυγχάνεται συνήθως μεγαλύτερη φυσικότητα.

4.9 Πρότυπα κωδικοποίησης ομιλίας

Η έννοια του προτύπου αφορά στην υιοθέτηση κάποιων κοινών αναφορών, από όλα τα εμπλεκόμενα στην αναζήτηση της λύσης για ένα πρόβλημα μέρη, προκειμένου να μπορεί να υπάρξει η καλύτερη δυνατή επικοινωνία. Ένα πρότυπο αναπτύσσεται από μία ομάδα ειδικών επί του εκάστοτε πεδίου, στη διάρκεια ενός μεγάλου χρονικού διαστήματος, μέσω δοκιμών και επαναξιολογήσεων κάθε σταδίου της διαδικασίας.

Έτσι, η μελέτη των προτυποποιημένων codec αφορά στην κατανόηση των πλέον επιδραστικών και επιτυχημένων ιδεών στο συγκεκριμένο πεδίο· ιδεών που εφαρμόστηκαν από οργανισμούς που είχαν τους πόρους ώστε να προβούν στο απαιτητικό, ούτως ή άλλως, εγχείρημα. Ένας μέσος όρος χρόνου που απαιτείται για την ολοκλήρωση των διαδικασιών σχεδίασης και εφαρμογής ενός τέτοιου προτύπου είναι περίπου 4.5 έτη.

Κάποιοι από τους γνωστότερους διεθνείς οργανισμούς που έχουν αναλάβει την επίβλεψη της ανάπτυξης και εφαρμογής σχετικών προτύπων είναι η International Telecommunications Union (ITU), η Telecommunications Industry Association (TIA), το European Telecommunications Standards Institute (ETSI), το United States Department of Defense (DoD), και το Research and Development Center for Radio Systems of Japan (RCR).

4.10 Αλγόριθμοι

Παρότι, όπως προαναφέρθηκε, η υλοποίηση ενός αλγορίθμου θεωρητικά μπορεί να γίνει είτε ως κώδικας ενός λογισμικού είτε ως κύκλωμα, σήμερα

το εν λόγω δίλημμα δεν υφίσταται, λόγω των σαφών πλεονεκτημάτων που έχει πια η πρώτη λύση.

4.10.1 Κώδικας αναφοράς

Η τάση που επικρατεί είναι να καταρτιστεί ένας πηγαίος κώδικας αναφοράς, γραμμένος σε κάποια προγραμματιστική γλώσσα υψηλού επιπέδου (π.χ. C ή C++). Επί αυτού του κώδικα αναφοράς εφαρμόζονται τα διάφορα τμήματα της κωδικοποίησης και αποκωδικοποίησης ομιλίας/ήχου, που αντιπροσωπεύουν τον κωδικοποιητή και τον αποκωδικοποιητή, αντίστοιχα.

Ο κώδικας αναφοράς, ακριβώς επειδή καταρτίζεται για γενική χρήση, συνήθως χρειάζεται βελτιστοποίηση ως προς την ταχύτητα και τις ανάγκες σε μνήμη, αλλά και ως προς τις συνθήκες και τις ανάγκες της εκάστοτε εφαρμογής. Ο σχεδιαστής καλείται σε κάθε περίπτωση να τροποποιήσει τμήματα του αλγορίθμου, ώστε να έχει το βέλτιστο αποτέλεσμα.

Προκειμένου να μπορεί να γίνει έλεγχος του πόσο επιτυχημένη είναι η εκάστοτε εφαρμογή του αλγορίθμου, οι οργανισμοί προτυποποίησης συχνά παρέχουν κάποια σύνολα διανυσμάτων ελέγχου· τροφοδοτώντας τον αλγόριθμο με ένα συγκεκριμένο διάνυσμα ελέγχου, θα πρέπει να ληφθεί ένα αντίστοιχο συγκεκριμένο διάνυσμα ελέγχου ως αποτέλεσμα. Η όποια απόκλιση θεωρείται ως αποτυχία, και ο σχεδιαστής καλείται να κάνει τις απαραίτητες παρεμβάσεις, έως ότου υπάρξει συμμόρφωση ως προς τα αναμενόμενα αποτελέσματα.

4.10.2 Η επιλογή της γλώσσας προγραμματισμού

Η καλή γνώση και χρήση της γλώσσας προγραμματισμού C είναι εδώ και πολλές δεκαετίες προαπαιτούμενο για την ανάπτυξη λογισμικού για εφαρμογές στο πεδίο της επεξεργασίας σήματος. Η δημοφιλία της οφείλεται

στις ευκολίες που παρέχει ως προς τη διαχείριση ενός μεγάλου εύρους εφαρμογών, και στην υψηλή αποτελεσματικότητά της ως προς την προσαρμογή σε οποιαδήποτε υπολογιστική πλατφόρμα –όπως και της μεταφοράς από πλατφόρμα σε πλατφόρμα.

Υπάρχουν, πάντως, και μειονεκτήματα στη χρήση της C -π.χ. η απουσία περιορισμών, που μπορεί να οδηγήσει σε προγραμματιστικές οδούς επιρρεπείς σε σφάλματα- και ομολογουμένως η χρήση της για λογισμικό εφαρμογών έχει μειωθεί αισθητά. Τη θέση της έχει πάρει σε μεγάλο βαθμό η γλώσσα C++.

Η C++ εφευρέθηκε αρχικά ως επέκταση της C, όμως στη συνέχεια επεκτάθηκε και ανεξαρτητοποιήθηκε σε μεγάλο βαθμό. Είναι μία αντικειμενοστραφής γλώσσα, η οποία προσφέρει αρκετά πλεονεκτήματα σε σχέση με τη C -όπως το ότι κώδικας C++ που υλοποιεί μεγάλα και σύνθετα συστήματα, με τα οποία εμπλέκεται μεγάλος αριθμός προγραμματιστών, είναι ευκολότερος στην κατανόηση.

Αυτό που ισχύει σε πολλές περιπτώσεις είναι ότι ο κώδικας αναφοράς για το κάθε codec διατίθεται σε γλώσσα C, και ότι η C++ χρησιμοποιείται για συντήρηση και επεκτασιμότητα. Η χρήση της, πάντως, οδηγεί και σε κάποιες αρνητικές επιπτώσεις ως προς την απόδοση, οι οποίες σχετίζονται με το πλήθος των επιπλέον χαρακτηριστικών της γλώσσας· γεγονός που κάνει τη χρήση της απαγορευτική για περιπτώσεις όπου η διαχείριση των πόρων είναι κρίσιμος παράγοντας.

4.10.3 Κόστη

Για την ανάλυση και τον σχεδιασμό οποιουδήποτε αλγορίθμου, θα πρέπει πάντα να υπάρχει πρόβλεψη των απαιτούμενων πόρων για την εφαρμογή του. Τα δύο κυριότερα κόστη είναι το *κόστος μνήμης* και το *υπολογιστικό κόστος*.

Το κόστος μνήμης αφορά σε μνήμη ROM, η οποία χρειάζεται για την αποθήκευση του κώδικα του αλγορίθμου, και σε μνήμη RAM, η οποία είναι εκείνη που κρατά τα δεδομένα εισόδου και εξόδου, καθώς και όποιες άλλες ενδιάμεσες τιμές μεταβλητών. Το περιεχόμενο της πρώτης παραμένει ουσιαστικά αμετάβλητο, ενώ της δεύτερης είναι μεταβλητό. Ο στόχος του σχεδιαστή είναι πάντοτε η βελτιστοποίηση του αλγορίθμου, ώστε να αποφεύγονται πλεονάζουσες πράξεις, και να μειώνεται το απαιτούμενο κόστος μνήμης. Κάποιες φορές, βέβαια, μείωση της απαιτούμενης μνήμης μπορεί να σημαίνει μικρότερη ταχύτητα εκτέλεσης του αλγορίθμου.

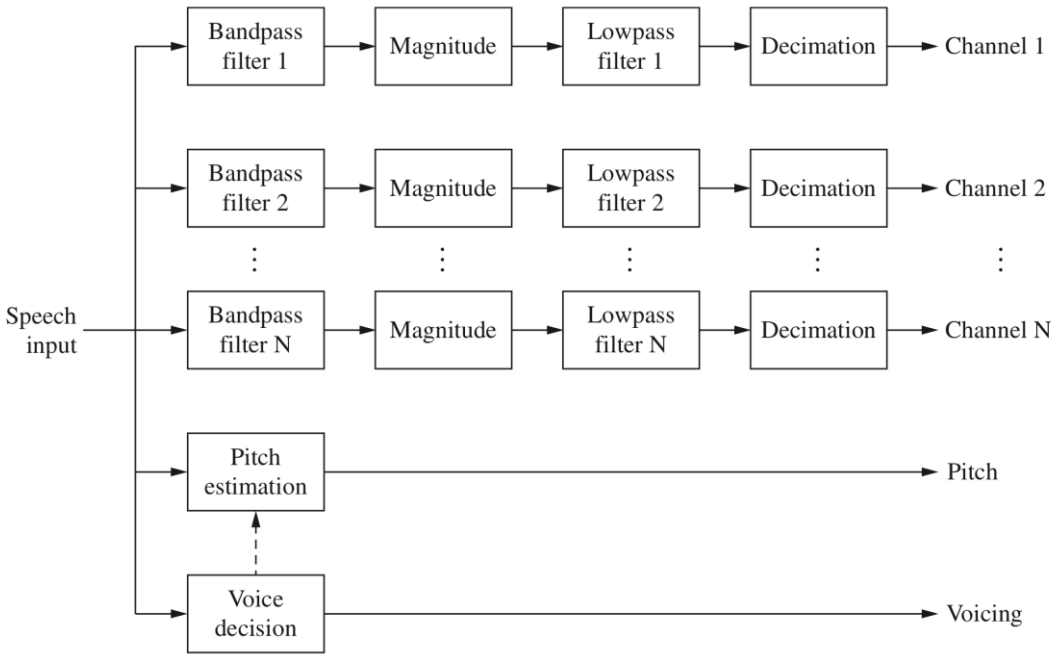
Με δεδομένη την είσοδο ενός αλγορίθμου, είναι πάντα επιθυμητό να λάβουμε την έξοδο σε όσο το δυνατόν λιγότερο χρόνο. Ο χρόνος αυτός μετράται με βάση τον αριθμό των πράξεων (αθροισμάτων, γινομένων κλπ.) που χρειάζεται να εκτελεστούν -αυτό είναι που αποκαλούμε υπολογιστικό κόστος του αλγορίθμου. Συγκρίνοντας διαφορετικούς αλγορίθμους για το ίδιο πρόβλημα, θεωρούμε πιο αποτελεσματικό εκείνον που απαιτεί μικρότερο αριθμό πράξεων.

Κεφάλαιο 5

Κωδικοποίηση φωνής: Μέθοδοι και αποτελέσματα

5.1 Αρχικά βήματα: Codec καναλιών, codec διαμορφωτών, και codec ημιτόνων

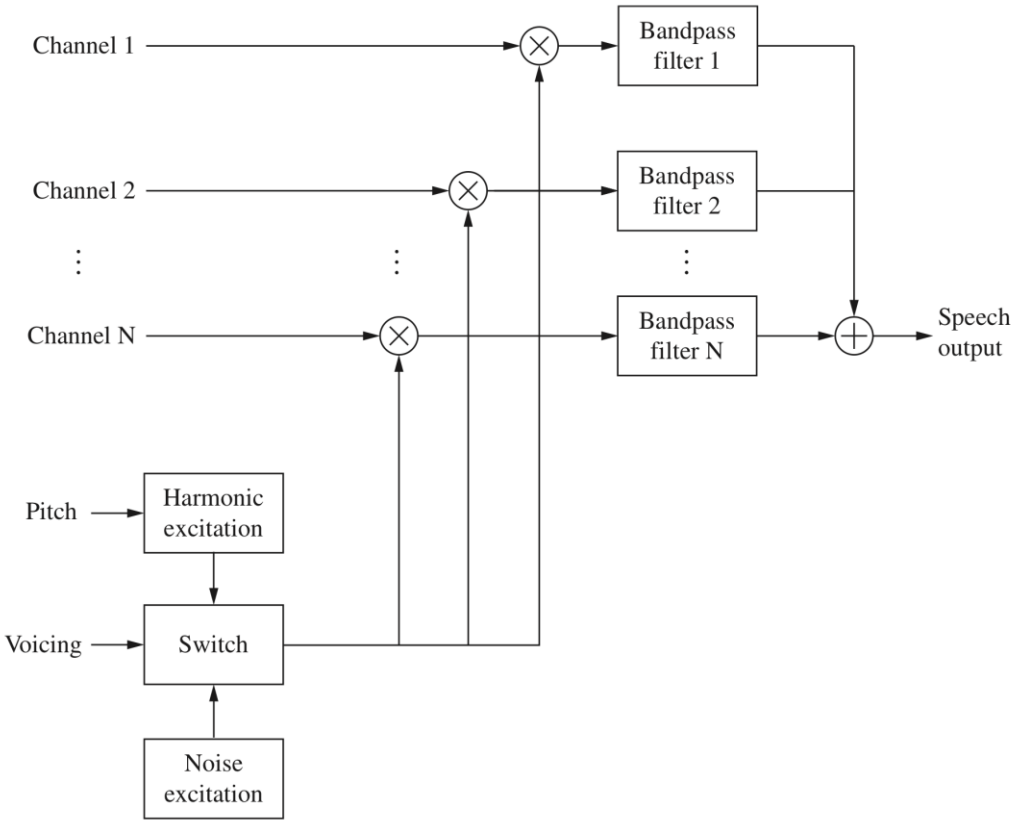
Ένα από τα πρώτα codec ομιλίας που υπήρξαν είναι ο *vocoder καναλιών*, όπως αποκαλείται συνήθως. Χρησιμοποιήθηκε για πρώτη φορά το 1939, και εισήγαγε την ιδέα ότι ένα σήμα ομιλίας μπορεί να εκφραστεί με τη μορφή συχνοτικών συνιστωσών.



Σχήμα 5.1 Κωδικοποιητής ανάλυσης ενός vocoder καναλιών [15].

Όπως εικονίζεται στο Σχήμα 5.1, το σήμα εισόδου λαμβάνεται ανά πλαίσιο (διάρκειας 10 έως 30 msec), και τεμαχίζεται σε συχνοτικές ζώνες με τη χρήση ζωνοπερατών φίλτρων. Οι ζώνες αυτές μπορεί να είναι στενότερες στις χαμηλές συχνότητες, για καλύτερη ανάλυση. Γίνεται εκτίμηση του

πλάτους της ενέργειας σε κάθε κανάλι, και στη συνέχεια γίνεται αποδεκατισμός της συχνότητας δειγματοληψίας. Επίσης, κάθε πλαίσιο χαρακτηρίζεται ως ηχηρό ή άηχο, και για τα ηχηρά γίνεται επιπλέον εκτίμηση της περιόδου του τονικού ύψους. Έτσι, στην έξοδο υπάρχουν δύο τύποι σημάτων: σήματα πλάτους και παράμετροι διέγερσης. Τα πρώτα μπορεί να κβαντιστούν λογαριθμικά, ή να κβαντιστεί η λογαριθμική διαφορά μεταξύ γειτονικών ζωνών, π.χ. με χρήση *Προσαρμοστικής Διαφορικής Παλμοκωδικής Διαμόρφωσης* (Adaptive Differential Pulse-Code Modulation, ADPCM). Οι παράμετροι διέγερσης των άηχων πλαισίων μπορεί να εκπροσωπηθούν από ένα μοναδικό bit. Τα ηχηρά πλαίσια σημειώνονται επίσης, και η περίοδος τονικού ύψους κβαντίζεται, π.χ. με μία λέξη των 8 bit.

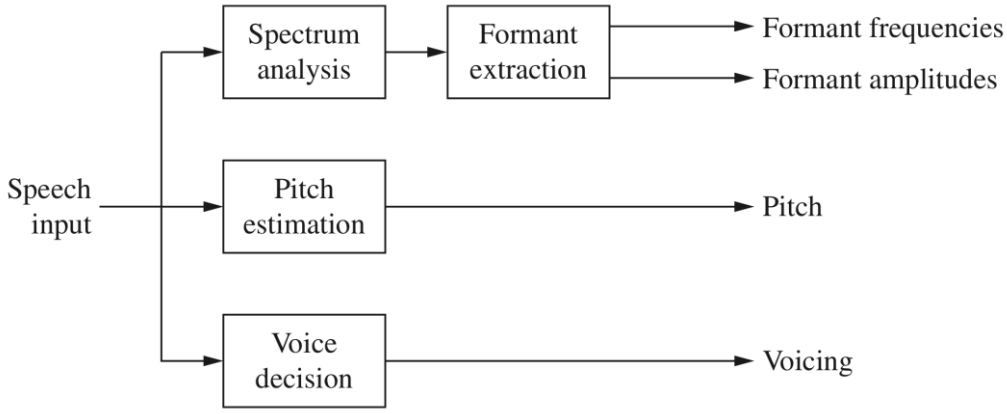


Σχήμα 5.2 Αποκωδικοποιητής σύνθεσης ενός vocoder καναλιών [15].

Στην πλευρά του αποκωδικοποιητή (Σχήμα 5.2) γίνεται χρήση των παραμέτρων διέγερσης ώστε να καθοριστεί ο τύπος του σήματος διέγερσης· για παράδειγμα, τα άηχα πλαίσια παράγονται από λευκό θόρυβο, ενώ τα

ηχηρά από αρμονική διέγερση που ρυθμίζεται από το τονικό ύψος. Αυτές οι πηγές εφαρμόζονται σε κάθε κανάλι, και τα μεγέθη τους προσαρμόζονται σε κάθε μία από τις υποζώνες.

Ένας τύπος codec που βελτίωσε την αποδοτικότητα του codec καναλιών, είναι εκείνος του *codec διαμορφωτών*. Εδώ αποφεύγεται η κωδικοποίηση όλης της φασματικής πληροφορίας από το σύνολο των συχνοτήτων, και επιλέγονται μόνο τα πλέον σημαντικά μέρη. Ο κωδικοποιητής αναγνωρίζει τους διαμορφωτές, εξετάζοντας διαδοχικά τα πλαίσια, και εντοπίζοντας τις κορυφές στην περιβάλλουσα του συχνοτικού φάσματος. Έπειτα, κωδικοποιεί τις συχνότητες των διαμορφωτών, και τα πλάτη τους. Όπως και στην περίπτωση των codec καναλιών, μεταφέρονται παράμετροι περί ηχηρότητας και τονικού ύψους. Όλα αυτά συνοψίζονται στο Σχήμα 5.3.

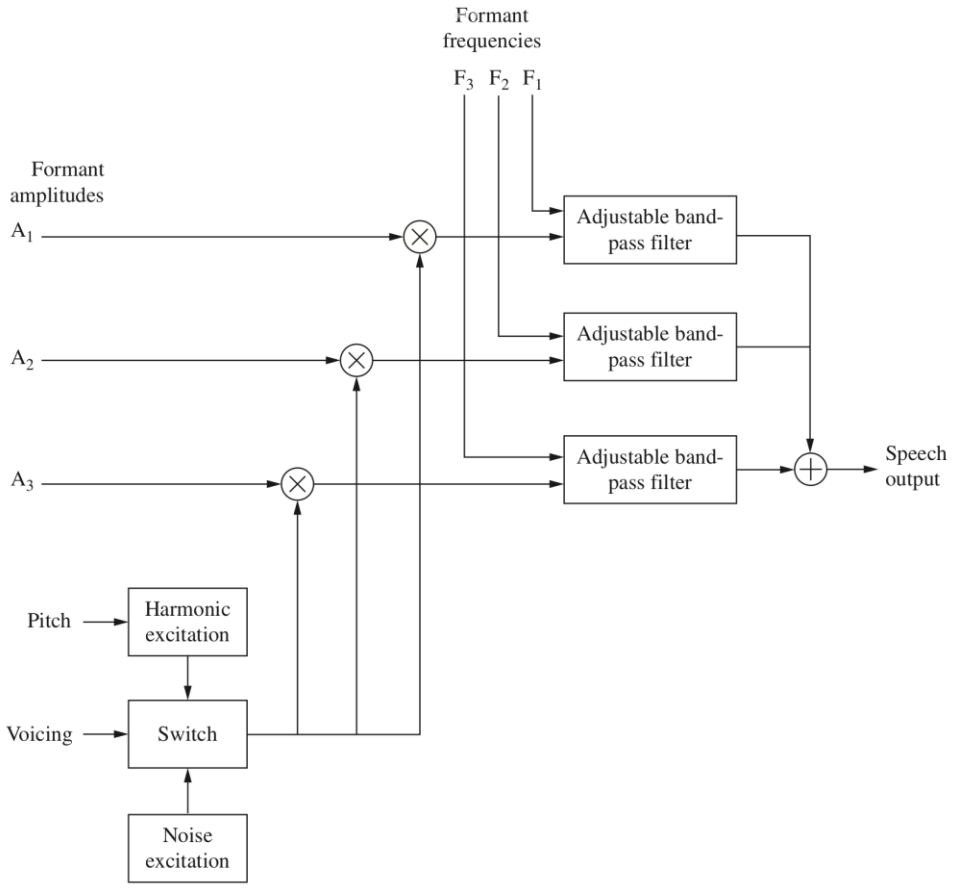


Σχήμα 5.3 Κωδικοποιητής ανάλυσης ενός vocoder διαμορφωτών [15].

Στον αποκωδικοποιητή (Σχήμα 5.4) το σήμα διέγερσης παράγεται με βάση τις παραμέτρους, και χρησιμοποιείται προκειμένου να ρυθμιστεί το πλάτος του διαμορφωτή. Αυτό εισάγεται σε μία συστοιχία ρυθμιζόμενων ζωνοπερατών φίλτρων, τα οποία είναι συντονισμένα μέσω της πληροφορίας που σχετίζεται με τις συχνότητες των διαμορφωτών, και οι έξοδοί τους αθροίζονται.

Λόγω του ότι υπάρχουν δυσκολίες στην ανίχνευση των συχνοτήτων των διαμορφωτών, σε περιπτώσεις, π.χ., όπου οι φασματικές κορυφές δεν σχετίζονται πλήρως με τους διαμορφωτές, η απόδοση τέτοιων μεθόδων

μπορεί να υπολείπεται αισθητά. Έτσι, η χρήση της κατηγορίας αυτής των codec αφορά κυρίως σε εφαρμογές σύνθεσης ομιλίας.



Σχήμα 5.4 Αποκωδικοποιητής σύνθεσης ενός vocoder διαμορφωτών [15].

Μία επόμενη κατηγορία είναι εκείνη των *codec ημιτόνων*. Σε αυτά, το σήμα διέγερσης αποτελείται από το άθροισμα ημιτονικών συνιστωσών, των οποίων το πλάτος, η συχνότητα και η φάση μεταβάλλονται, έτσι ώστε να προσεγγίσουν το αρχικό σήμα. Για παράδειγμα, για άηχο σήμα χρησιμοποιούνται διάφορες συνιστώσες τυχαίας φάσης, ενώ για ηχηρό σήμα οι συνιστώσες πρέπει να είναι συμφασικές, αρμονικά συσχετισμένες, και με συχνότητα που αποτελεί ακέραιο πολλαπλάσιο εκείνης του τονικού ύψους.

Το σήμα ομιλίας στην είσοδο του κωδικοποιητή μπορεί να υποστεί μετασχηματισμό Fourier, ώστε να προσδιοριστούν οι κορυφές του, και να ληφθούν πληροφορίες ηχηρότητας και τονικού ύψους. Τα πλάτη περνούν από αντίστροφο λογαριθμικό μετασχηματισμό για να κωδικοποιηθούν. Στον

αποκωδικοποιητή, όλες οι παράμετροι μετατρέπονται σε ημιτονικές, μέσω αντίστροφου μετασχηματισμού, και τα πλάτη, οι συχνότητες, και οι φάσεις τους χρησιμοποιούνται ώστε να συντεθεί το σήμα ομιλίας της εξόδου.

5.2 Προβλεπτική κωδικοποίηση: Η ιδέα πίσω από τα σύγχρονα codec ομιλίας

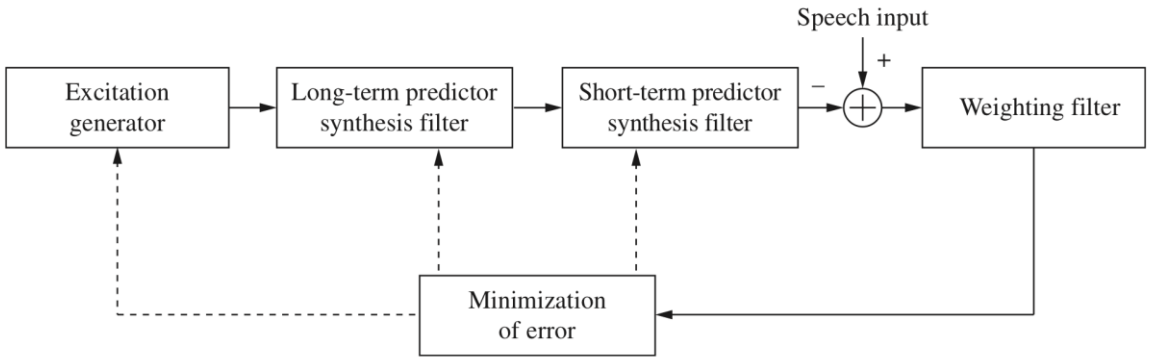
Η κυματομορφή ενός σήματος ομιλίας είναι, γενικά, απλή: αποτελείται από περιοδικές θεμελιώδεις συχνότητες, διακοπτόμενες από ριπές θορύβου. Μπορεί, ως εκ τούτου, να μοντελοποιηθεί ως παλμοσειρά, ως θόρυβος, ή ως συνδυασμός των δύο. Στην κωδικοποίηση ομιλίας, το σήμα στην έξοδο υποτίθεται ότι είναι η απόκριση σε ένα σήμα διέγερσης στην είσοδο. Μεταδίδοντας τις παραμέτρους που χαρακτηρίζουν τη διέγερση και τις επιδράσεις της φωνητικής οδού, το σήμα ομιλίας μπορεί να συντεθεί στον αποκωδικοποιητή.

Ένα σήμα ομιλίας αλλάζει με τον χρόνο, και είναι αυτές οι αλλαγές, των φθόγγων και του συχνοτικού περιεχομένου, που μεταφέρουν την περιεχόμενη πληροφορία. Εντούτοις, αν κανείς εστιάσει σε μικρές χρονικές περιόδους, θα παρατηρήσει ότι οι αλλαγές δεν είναι τόσο σημαντικές· τα σήματα ομιλίας, μάλιστα, χαρακτηρίζονται, υπό αυτό το πρίσμα, ως οιονεί στάσιμα. Κάποια σήματα, όπως π.χ. των φωνηέντων, έχουν πολύ υψηλή συσχέτιση από δείγμα σε δείγμα. Έτσι, πολλά codec ομιλίας λειτουργούν στη βάση μικρών χρονικών διαστημάτων (π.χ. μέχρι 20 msec), και χρησιμοποιούν *προβλεπτική κωδικοποίηση* προκειμένου να αφαιρέσουν πλεονασμούς. Στην εν λόγω κωδικοποίηση, η τιμή του τρέχοντος δείγματος εισόδου προβλέπεται με βάση τις τιμές των προηγούμενων αναδομημένων δειγμάτων. Έπειτα γίνεται κβάντιση της διαφοράς ανάμεσα στην πραγματική τρέχουσα τιμή και στην προβλεφθείσα τιμή, πράγμα που εξοικονομεί περισσότερα bit σε σχέση με το να κωδικοποιούνταν οι ίδιες οι τιμές. Στη συνέχεια, ο αποκωδικοποιητής χρησιμοποιεί την τιμή της διαφοράς για να αναδομήσει το σήμα. Πρόκειται για τεχνική που χρησιμοποιούν τα ADPCM

codec.

Σε χαμηλούς ρυθμούς δεδομένων, αντί της απευθείας κωδικοποίησης του σήματος της διαφοράς, είναι αποτελεσματικότερο να κωδικοποιηθεί ένα σήμα διέγερσης, το οποίο στη συνέχεια θα χρησιμοποιήσει ο αποκωδικοποιητής προκειμένου να συνθέσει ένα σήμα που θα είναι κοντά στο αρχικό. Ένας βραχυπρόθεσμος προβλέπτης περιγράφει τη φασματική περιβάλλουσα του σήματος ομιλίας, ενώ ένας μακροπρόθεσμος περιγράφει τη λεπτομερειακή φασματική δομή. Σε κάποιες υλοποιήσεις, ο μακροπρόθεσμος προβλέπτης παραλείπεται, ή αλλάζει η σειρά των προβλεπτών στην αλυσίδα.

Οι προβλέπτες και των δύο τύπων, βραχυπρόθεσμος και μακροπρόθεσμος, τροποποιούνται συνεχώς -αλλάζουν οι παράμετροι των φίλτρων τους, με βάση το σήμα ομιλίας- ώστε να βελτιώνεται η αποτελεσματικότητα της πρόβλεψης. Λόγω του ότι η μορφή της φωνητικής οδού αλλάζει σχετικά αργά, άρα εξίσου αργά αλλάζει και ο ήχος που εξέρχεται αυτής, οι παράμετροι μπορεί να τροποποιούνται με αντίστοιχα αργή ταχύτητα (π.χ. ο ρυθμός για τον βραχυπρόθεσμο προβλέπτη μπορεί να είναι από 30 έως 500 φορές/sec). Επειδή τυχόν απότομες αλλαγές στις συνιστώσες των φίλτρων μπορεί να προκαλέσουν ανεπιθύμητες ατέλειες που θα είναι ανιχνεύσιμες στο τελικό ακουστικό σήμα, υπάρχει η δυνατότητα χρήσης παρεμβαλλόμενων τιμών.



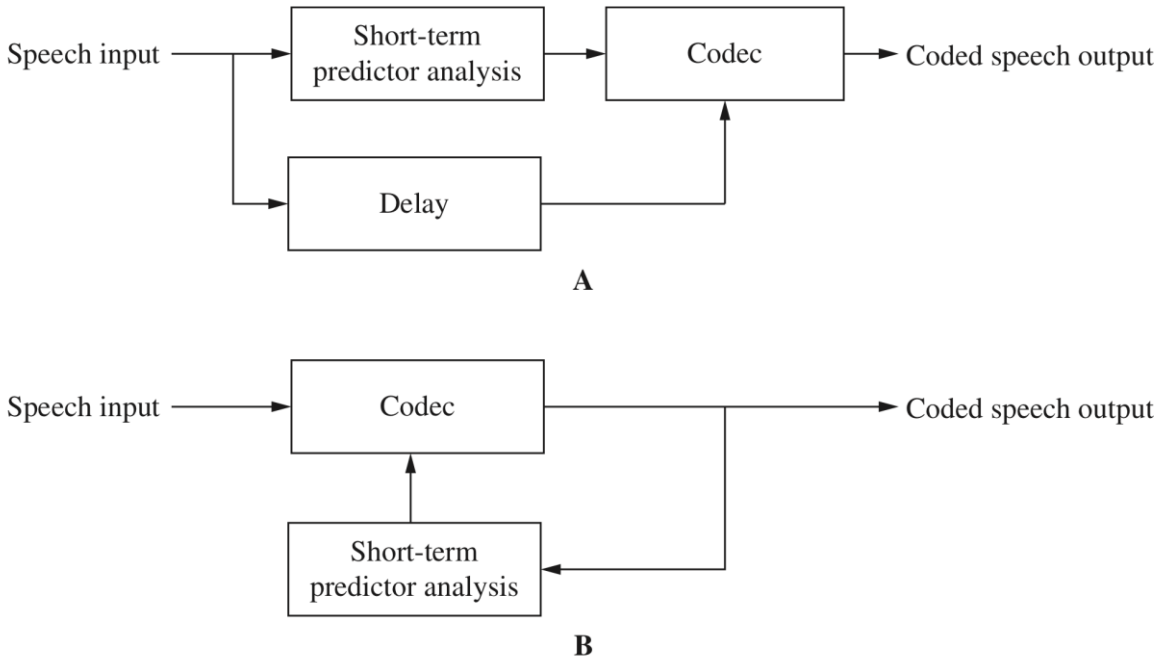
Σχήμα 5.5 Κωδικοποιητής ανάλυσης-μέσω-σύνθεσης, με βραχυπρόθεσμη και μακροπρόθεσμη πρόβλεψη [15].

Σε codec με διέγερση κώδικα, προκειμένου να μοντελοποιηθεί η περιοδικότητα του σήματος διέγερσης, χρησιμοποιείται ένα φίλτρο μακροπρόθεσμης πρόβλεψης, αφού ένα μικρό βιβλίο κωδικών (codebook) δεν μπορεί να αναπαραστήσει αποτελεσματικά την περιοδικότητα. Μία χρονική καθυστέρηση από 2.5 έως 18 msec αντιστοιχεί σε περιοδικότητα τονικού ύψους από 50 έως 400 Hz, που είναι το εύρος της πλειοψηφίας των ομιλητών. Οι συνιστώσες αναθεωρούνται σε ρυθμούς από 50 έως 200 φορές/sec.

Οι παράμετροι του προβλέπτη μπορεί να προσδιοριστούν από ένα τμήμα του σήματος ομιλίας, το οποίο θα έχει διάρκεια, π.χ., από 10 έως 30 msec. Στην προσαρμοστική προς τα εμπρός πρόβλεψη χρησιμοποιείται το αρχικό σήμα ομιλίας εισόδου, όπως φαίνεται και στο Σχήμα 5.6Α. Εφαρμόζεται χρονική υστέρηση προκειμένου να συναχθεί το απαιτούμενο τμήμα του σήματος, και οι παράμετροι μεταφέρονται ως τμήμα της ροής bit εξόδου. Στην προσαρμοστική προς τα πίσω πρόβλεψη, το αναδομημένο σήμα χρησιμοποιείται για την εκτίμηση των παραμέτρων, όπως δείχνει το Σχήμα 5.6Β. Λόγω ακριβώς της χρήσης του ανακτημένου σήματος, δεν είναι απαραίτητη η μεταφορά των παραμέτρων στη ροή bit εξόδου. Και οι δύο αυτές μέθοδοι μπορεί να χρησιμοποιηθούν για την εκτίμηση των παραμέτρων του βραχυπρόθεσμου προβλέπτη.

Οι παράμετροι του μακροπρόθεσμου προβλέπτη μπορεί να εκτιμηθούν με χρήση της μεθόδου ανάλυσης-μέσω-σύνθεσης, όπου οι γραμμικές εξισώσεις επιλύονται για μία βέλτιστη χρονική περίοδο υστέρησης. Εδώ δεν χρησιμοποιείται προσαρμοστική προς τα πίσω πρόβλεψη, διότι οι τιμές έχουν ευαισθησία σε σφάλματα στη μετάδοση. Κάποιες υλοποιήσεις, αντί του φίλτρου μακροπρόθεσμης σύνθεσης, χρησιμοποιούν ένα προσαρμοστικό βιβλίο κωδικών, το οποίο περιέχει εκδοχές της προηγούμενης διέγερσης. Σε ένα codec με διέγερση κώδικα, οι παράμετροι της διέγερσης μπορεί να προσδιοριστούν από τη δομή της διέγερσης, και αναζητώντας εκείνη τη σειρά από το βιβλίο κωδικών, η οποία δίνει το ελάχιστο σφάλμα με το σταθμισμένο σήμα ομιλίας.

Η απόκλιση μεταξύ του αρχικού σήματος και του ανακτημένου μπορεί να ελαχιστοποιηθεί κάνοντας χρήση ενός κριτηρίου όπως το μέσο τετραγωνικό σφάλμα. Με εξαίρεση τους πολύ χαμηλούς ρυθμούς bit, κάτι τέτοιο έχει ως αποτέλεσμα θόρυβο κβάντισης, με ίση ενέργεια σε όλες τις συχνότητες. Σαφέστατα, όσο χαμηλότερος είναι ο ρυθμός δεδομένων τόσο υψηλότερα

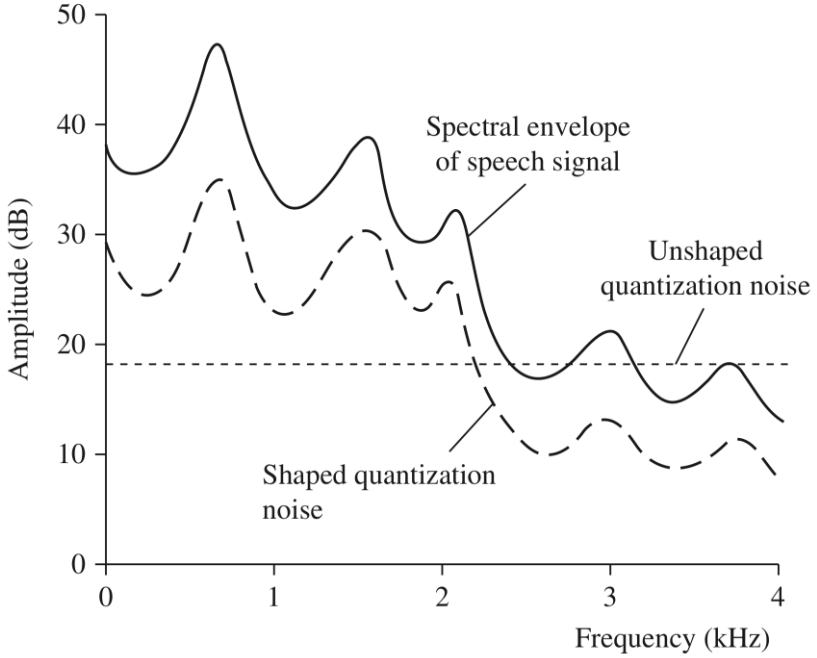


Σχήμα 5.6 Α. Προσαρμοστική προς τα εμπρός πρόβλεψη. **Β.** Προσαρμοστική προς τα πίσω πρόβλεψη [15].

είναι τα επίπεδα θορύβου. Η ακουστότητα του θορύβου, πάντως, μπορεί να ελαχιστοποιηθεί εάν μελετηθούν οι ιδιότητες του φαινομένου της κάλυψης. Για παράδειγμα, το φάσμα του θορύβου κβάντισης μπορεί να διαμορφωθεί έτσι ώστε η κατανομή του να συμπέσει με τις κορυφές των διαμορφωτών στη φασματική περιβάλλουσα του σήματος ομιλίας, όπως συμβαίνει στην περίπτωση του Σχήματος 5.7. Αυτού του είδους η διαμόρφωση θορύβου μπορεί να πραγματοποιηθεί με χρήση κατάλληλου φίλτρου στάθμισης σφάλματος. Σημειώνεται ότι η εν λόγω τεχνική αυξάνει το μέσο τετραγωνικό σφάλμα, αλλά μειώνει την ακουστότητα του θορύβου.

Σε codec με διέγερση κώδικα ή με διέγερση διανύσματος, κωδικοποιητής και αποκωδικοποιητής περιέχουν βιβλίο κωδικών με πιθανές αλληλουχίες bit. Ο

κωδικοποιητής μεταφέρει έναν δείκτη (επιλεγμένο ώστε να δώσει το μικρότερο σταθμισμένο μέσο τετραγωνικό σφάλμα) σε ένα διάνυσμα στο βιβλίο κωδικών. Αυτό χρησιμοποιείται για να καθοριστεί το σήμα διέγερσης για κάθε πλαίσιο. Η εν λόγω τεχνική διέγερσης από κώδικα είναι ιδιαίτερα αποτελεσματική, αφού ένας δείκτης απαιτεί μόλις 0.2 έως 2 bit/δείγμα.



Σχήμα 5.7 Διαμόρφωση του φάσματος του θορύβου κβάντισης, έτσι ώστε αυτός να μην είναι αντιληπτός [15].

Σε κάποιες περιπτώσεις, γίνεται χρήση ενίσχυσης της ομιλίας στο στάδιο μετά το φίλτρο, ώστε να τονιστούν εκείνες οι κορυφές διαμορφωτών του σήματος, οι οποίες είναι κρίσιμες ως προς την αντιληπτότητα. Κάτι τέτοιο μπορεί να βελτιώσει τη διακριτότητα, και να μειώσει την ακουστότητα του θορύβου. Οι μετά το φίλτρο παράμετροι μπορεί να προκύψουν από τις παραμέτρους της βραχυπρόθεσμης ή της μακροπρόθεσμης πρόβλεψης. Παρ' όλα αυτά, η συγκεκριμένη διαδικασία μπορεί να εισαγάγει θόρυβο και να χειροτερεύσει την ποιότητα άλλων σημάτων (μη ομιλίας).

5.3 Γραμμική Προβλεπτική Κωδικοποίηση (LPC)

Η Γραμμική Προβλεπτική Κωδικοποίηση (Linear Predictive Coding, LPC), ή απλώς *Γραμμική Πρόβλεψη*, χρησιμοποιείται συχνά για κωδικοποίηση ομιλίας, καθώς παρέχει μία αποτελεσματική τεχνική για ανάλυση ομιλίας. Λειτουργεί σε χαμηλούς ρυθμούς bit και είναι υπολογιστικά αποδοτική. Παράγει εκτιμήσεις παραμέτρων της ομιλίας κάνοντας χρήση ενός μοντέλου πηγής - φίλτρου, προκειμένου να αναπαραστήσει τη φωνητική οδό. Χρησιμοποιεί το άθροισμα των προηγούμενων σταθμισμένων δειγμάτων ομιλίας από το πεδίο του χρόνου για να προβλέψει το τρέχον σήμα ομιλίας. Κάτι τέτοιο μειώνει εγγενώς τον πλεονασμό στη βραχυπρόθεσμη συσχέτιση μεταξύ δειγμάτων.

Σε ένα απλό codec, η LPC χρησιμοποιείται στον κωδικοποιητή για ανάλυση του σήματος ομιλίας σε κάθε πλαίσιο, προκειμένου να εκτιμηθούν οι συνιστώσες του χρονομεταβλητού φίλτρου. Στη συνέχεια, αυτές οι συνιστώσες μεταδίδονται, μαζί με έναν παράγοντα κλίμακας. Ο αποκωδικοποιητής παράγει ένα σήμα λευκού θορύβου, το πολλαπλασιάζει με τον παράγοντα κλίμακας, και στη συνέχεια το περνάει από ένα φίλτρο που έχει ρυθμιστεί με τις προαναφερθείσες συνιστώσες. Η εν λόγω διαδικασία επικαιροποιείται και επαναλαμβάνεται για κάθε επόμενο πλαίσιο, παράγοντας ένα συνεχώς μεταβαλλόμενο σήμα ομιλίας.

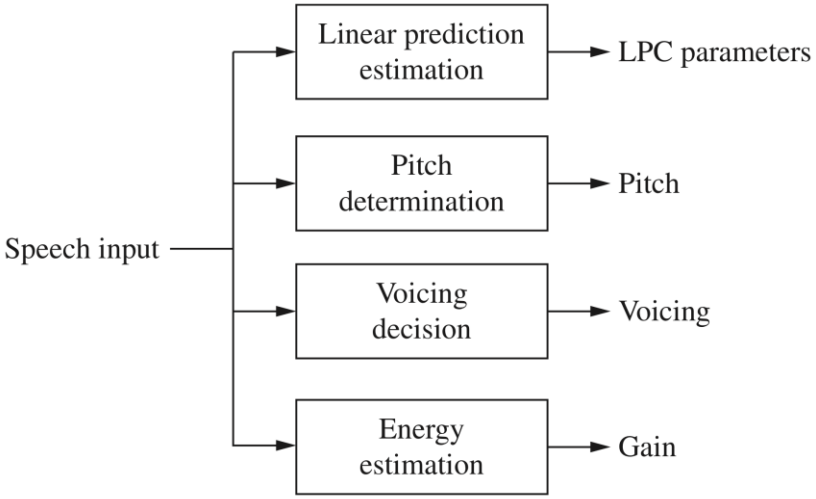
Η LPC είναι τεχνική εκτίμησης φάσματος: τα σήματα εισόδου και εξόδου ίσως διαφέρουν αισθητά αν ιδωθούν στο πεδίο του χρόνου, όμως τα συχνοτικά φάσματά τους θα μοιάζουν. Καθώς οι διαφορές στη φάση δεν γίνονται αντιληπτές, το σήμα εξόδου θα ακούγεται παρόμοιο με το αρχικό. Το γεγονός δε ότι μεταδίδονται μόνο οι συνιστώσες του φίλτρου και ο παράγοντας κλίμακας, οδηγεί σε μεγάλη μείωση του ρυθμού bit σε σχέση με το σήμα του οποίου έγινε αρχικά δειγματοληψία. Για παράδειγμα, για ένα πλαίσιο με 256 δείγματα μήκους 16 bit (δηλαδή 4096 bit συνολικά) χρειάζονται 40 bit για τις συνιστώσες και 5 bit για τον παράγοντα κλίμακας.

Στην πράξη, βέβαια, εφαρμόζονται επιπλέον τροποποιήσεις, οι οποίες βελτιώνουν περαιτέρω την αποτελεσματικότητα και την απόδοση.
Στα codec που χρησιμοποιούν γραμμική πρόβλεψη, το σήμα ομιλίας χωρίζεται σε δύο κομμάτια: μία φασματική περιβάλλουσα, και ένα υπολειπόμενο σήμα. Σε κάποιες υλοποιήσεις, η φασματική περιβάλλουσα εκπροσωπείται από συνιστώσες LP (ή συνιστώσες ανάκλασης), ενώ σε άλλες, το σήμα αντιπροσωπεύεται από γραμμικές φασματικές συχνότητες (line spectral frequencies, LSF). Το υπολειπόμενο σήμα (σήμα σφάλματος ή διαφοράς) του κωδικοποιητή LPC μπορεί να αναπαρασταθεί με διάφορους τρόπους: τα codec LPC χρησιμοποιούν αναπαράσταση παλμοσειράς (για ηχηρό φθόγγο) και θορύβου (για άηχο), τα codec ADPCM χρησιμοποιούν αναπαράσταση κυματομορφής, και τα παραμετρικά codec κάνουν χρήση παραμετρικής αναπαράστασης, με αρμονικές ή ημιτονικές κυματομορφές. Η όποια πληροφορία φάσης στο υπολειπόμενο σήμα δεν μεταφέρεται. Η επιλογή του codec εξαρτάται από την εφαρμογή και τον ρυθμό bit.

Κατ’ ουσία, τα LPC codec ομιλίας μοντελοποιούν το ανθρώπινο σύστημα ομιλίας σαν ένα κουδούνι που έχει τοποθετηθεί στο ένα άκρο ενός σωλήνα. Πιο συγκεκριμένα, το μοντέλο συνελίσσει τη γλωττιδική δόνηση με την απόκριση της φωνητικής οδού. Οι φωνητικές χορδές και το διάστημα μεταξύ τους, η γλωττίδα δηλαδή, παράγουν το «κουδούνισμα», το οποίο χαρακτηρίζεται από αλλαγές στην ένταση και το τονικό ύψος. Ο λαιμός και το στόμα της φωνητικής οδού μοντελοποιούνται σαν ένας σωλήνας, ο οποίος παράγει αντηχήσεις, δηλαδή διαμορφωτές. Το εν λόγω μοντέλο λειτουργεί καλά για τα φωνήεντα, αλλά λιγότερα καλά για ρινικούς ήχους· μπορεί, πάντως, να βελτιωθεί με την προσθήκη ενός παρακλαδιού το οποίο θα αντιπροσωπεύει τη ρινική κοιλότητα, αυξάνοντας όμως την υπολογιστική πολυπλοκότητα. Στην πράξη, οι ρινικοί ήχοι μπορεί να συνυπολογιστούν στο υπολειπόμενο σήμα.

Ένας κωδικοποιητής LPC απεικονίζεται στο Σχήμα 5.8. Όλη η ανάλυση γίνεται επί δεδομένων δειγματοληψίας, τα οποία έχουν χωριστεί σε πλαίσια. Η Γραμμική Πρόβλεψη εκτιμά τους διαμορφωτές στη φασματική περιβάλλουσα, και χρησιμοποιεί αντίστροφα φιλτράρισμα για να αφαιρέσει

την επίδρασή τους από το σήμα, δίνοντας ένα υπολειπόμενο σήμα. Στη συνέχεια, ο κωδικοποιητής εκτιμά την περίοδο τονικού ύψους και το κέρδος του υπολειπόμενου σήματος. Το πλαίσιο κατηγοριοποιείται ως ηχηρό ή άηχο· ως προς αυτό χρησιμοποιείται πληροφορία που αφορά στη σημασία της αρμονικής δομής του πλαισίου, στον αριθμό των διελεύσεων της κυματομορφής πεδίου του χρόνου από το μηδέν (η ηχηρή ομιλία έχει μικρότερο αριθμό διελεύσεων σε σχέση με την άηχη), ή στην ενέργεια στο πλαίσιο (η ηχηρή ομιλία έχει υψηλότερη ενέργεια). Οι διαμορφωτές, το υπολειπόμενο σήμα και λοιπές πληροφορίες αναπαρίστανται από τιμές που δύναται να αποθηκευθούν ή να μεταδοθούν.

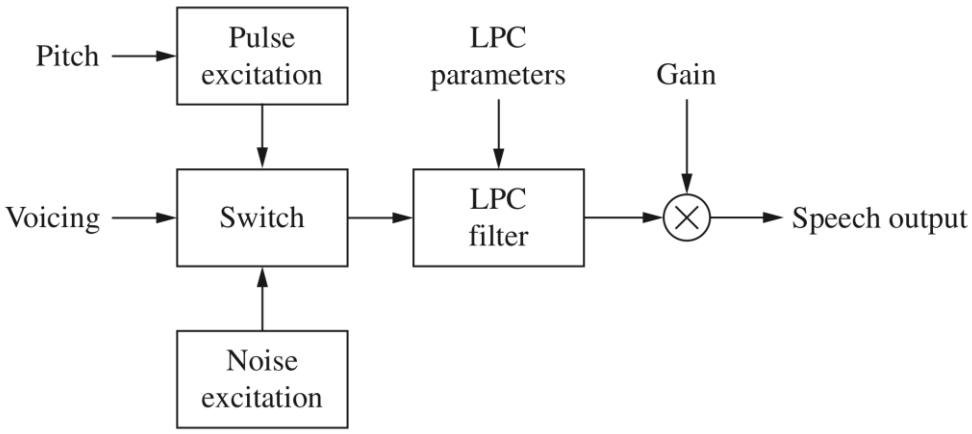


Σχήμα 5.8 Κωδικοποιητής LPC [15].

Στην πλευρά του αποκωδικοποιητή (Σχήμα 5.9) αντιστρέφεται η διαδικασία κωδικοποίησης και συντίθεται το σήμα ομιλίας. Χρησιμοποιείται το υπολειπόμενο σήμα για να σχηματιστεί ένα σήμα πηγής, και έπειτα χρησιμοποιούνται οι διαμορφωτές για φιλτράρισμα της πηγής και δημιουργία ομιλίας, λειτουργώντας έτσι ως το αρχικό μοντέλο του ανθρώπινου συστήματος ομιλίας. Το σήμα φωνής στην είσοδο ορίζει αν το πλαίσιο θα αντιμετωπιστεί ως ηχηρό ή ως άηχο. Όταν είναι ηχηρό, το σήμα διέγερσης είναι περιοδικό και παράγονται παλμοί με την κατάλληλη περίοδο τονικού ύψους. Όταν πρόκειται για άηχο πλαίσιο, χρησιμοποιείται τυχαία διέγερση θορύβου. Σε κάθε περίπτωση, το σήμα διέγερσης μορφοποιείται

από το φίλτρο LPC και το κέρδος αυτού προσαρμόζεται ώστε να ρυθμίσει το πλάτος του σήματος εξόδου.

Κατά τη διαδικασία εκτίμησης των διαμορφωτών ενός σήματος ομιλίας, κάθε δείγμα εκφράζεται ως γραμμικός συνδυασμός προηγούμενων δειγμάτων. Χρησιμοποιείται, ως προς αυτό, μία εξίσωση διαφοράς που ονομάζεται *γραμμικός προβλέπτης* (εξ ου και το όνομα της μεθόδου). Παράγονται έτσι συνιστώσες πρόβλεψης που χαρακτηρίζουν τους διαμορφωτές. Ο κωδικοποιητής εκτιμά της συνιστώσες μέσω της ελαχιστοποίησης του μέσου τετραγωνικού σφάλματος μεταξύ του πραγματικού σήματος και του προβλεπόμενου. Ως προς αυτό μπορεί να χρησιμοποιηθούν μέθοδοι όπως της αυτοσυσχέτισης, της συμμεταβλητότητας, και της αναδρομικής διαμέρισης πλέγματος. Η όλη διαδικασία τρέχει ανά πλαίσιο, με τον ρυθμό πλαισίων να κυμαίνεται μεταξύ 30 και 50 πλαισίων/sec. Ο συνολικός ρυθμός bit μπορεί να είναι 2.4 kbps.



Σχήμα 5.9 Αποκωδικοποιητής LPC [15].

Η απόδοση της LPC τεχνικής εξαρτάται εν μέρει από το ίδιο το σήμα ομιλίας. Οι ήχοι των φωνηέντων, οι οποίοι μοντελοποιούνται ως «κουδούνισμα», μπορεί να αναπαρασταθούν ως ακριβείς συνιστώσες του προβλέπτη: η συχνότητα και το πλάτος σημάτων αυτού του τύπου είναι εύκολο να αναπαρασταθούν. Αντίθετα, άλλα σήματα, ήχοι που παράγονται με βίαιη ροή αέρα μέσα από τη φωνητική οδό, και αφορούν σε σύμφωνα, τριβόμενα (κατά την προφορά τους στενεύει η φωνητική οδός) και κλειστά

(κατά την προφορά τους φράσσεται η φωνητική οδός), είναι πολύ διαφορετικά από ένα περιοδικό «κουδούνισμα». Ο κωδικοποιητής πρέπει να διαφοροποιεί την αντιμετώπισή του ανάλογα με το σήμα: για έναν «κουδουνιστό» (buzzing) ήχο μπορεί να εκτιμά τη συχνότητα και την ένταση, ενώ για έναν «συριστικό» (hissing) μπορεί να εκτιμά μόνο την ένταση.

Η απόδοση του αλγορίθμου μπορεί να μειώνεται για κάποιους ήχους που δεν εμπίπτουν στο μοντέλο LPC, όπως, π.χ., για ήχους που περιλαμβάνουν «κουδούνισμα» και «συριγμό» ταυτόχρονα, για κάποιους ρινικούς, με συγκεκριμένες θέσεις της γλώσσας, και για τραχειακές διεγέρσεις. Σε τέτοιες περιπτώσεις προκύπτουν ανακρίβειες στην εκτίμηση των διαμορφωτών, οπότε απαιτείται κωδικοποίηση επιπλέον πληροφορίας στο υπολειπόμενο σήμα. Κάτι τέτοιο μπορεί να αποφέρει βελτίωση της ποιότητας, αλλά σε βάρος του στόχου της μείωσης του όγκου των προς μετάδοση δεδομένων.

Όπως προαναφέρθηκε, οι κωδικοποιητές Γραμμικής Πρόβλεψης παράγουν ένα υπολειπόμενο σήμα μέσω φιλτραρίσματος του σήματος με ένα αντίστροφο βαθυπερατό φίλτρο. Αν αυτό το σήμα μεταφερόταν στον αποκωδικοποιητή και χρησιμοποιούνταν ως σήμα διέγερσης, το σήμα εξόδου θα ήταν ταυτόσημο του αρχικού. Εντούτοις, κάτι τέτοιο θα απαιτούσε ρυθμό δεδομένων που κρίνεται απαγορευτικός. Έτσι, απαιτείται μείωση του ρυθμού δεδομένων για τη μετάδοση του υπολειπόμενου σήματος. Όσο πιο κοντά είναι το κωδικοποιημένο υπολειπόμενο σήμα στο αρχικό, τόσο καλύτερη είναι η απόδοση του codec.

5.4 Γραμμική Πρόβλεψη Με Διέγερση Κώδικα (CELP)

Τα codec Γραμμικής Πρόβλεψης Με Διέγερση Κώδικα (Code-Excited Linear Prediction, CELP) είναι τα πλέον ευρέως χρησιμοποιούμενα για κωδικοποίηση ομιλίας. Ο σχετικός αλγόριθμος σχεδιάστηκε αρχικά από τους

Manfred Schroeder και Bishnu Atal, το 1983, και εκδόθηκε δύο χρόνια αργότερα. Αρχικά ο όρος CELP αναφερόταν σε έναν συγκεκριμένο αλγόριθμο, όμως πλέον χρησιμοποιείται για μία ευρεία γκάμα παραλλαγών (όπως ACELP, CS-ACELP, RCELP κλπ.). Codec που βασίζονται στην CELP χρησιμοποιούνται σε πολλά πρότυπα δικτύων κινητών επικοινωνιών, συμπεριλαμβανομένου του GSM, αλλά και στην τεχνολογία VoIP, καθώς και σε εφαρμογές λογισμικού όπως ο Windows Media Player. Μέρος της δημοφιλίας της τεχνικής οφείλεται στην ικανότητά της να αποδίδει καλά σε μεγάλο εύρος ρυθμών δυαδικών ψηφίων.

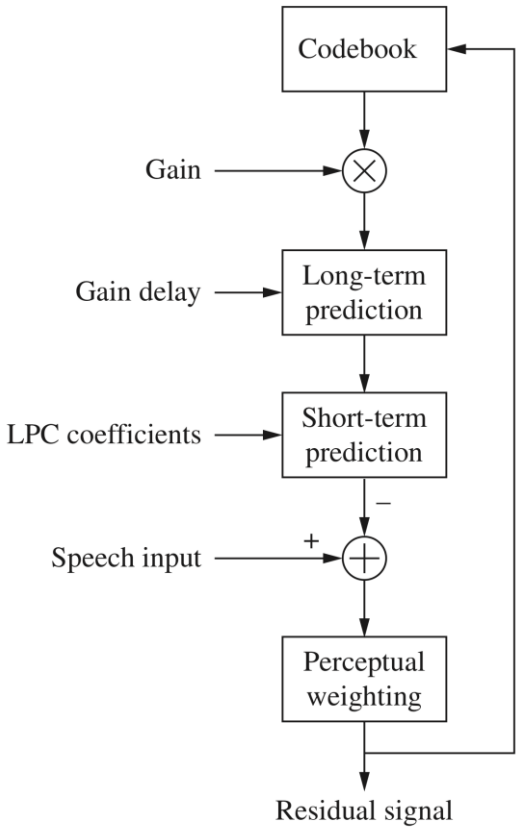
Ο αλγόριθμος CELP βασίζεται στη Γραμμική Πρόβλεψη. Κατ' αρχάς, βελτιώνει την απόδοση σε σχέση με την LPC, με το να διατηρεί βιβλία κωδικών των διαφόρων σημάτων διέγερσης, στον κωδικοποιητή και στον αποκωδικοποιητή. Ο κωδικοποιητής αναλύει το σήμα εισόδου, βρίσκει ένα αντίστοιχο σήμα διέγερσης στον κατάλογό του, και δίνει έναν κωδικό ταυτοποίησης. Ο αποκωδικοποιητής χρησιμοποιεί αυτόν τον κωδικό ταυτοποίησης για να βρει το σήμα διέγερσης στον δικό του κατάλογο, και χρησιμοποιεί αυτό το σήμα για να διεγείρει το φίλτρο διαμορφωτή.

Ως ολοκληρωμένο σύστημα, το codec CELP συνδυάζει ένα σύνολο τεχνικών κωδικοποίησης σε μία καινοτόμα αρχιτεκτονική. Η φωνητική οδός μοντελοποιείται ως φίλτρο με χρήση Γραμμικής Πρόβλεψης, με σήματα διέγερσης που περιέχονται στον κατάλογο κωδικών να χρησιμοποιούνται στην είσοδό του. Ο κωδικοποιητής χρησιμοποιεί μία αναζήτηση κλειστού βρόχου ανάλυσης-μέσω-σύνθεσης, η οποία πραγματοποιείται σε ένα αντιληπτικά σταθμισμένο πεδίο, ενώ χρησιμοποιείται και κβάντιση διανύσματος, για βελτίωση της αποτελεσματικότητας της κωδικοποίησης.

Ο πυρήνας του αλγορίθμου CELP απεικονίζεται στο Σχήμα 5.10. Ο κατάλογος κωδικών περιέχει εκατοντάδες τυπικών κυματομορφών υπολειπόμενου σήματος, οι οποίες είναι αποθηκευμένες σε μορφή διανύσματος. Κάθε υπολειπόμενο σήμα αντιστοιχεί σε ένα πλαίσιο του αρχικού σήματος, διάρκειας, π.χ., 5 έως 10 msec, και αντιπροσωπεύεται από μία καταγραφή στον κατάλογο κωδικών. Με χρήση μίας μεθόδου ανάλυσης-

μέσω-σύνθεσης επιλέγεται εκείνη η καταγραφή στον κατάλογο η οποία αντιστοιχεί πλησιέστερα στο υπολειπόμενο σήμα. Η έξοδος αυτή του καταλόγου κωδικών οδηγείται σε διαδικασία σύνθεσης σε έναν τοπικό αποκωδικοποιητή που υπάρχει στον κωδικοποιητή. Η επιλογή γίνεται με το ταίριασμα του συνθετικού σήματος ομιλίας που προκύπτει από το επιλεγμένο διάνυσμα, με το αρχικό σήμα ομιλίας· συγκεκριμένα, ελαχιστοποιείται το αντιληπτικά σταθμισμένο σφάλμα.

Όπως γίνεται εύκολα αντιληπτό, η χρήση καταλόγου κωδικών σημαίνει ότι ο αριθμός των πιθανών υπολειπόμενων σημάτων είναι πεπερασμένος. Ο



Σχήμα 5.10 Στον αλγόριθμο CELP πραγματοποιείται αναζήτηση τύπου ανάλυση-μέσω-σύνθεσης, σε κλειστό βρόχο, στο αντιληπτικά σταθμισμένο πεδίο. Το υπολειπόμενο σήμα στην έξοδο ελαχιστοποιείται αναζητώντας και επιλέγοντας τη βέλτιστη καταχώρηση στον κατάλογο κωδικών [15].

αποκωδικοποιητής περιλαμβάνει ένα στάδιο κέρδους για τη στάθμιση του υπολειπόμενου σήματος, ένα φίλτρο μακροπρόθεσμης πρόβλεψης, και ένα

φίλτρο βραχυπρόθεσμης. Το φίλτρο μακροπρόθεσμης πρόβλεψης περιέχει χρονικά μετατοπισμένες εκδοχές προηγούμενων διεγέρσεων, και λειτουργεί ως προσαρμοστικός κατάλογος κωδικών. Το συντεθειμένο σήμα αφαιρείται από το αρχικό, και η διαφορά εφαρμόζεται σε ένα φίλτρο αντιληπτικής στάθμισης. Η συγκεκριμένη αλληλουχία βημάτων επαναλαμβάνεται ώσπου ο αλγόριθμος να εντοπίσει την καταχώριση εκείνη του καταλόγου η οποία παράγει το ελάχιστο υπολειπόμενο σήμα· αυτή θα είναι και η τελική έξοδος. Στην πλευρά του δέκτη υπάρχει ο ίδιος κατάλογος κωδικών και ο ίδιος αποκωδικοποιητής, και χρησιμοποιούνται για να συντεθεί το σήμα ομιλίας εξόδου.

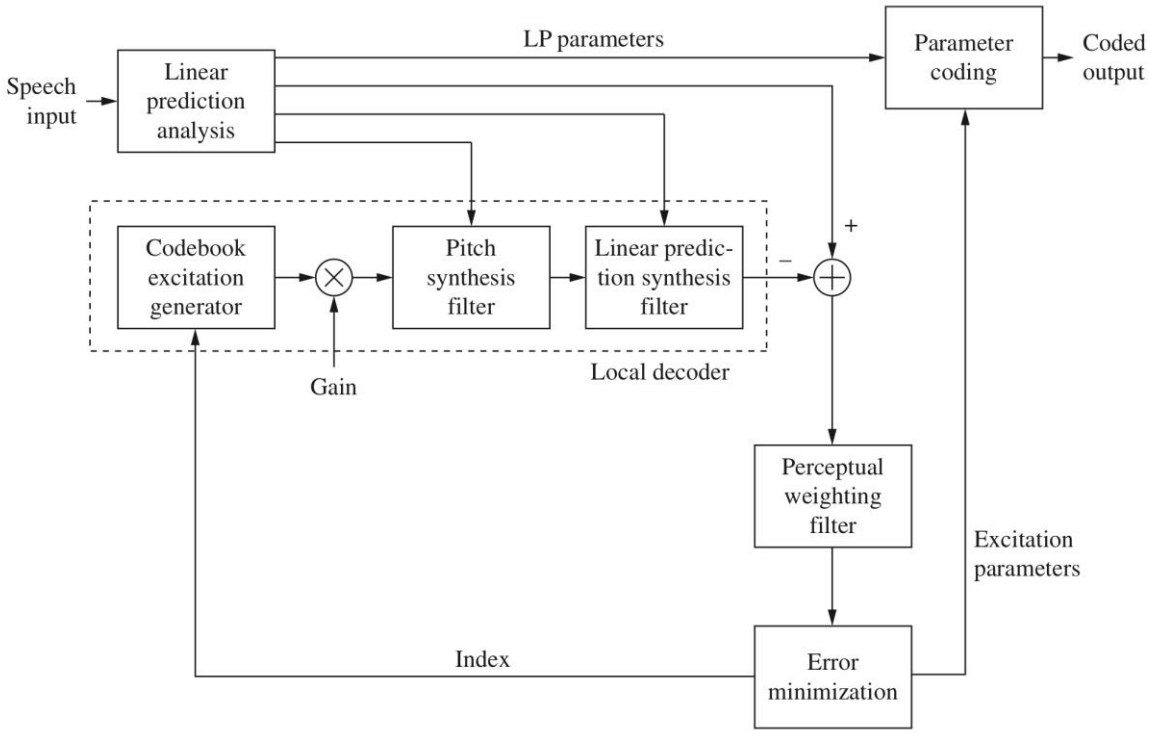
Το τελικό σήμα διέγερσης λαμβάνεται αθροίζοντας την πρόβλεψη τονικού ύψους με ένα νέο σήμα από έναν σταθερό κατάλογο κωδικών, ο οποίος είναι, κατ' ουσία, ένα λεξικό διανυσματικής κβάντισης που έχει ενσωματωθεί στο codec. Αυτός ο κατάλογος μπορεί να αποθηκευθεί αναλυτικά, ή μπορεί να είναι αλγεβρικός, όπως στην *Αλγεβρική Γραμμική Πρόβλεψη Με Διέγερση Κώδικα* (Algebraic Code-Excited Linear Prediction, ACELP). Το νέο σήμα είναι το μέρος εκείνο του σήματος που δεν θα μπορούσε να ληφθεί από τη Γραμμική Πρόβλεψη ή από την πρόβλεψη τονικού ύψους, και η κβάντισή του αφορά στα περισσότερα από τα bit ενός κωδικοποιημένου σήματος.

Τα CELP codec διαμορφώνουν επίσης το σήμα σφάλματος (θορύβου), έτσι ώστε η ενέργειά του να κατανέμεται κυρίως σε περιοχές όπου θα γίνεται λιγότερο ανιχνεύσιμο από την ανθρώπινη ακοή· ελαχιστοποιώντας έτσι την αντιληπτότητα του θορύβου. Επίσης, σε κάποιες περιπτώσεις, η απόδοση των σχετικών codec μπορεί να βελτιωθεί με αναπαραγωγή επιπλέον αρμονικού περιεχομένου, που χάθηκε κατά την κωδικοποίηση.

5.4.1 Η δομή του κωδικοποιητή και του αποκωδικοποιητή CELP

Το Σχήμα 5.11 απεικονίζει μία απλοποιημένη εκδοχή του κωδικοποιητή CELP.

Η διαδικασία της ανάλυσης-μέσω-σύνθεσης ξεκινά με ένα αρχικό σετ παραμέτρων κωδικοποίησης, το οποίο παράγεται από ανάλυση γραμμικής πρόβλεψης για να προσδιοριστεί η κρουστική απόκριση του φωνητικού συστήματος. Η γεννήτρια καταλόγου κωδικών διέγερσης και το φίλτρο γραμμικής πρόβλεψης αποτελούν έναν τοπικό αποκωδικοποιητή. Η έξοδος του τελευταίου δίνει ένα συντεθειμένο σήμα ομιλίας, το οποίο αφαιρείται από το αρχικό. Η προκύπτουσα διαφορά αποτελεί το σήμα σφάλματος, το οποίο χρησιμοποιείται για τη βελτίωση των παραμέτρων, μέσω της ελαχιστοποίησης του αντιληπτικά σταθμισμένου σφάλματος. Ο βρόχος εφαρμόζει επαναλαμβανόμενα το ελαχιστοποιημένο σφάλμα στον τοπικό

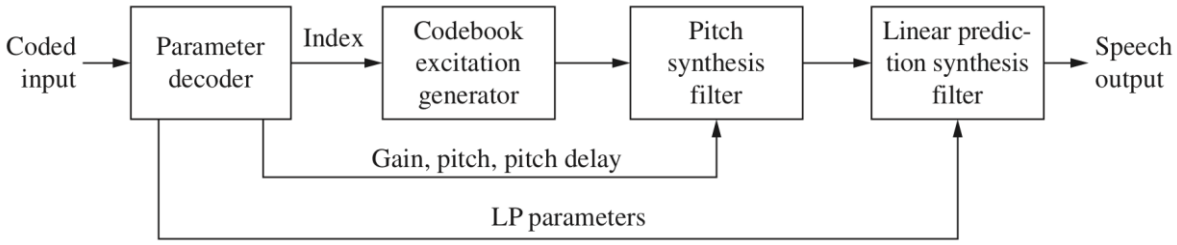


Σχήμα 5.11 Ο κωδικοποιητής CELP χρησιμοποιεί γραμμική πρόβλεψη και έναν τοπικό αποκωδικοποιητή ανάλυσης-μέσω-σύνθεσης, ο οποίος ελαχιστοποιεί το σήμα σφάλματος και δίνει στην έξοδό του παραμέτρους κωδικοποίησης [15].

αποκωδικοποιητή, ώσπου η έξοδός του να δώσει παραμέτρους διέγερσης οι οποίες να ελαχιστοποιούν βέλτιστα την ενέργεια του σήματος σφάλματος· αυτές περνούν στην έξοδο του κωδικοποιητή. Οι LPC παράμετροι ενημερώνονται για κάθε επόμενο πλαίσιο, ενώ οι παράμετροι διέγερσης ενημερώνονται συχνότερα, σε διαστήματα μικρότερα της διάρκειας

πλαισίου. Σε κάποιες περιπτώσεις, ο υπολογισμός αντιληπτικής στάθμισης εφαρμόζεται στο αρχικό σήμα ομιλίας εισόδου, αντί στο υπολειπόμενο σήμα. Με αυτόν τον τρόπο, ο υπολογισμός πραγματοποιείται μία φορά, και όχι σε κάθε επανάληψη του βρόχου.

Ο αποκωδικοποιητής CELP, που απεικονίζεται στο Σχήμα 5.12, μπορεί να ιδωθεί ως συνθετητής, καθώς χρησιμοποιεί μία τεχνική σύνθεσης ομιλίας για να δώσει σήμα στην έξοδό του. Στην είσοδό του λαμβάνει κβαντισμένη διέγερση και παραμέτρους LPC. Πληροφορίες κέρδους, τονικού ύψους, και υστέρησης τονικού ύψους εφαρμόζονται στο φίλτρο διαμορφωτή. Γίνεται χρήση LPC συνιστωσών, που αντιπροσωπεύουν τη φωνητική οδό, προκειμένου να αναδομηθεί το φίλτρο Γραμμικής Πρόβλεψης. Ο κατάλογος κωδικών χρησιμοποιεί τον ληφθέντα κωδικό - δείκτη για να βρει τα αντίστοιχα υπολειπόμενα διέγερσης, και χρησιμοποιεί αυτά τα σήματα για να διεγείρει το φίλτρο διαμορφωτή. Η έξοδος του τελευταίου εφαρμόζεται στο φίλτρο Γραμμικής Πρόβλεψης για να παραγάγει το συντεθειμένο σήμα ομιλίας εξόδου· κάτι που γίνεται με πρόβλεψη του σήματος εισόδου, κάνοντας χρήση του γραμμικού συνδυασμού προηγούμενων δειγμάτων. Όπως και με οποιαδήποτε πρόβλεψη, υπάρχει ένα σχετικό σφάλμα, το οποίο γίνεται προσπάθεια να ελαχιστοποιείται μέσω της όσο το δυνατό πιο ακριβούς επιλογής των συνιστωσών πρόβλεψης. Για την πραγματοποίηση του υπολογισμού χρησιμοποιείται ο αλγόριθμος Levinson - Durbin[4].



Σχήμα 5.12 Ο αποκωδικοποιητής CELP χρησιμοποιεί τις παραμέτρους που λαμβάνει για να εντοπίσει το αντίστοιχο υπολειπόμενο σήμα διέγερσης, και να το χρησιμοποιήσει για να διεγείρει το φίλτρο διαμορφωτή, δίνοντας στην έξοδό του, μέσω του LPC φίλτρου, το ανακατασκευασμένο σήμα ομιλίας [15].

[4] Προς τιμή του Αμερικανού μαθηματικού Norman Levinson (1912-1975) και του Βρετανού στατιστικολόγου James Durbin (1923-2012).

5.4.2 Βιβλία κωδικών

Όπως περιγράφηκε παραπάνω, η χρήση βιβλίων κωδικών στον κωδικοποιητή CELP προσφέρει υψηλή ποιότητα σήματος ομιλίας, αλλά και χαμηλό ρυθμό bit.

Όπως γίνεται εύκολα κατανοητό, η δοκιμή κάθε καταχώρησης σε έναν μακρύ κατάλογο κωδικών θα απαιτούσε μεγάλους χρόνους και απαγορευτικό υπολογιστικό κόστος. Ειδικά για τη μείωση του τελευταίου, αφιερώνεται μεγάλη προσοχή κατά τη σχεδίαση του αλγορίθμου CELP, όπως και στη δομή των βιβλίων κωδικών, ώστε η αναζήτηση εντός τους να γίνεται με όσο το δυνατό μεγαλύτερη αποτελεσματικότητα και ταχύτητα. Εντός των βιβλίων μπορεί να υπάρχει αλληλοεπικάλυψη: αντί ο κάθε κωδικός να αποθηκεύεται ξεχωριστά, όλες οι καταχωρήσεις αποθηκεύονται σε έναν πίνακα, όπου η κάθε κωδική λέξη αλληλεπικαλύπτεται με την επόμενη.

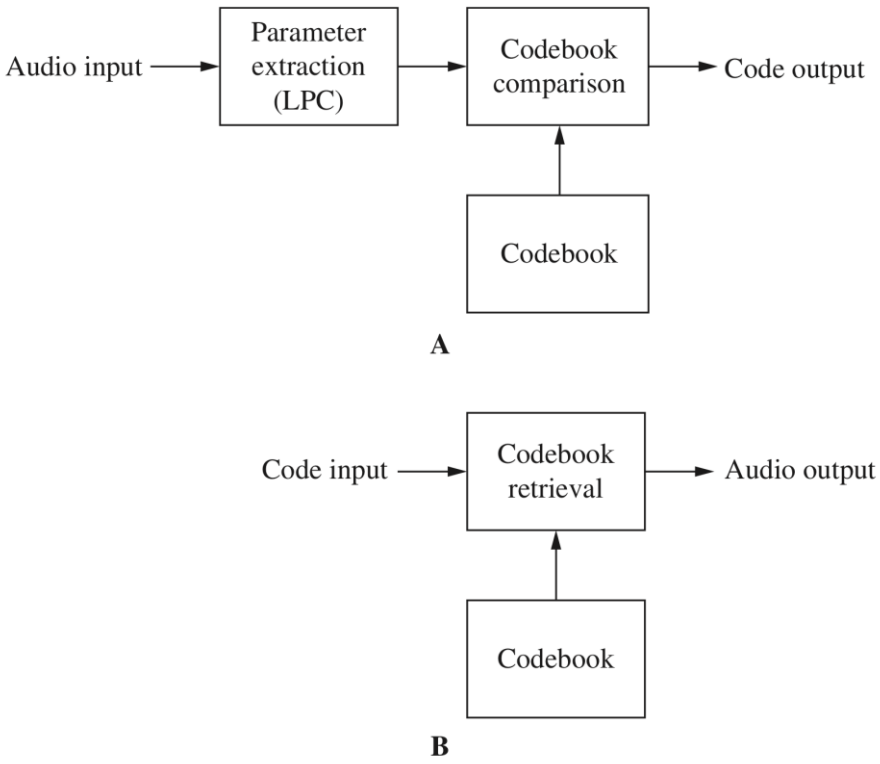
Γενικά, το μέγεθος ενός καταλόγου κυμαίνεται από 256 (κωδικός μήκους 8 bit) έως 4096 (μήκος 12 bit) καταχωρήσεις. Προς αποφυγή του προβλήματος των μεγάλων καταλόγων, υπάρχει η επιλογή της χρήσης δύο μικρότερων καταλόγων αντί ενός μεγάλου. Πιο συγκεκριμένα, ο ένας κατάλογος μπορεί να είναι σταθερός και ο άλλος προσαρμοστικός. Ο πρώτος ρυθμίζεται από πριν, και περιέχει συνιστώσες σήματος, τυχαίες και στοχαστικές, οι οποίες δεν μπορούν να προκύψουν από προηγούμενα πλαίσια. Ο δεύτερος κατάλογος είναι άδειος κατά την εκκίνηση, και χρησιμοποιείται κατά τα διαστήματα των ηχηρών ήχων, όπου το σήμα είναι περιοδικό, οπότε και ενημερώνεται με χρονικά μετατοπισμένες εκδοχές σημάτων διέγερσης που κωδικοποιήθηκαν σε προηγούμενα πλαίσια.

5.4.3 Κβάντιση Διανύσματος (VQ)

Η *Κβάντιση Διανύσματος* (Vector Quantization, VQ) είναι μία τεχνική που χρησιμοποιείται στους αλγόριθμους CELP (αλλά και σε άλλες περιπτώσεις).

Αντί για την κωδικοποίηση ξεχωριστών σημείων των δεδομένων, με την VQ επιχειρείται η κωδικοποίηση ολόκληρων ομάδων τέτοιων σημείων. Επιπλέον, αντί της κωδικοποίησης δειγμάτων του σήματος, γίνεται κωδικοποίηση παραμέτρων των εν λόγω δειγμάτων.

Το Σχήμα 5.13 δείχνει τη βασική δομή του κωδικοποιητή και του αποκωδικοποιητή VQ. Αρχικά γίνεται άντληση των παραμέτρων σε βάση πλαισίου, π.χ. με χρήση κωδικοποίησης Γραμμικής Πρόβλεψης. Αυτές συγκρίνονται με τις καταχωρήσεις που υπάρχουν σε έναν σταθερό κατάλογο κωδικών, εντοπίζεται εκείνη που προσομοιάζει καλύτερα, και ο κωδικοποιητής δίνει στην έξοδό του τον αντίστοιχο κωδικό - δείκτη.



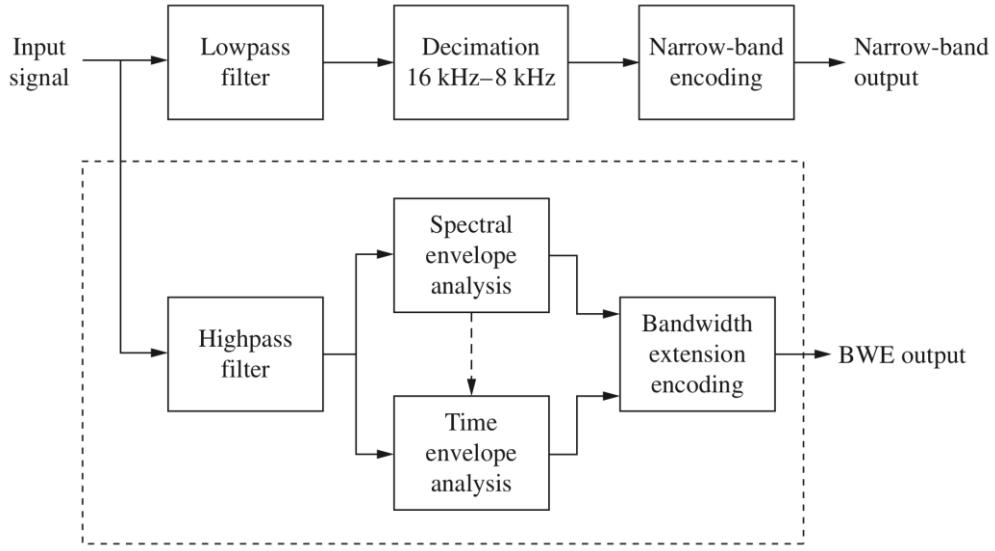
Σχήμα 5.13 Α. Κωδικοποιητής VQ. **Β.** Αποκωδικοποιητής VQ [15].

Στην πλευρά του αποκωδικοποιητή, ο οποίος επίσης διαθέτει τον ίδιο κατάλογο, εντοπίζεται το αντίστοιχο διάνυσμα, και αυτό χρησιμοποιείται προκειμένου να συντεθεί η ομιλία εξόδου.

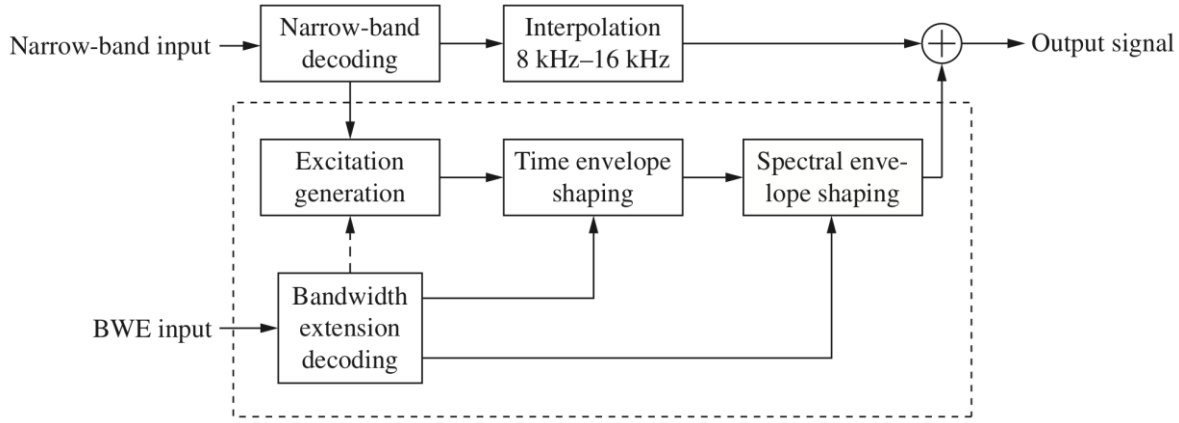
5.5 Επέκταση Εύρους Ζώνης (BWE)

Η *Επέκταση Εύρους Ζώνης* (Bandwidth Extension, BWE) είναι μία τεχνική που βελτιώνει την ποιότητα ήχου στα codec ομιλίας. Αυτό που ουσιαστικά επιτυγχάνει είναι να εμπλουτίσει τα δεδομένα ενός σήματος στενής ζώνης (NB) με επιπλέον δεδομένα, μετατρέποντάς το σε σήμα ευρείας ζώνης (WB), με μικρή μόνο αύξηση του αριθμού των απαιτούμενων bit. Ειδικά η επέκταση του εύρους προς την πλευρά των υψηλών συχνοτήτων βελτιώνει τη διακριτότητα, κυρίως σε ό,τι αφορά τα τριβόμενα σύμφωνα. Σημαντικό πλεονέκτημα των συστημάτων WB BWE είναι ότι έχουν προς τα πίσω συμβατότητα με τα NB συστήματα.

Η εφαρμογή της εν λόγω τεχνικής μπορεί να γίνει με διάφορους τρόπους. Μία εκδοχή είναι οι παράμετροι του NB σήματος να χρησιμοποιηθούν για να γίνει εκτίμηση του WB σήματος. Μπορεί να γίνει χρήση ενός μοντέλου πηγής – φίλτρου, παρόμοιου με εκείνο που χρησιμοποιείται στα codec που παρουσιάστηκαν νωρίτερα· όμως τώρα πηγή και φίλτρο αντιμετωπίζονται χωριστά. Το εύρος ζώνης του σήματος διέγερσης αυξάνεται, και το NB σήμα εφαρμόζεται σε φίλτρα ανάλυσης για ανασύνθεσή του, με προσθήκη χαμηλότερων και υψηλότερων συχνοτήτων. Οι αρμονικές των επιπλέον



Σχήμα 5.14 Κωδικοποιητής NB με επέκταση εύρους ζώνης. Τα σχετικά με την BWE μεταδίδονται ως πλευρική πληροφορία [15].



Σχήμα 5.15 Αποκωδικοποιητής NB με επέκταση εύρους ζώνης. Το σήμα στενής ζώνης αποκωδικοποιείται και επεκτείνεται στην άνω υποζώνη συχνοτήτων, μέσω παρεμβολής [15].

συχνοτήτων μπορεί να παραχθούν με διάφορους τρόπους: παραμόρφωση, φιλτράρισμα, κατοπτρισμό, μετατόπιση κλπ. Επιπλέον, μπορεί να εφαρμοστεί άντληση χαρακτηριστικών στο NB σήμα, έτσι ώστε να εκτιμηθεί η συχνοτική περιβάλλουσα.

5.6 Τεχνικές ακύρωσης ηχούς

Προκειμένου να βελτιωθούν οι επιδόσεις ενός codec ομιλίας, ο σχεδιαστής καλείται να μεριμνήσει για την αντιμετώπιση παραμορφώσεων που συχνά παρατηρούνται στην επικοινωνία μέσω δικτύων κινητών επικοινωνιών. Μία από τις σημαντικότερες (και πιο ενοχλητικές) παραμορφώσεις είναι η ηχώ (echo), η οποία προκύπτει είτε ακουστικά είτε ηλεκτρικά. Ακουστική ηχώ μπορεί, για παράδειγμα, να προκύψει κατά τη χρήση συσκευών hands-free, λόγω ακουστικής ζεύξης μεταξύ του ηχείου και του μικροφώνου. Ηλεκτρική ηχώ, από την άλλη, παράγεται π.χ. λόγω ηλεκτρομαγνητικής παρεμβολής ή φθαρμένου εξοπλισμού, ή -αν πρόκειται για εφαρμογή VoIP- λόγω αργής ταχύτητας σύνδεσης. Σε κάθε περίπτωση, η ηχώ με μεγάλο χρόνο επανάληψης είναι ιδιαίτερα ενοχλητική, και υποβαθμίζει δραστικά τη διακριτότητα της ομιλίας.

Μία τεχνική για τη μείωση του φαινομένου είναι εκείνη η οποία βασίζεται

σε «μεταγωγή κυκλώματος», που ελέγχεται από τη φωνή. Πιο συγκεκριμένα, όποτε ανιχνεύεται σήμα στο άλλο άκρο της σύνδεσης, γίνεται εξασθένηση του σήματος του μικροφώνου. Αντιστρόφως, όταν υπάρχει δραστηριότητα στην από εδώ πλευρά της σύνδεσης, γίνεται εξασθένηση του σήματος στο ηχείο. Βέβαια, όπως γίνεται εύκολα αντιληπτό, η συγκεκριμένη τεχνική δημιουργεί μία κατ' ουσία half-duplex επικοινωνία, ενώ εμφανίζονται και αφύσικες αλλαγές του θορύβου -οι οποίες, εντούτοις, μπορεί να εξομαλυνθούν μέσω μίας τεχνικής που ονομάζεται *Παραγωγή Θορύβου Ανακούφισης* (Comfort Noise Generation, CNG).

Υπάρχει επίσης η δυνατότητα χρήσης περιοριστή ηχούς, ο οποίος επικεντρώνεται στην ανίχνευση της ύπαρξης ηχούς, και όχι σήματος φωνής. Όταν συμβεί αυτό, εφαρμόζεται στη σύνδεση μεγάλη απώλεια ισχύος, έτσι ώστε αυτή στιγμιαία να διακοπεί. Σε περιπτώσεις όπου η στάθμη του σήματος ομιλίας είναι χαμηλή, ή όπου η στάθμη του σήματος ηχούς είναι υψηλή, μπορεί να προκύψουν σφάλματα λειτουργίας.

Μία άλλη προσέγγιση, πολύ πιο αποτελεσματική, είναι η προσαρμοστική ακύρωση ηχούς. Σε αυτήν την περίπτωση, δεν υπάρχει διακοπή της επικοινωνίας, ούτε τροποποίηση της οδού της σύνδεσης. Αντίθετα, παράγεται μία συνθετική ηχώ του εισερχόμενου σήματος ομιλίας, η οποία αργότερα αφαιρείται από το σήμα εξόδου, έτσι ώστε να ακυρώσει την ηχώ. Ως προς αυτό, χρησιμοποιείται ένα προσαρμοστικό φίλτρο, το οποίο εφαρμόζεται παράλληλα στο μονοπάτι του πραγματικού σήματος ηχούς, και το οποίο χρησιμοποιεί το σήμα του ηχείου προκειμένου να παραγάγει ένα αντίγραφο της ακουστικής ηχούς. Το τελευταίο αφαιρείται από το σήμα του μικροφώνου για να προκύψει το επιθυμητό σήμα ομιλίας εξόδου.

Η δυσκολία στην επιτυχία της εν λόγω μεθόδου έγκειται στον εντοπισμό του μονοπατιού της αυθεντικής ηχούς, και στη δημιουργία ενός ακριβούς αντιγράφου της, καθώς πάντα υπάρχει μία υπολειπόμενη ηχώ. Πολλοί προσαρμοστικοί αλγόριθμοι κάνουν επαναληπτική ενημέρωση των συνιστωσών του φίλτρου, χρησιμοποιώντας μηχανισμούς ελέγχου προκειμένου να βελτιστοποιήσουν το μέγεθος του βήματος κάθε

επανάληψης. Μικρό βήμα βελτιώνει την ακρίβεια, αλλά μπορεί να οδηγήσει σε αποτυχία ανίχνευσης του μονοπατιού της ακουστικής ηχούς, ενώ μεγάλο βήμα μπορεί να οδηγήσει σε απόκλιση του αλγορίθμου, σε περίπτωση θορυβωδών συνθηκών.

Κάποιες υλοποιήσεις χρησιμοποιούν συνδυασμό των προαναφερθέντων μεθόδων, ενώ γενικά υπάρχει η δυνατότητα χρήσης ενός προσαρμοστικού φίλτρου επιλογής συχνότητας στο τελικό στάδιο, το οποίο θα επιχειρεί την ελαχιστοποίηση της υπολειπόμενης ηχούς, κατόπιν της ακύρωσης, κάνοντας χρήση της φασματικής πυκνότητας ισχύος του υπολειπόμενου σήματος. Η λειτουργία του είναι παρόμοια με εκείνη ενός φίλτρου περιορισμού του θορύβου -μάλιστα, υπάρχει η δυνατότητα της ταυτόχρονης εφαρμογής περιορισμού του θορύβου. Η εκτίμηση, άλλωστε, της φασματικής πυκνότητας ισχύος της υπολειπόμενης ηχούς είναι δύσκολη εν μέσω θορύβου υποβάθρου.

5.7 Λοιπές τεχνικές

Στα πλαίσια των codec ομιλίας, και των προαναφερθεισών μεθόδων, χρησιμοποιούνται κάποιες ακόμα τεχνικές, που αφορούν στην εξομάλυνση της επικοινωνίας, αλλά και στην εξοικονόμηση πόρων.

Μία πρώτη τέτοια τεχνική είναι η *Ανίχνευση Φωνητικής Δραστηριότητας* (Voice Activity Detection, VAD), δηλαδή η ανίχνευση δραστηριότητας σε κάποια από τις πλευρές που συμμετέχουν σε μία συνδιάλεξη. Ουσιαστικά αναφερόμαστε στην παρακολούθηση του σήματος εισόδου, προκειμένου να διαπιστωθεί εάν το σήμα αφορά σε ομιλία ή σε θόρυβο υποβάθρου. Στη δεύτερη περίπτωση, το σήμα κωδικοποιείται με το χαμηλότερο δυνατό bit rate ή εναλλακτικά ενεργοποιείται λειτουργία σίγασης. Κατά τη διάρκεια αυτών των χρονικών διαστημάτων αδράνειας, ο αποκωδικοποιητής δύναται να παράγει θόρυβο ανακούφισης. Η VAD μπορεί να ενεργοποιείται αυτόματα, εάν πρόκειται για λειτουργία μεταβλητού bit rate.

Μία ακόμη τεχνική είναι η *Ασυνεχής Μετάδοση* (Discontinuous Transmission, DTX). Εδώ, κατά την περίπτωση που ανιχνευτεί ότι το σήμα εισόδου είναι θόρυβος, η μετάδοση δεδομένων διακόπτεται εντελώς. Με αυτόν τον τρόπο, γίνεται περαιτέρω εξοικονόμηση, λόγω της μείωσης του bit rate [15].

Τέλος, υπάρχει η τεχνική CNG, η οποία αναφέρθηκε στην προηγούμενη παράγραφο. Αυτή παράγει τον λεγόμενο θόρυβο ανακούφισης, δηλαδή συνθετικό θόρυβο, ο οποίος χρησιμοποιείται όχι μόνο στις ασύρματες επικοινωνίες, αλλά και στη μετάδοση ραδιοφωνικών εκπομπών. Η παραγωγή του έχει σκοπό να «γεμίζει» τα διαστήματα απόλυτης σιωπής, τα οποία προκαλούνται από την εφαρμογή της τεχνικής VAD, και την απόλυτη εξάλειψη θορύβου που αυτή επιτυγχάνει, ώστε να δημιουργείται μία οικεία αίσθηση στους συνδιαλεγόμενους [26].

Κεφάλαιο 6

Κωδικοποίηση μουσικής

Όπως περιγράφηκε σε προηγούμενα κεφάλαια, η λογική που ως επί το πλείστον ακολουθείται για την κωδικοποίηση σημάτων φωνής αξιοποιεί τεχνικές μοντελοποίησης της φωνητικής οδού. Τα codec που επικεντρώνονται σε κωδικοποίηση μουσικής, από την άλλη, παρότι έχουν τον ίδιο στόχο, ήτοι την υψηλότερη δυνατή ποιότητα του σήματος που θα μεταδοθεί, σε συνδυασμό με τον χαμηλότερο δυνατό ρυθμό δυαδικών ψηφίων, ακολουθούν αρκετά διαφορετικές μεθόδους. Πιο συγκεκριμένα, στην περίπτωση της μουσικής χρησιμοποιούνται είτε η PCM, είτε αλγόριθμοι *αντιληπτικής κωδικοποίησης* (perceptual coding), οι οποίοι αξιοποιούν τεχνικές μοντελοποίησης της ανθρώπινης ακοής.

6.1 Αντιληπτική κωδικοποίηση και ψυχοακουστικά μοντέλα

Σε ό,τι αφορά μουσικό περιεχόμενο, τα αναλογικά μέσα, αλλά και αρκετά ψηφιακά, στοχεύουν στην αποθήκευση της ακουστικής κυματομορφής με μια λογική αναλόγου, δηλαδή μίμησής της. Με στόχο την ανακατασκευή της με όσο το δυνατό πιο ακριβή, ως προς τη μορφή και τα μεγέθη, τρόπο.

Η αντιληπτική κωδικοποίηση είναι μία απολεστική (lossy) τεχνική, κατά την οποία η φυσική ομοιότητα υποχωρεί για χάρη της αντιληπτικής ομοιότητας. Ως προς αυτό, αξιοποιούνται ψυχοακουστικά μοντέλα του ανθρώπινου συστήματος ακοής, προκειμένου να εντοπιστεί μη αντιληπτό περιεχόμενο του σήματος, και να εξαιρεθεί από την ανάθεση δυαδικών ψηφίων. Έπειτα η κωδικοποίηση γίνεται με μέριμνα για αποφυγή πλεονασμού, και έτσι γίνεται εξοικονόμηση στον ρυθμό δυαδικών δεδομένων· ταυτόχρονα, όμως, αυξάνεται και ο θόρυβος κβάντισης. Ως προς την αντιμετώπιση αυτής της

αρνητικής επίπτωσης, εφαρμόζεται διαμόρφωση του θορύβου, προκειμένου αυτός να εντοπίζεται κάτω από το κατώφλι ακουστότητας.

6.1.1 Βασικά στοιχεία Ψυχοακουστικής

Η Ψυχοακουστική είναι η επιστήμη η οποία, σε συνεργασία με τη Φυσιολογία, ερευνά το πώς ο ήχος γίνεται αντιληπτός από τον άνθρωπο. Ή, διαφορετικά, είναι η επιστήμη η οποία μελετά τον ακροατή, και αναλύει το πώς οι βιολογικοί μηχανισμοί της ακοής, του νευρικού συστήματος και του εγκεφάλου, αντιδρούν στα διάφορα ηχητικά ερεθίσματα, και πώς γίνεται η ψυχολογική ερμηνεία αυτών, που οδηγεί στη δημιουργία συναισθημάτων, αντιλήψεων, και εμπειριών. Η Ψυχοακουστική εξηγεί την υποκειμενική μας απόκριση σε ό,τι ακούμε, και είναι ο τελικός κριτής αυτού· είναι, άλλωστε η δική μας αντίδραση που μετράει τελικά.

Η ακοή είναι αναντίρρητα η πλέον ανεπτυγμένη από τις ανθρώπινες αισθήσεις. Το σύστημα που την υποστηρίζει είναι εξαιρετικά σύνθετο, και διαθέτει εντυπωσιακές ικανότητες αντίληψης, αλλά διέπεται και από σαφείς περιορισμούς. Το αφτί είναι ιδιαίτερα ικανό να αντιληφθεί μικρές διαφοροποιήσεις στα ηχητικά κύματα, αλλά και εξίσου ανυποψίαστο για άλλες πλευρές των ίδιων ερεθισμάτων. Αυτό το τελευταίο χαρακτηριστικό σημαίνει ότι μπορούμε να μειώσουμε την ακρίβεια κάποιων πληροφοριών εντός του κωδικοποιημένου σήματος, σε βαθμό που, πάντως, εξαρτάται πολύ από τη συχνότητα και τον χρόνο.

Το δυναμικό εύρος των ήχων που το αφτί μπορεί να ανιχνεύσει είναι ιδιαίτερα μεγάλο. Για παράδειγμα, το κατώφλι αντίληψης ήχων στα 120 dB SPL έχει ηχητική ένταση που είναι 10^{12} φορές μεγαλύτερη από εκείνη που ισχύει για 0 dB SPL. Είναι, επίσης, αξιοσημείωτη η ευαισθησία του: στα 3 kHz ο ελάχιστος αντιληπτός ήχος μετατοπίζει το τύμπανο κατά διάστημα που αντιστοιχεί μόλις στο 1/10 της διαμέτρου ενός ατόμου υδρογόνου. Κάτι τέτοιο εξηγεί την ανάγκη για χρήση λογαριθμικής κλίμακας για την απεικόνιση των σχετικών μεγεθών.

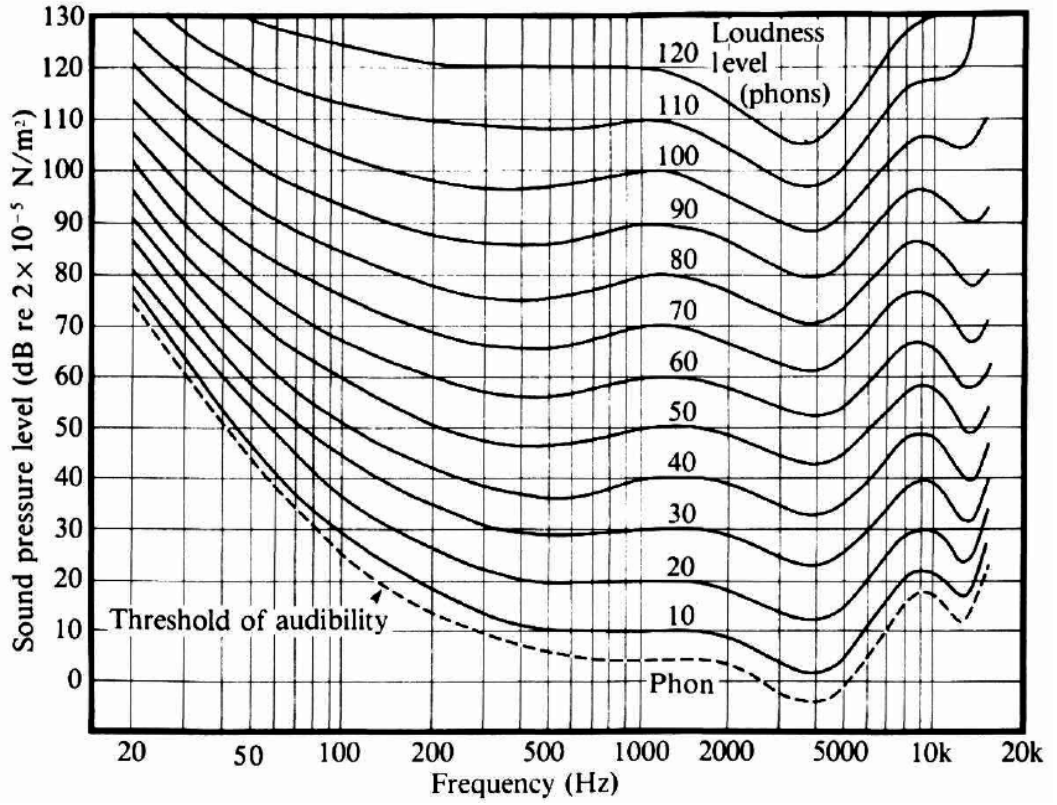
Το αφτί είναι επίσης πολύ γρήγορο στην απόκρισή του: εντός 500 msec από την ακρόαση ενός ήχου μέγιστης ισχύος, μπορεί να ανιχνεύσει έναν άλλο, ελάχιστης ισχύος. Το μάτι, αντίθετα, αργεί να προσαρμοστεί σε αλλαγές φωτισμού, και λειτουργεί σε κάθε στιγμή σε ένα μικρό μέρος του μέγιστου δυνατού εύρους του· ενώ το αφτί λειτουργεί σχεδόν ακαριαία, για οποιονδήποτε ήχο εντός του συνολικού εύρους του. Τέλος, ενώ το μάτι μπορεί να αντιληφθεί εναλλαγές του φωτός που διαρκούν τουλάχιστον 1/60 sec, η αντίστοιχη διάρκεια ήχου για το αφτί είναι 1/500 sec.

Η ανθρώπινη ακοή είναι αμφιωτική: οι ήχοι συλλαμβάνονται και από τα δύο αφτιά. Τα σήματα που φτάνουν στο κάθε αφτί δεν είναι πανομοιότυπα: υπεισέρχονται διαφορές στη φάση, αλλά και στην ένταση, λόγω της σκίασης από το κεφάλι και τα ωτικά πτερύγια. Όμως, ο άνθρωπος τελικά αντιλαμβάνεται έναν μόνο ήχο, καθώς στον εγκέφαλο γίνεται επεξεργασία των δύο ηχητικών κυμάτων· μέσω αυτής γίνονται αντιληπτά τα χαρακτηριστικά του ήχου, και η κατεύθυνση από την οποία αυτός έρχεται. Σε ό,τι αφορά τα στερεοφωνικά σήματα, αυτά σχεδιάζονται ως η συνήχηση δύο διαφορετικών μονοφωνικών καναλιών· η αίσθηση ότι ο προκύπτων ήχος έρχεται από το μέσο της απόστασης ανάμεσα στα δύο ηχεία είναι απλώς μία αυταπάτη.

Τα δύο αφτιά δεν διαφέρουν ως προς τη φυσιολογία τους, άρα και ως προς την ικανότητά τους να ανιχνεύσουν τους ήχους. Όμως, τα δύο ημισφαίρια του εγκεφάλου έχουν διακριτές διαφορές ως προς αυτό: το αριστερό διαχειρίζεται τη λεκτική πληροφορία, ενώ το δεξί διαχειρίζεται τη μελωδική (και κάθε άλλη, πλην της λεκτικής). Έτσι, και καθώς το δεξί αφτί συνδέεται με το αριστερό ημισφαίριο, και το αριστερό αφτί με το δεξί ημισφαίριο, είναι πιθανότερο ένας ακροατής να έχει καλύτερη αντιληπτική ικανότητα ως προς τη μουσική μέσω του αριστερού του αφτιού.

Η ένταση, και τα συγγενή της μεγέθη, τα οποία αναλύθηκαν στο Κεφάλαιο 3, μπορούν να μετρηθούν αντικειμενικά, με χρήση σχετικών οργάνων. Η *ακουστότητα* (loudness), από την άλλη, είναι ένα υποκειμενικό μέγεθος, το οποίο χαρακτηρίζεται από την ένταση και τη συχνότητα, ταυτόχρονα. Με

βάση αυτήν, διακρίνουμε εάν ένας ήχος είναι ισχυρός ή ασθενής. Η μέτρησή της δεν μπορεί να γίνει αντικειμενικά· αντί αυτού, καθορίζεται από σειρά μετρήσεων, στις οποίες συμμετέχουν εθελοντές ακροατές, και όπου καταγράφονται οι κρίσεις τους. Από μία τέτοια σειρά μετρήσεων, προέκυψαν και οι καμπύλες ίσης ακουστότητας που απεικονίζονται στο Σχήμα 6.1. Μονάδα μέτρησης της στάθμης ακουστότητας είναι το phon (αντικειμενική μονάδα), αλλά και το sone (υποκειμενική μονάδα).



Σχήμα 6.1 Καμπύλες ίσης ακουστότητας, που προέκυψαν από ψυχοακουστικές μελέτες των Fletcher και Munson, τις οποίες βελτίωσαν στη συνέχεια οι Robinson και Dudson· από αυτές προκύπτει ότι το ανθρώπινο αφτί έχει μη γραμμική απόκριση στη συχνότητα και στην SPL [28].

Παρότι το δυναμικό εύρος του αφτιού είναι ευρύ, η ευαισθησία του εξαρτάται από τη συχνότητα -όπως προαναφέρθηκε στο Κεφάλαιο 4, η μέγιστη ευαισθησία παρατηρείται στην περιοχή από 1 έως 5 kHz. Κάθε μία από τις καμπύλες του Σχήματος 6.1 αντιπροσωπεύει ένα εύρος συχνοτήτων, οι οποίες γίνονται αντιληπτές ως έχουσες την ίδια ακουστότητα. Η χαμηλότερη εξ αυτών ορίζει το κατώφλι ακουστότητας, την ελάχιστη,

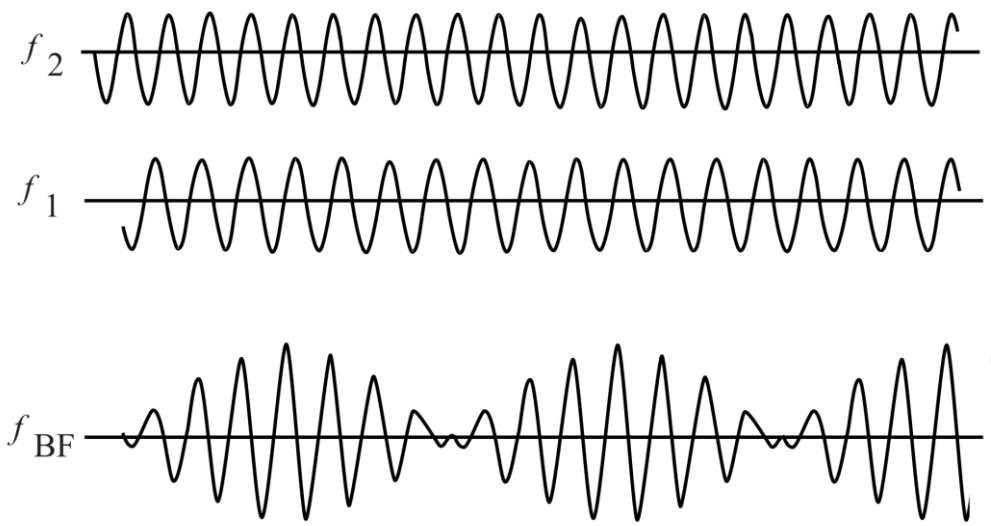
δηλαδή, στάθμη ακουστικής πίεσης που μπορεί ένας άνθρωπος με κανονική ακοή να αντιληφθεί.

Ένα άλλο ζεύγος αντικειμενικού - υποκειμενικού μεγέθους είναι εκείνο της συχνότητας και του τονικού ύψους (pitch). Το δεύτερο είναι ένα σύνθετο ηχητικό χαρακτηριστικό, το οποίο βασίζεται στη συχνότητα, αλλά και σε άλλους παράγοντες, όπως η κυματομορφή και η ένταση. Αν, για παράδειγμα, ενα ημιτονικό κύμα συχνότητας 200 Hz, ηχήσει με σταδιακά αυξανόμενη ένταση, ο ακροατής θα αντιληφθεί τον ισχυρότερο ήχο ως έχοντα χαμηλότερο τονικό ύψος. Δεν θα συμβεί το ίδιο, όμως, εντός της ζώνης 1 έως 5 kHz, όπου το τονικό ύψος θα μοιάζει σταθερό και ανεξάρτητο της ακουστότητας. Εν ολίγοις, είναι το τονικό ύψος εκείνο που δίνει στους ήχους τη μουσική διάστασή τους.

Η απόκριση της ανθρώπινης ακοής στη συχνότητα είναι λογαριθμική, και αυτό είναι κάτι που διαπιστώνεται μέσω της μελέτης της αντίληψής της στα μουσικά διαστήματα. Η απόσταση ανάμεσα στα 100 και στα 200 Hz γίνεται αντιληπτή ως μία οκτάβα· το ίδιο και η απόσταση ανάμεσα στα 1000 και στα 2000 Hz.

Ένα άλλο φαινόμενο που σχετίζεται με την ψυχοακουστική είναι το *διακρότημα* (beat). Πρόκειται για ταλαντώσεις στην ακουστότητα, οι οποίες προκύπτουν από τη συνήχηση δύο τόνων με ελάχιστα διαφορετικές συχνότητες, και οι οποίες έχουν συχνότητα ίση με τη διαφορά συχνοτήτων των δύο τόνων. Το διακρότημα γίνεται ιδιαίτερα αντιληπτό σε υψηλές συχνότητες, σε υψηλές εντάσεις, και όταν οι τόνοι δεν απέχουν παραπάνω από ένα διάστημα πέμπτης (δηλαδή 3:2).

Ένα ακόμα μέγεθος που ενδιαφέρει είναι η *αντιληπτική εντροπία* (perceptual entropy), η οποία εκφράζει το μέρος του ηχητικού σήματος το οποίο γίνεται αντιληπτό από την ανθρώπινη ακοή. Η αντιληπτική εντροπία μετράται σε bit/δείγμα, και μπορεί να έχει τιμές που ξεκινούν από 1.5 bit/δείγμα. Σήματα με χαμηλή εντροπία μπορεί να κωδικοποιηθούν με αποτελεσματική εξοικονόμηση, ενώ εκείνα με υψηλή όχι. Σήματα με διαφορές στην εντροπία



Σχήμα 6.2 Διακρότημα (BF) που δημιουργείται από τη συνήχηση δύο τόνων (f_1, f_2) με ελάχιστη συχνοτική διαφορά μεταξύ τους [9].

απαιτούν codec μεταβλητού bit rate, προκειμένου να κωδικοποιηθούν με τη μέγιστη δυνατή εξοικονόμηση.

Όπως, ίσως, προκύπτει από τα παραπάνω στοιχεία, το πεδίο της Ψυχοακουστικής είναι σύνθετο και ευρύ. Εκείνο που είναι βέβαιο είναι ότι ο σχεδιαστής codec χρειάζεται να λαμβάνει υπόψη του ότι ο τελικός αποδέκτης του σήματος είναι το ανθρώπινο σύστημα ακοής και ερμηνείας, και ότι η πρόσληψη της μουσικής εμπειρίας διέπεται από υποκειμενικότητα.

6.1.2 Κρίσιμες ζώνες

Έχει αποδειχθεί μέσω πειραμάτων ότι, όποτε κάποιο σήμα θορύβου καλύπτει έναν καθαρό τόνο, οι συχνοτικές συνιστώσες του θορύβου οι οποίες παίζουν ρόλο στην κάλυψη είναι εκείνες οι οποίες βρίσκονται κοντά στη συχνότητα του τόνου· όλο το υπόλοιπο συχνοτικό περιεχόμενο δεν επηρεάζει την κάλυψη. Το εύρος συχνοτήτων εντός του οποίου συμβαίνει η κάλυψη ονομάζεται *κρίσιμη ζώνη*[5]. Οι κρίσιμες ζώνες είναι ουσιαστικά οι ζώνες

[5] Η έννοια των κρίσιμων ζωνών εισήχθη από τον Αμερικανό φυσικό Harvey Fletcher (1884-1981).

συχνοτήτων εντός των οποίων ενεργοποιούνται τα τριχοειδή κύτταρα του αφτιού.

Το εύρος των εν λόγω ζωνών είναι στενότερο στις χαμηλές συχνότητες, και το μεγαλύτερο ποσοστό του πλήθους τους εντοπίζεται κάτω από τα 5 kHz. Το εύρος τους είναι 100 Hz στην περιοχή από 20 έως 500 Hz, και περίπου 1.5 οκτάβες για συχνότητες από 1 έως 7 kHz. Μία γνωστή μοντελοποίηση[6] του αφτιού περιλαμβάνει 24 αυθαίρετα επιλεγόμενες κρίσιμες ζώνες για την περιοχή κάτω από τα 15 kHz, και μία ακόμη για την περιοχή 15 έως 20 kHz. Ένα σχετικό παράδειγμα αναλύεται στον Πίνακα 6.1.

Από άποψη φυσιολογίας, η κάθε κρίσιμη ζώνη εκτείνεται σε 1.3 mm της βασικής μεμβράνης, και αποτελείται από 1,300 πρωτεύοντα τριχοειδή κύτταρα. Ως μοντέλο, περιγράφει μία διαδικασία φιλτραρίσματος που πραγματοποιείται στο αφτί: ένα σύστημα ανάλογο ενός αναλυτή φάσματος, ο οποίος απεικονίζει τις αποκρίσεις αλληλεπικαλυπτόμενων ζωνοπερατών φίλτρων, με μεταβλητές κεντρικές συχνότητες. Είναι κρίσιμο να τονιστεί αυτό το τελευταίο: οι κρίσιμες ζώνες δεν είναι σταθερές, και κάθε αντιληπτός ήχος δημιουργεί γύρω του μία κρίσιμη ζώνη. Σε κάθε περίπτωση, μιλάμε για ένα ακόμα εμπειρικό φαινόμενο, στα πλαίσια της Ψυχοακουστικής.

Μονάδα μέτρησης της αντιληπτικής συχνότητας είναι το Bark[7], το οποίο υποδιαιρείται σε 100 mel. Το εύρος μίας κρίσιμης ζώνης είναι 1 Bark. Η κλίμακα Bark συνδέει την απόλυτη συχνότητα (σε Hz) με συχνότητες που μετρώνται αντιληπτικά, όπως το τονικό ύψος και οι κρίσιμες ζώνες. Η μετατροπή γίνεται μέσω της σχέσης:

$$z(f) = 13\arctan(0.00076 f) + 3.5\arctan[(f/7500)^2] \text{ Bark} \quad [6.1]$$

[6] Από τον Γερμανό επιστήμονα Karl Eberhard Zwicker (1924-1990).
[7] Προς τιμή του Γερμανού φυσικού Heinrich Georg Barkhausen (1881-1956), ο οποίος υπήρξε ο πρώτος που εισήγαγε την έννοια της ακουστότητας.

Κρίσιμη ζώνη	Κεντρική συχνότητα (Hz)	Εύρος (Hz)	Κάτω συχνότητα αποκοπής (Hz)	Άνω συχνότητα αποκοπής (Hz)
1	50	-	-	100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
6	570	120	510	630
7	700	140	630	770
8	840	150	770	920
9	1000	160	920	1080
10	1170	190	1080	1270
11	1370	210	1270	1480
12	1600	240	1480	1720
13	1850	280	1720	2000
14	2150	320	2000	2320
15	2500	380	2320	2700
16	2900	450	2700	3150
17	3400	550	3150	3700
18	4000	700	3700	4400
19	4800	900	4400	5300
20	5800	1100	5300	6400
21	7000	1300	6400	7700
22	8500	1800	7700	9500
23	10500	2500	9500	12000
24	13500	3500	12000	15500
25	19500	6550	15500	22050

Πίνακας 6.1 Παράδειγμα κατανομής των κρίσιμων ζωνών, εντός του συχνοτικού εύρους της ανθρώπινης ακοής [15].

Μέσω της κλίμακας Bark, το φάσμα των απόλυτων συχνοτήτων μπορεί να μετατραπεί σε ένα υποκειμενικό φάσμα, κατά μήκος της βασικής μεμβράνης.

6.1.3 Ψυχοακουστικά μοντέλα

Προκειμένου ένα ηχητικό σήμα να κωδικοποιηθεί αποτελεσματικά, χρειάζεται ο θόρυβος κβάντισης να καταστεί όσο το δυνατό λιγότερο

αντιληπτός. Κάτι τέτοιο συνεπάγεται τον υπολογισμό των κατωφλίων κάλυψης για κάθε κρίσιμη ζώνη, ή με άλλα λόγια, τον προσδιορισμό του μέγιστου επιτρεπτού θορύβου ανά κρίσιμη ζώνη. Ως προς αυτό, γίνεται χρήση ψυχοακουστικών μοντέλων, τα οποία προσομοιώνουν το ανθρώπινο σύστημα ακοής, και αναλύουν τα σχετικά φασματικά δεδομένα.

Κάποια μεγέθη που ενδιαφέρουν στην εν λόγω διαδικασία είναι ο *λόγος σήματος-προς-κάλυψη* (signal-to-mask ratio, *SMR*), ο *λόγος θορύβου-προς-κάλυψη* (noise-to-mask ratio, *NMR*), και βέβαιο ο *SNR*. Η μεταξύ τους σχέση έχεις ως εξής:

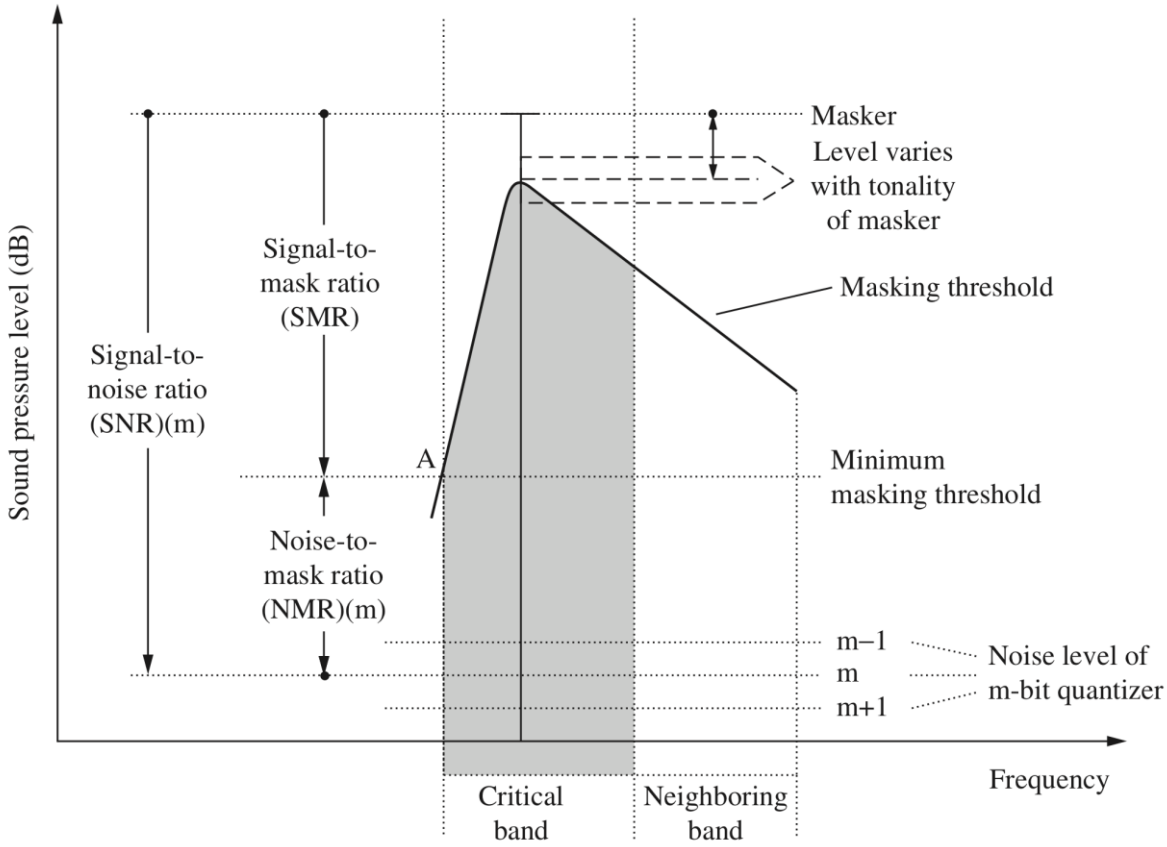
$$NMR = SMR - SNR \text{ dB} \quad [6.2]$$

Στα περισσότερα codec, ο στόχος κατά τη διαδικασία ανάθεσης bit είναι η ελαχιστοποίηση του συνολικού *NMR* ενός εκάστου πλαισίου. Εντός μίας συγκεκριμένης κρίσιμης ζώνης, όσο μεγαλύτερος είναι ο *SNR* σε σχέση με τον *SMR*, τόσο λιγότερο αντιληπτός γίνεται ο θόρυβος κβάντισης. Οπότε, ο στόχος του codec είναι η ελαχιστοποίηση της τιμής του *NMR*: τιμές υπό το 0 σημαίνουν κωδικοποίηση χωρίς αντιληπτή υποβάθμιση της ποιότητας. Τα σχετικά μεγέθη και η σχέση τους εντός μίας κρίσιμης ζώνης απεικονίζονται στο Σχήμα 6.3.

Πολλά ψυχοακουστικά μοντέλα χρησιμοποιούν μία *συνάρτηση εξάπλωσης* (spreading function), προκειμένου να υπολογίσουν τα επίπεδα κάλυψης, όχι για μία συγκεκριμένη κρίσιμη ζώνη, αλλά για το συνολικό φάσμα της βασικής μεμβράνης· να λάβουν υπόψη, δηλαδή, κάλυψη που ενδεχομένως προκύπτει αρκετά Bark πέρα από το σήμα που την προκαλεί. Η μορφή της καμπύλης της συνάρτησης εξάπλωσης είναι εκείνη ενός ασύμμετρου τριγώνου, ενώ μία εκλεπτυσμένη εκδοχή της εξίσωσης που την περιγράφει δίνεται από τη σχέση:

$$10\log SF(dz) = 15.81 + 7.5(dz + 0.474) - 17.5[1 + (dz + 0.474)^2]^{1/2} \text{ dB} \quad [6.3]$$

όπου *dz* είναι η συχνοτική απόσταση σε Bark, μεταξύ της συχνότητας που



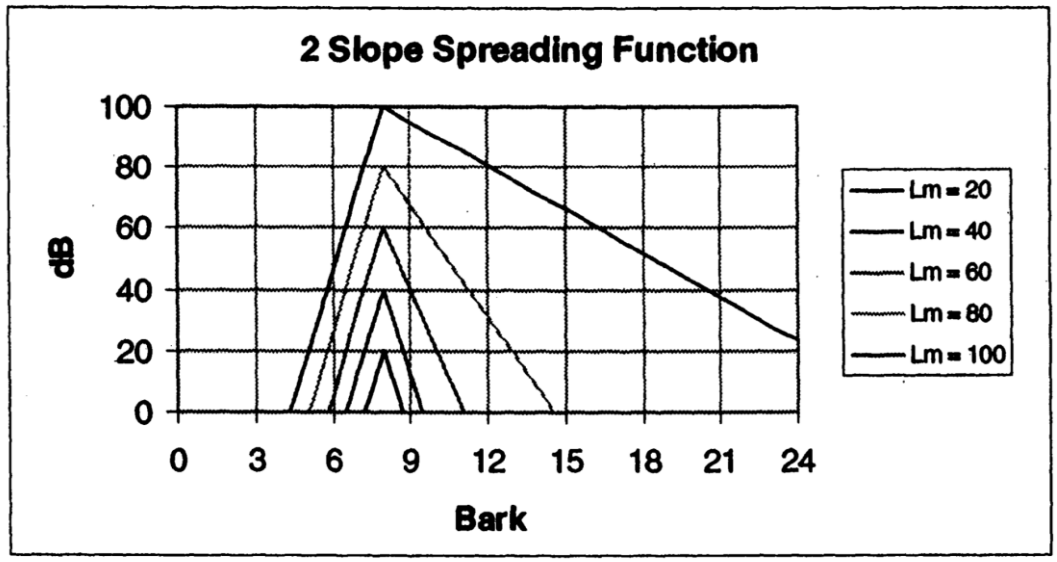
Σχήμα 6.3 Η σχέση των μεγεθών *NMR*, *SMR* και *SNR*, εκφρασμένη σε dB. Το κατώφλι κάλυψης κυμαίνεται, ανάλογα με την τονικότητα του σήματος [15].

προκαλεί την κάλυψη και της συχνότητας που καλύπτεται. Στο Σχήμα 6.4 απεικονίζεται μία σειρά από καμπύλες της άνωθεν συνάρτησης εξάπλωσης.

Για την εφαρμογή της συνάρτησης εξάπλωσης, απαιτείται αρχικά ο υπολογισμός της ενέργειας κάθε κρίσιμης ζώνης. Στη συνέχεια, λαμβάνεται η συνέλιξη αυτής με τη συνάρτηση εξάπλωσης, οπότε και προκύπτει το ακουστικό φάσμα. Λαμβάνοντας υπόψη τις εκάστοτε αποκλίσεις και το απόλυτο κατώφλι ακουστότητας, παράγονται τα τελικά κατώφλια κάλυψης. Για τον υπολογισμό του καθολικού κατωφλίου κάλυψης, πρέπει να ληφθεί υπόψη η επίδραση πολλών διαφορετικών συχνοτήτων κάλυψης.

Μία ακόμα λειτουργία που υπεισέρχεται στη σχεδίαση ψυχοακουστικών μοντέλων είναι η διάκριση μεταξύ τονικών και μη τονικών συνιστωσών,

καθώς η κάθε περίπτωση απαιτεί διαφορετική προσομοίωση κάλυψης. Για



Σχήμα 6.4 Καμπύλες της συνάρτησης εξάπλωσης [6.3], για διαφορετικές στάθμες του σήματος κάλυψης [18].

παράδειγμα, ο θόρυβος προκαλεί μεγαλύτερη κάλυψη σε σχέση με έναν καθαρό τόνο. Ως προς αυτό, έχουν εφευρεθεί διάφορες μέθοδοι, όπως, π.χ., η ανίχνευση τοπικών μεγίστων στο ηχητικό φάσμα, η οποία υποδηλώνει την ύπαρξη τονικού περιεχομένου. Η εν λόγω ικανότητα διάκρισης είναι εξίσου χρήσιμο να ενσωματωθεί και στον αποκωδικοποιητή, ειδικά αν το κανάλι μετάδοσης χαρακτηρίζεται από υψηλό ρυθμό σφαλμάτων.

Σε ό,τι αφορά την κατανομή των bit, υπάρχουν δύο διαφορετικές στρατηγικές. Στην προσαρμοστική προς τα εμπρός κατανομή, όλη η διαδικασία κατανομής πραγματοποιείται στον κωδικοποιητή, και η εν λόγω πληροφορία ενυπάρχει στη ροή των bit. Με την προϋπόθεση ότι ο κωδικοποιητής θα είναι σχεδιασμένος κατάλληλα, η συγκεκριμένη μέθοδος επιτρέπει μεγάλη ακρίβεια κατανομής. Το σημαντικότερο πλεονέκτημά της είναι ότι το ψυχοακουστικό μοντέλο βρίσκεται στον κωδικοποιητή, και ο αποκωδικοποιητής δεν χρειάζεται κάποιο αντίστοιχο μοντέλο προκειμένου να ανακατασκευάσει το σήμα. Έτσι, η όποια βελτίωση του ψυχοακουστικού μοντέλου θα επιφέρει βελτίωση της ποιότητας χωρίς ανάγκη επεμβάσεων στον αποκωδικοποιητή. Ένα μειονέκτημα της μεθόδου είναι ότι μέρος του

διαθέσιμου ρυθμού δυαδικών ψηφίων χρειάζεται να αφιερωθεί στη μετάδοση της πληροφορίας κατανομής bit στον αποκωδικοποιητή.

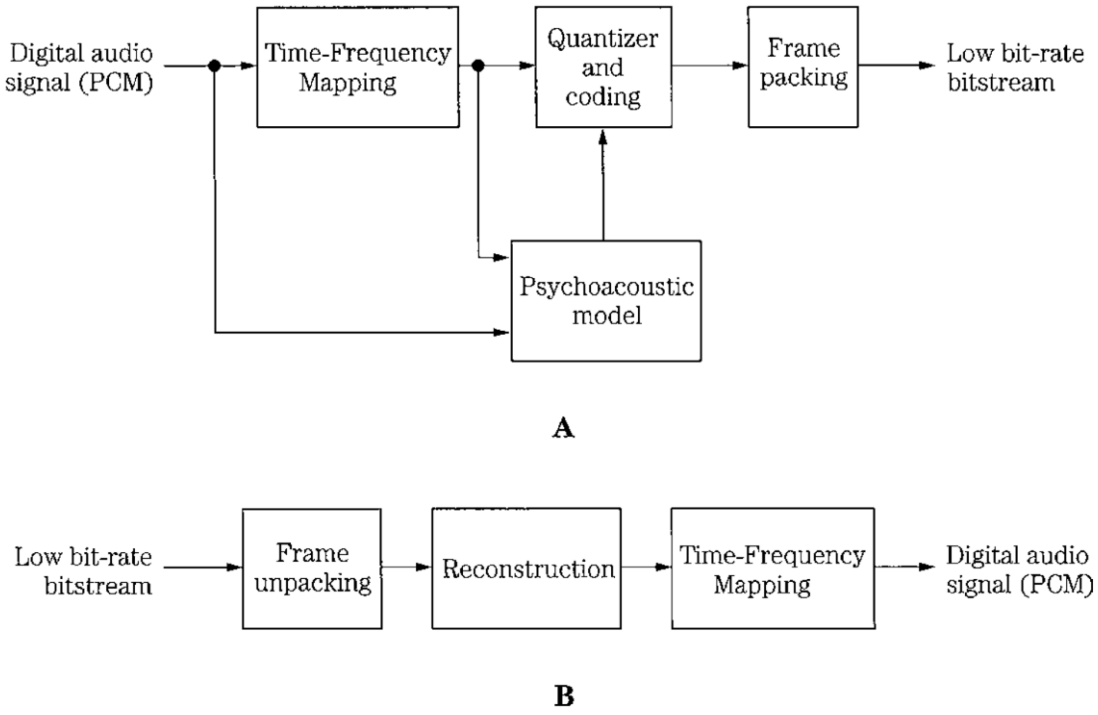
Στην προσαρμοστική προς τα πίσω κατανομή, η πληροφορία που αφορά στην κατανομή των bit εξάγεται από τα ίδια τα κωδικοποιημένα audio δεδομένα· έτσι, δεν σπαταλιέται ρυθμός δεδομένων για αυτό. Από την άλλη, η ακρίβεια της μεθόδου είναι σαφώς μικρότερη, καθώς ο αποκωδικοποιητής πρέπει να συμπεράνει ό,τι αφορά στην κατανομή των bit μέσα από περιορισμένη ποσότητα πληροφορίας. Επιπλέον, ο αποκωδικοποιητής παρουσιάζει αυξημένη πολυπλοκότητα, και το ψυχοακουστικό μοντέλο δεν είναι δυνατό να βελτιωθεί, στην περίπτωση που εισαχθούν νέα codec.

Σε κάθε περίπτωση, η ευστάθεια και η αποτελεσματικότητα του εκάστοτε χρησιμοποιούμενου ψυχοακουστικού μοντέλου είναι παράγοντες κρίσιμοι για την επιτυχία στη σχεδίαση οποιουδήποτε αλγορίθμου αντιληπτικής κωδικοποίησης. Εξίσου σημαντικός, όμως, είναι και ο τρόπος χρήσης των αποτελεσμάτων του μοντέλου, κατά τη διαδικασία κατανομής των bit και κατά την κβάντιση, οπότε και καθορίζεται ο βαθμός στον οποίο θα γίνεται αντιληπτός ο εμπλεκόμενος θόρυβος. Είναι, δηλαδή, κυρίως η διασύνδεση μεταξύ του μοντέλου και του κβαντιστή, που αναδεικνύεται ως το πλέον κρίσιμο κομμάτι ενός codec αυτής της κατηγορίας.

6.2 Διαφορετικές προσεγγίσεις στο πεδίο της συχνότητας

Όπως προκύπτει από την ανάλυση που έχει προηγηθεί, τα codec που λειτουργούν με εστίαση στο πεδίο της συχνότητας, και αξιοποιώντας τα ψυχοακουστικά χαρακτηριστικά της ανθρώπινης ακοής, δεν κωδικοποιούν όλο το περιεχόμενο ενός σήματος, αλλά μόνο εκείνα τα τμήματά του που κρίνονται σημαντικά. Παράλληλα, η κβάντιση προσαρμόζεται δυναμικά, έτσι ώστε τα σφάλματα να καλύπτονται από το σήμα. Αποτέλεσμα όλων αυτών είναι η σημαντική μείωση της απαιτούμενης προς μετάδοση πληροφορίας.

Υπάρχουν δύο προσεγγίσεις στην εν λόγω κατηγορία: τα *codec υποζωνών* και τα *codec μετασχηματισμού*. Τα codec υποζωνών χρησιμοποιούν έναν μικρό αριθμό υποζωνών και επεξεργάζονται τα ηχητικά δείγματα με βάση τη χρονική γειτνίασή τους· τα codec μετασχηματισμού χρησιμοποιούν μεγάλο αριθμό υποζωνών και επεξεργάζονται τα ηχητικά δείγματα με βάση τη συχνοτική γειτνίαση. Γενικά, τα πρώτα παρέχουν καλή ανάλυση στον χρόνο και φτωχή ανάλυση στη συχνότητα, ενώ τα δεύτερα καλή ανάλυση στη συχνότητα και φτωχή στον χρόνο.

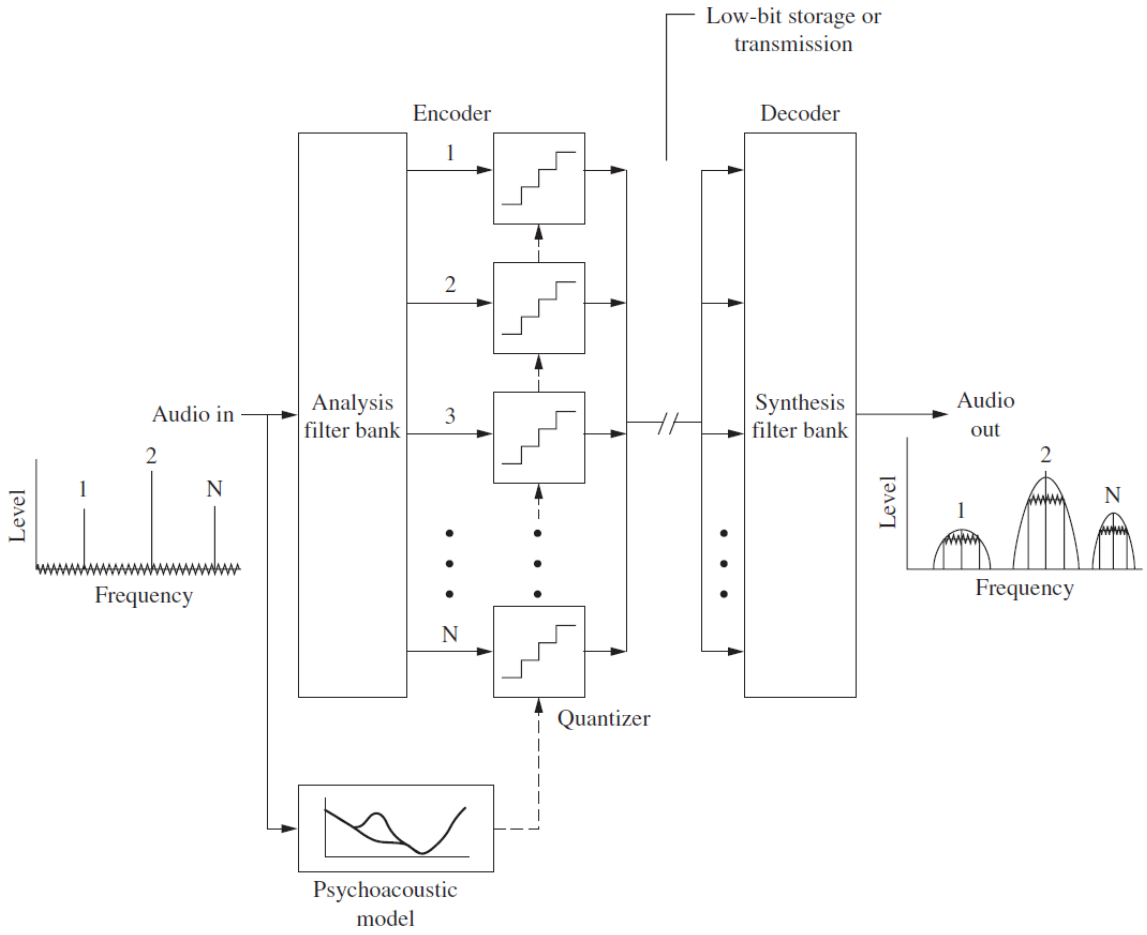


Σχήμα 6.5 Βασική αρχιτεκτονική κωδικοποιητή (**A**) και αποκωδικοποιητή (**B**) για codec υποζωνών και codec μετασχηματισμού, τα οποία κβαντίζουν συχνοτικές συνιστώσες [15].

Η προαναφερθείσα διάκριση, βέβαια, αφορά κυρίως στο ιστορικό πλαίσιο της ανάπτυξής τους, αφού από μαθηματική σκοπιά όλες οι μετατροπές που γίνονται στα πλαίσια ενός codec μπορεί να ιδωθούν ως συστοιχίες φίλτρων, ενώ η μόνη πρακτική διαφορά είναι στον αριθμό των ζωνών που επεξεργάζονται. Και οι δύο κατηγορίες, άλλωστε, ακολουθούν την αρχιτεκτονική που απεικονίζεται στο Σχήμα 6.5.

6.2.1 Κωδικοποίηση σε υποζώνες

Η κωδικοποίηση σε υποζώνες αναπτύχθηκε πρώτη φορά στα Bell Labs, στις αρχές της δεκαετίας του 1980, ενώ η εξέλιξή της στα επόμενα χρόνια πέρασε από ευρωπαϊκά εργαστήρια.



Σχήμα 6.6 Ο κωδικοποιητής με υποζώνες αναλύει το ευρυζωνικό audio σήμα σε στενές υποζώνες. Με βάση την πληροφορία, που του παρέχεται από το ψυχοακουστικό μοντέλο, σχετικά με την κάλυψη, κβαντίζει τα δείγματα της κάθε υποζώνης -διαδικασία που ανυψώνει το επίπεδο θορύβου. Όταν τα δείγματα ανακατασκευάζονται στον αποκωδικοποιητή, το φίλτρο σύνθεσης περιορίζει το επίπεδο θορύβου στην κάθε υποζώνη, οπότε και αυτό καλύπτεται από το audio σήμα [15].

Τα μπλοκ διαγράμματα του κωδικοποιητή και του αποκωδικοποιητή έχουν όπως στο Σχήμα 6.6. Το σήμα εισόδου στον κωδικοποιητή αναλύεται σε έως 32 κανάλια στενής ζώνης, καθώς περνά από μία ψηφιακή τράπεζα φίλτρων· τα κανάλια αυτά προσεγγίζουν τις κρίσιμες ζώνες της ανθρώπινης ακοής. Τα

δείγματα σε κάθε υποζώνη αναλύονται και συγκρίνονται με ένα ψυχοακουστικό μοντέλο. Το codec κβαντίζει τα δείγματα σε κάθε υποζώνη, ανάλογα με το κατώφλι κάλυψης της κάθε μίας. Η κάθε υποζώνη κωδικοποιείται ανεξάρτητα, με διαφορετικό αριθμό bit ανά δείγμα, από τις υπόλοιπες. Υποζώνες που δεν έχουν συχνοτικό περιεχόμενο κβαντίζονται στη μηδενική στάθμη.

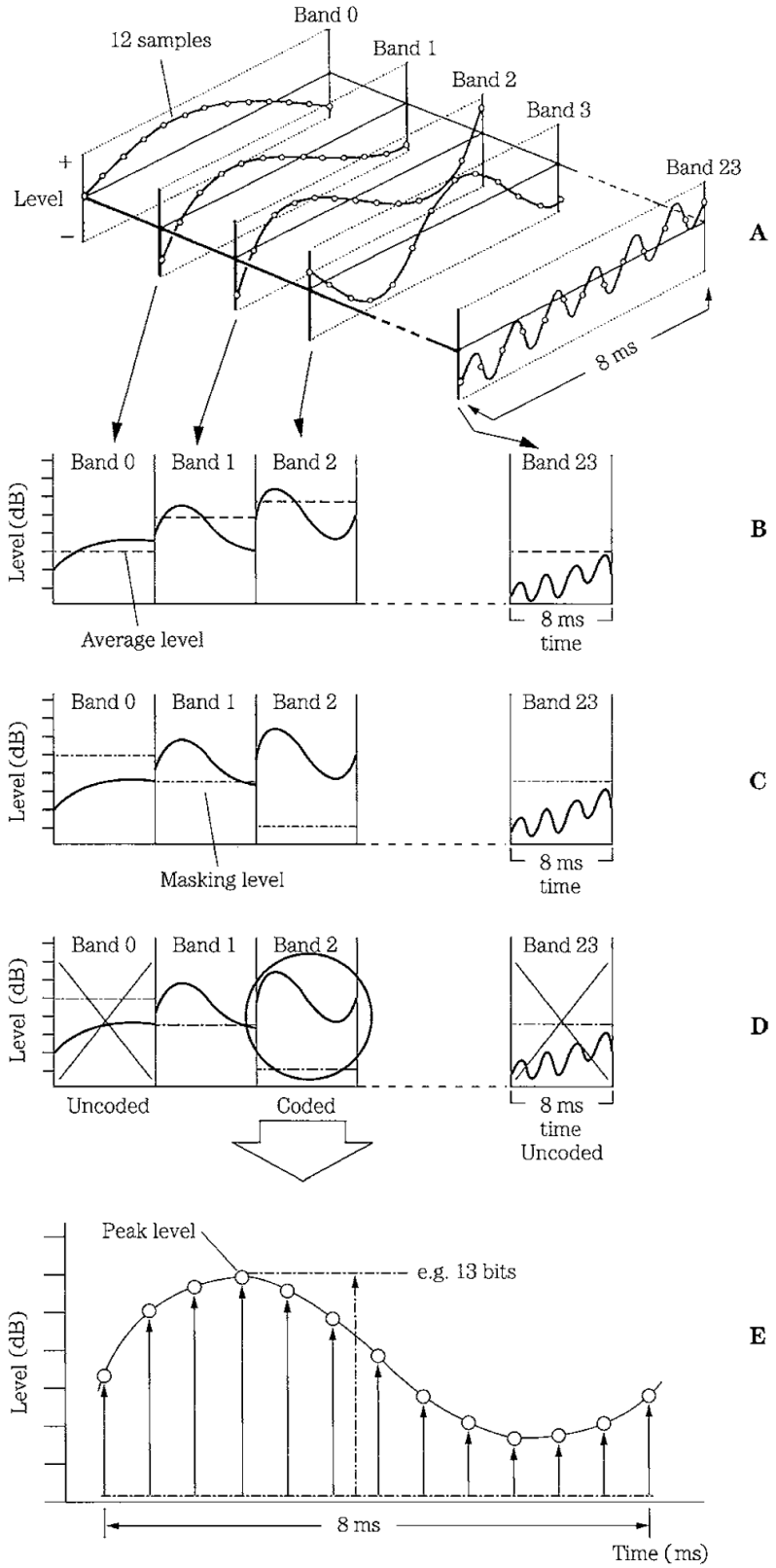
Στο Σχήμα 6.7 φαίνεται αναλυτικά η διαδικασία που ακολουθείται μετά την έξοδο της τράπεζας φίλτρων, ενώ στο Σχήμα 6.8 εξηγείται η κατανομή των bit. Ως κριτήριο χρησιμοποιείται ο *SMR*: τα σήματα που έχουν μεγαλύτερες απαιτήσεις ως προς αυτόν λαμβάνουν και τα περισσότερα bit.

6.2.2 Κωδικοποίηση με μετασχηματισμό

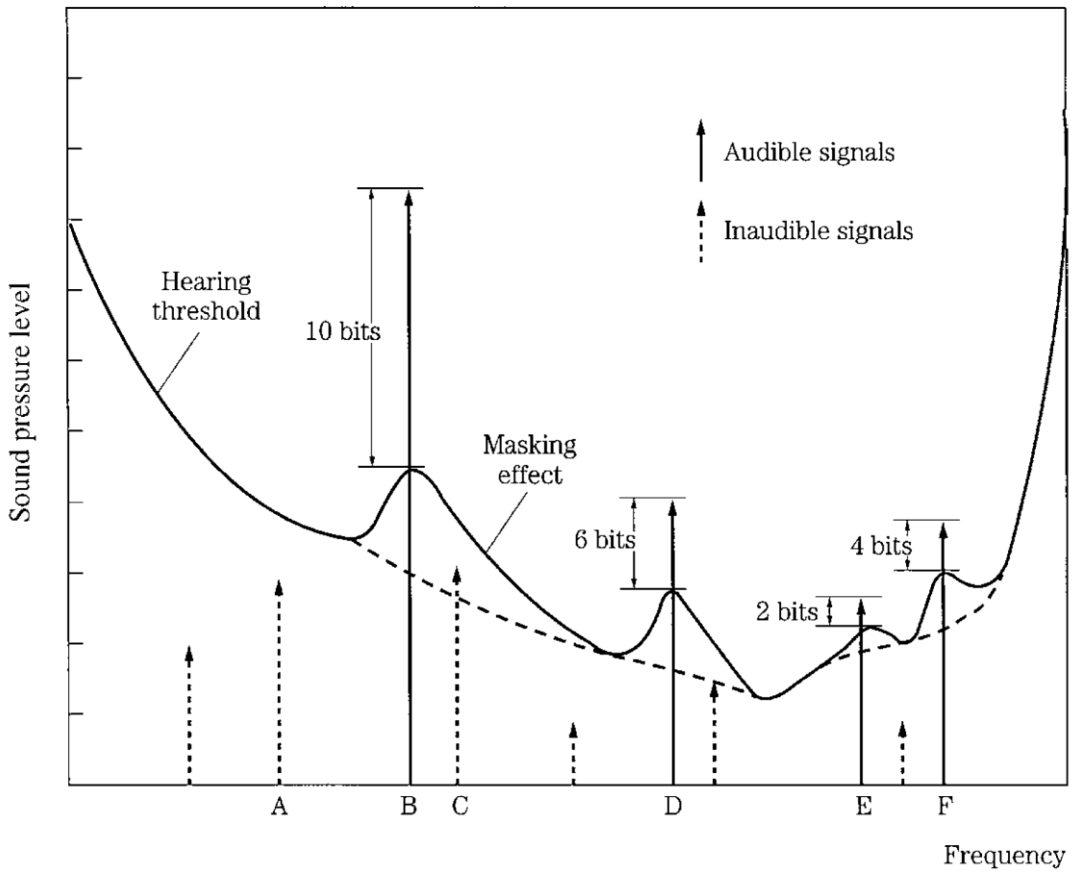
Στην κωδικοποίηση με μετασχηματισμό, το σήμα audio θεωρείται ως οιονεί σταθερό, εντός ενός μικρού χρονικού διαστήματος. Πλαίσια δειγμάτων του στο πεδίο του χρόνου μετασχηματίζονται στο πεδίο της συχνότητας, και οι συχνοτικές συνιστώσες κβαντίζονται. Στον αποκωδικοποιητή ακολουθείται ο αντίστροφος μετασχηματισμός.

Τα codec αυτής της κατηγορίας χρησιμοποιούν *Διακριτό Συνημιτονικό Μετασχηματισμό* (Discrete Cosine Transform, DCT) ή *Τροποποιημένο Διακριτό Συνημιτονικό Μετασχηματισμό* (Modified Discrete Cosine Transform, MDCT), για το πέρασμα από το πεδίο του χρόνου σε εκείνο της συχνότητας, και αντίστροφα. Επίσης, τα περισσότερα από αυτά φροντίζουν να υπάρχει μία αλληλοκάλυψη, της τάξης του 50% περίπου, ανάμεσα στα διαδοχικά πλαίσια· έτσι μειώνεται η διαφορά στο περιεχόμενο από φάσμα σε φάσμα και επιτυγχάνεται καλύτερη ανάλυση στο πεδίο του χρόνου. Επίσης, χρησιμοποιείται *Γρήγορος Μετασχηματισμός Fourier* (Fast Fourier Transform, FFT), από τον οποίον προκύπτουν οι συνιστώσες που απαιτούνται για την αντιληπτική μοντελοποίηση.

Λόγω του ότι το μικρό μήκος των πλαισίων περιορίζει την καλή ανάλυση



Σχήμα 6.7 A. Έξοδος φίλτρου με 24 υποζώνες. **B.** Υπολογισμός της μέσης στάθμης κάθε υποζώνης. **C.** Υπολογισμός της στάθμης κάλυψης κάθε υποζώνης. **D.** Κωδικοποιούνται μόνο οι υποζώνες που έχουν περιεχόμενο πάνω από το κατώφλι ακουστότητας. **E.** Τα bit κατανέμονται ανάλογα με τη στάθμη κορυφής πάνω από το κατώφλι κάλυψης [15].

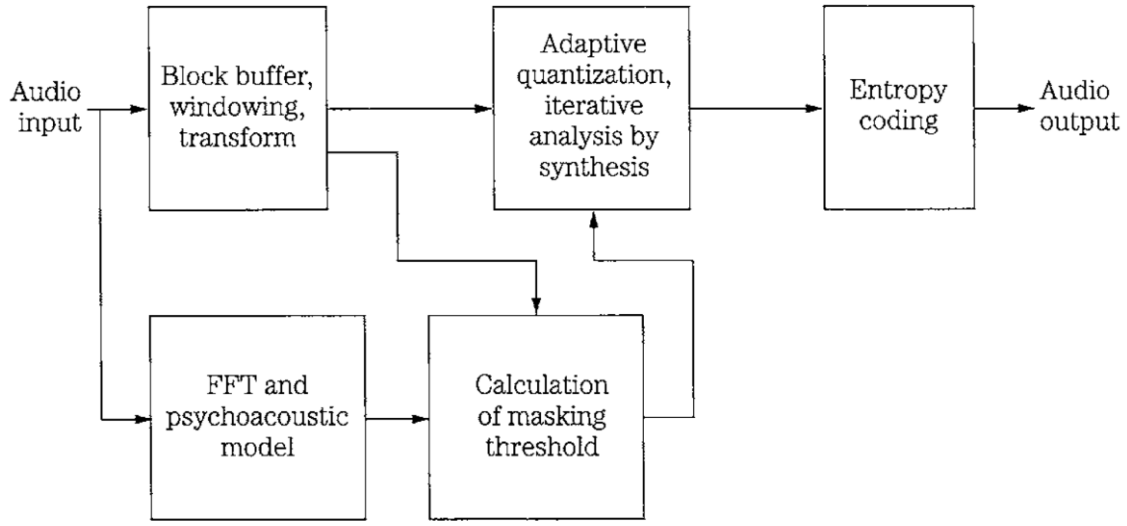


Σχήμα 6.8 Ο αλγόριθμος κατανομής των bit βασίζεται στην ακουστότητα των σημάτων της κάθε υποζώνης. Δεν ανατίθενται bit σε τόνους που καλύπτονται ή δεν γίνονται αντιληπτοί [15].

στο πεδίο της συχνότητας, αλλά, από την άλλη, μειώνει την πιθανότητα εμφάνισης παραμορφώσεων, στα περισσότερα codec της κατηγορίας αυτής το μήκος του πλαισίου προσαρμόζεται δυναμικά, με βάση την κατάσταση του σήματος. Δυστυχώς, δεν είναι δυνατό να αποφευχθούν κοστοβόροι συμβιβασμοί, καθώς το μουσικό περιεχόμενο πολύ συχνά θέτει αντικρουόμενες απαιτήσεις στον σχεδιαστή του εκάστοτε codec.

Ένα παράδειγμα προσαρμοστικού codec με μετασχηματισμό φαίνεται στο Σχήμα 6.9. Το σήμα μεταφέρεται στο πεδίο της συχνότητας, μέσω MDCT, και υπολογίζεται η ενέργειά του σε κάθε κρίσιμη ζώνη, με χρήση των φασματικών συνιστωσών. Δύο επαναληπτικοί βρόχοι πραγματοποιούν κβάντιση και κωδικοποίηση, με χρήση της τεχνικής ανάλυση-μέσω-σύνθεσης. Ένας εξωτερικός βρόχος υπολογίζει το σφάλμα κβάντισης που θα

εμφανίζεται στο αναδομημένο σήμα. Οι επαναλήψεις συνεχίζονται, ώσπου να επιτευχθούν τα επιθυμητά αποτελέσματα.



Σχήμα 6.9 Προσαρμοστικό codec με μετασχηματισμό, που χρησιμοποιεί έναν πλευρικό κλάδο FFT, επαναληπτική κβάντιση, και κωδικοποίηση εντροπίας [15].

6.3 Κωδικοποίηση πολυκαναλικών σημάτων

Στην έως τώρα ανάλυσή μας αναφερθήκαμε σε audio σήματα, υποθέτοντας ότι αυτά είναι μονοφωνικά. Σε ό,τι αφορά στην κωδικοποίηση μουσικού περιεχομένου, όμως, η πληροφορία είναι σχεδόν πάντα πολυκαναλική: χωρίζεται είτε σε δύο κανάλια (στερεοφωνία), είτε σε περισσότερα (5.1 Surround, Dolby Atmos κλπ.). Η πολυκαναλική κωδικοποίηση, όπως εύκολα συμπεραίνει κανείς, παρουσιάζει επιπλέον ευκαιρίες, αλλά και περιορισμούς, για τον σχεδιαστή των σχετικών αλγορίθμων.

Ένας παράγοντας που θα πρέπει να ληφθεί υπόψη σε μία τέτοια περίπτωση είναι, οι διακαναλικοί πλεονασμοί που προκύπτουν, οι οποίοι ενδέχεται να δίνουν επιπλέον χώρο για εξοικονόμηση στον ρυθμό δυαδικών ψηφίων. Από την άλλη, η αξιοποίηση του φαινομένου της κάλυψης γίνεται τώρα πιο σύνθετο ζήτημα, καθώς, π.χ., ο θόρυβος κβάντισης μπορεί να καλύπτεται στο ένα κανάλι, αλλά να αποκαλύπτεται στο άλλο. Επιπλέον, η

αντιληπτικότητα της ακοής σε σήματα που προέρχονται από ηχεία είναι διαφορετική σε σχέση με σήματα που προέρχονται από ακουστικά.

Ειδικά για τα στερεοφωνικά σήματα, υπάρχει μία σειρά τεχνικών που χρησιμοποιούνται. Η *dual-mono* κωδικοποίηση χρησιμοποιεί δύο codec που λειτουργούν ανεξάρτητα, η *joint-mono* χρησιμοποιεί δύο μονοφωνικά codec, τα οποία λειτουργούν με τον περιορισμό του κοινού ρυθμού δεδομένων, και η *joint-stereo* κωδικοποίηση αξιοποιεί διακαναλικές ιδιότητες και πλεονασμούς, προκειμένου να δράσει αποτελεσματικά. Οι κοινές ιδιότητες ανάμεσα στα κανάλια συνήθως δεν είναι εύκολο να εντοπιστούν στο πεδίο του χρόνου, αλλά γίνονται προφανείς στο πεδίο της συχνότητας. Η joint-stereo προσέγγιση κωδικοποιεί την κοινή πληροφορία μόνο μία φορά· ένα κανάλι joint-stereo των 256 kbps αποδίδει καλύτερα από δύο κανάλια των 128 kbps.

Τα παραπάνω επεκτείνονται αντίστοιχα σε περιπτώσεις με περισσότερα κανάλια. Πολύ γενικά, ισχύει ο κανόνας ότι ο αριθμός των bit που απαιτούνται για την κωδικοποίηση ενός πολυκαναλικού σήματος προκύπτει ως γινόμενο του αριθμού των bit που απαιτούνται για το ένα κανάλι, επί την τετραγωνική ρίζα του συνολικού αριθμού των καναλιών. Για ένα σύστημα 5.1 ηχείων, δηλαδή, απαιτούνται 2.45 φορές περισσότερα bit από όσα απαιτούνται για το ένα κανάλι.

6.4 Κωδικοποίηση χωρίς απώλειες (lossless)

Έως εδώ έχουμε αναφερθεί σε αλγόριθμους κωδικοποίησης οι οποίοι δίνουν στην έξοδό τους ένα σήμα το οποίο είναι αμετάκλητα αλλαγμένο σε σχέση με το αρχικό, έστω κι αν υπάρχει μέριμνα ώστε αυτές οι αλλαγές να είναι κατά το δυνατό μη αντιληπτές από το ανθρώπινο σύστημα ακοής.

Υπάρχει, όμως, κάποιες φορές μία απαίτηση, η οποία γίνεται ολοένα και μεγαλύτερη όσο οι τεχνολογία εξελίσσεται, για codec τα οποία, ενώ θα

δημιουργούν ένα ενδιάμεσο σήμα, το οποίο θα έχει μικρότερο αποτύπωμα ως προς το κόστος αποθήκευσης και μετάδοσής του, το τελικό σήμα στην έξοδο του αποκωδικοποιητή θα είναι απαράλλαχτο, bit προς bit, σε σχέση με το αρχικό. Μιλάμε σε αυτήν την περίπτωση για *κωδικοποίηση χωρίς απώλειες* (lossless coding).

6.4.1 Κωδικοποίηση εντροπίας

Για την επίτευξη του στόχου της lossless κωδικοποίησης ο σχεδιαστής του codec χρειάζεται να αξιοποιήσει την *εντροπία H* ενός σήματος. Πρόκειται για ένα μέγεθος το οποίο δηλώνει τη μέση ποσότητα πληροφορίας που προκύπτει σε ένα σήμα συναρτήσει του χρόνου. Με άλλα λόγια, η εντροπία αποτελεί μέτρο της τυχαιότητας ενός «γεγονότος» εντός του σήματος· του πόση πληροφορία απαιτείται προκειμένου αυτό να περιγραφεί. Εάν όλα τα «γεγονότα» έχουν την ίδια πιθανότητα να εμφανιστούν, τότε η εντροπία είναι μέγιστη και συμβολίζεται ως H_{max}. Ο πλεονασμός στο περιεχόμενο ενός σήματος δίνεται από τον όρο $1-\frac{H}{H_{\max}}$, και όσο αυτός αυξάνεται, τόσο αυξάνεται και ο bit rate.

Στην *κωδικοποίηση εντροπίας* (ή *κωδικοποίηση Huffman*[8], ή *κωδικοποίηση μεταβλητού μήκους*, ή *βέλτιστη κωδικοποίηση*), μία τεχνική που χρησιμοποιείται ευρέως για audio και video σήματα, αξιοποιείται η πιθανότητα εμφάνισης κάθε δείγματος. Τα δείγματα που εμφανίζονται πιο συχνά λαμβάνουν κωδική λέξη με τον μικρότερο αριθμό bit, ενώ όσο μικρότερη συχνότητα εμφάνισης έχει ένα δείγμα, τόσο μεγαλύτερο μήκος λέξης του αντιστοιχίζεται. Στον αποκωδικοποιητή η διαδικασία αντιστρέφεται, και έτσι δεν υπάρχει απώλεια πληροφορίας.

Γενικά, η κωδικοποίηση Huffman είναι μία τεχνική χωρίς θόρυβο, η οποία προσφέρει εξοικονόμηση όταν τα προς κωδικοποίηση σύμβολα εμφανίζονται

[8] Προς τιμή του εφευρέτη της, Αμερικανού επιστήμονα David Albert Huffman (1925-1999).

με διαφορετική πιθανότητα. Η βάση της βρίσκεται στα προθήματα: χρησιμοποιείται ένα σύστημα όπου τα σύμβολα με τη μεγαλύτερη συχνότητα εμφάνισης κωδικοποιούνται με λέξεις μικρού μήκους, οι οποίες, όμως, δεν μπορεί να επαναχρησιμοποιηθούν ως πρόθημα μίας μεγαλύτερης λέξης. Για παράδειγμα, οι σειρές 101 και 101011 δεν μπορεί να ανήκουν και οι δύο στο σύνολο κωδικών λέξεων του ίδιου συστήματος.

6.4.2 Συμπίεση δεδομένων audio

Οι lossless τεχνικές συμπίεσης audio σημάτων δεν προσφέρουν τον βαθμό εξοικονόμησης δεδομένων που επιτυγχάνουν οι αλγόριθμοι με απώλειες. Επίσης, η υπολογιστική πολυπλοκότητά τους, καθώς και η χρονική υστέρηση που εισάγουν, δεν είναι διόλου αμελητέες. Εντούτοις, είναι δυνατό να επιτευχθούν λόγοι συμπίεσης από 1.5:1 έως και 3.5:1.

Ο στόχος της lossless κωδικοποίησης είναι να λάβει ένα σήμα PCM και να το επεξεργαστεί, ώστε να οργανώσει πιο αποτελεσματικά τα πακέτα δεδομένων που θα αποθηκευτούν ή/και θα μεταδοθούν. Η αποτελεσματικότητα αυτή εξαρτάται σε μεγάλο βαθμό από το περιεχόμενο του σήματος PCM: όσο μεγαλύτερος πλεονασμός εμφανίζεται σε αυτό, τόσο μεγαλύτερος είναι και ο δυνατός βαθμός συμπίεσης. Επομένως, πιο αποτελεσματικό ως προς τη lossless διαδικασία θα είναι ένα σύστημα με μεταβλητό ρυθμό δυαδικών δεδομένων εξόδου. Σημειώνεται ότι ο σχεδιαστής της εκάστοτε μεθόδου θα πρέπει να λάβει υπόψη τον μέγιστο δυνατό bit rate του συστήματος, ο οποίος δεν είναι δυνατό να παρακαμφθεί, ακόμα και κατά την κωδικοποίηση δεδομένων με πολύ μικρό πλεονασμό.

Γενικά, οι σχετικές τεχνικές λειτουργούν σε δύο βήματα. Στο πρώτο βήμα παράγεται ένα *στατιστικό μοντέλο* για τα δεδομένα εισόδου, και στο δεύτερο γίνεται η αντιστοίχηση της κατάλληλης λέξης (μικρού ή μεγάλου μήκους) σε κάθε δείγμα. Σε ό,τι αφορά στο στατιστικό μοντέλο, αυτό μπορεί να είναι *στατικό* ή *προσαρμοζόμενο*. Πλέον χρησιμοποιείται σχεδόν αποκλειστικά η δεύτερη κατηγορία, αφού δίνει τη δυνατότητα βελτίωσης της απόδοσης,

καθώς το μοντέλο «μαθαίνει» περισσότερα για το σήμα εισόδου κατά τη διάρκεια της διαδικασίας συμπίεσης.

Κεφάλαιο 7

Σύντομη επισκόπηση επιλεγμένων codec

Όπως εύκολα θα διαπιστώσει ο μελετητής που θα ανατρέξει στη σχετική βιβλιογραφία και διαδικτυογραφία, ο κατάλογος των διαθέσιμων codec είναι ιδιαίτερα μακρύς, σε σημείο που κάποτε μπορεί να μοιάζει ανεξάντλητος. Ακόμα και αν περιορίσουμε την αναζήτηση σε εκείνα που εξακολουθούν να χρησιμοποιούνται στα διάφορα πρότυπα δικτύων κινητών επικοινωνιών, και πάλι η διαθέσιμη πληροφορία είναι ιδιαίτερα ογκώδης.

Στο παρόν κεφάλαιο θα ανατρέξουμε σε κάποια επιλεγμένα codec, από εκείνα που εφαρμόστηκαν σε πρότυπα της τρίτης γενιάς δικτύων κινητών επικοινωνιών, και έπειτα, και αφορούν στις δύο μεγάλες κατηγορίες codec, εκείνα που επικεντρώνονται στην ομιλία και εκείνα που ειδικεύονται σε μουσικό περιεχόμενο.

7.1 Adaptive Multi-Rate (AMR)

Το *Adaptive Multi-Rate* (AMR), που αναφέρεται και με τις συντομογραφίες AMR-NB ή GSM-AMR, είναι ένα codec που λειτουργεί βέλτιστα για κωδικοποίηση ομιλίας. Η πρώτη του έκδοση έγινε τον Ιούνιο του 1999, και την ίδια χρονιά υιοθετήθηκε από την 3GPP ως το στάνταρ codec ομιλίας. Σήμερα χρησιμοποιείται ευρέως στο GSM/UMTS πρότυπο, ενώ χρησιμοποιείται και από software εφαρμογές όπως οι QuickTime και RealPlayer, καθώς και στα ευρύτερα πλαίσια των λειτουργικών συστημάτων Android (Google) και iOS και MacOS (Apple).

Το AMR χρησιμοποιεί πολλαπλό bit rate και εστιάζει στη στενή ζώνη (NB), από 200 Hz έως 3,400 Hz. Προκειμένου να επιλέξει ένα από τα οχτώ

διαθέσιμα bit rate, χρησιμοποιεί προσαρμογή ζεύξης -ελέγχει δηλαδή τις συνθήκες του καναλιού. Αν οι συνθήκες είναι κακές, μειώνεται η κωδικοποίηση πηγής και αυξάνεται η κωδικοποίηση καναλιού, γεγονός που βελτιώνει την ανθεκτικότητα της δικτυακής σύνδεσης, θυσιάζοντας, όμως, κάποιο ποσοστό καθαρότητας της φωνής.

Στον Πίνακα 7.1 καταγράφονται οι διαφορετικοί τρόποι λειτουργίας του AMR, με οχτώ από αυτούς να είναι διαθέσιμοι σε full rate κανάλι (FR), και έξι να αφορούν σε half rate (HR).

Τρόπος λειτουργίας	**Bit rate**	**Κανάλι**
AMR_12.20	12.20	FR
AMR_10.20	10.20	FR
AMR_7.95	7.95	FR/HR
AMR_7.40	7.40	FR/HR
AMR_6.70	6.70	FR/HR
AMR_5.90	5.90	FR/HR
AMR_5.15	5.15	FR/HR
AMR_4.75	4.75	FR/HR
AMR_SID	1.80	FR/HR

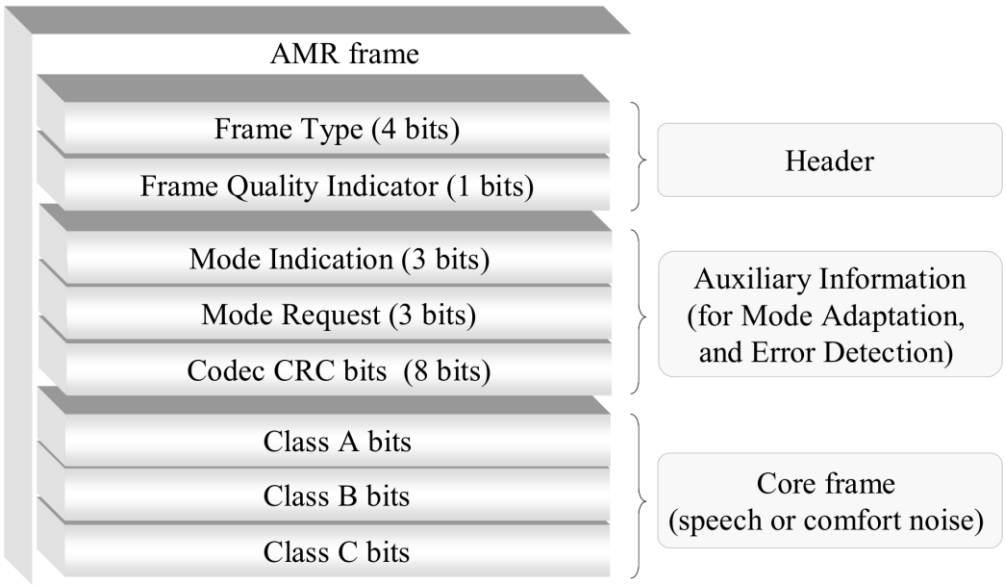
Πίνακας 7.1 Τρόποι λειτουργίας του AMR [26].

Η συχνότητα δειγματοληψίας που χρησιμοποιεί το AMR είναι 8 kHz στα 13 bit. Κάποιες από τις τεχνολογίες που χρησιμοποιεί προκειμένου να μειώσει το εύρος ζώνης κατά τις περιόδους απουσίας σήματος ομιλίας είναι οι DTX, VAD και CNG.

Το AMR είναι υβριδικό codec· μεταδίδει δηλαδή και παραμέτρους της ομιλίας, και σήματα κυματομορφής. Χρησιμοποιείται LPC για τη σύνθεση της ομιλίας από την υπολειπόμενη κυματομορφή. Οι παράμετροι της LPC κωδικοποιούνται ως γραμμικά φασματικά ζεύγη (line spectral pairs). Η υπολειπόμενη κυματομορφή κωδικοποιείται με χρήση ACELP.

Τα σαφή πλεονεκτήματα του AMR -η προσαρμοστικότητά του στις συνθήκες του εκάστοτε καναλιού επικοινωνίας, η καλή ποιότητα φωνής με σχετικά χαμηλό κόστος, η συμβατότητά του με όλα τα λειτουργικά συστήματα υπολογιστών κλπ.- το έχουν αναδείξει σε ένα από τα πλέον χρησιμοποιούμενα codec για επικοινωνίες φωνής. Στα μειονεκτήματα συγκαταλέγονται οι πολλές χρονικές υστερήσεις που υπεισέρχονται, καθώς και η αδυναμία υποστήριξης αναπαραγωγής μουσικής.

Το πλαίσιο του AMR codec περιέχει 160 δείγματα και έχει διάρκεια 20 msec. Η γενική δομή του εικονίζεται στο Σχήμα 7.1.



Σχήμα 7.1 Γενική δομή του πλαισίου δεδομένων του AMR [23].

Μία άλλη εκδοχή του εν λόγω codec είναι η Adaptive Multi-Rate - Wideband (AMR-WB). Εδώ χρησιμοποιείται η τεχνική BWE, δίνοντας ένα σαφώς διευρυμένο εύρος (50 έως 7,000 Hz). Χρησιμοποιεί ACELP. Μία επέκταση του AMR-WB αποτελεί το Extended Adaptive Multi-Rate - Wideband (AMR-WB+), το οποίο υποστηρίζει στερεοφωνικά σήματα και υψηλότερες συχνότητες δειγματοληψίας.

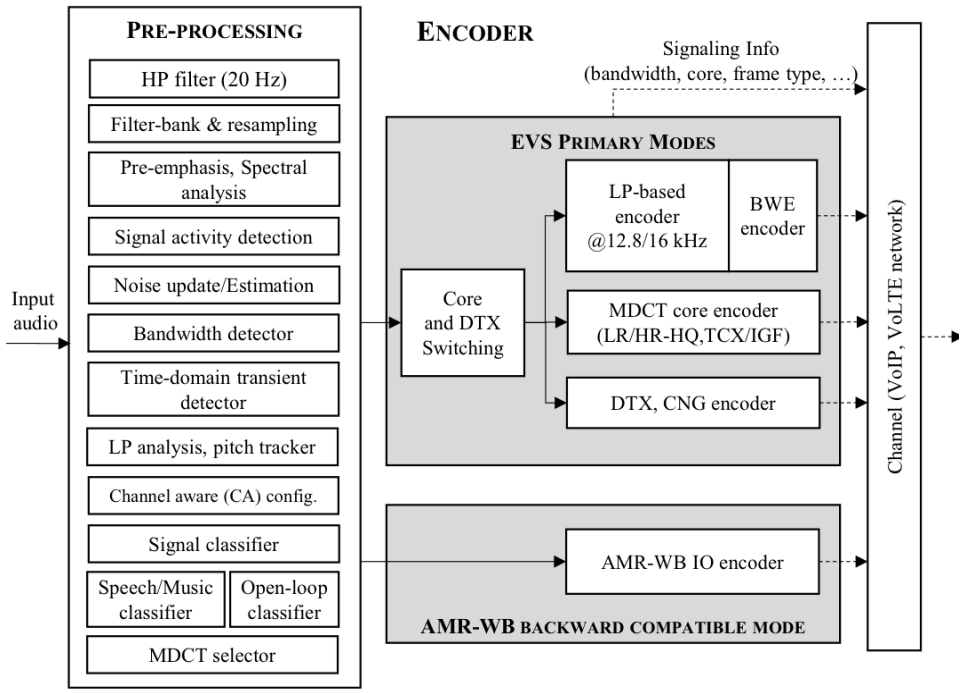
7.2 Enhanced Voice Services (EVS)

Το EVS είναι υπερευρυζωνικό codec φωνής, το οποίο αναπτύχθηκε κυρίως για την τεχνολογία Voice over LTE (VoLTE). Σχεδιάστηκε για το 3GPP και προτυποποιήθηκε το 2014. Σήμερα έχει υιοθετηθεί από δεκάδες παρόχους υπηρεσιών VoLTE, ενώ υπάρχουν στην αγορά περισσότερα από 200 μοντέλα, από κατασκευαστές έξυπνων συσκευών (Apple, Samsung, Google κ.ά.), που το υποστηρίζουν.

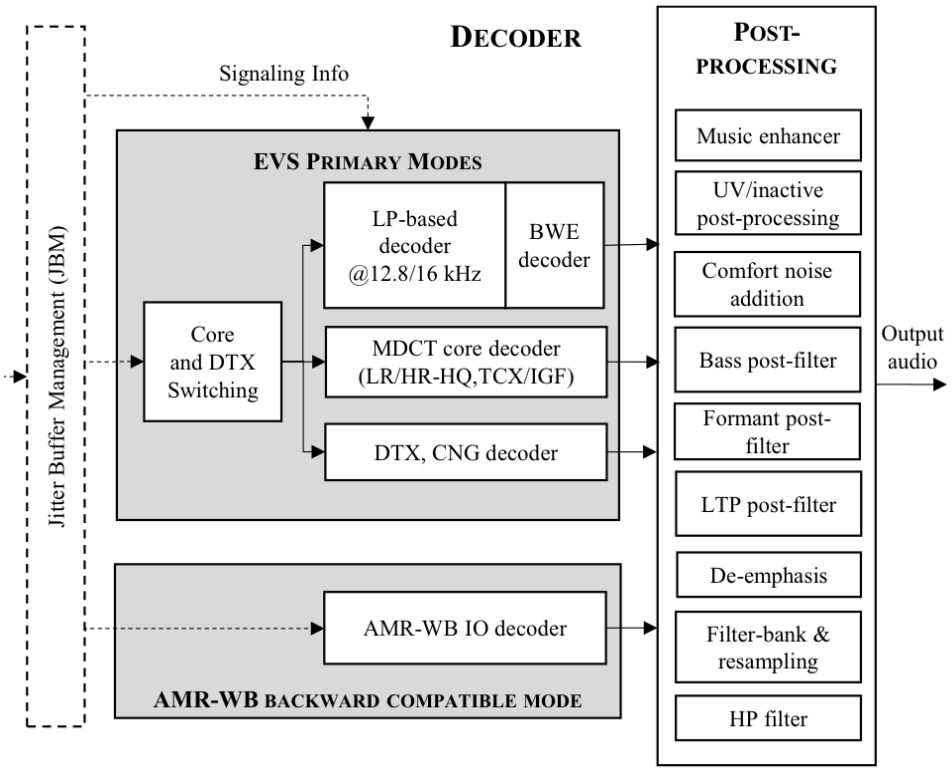
Το EVS προσφέρει εύρος ζώνης έως και 20 kHz, και διαθέτει ευρωστία έναντι του delay jitter και των απωλειών πακέτων δεδομένων. Διατηρεί αρκετά από τα χαρακτηριστικά προκατόχων του, όπως του AMR-WB, με το οποίο διατηρεί προς τα πίσω συμβατότητα. Χρησιμοποιεί μία βελτιωμένη εκδοχή του ACELP για ομιλία, MDCT για λοιπό audio περιεχόμενο, και εναλλάσσει την κωδικοποίησή του ανάλογα με το αν έχει να χειριστεί ομιλία, ή μουσική και άλλο audio περιεχόμενο. Είναι το πρώτο codec που εφάρμοσε «εν κινήσει» εναλλαγή μεταξύ συμπίεσης φωνής και ήχου σε χαμηλές αλγοριθμικές υστερήσεις των 32 msec, και χαμηλά bit rate της τάξης των 5.9 kbps (μέσο) ή 7.2 kbps (σταθερό). Έτσι, η κωδικοποίηση γενικού περιεχομένου, όπως φυσικών ήχων υποβάθρου ή μουσικής, εμφανίζει μεγάλη βελτίωση σε σχέση με παλαιότερα codec.

Το EVS είναι επίσης το πρώτο codec που προσέφερε υπερευρυζωνική κωδικοποίηση ομιλίας, έως και 16 kHz εύρος ζώνης από bit rate που ξεκινούν από 9.6 kbps, σε συνδυασμό με χαρακτηριστικά όπως υποστήριξη ασυνεχούς μετάδοσης (DTX) και αυξημένη αποτελεσματικότητα στον περιορισμό της απώλειας πακέτων. Μπορεί επίσης να προσφέρει full band (FB) κωδικοποίηση, εύρους ζώνης μέχρι και 20 kHz, ξεκινώντας από τα 16.4 kbps.

Συνοπτικά, το EVS προσφέρει συχνότητες δειγματοληψίας των 8, 16, 32 και 48 kHz, και bit rate από 7.2 έως 24.4 kbps για NB, από 7.2 έως 128 kbps για WB, από 9.6 έως 128 kbps για SWB, και από 16.4 έως 128 kbps για FB.



Σχήμα 7.2 Μπλοκ διάγραμμα του κωδικοποιητή του EVS [24].



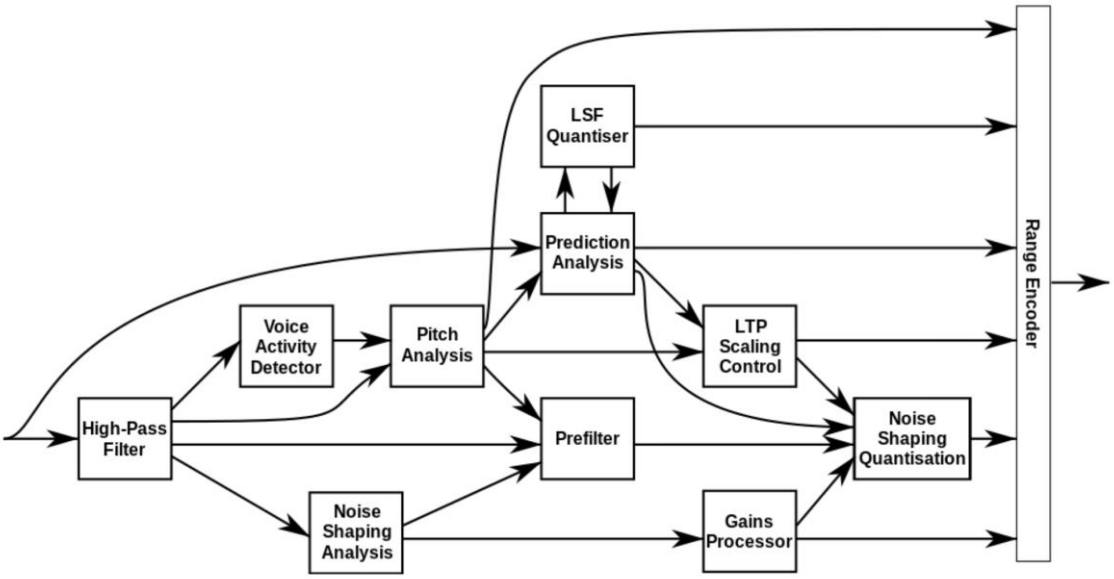
Σχήμα 7.3 Μπλοκ διάγραμμα του αποκωδικοποιητή του EVS [24].

Κάνει, επίσης, χρήση CNG, και Source-Controlled Variable Bit Rate (SC-

VBR), σε μέσο bit rate 5.9 kbps για NB και WB.

7.3 SILK

Το SILK είναι ένα ευρυζωνικό audio codec, το οποίο αναπτύχθηκε από την εταιρεία Skype Limited -που πλέον ανήκει στη Microsoft. Σχεδιάστηκε προκειμένου να χρησιμοποιηθεί για τη δημοφιλή εφαρμογή VoIP και τηλεδιασκέψεων Skype, αντικαθιστώντας το παλαιότερο SVOCP. Είναι codec ανοιχτού κώδικα, γραμμένου σε γλώσσα ANSI C, με χρήση αριθμητικής με σταθερή υποδιαστολή. Το SILK τέθηκε σε εφαρμογή για πρώτη φορά το 2009, στην τέταρτη έκδοση του Skype, ενώ αργότερα χρησιμοποιήθηκε και αλλού, όπως, για παράδειγμα, στο Zoom (την επίσης δημοφιλή πλατφόρμα τηλεδιασκέψεων), στα πλαίσια του voice chat μεταξύ παικτών βιντεοπαιχνιδιών (όπως το *Team Fortress 2*), αλλά και μεταξύ χρηστών σχετικών πλατφορμών (όπως η Steam).



Σχήμα 7.4 Μπλοκ διάγραμμα του κωδικοποιητή του SILK [26].

Η λειτουργία του SILK είναι βασισμένη σε τεχνικές LPC. Προσφέρει μεταβλητότητα σε πραγματικό χρόνο, σε ό,τι αφορά το εύρος ζώνης, το bit rate, και την πολυπλοκότητα. Μπορεί να λειτουργήσει σε τέσσερις

διαφορετικές συχνότητες δειγματοληψίας: 8 (NB), 12 (MB), 16 (WB), και 24 (SWB) kHz. Τα αντίστοιχα bit rate είναι 6 έως 20 kbps, 7 έως 25 kbps, 8 έως 30 kbps, και 12 έως 40 kbps. Η αλγοριθμική χρονική υστέρηση που εισάγει είναι χαμηλή, της τάξης των 25 msec, από τα οποία τα 20 msec αφορούν στο μέγεθος του πλαισίου δειγμάτων, και τα υπόλοιπα 5 msec στην πρόβλεψη. Το μήκος των πακέτων μπορεί να είναι 20, 40, 60, 80, ή 100 msec.

Βασικά πλεονεκτήματα του SILK είναι η προσαρμοστικότητά του, και η βελτιστοποίηση της απόδοσής του, ακόμα και σε ιδιαίτερα αντίξοες συνθήκες, είτε αυτές αφορούν σε κακή κατάσταση του καναλιού, είτε σε περιορισμένες δυνατότητες του υλικού εξοπλισμού.

7.4 Opus

Το Opus είναι ένα απολεστικό audio codec ανοιχτού κώδικα, το οποίο αναπτύχθηκε από την Xiph.Org Foundation, και προτυποποιήθηκε το 2012 από την Internet Engineering Task Force (IETF).

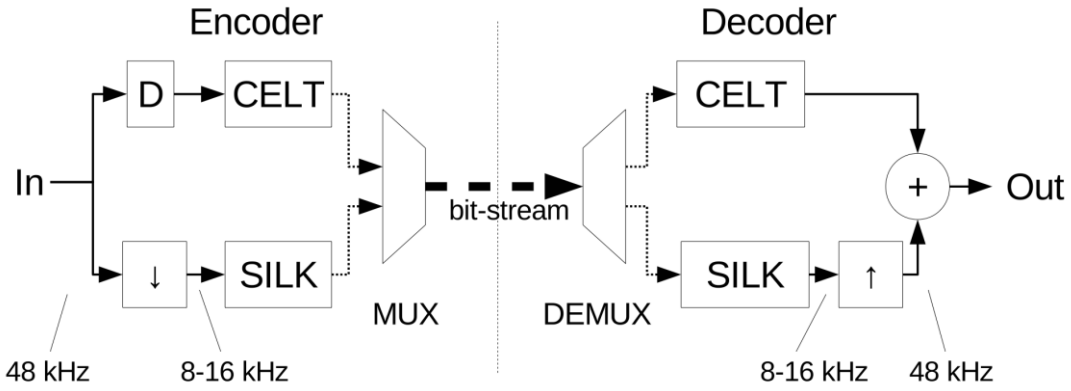
Στα δίκτυα κινητών επικοινωνιών το Opus χρησιμοποιείται σε πληθώρα εφαρμογών: στο Spotify (που είναι η μεγαλύτερη πλατφόρμα για streaming μουσικού audio περιεχομένου σήμερα), αλλά και σε VoIP πλαίσια, όπως στις περιπτώσεις των Discord και WhatsApp.

Το Opus είναι εξαιρετικά ευέλικτο, καθώς έχει σχεδιαστεί για διαδραστικές διαδικτυακές εφαρμογές. Υποστηρίζει ομιλία και μουσική, εναλλασσόμενο bit rate, σωστή συνεργασία με το Real-Time Protocol (RTP), και καλή αναπλήρωση ως προς τις απώλειες πακέτων -όλα αυτά με χαμηλή αλγοριθμική υστέρηση, που μπορεί να αγγίξει τα μόλις 5 msec.

Συνοπτικά, το Opus υποστηρίζει ρυθμούς δυαδικών ψηφίων από 6 έως 510 kbps, πέντε ζώνες λειτουργίας, από 8 kHz (NB) έως 48 kHz (FB), μεγέθη

πλαισίου από 2.5 έως 60 msec, υποστήριξη μονοφωνικών και στερεοφωνικών σημάτων, Constant Bit Rate (CBR) και Variable Bit Rate (VBR), αριθμητική κινητής αλλά και σταθερής υποδιαστολής, και μεταβλητή πολυπλοκότητα κωδικοποίησης. Όλα τα παραπάνω μπορεί να εναλλάσσονται δυναμικά, εντός της εκάστοτε ζώνης, χωρίς αντιληπτές παραμορφώσεις.

Το Opus αναπτύχθηκε ως συνδυασμός δύο παλαιότερων τεχνολογιών codec: του SILK (που βασίζεται στην LPC), και του Constrained Energy Lapped Transform (CELT), ενός απολεστικού codec ανοιχτού κώδικα, που βασίζεται στον MDCT. Λειτουργεί με τρεις διαφορετικούς τρόπους: ως SILK (για σήματα ομιλίας μέχρι και WB), ως CELT (για μουσική και ομίλα μεγάλων bit rate), και υβριδικά, δηλαδή ως SILK και CELT ταυτόχρονα (για SWB και FB ομιλία).



Σχήμα 7.5 Κωδικοποιητής και αποκωδικοποιητής του Opus [25].

Το κομμάτι του CELT λειτουργεί πάντα σε συχνότητα δειγματοληψίας 48 kHz, ενώ το κομμάτι του SILK μπορεί να λειτουργεί στα 8, 12 και 16 kHz. Στον υβριδικό τρόπο λειτουργίας, η συχνότητα δειγματοληψίας μετάβασης είναι 8 kHz, με το SILK να λειτουργεί στα 16 kHz, και το CELT να απορρίπτει όλες τις συχνότητες δειγματοληψίας κάτω από τα 8 kHz.

Βιβλιογραφία - Διαδικτυογραφία

[1] Χ. Γ. Χαραλαμπάκης (επιμ.), *Χρηστικό Λεξικό Της Νεοελληνικής Γλώσσας Τόμος 2*, ειδική έκδ., Αθήνα: Δημοσιογραφικός Οργανισμός Λαμπράκη, 2016, σσ. 649-650

[2] A. A. Huurdeman, *The Worldwide History Of Telecommunications*, New Jersey: John Wiley & Sons, 2003, σσ. 1-217

[3] Scholarly Community Encyclopedia. (2024). *History Of Telecommunication* [Online]. Διαθέσιμο στο: https://encyclopedia.pub/entry/35081

[4] F. Di Trocchio, *Αλλοπαρμένες Μεγαλοφυΐες: Ιδέες Και Άνθρωποι Που Δεν Έγιναν Κατανοητοί Από Την Επιστήμη*, 5η έκδ., Αθήνα: Π. Τραυλός, 2003, σσ. 373-378

[5] R. Mashayekki. (2020). *What are 0G, 1G, 2G, 3G, 4G, 5G Cellular Mobile Networks - History of Wireless Telecommunications* [Online]. Διαθέσιμο στο: https://www.youtube.com/watch?v=m8YkIcDVbGQ&ab_channel=VisionAcademy

[6] A. Khanna, A. Bengani, A. Bhatt, A. Bhardawaj, "A Critical Review Of Various Generations Of Mobile Network Technologies", *International Journal of Information & Computational Technology*, Vol. 4, Number 11, σσ. 1023-1028, 2014

[7] Λ. Χαδέλλης, *Ήχος - Μουσική & Τεχνολογία Τόμος Α΄*, Αθήνα: Σύγχρονη Μουσική, 1992, σσ. 10-62

[8] Δ. Δώδης, *Ηχοληψία: Η Δημιουργία Με Τη Σύγχρονη Τεχνολογία*, 3η έκδ., Περιστέρι: Ίων, 2001, σσ. 61-109

[9] B. Parker, *Good Vibrations: The Physics Of Music*, Baltimore: The Johns Hopkins University Press, 2009, σσ. 13-76

[10] R. E. Berg και D. G. Stork, *The Physics Of Sound*, 3rd ed., San Francisco: Pearson Education, Inc., 2005, σσ. 1-91, 145-180

[11] D. Hosken, *An Introduction To Music Technology*, New York: Routledge, 2011, σσ. 7-31

[12] J. Watkinson, *The Art Of Digital Audio*, 3rd ed., Oxford: Focal Press, 2001, σσ. 32-80

[13] W. C. Chu, *Speech Coding Algorithms: Foundation And Evolution Of Standardized Coders*, New Jersey: John Wiley & Sons, 2003, σσ. xii-xviii, 1-32

[14] D. O'Shaughnessy, "Review Of Methods For Coding Of Speech Signals", *EURASIP Journal On Audio, Speech And Music Processing*, 2023

[15] K. C. Pohlmann, *Principles Of Digital Audio*, 6th ed., New York: McGraw-Hill, 2011, σσ. 19-36, 335-391, 451-484

[16] M. Hasegawa - Johnson και A. Alwan, "Speech Coding: Fundamentals And Applications" στο *Wiley Encyclopedia Of Telecommunications*, J. G. Proakis, Ed. New Jersey: John Wiley & Sons, 2003

[17] L. R. Rabiner και R. W. Schafer, "Introduction To Digital Speech Processing", *Foundations and Trends in Signal Processing*, Vol. 1, Nos. 1-2, 2007

[18] M. Bosi και R. E. Goldberg, *Introduction To Digital Audio Coding And Standards*, New York: Springer Science+Business Media, 2003, σσ. 149-200

[19] Χ. Θεμιστοκλέους, *Εισαγωγή Στη Φωνητική Και Στη Φωνολογία*, Κύπρος: αυτοέκδοση, 2011, σσ. 1-46

[20] Α. Ρεβυθιάδου και Β. Σπυρόπουλος (επιμ.), *Αντιπαραβολική Μελέτη Γραμματικών Δομών Αλβανικής – Ελληνικής*, Θεσσαλονίκη: Ειδικός Λογαριασμός Κονδυλίων Έρευνας Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, 2013, σσ. 1-32

[21] Ε. Βαρβαρίγος και Κ. Μπερμπερίδης, *Κινητά Δίκτυα Επικοινωνιών (Πανεπιστημιακές Σημειώσεις)*, Πάτρα: Πανεπιστήμιο Πατρών, 2008, σσ. 108-131

[22] R. Drew. (2023, Μαΐου 19). *What Is Phone Echoing: Causes & How To Fix It* [Online]. Διαθέσιμο στο: https://getvoip.com/blog/phone-echoing/

[23] J. Sanchez και M. Thioune, "Appendix 1: AMR Codec in UMTS" στο *UMTS*, London: ISTE Ltd., 2007, σσ. 375-381

[24] M. Dietz, M. Multrus, V. Eksler, V. Malenovsky, E. Norvell, et al., "Overview Of The EVS Codec Architecture", *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, σσ. 5698-5702, 2015

[25] J.-M. Valin, G. Maxwell, T. B. Terriberry, K. Vos, "High-Quality, Low-Delay Music Coding In The Opus Codec", *135th AES Convention*, 2013

[26] Anon. (2024). Wikipedia [Online]. Διαθέσιμο στο: https://www.wikipedia.org

[27] Anon. (2024), Geeks For Geeks [Online]. Διαθέσιμο στο: https://www.geeksforgeeks.org

[28] ResearchGate. (2024). ResearchGate [Online]. Διαθέσιμο στο: https://www.researchgate.net

[29] R. Slavov. (2013). *Beyond LTE (Part 1): 0G and the birth of the mobile radio phone* [Online]. Διαθέσιμο στο: https://www.phonearena.com/news/Beyond-LTE-Part-1-0G-and-the-birth-of-the-mobile-radio-phone_id37843

[30] R. H. Colton και J. K. Casper και R. Leonard, *Κατανοώντας Τις Διαταραχές Φώνησης*, Πάτρα: Gotsis, 2015, σσ. 695-712