

## Απωλεστική συμπίεση

Στην απωλεστική συμπίεση μπορούμε να πετύχουμε συμπίεση άνω του 90% της ηχητικής πληροφορίας (με σημαντική απώλεια ηχητικής ποιότητας) (Salomon, 2000).

Η απωλεστική συμπίεση χρησιμοποιείται κυρίως για την συμπίεση μουσικής αλλά και ομιλίας. Οι βασικές τεχνικές απωλεστικής συμπίεσης είναι δύο:

- προγνωστική κωδικοποίηση (Predictive encoding)
- αντιληπτική κωδικοποίηση (Perceptual encoding)

Η προγνωστική κωδικοποίηση είναι μία από τις πιο ισχυρές τεχνικές ανάλυσης ομιλίας, και μία από τις πιο χρήσιμες μεθόδους για την κωδικοποίηση/συμπίεση της ομιλίας και συνδυάζει καλή ποιότητα ήχου σε ένα χαμηλό ρυθμό μετάδοσης και παρέχει εξαιρετικά ακριβείς εκτιμήσεις των παραμέτρων της ομιλίας (Li & O'Shaughnessy, 2003).

Η προγνωστική κωδικοποίηση βασίζεται στην παραδοχή ότι ένα σήμα ομιλίας παράγεται από ένα βομβητή (buzzer) στο τέλος ενός σωλήνα, με περιστασιακά προστιθέμενα σφυρίγματα και σκασίματα ήχου (popping sounds). Αν και φαινομενικά ακατέργαστο, το μοντέλο αυτό είναι στην πραγματικότητα μια στενή προσέγγιση της πραγματικότητας της παραγωγής της ομιλίας. Η γλωττίδα (ο χώρος μεταξύ των φωνητικών χορδών) παράγει το βόμβο, ο οποίος χαρακτηρίζεται από την ένταση και τη συχνότητα (pitch). Η φωνητική οδός (ο λαιμός και το στόμα) σχηματίζει το σωλήνα, ο οποίος σωλήνας χαρακτηρίζεται από συντονισμούς του, οι οποίοι συντονισμοί προκαλούν φορμάντα, ή ενισχυμένες ζώνες συχνοτήτων στον ήχο που παράγεται. Τα σφυρίγματα και τα σκασίματα δημιουργούνται από τη δράση της γλώσσας, των χειλιών για την παραγωγή των άφωνων συμφώνων (Σπυρίδης, 1989).

Η προγνωστική κωδικοποίηση αναλύει το σήμα ομιλίας με την εκτίμηση των φορμάντων, αφαιρώντας τα αποτελέσματά τους από το σήμα ομιλίας, και στη συνέχεια γίνεται εκτίμηση της έντασης και της συχνότητας του εναπομένου βόμβου. Η διαδικασία της απομάκρυνσης των φορμάντων ονομάζεται αντίστροφο φιλτράρισμα, και το υπόλοιπο του σήματος μετά την αφαίρεση του φιλτραρισμένου σήματος, ονομάζεται το υπόλειμμα.

Η προγνωστική κωδικοποίηση συνθέτει το σήμα της ομιλίας από την αντιστροφή αυτής της διαδικασίας. Χρησιμοποιεί τον βόμβο και το υπόλειμμα για να

δημιουργήσει ένα σήμα πηγής, και στη συνέχεια χρησιμοποιεί τα φορμάντα για να δημιουργήσει ένα φίλτρο (το οποίο αντιπροσωπεύει το σωλήνα), και περνάει το σήμα της πηγής μέσα από το φίλτρο, με αποτέλεσμα την αναπαραγωγή της ομιλίας.

Επειδή σήματα ομιλίας μεταβάλλεται με το χρόνο, αυτή η διαδικασία γίνεται σε σύντομο κομμάτια του σήματος ομιλίας, τα οποία ονομάζονται πλαίσια. Χρησιμοποιώντας 30 έως 50 πλαίσια ανά δευτερόλεπτο δίνουν σαν αποτέλεσμα κατανοητή ομιλία με πολύ καλή συμπίεση. Χαρακτηριστικά μπορεί να δώσει συμπίεση της τάξης του 98% ή και μεγαλύτερη (Olympus Imaging Corporation, 2012).

Η αντιληπτική κωδικοποίηση χρησιμοποιείται για την κωδικοποίηση/συμπίεση κυρίως μουσικών σημάτων και βασίζεται στην αντίληψη του ήχου από τον άνθρωπο. Η αντιληπτική κωδικοποίηση εκμεταλλεύεται ορισμένες ατέλειες του ανθρώπινου ακουστικού συστήματος για την κωδικοποίηση ενός σήματος κατά τέτοιο τρόπο ώστε να δίνει την ίδια αντίληψη της ακοής, ακόμη και αν αυτό φαίνεται έχει διαφορετική κυματομορφή από το πρωτότυπο.

Η βασική ιδιότητα της αντιληπτικής κωδικοποίησης είναι η ακουστική συγκάλυψη δηλαδή το γεγονός ότι η αντίληψη ενός ήχου μπορεί να καλύψει άλλους ήχους (Auditory masking). Η ακουστική συγκάλυψη στο πεδίο των συχνοτήτων είναι γνωστή ως ταυτόχρονη συγκάλυψη, συγκάλυψη συχνότητας ή φασματική συγκάλυψη. Η ακουστική συγκάλυψης στο πεδίο του χρόνου είναι γνωστή ως χρονική συγκάλυψη ή μη ταυτόχρονη συγκάλυψη (Wikipedia, Auditory masking, 2013).

Ένα άλλο φαινόμενο που εκμεταλλεύεται η αντιληπτική κωδικοποίηση εκμεταλλεύεται είναι το φαινόμενο του απόλυτου κατώφλιου της ακοής (ATH) που είναι η ελάχιστη ηχοστάθμη ενός καθαρού τόνου που ένα μέσο αυτί με φυσιολογική ακοή μπορεί να ακούσει με την προϋπόθεση της απουσίας άλλων ήχων. Το κατώφλι αυτό δεν είναι ένα διακριτό σημείο, και επομένως χαρακτηρίζεται ως το σημείο στο οποίο μία απόκριση προκλήθηκε για ένα ελάχιστο χρονικό διάστημα (Durrant & Lovrinic, 1984).

Τέλος η αντιληπτική κωδικοποίηση εκμεταλλεύεται το ότι τα όρια της ανθρώπινης ακοής είναι στο διάστημα 20Hz έως 20KHz, αν και υπάρχουν σημαντικές διαφορές μεταξύ των ατόμων, ιδιαίτερα στις υψηλές συχνότητες, όπου μια σταδιακή μείωση με

την ηλικία θεωρείται φυσιολογική. Η ευαισθησία ποικίλλει επίσης με τη συχνότητα (Σπυρίδης, Μουσική Ακουστική, 1990).

Ο πιο γνωστός αλγόριθμος αντιληπτικής κωδικοποίησης είναι ο αλγόριθμος MP3 (MPEG audio layer 3), που είναι ένα κομμάτι του προτύπου MPEG. Ο αλγόριθμός αυτός αναλύει και διαιρεί το ηχητικό φάσμα σε διάφορα τμήματα και ανιχνεύει φαινόμενα ακουστικής συγκάλυψης. Όταν τα βρει μηδενίζει τις τιμές στις συχνότητες αυτές. Αν έχουμε μερική συγκάλυψη αφήνει ένα μικρό αριθμό από bits ανεπηρέαστα και μηδενίζει τα υπόλοιπα. Τα υπόλοιπα bits μένουν ανεπηρέαστα.

Το πρότυπο MPEG-1 δεν περιλαμβάνει ακριβείς προδιαγραφές για έναν κωδικοποιητή MP3, αλλά παρέχει παραδείγματα ψυχοακουστικών μοντέλων. Ως εκ τούτου, υπάρχουν πολλοί διαφορετικοί κωδικοποιητές MP3 διαθέσιμοι, οι οποίοι παράγουν αρχεία διαφορετικής ποιότητας. Ένας κωδικοποιητής όμως που είναι ικανός να παράγει καλά αποτελέσματα σε κωδικοποίηση σε υψηλότερους ρυθμούς bit (όπως ο codec LAME) δεν είναι απαραίτητα τόσο καλός σε κωδικοποίηση με χαμηλότερους ρυθμούς bit (Wikipedia, MP3, 2013).

## Βιβλιογραφία

1. Durrant, J. D., & Lovrinic, J. H. (1984). *Bases of Hearing Sciences. Second Edition*. Williams & Wilkins.
2. Li, D., & O'Shaughnessy, D. (2003). *Speech processing: a dynamic and optimization-oriented approach pp. 41–48*. Marcel Dekker.
3. Olympus Imaging Corporation. (2012). *WS-400S User Manual*. Tokyo: Olympus.
4. Salomon, D. (2000). *Data Compression the Complete Reference 2nd Edition*. New York: Spring Verlag.
5. *Wikipedia, Auditory masking*. (2013, Ιούνιος). Ανάκτηση Ιούνιος 2013, από Wikipedia: [http://en.wikipedia.org/wiki/Auditory\\_masking](http://en.wikipedia.org/wiki/Auditory_masking)
6. *Wikipedia, MP3*. (2013, Ιούνιος). Ανάκτηση Μάιος 2013, από Wikipedia: <http://en.wikipedia.org/wiki/MP3>
7. Σπυρίδης, Χ. (1989). *Στοιχεία Ηλεκτρακουστικής*. Θεσσαλονίκη: ΑΠΘ, Υπηρεσία Δημοσιευμάτων.
8. Σπυρίδης, Χ. (1990). *Μουσική Ακουστική*. Θεσσαλονίκη: ΑΠΘ, Υπηρεσία δημοσιευμάτων.

