

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Πολυτεχνική Σχολή,
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών,
Τομέας Ηλεκτρονικής & Υπολογιστών

Συστήματα Πολυμέσων και Εικονική Πραγματικότητα

9ο Εξάμηνο

ΑΑC Κωδικοποίηση και Αποκωδικοποίηση



Διδάσκων: Χρήστος Δίου

Φοιτητές: Τζιρογλου Γεώργιος, 7990, gtziroga@ece.auth.gr

Χρυσοβέργης Ηλίας, 8009, iliachry@ece.auth.gr

Συστήματα Πολυμέσων και Εικονική Πραγματικότητα

Εισαγωγή

Η εργασία στοχεύει στην υλοποίηση ενός κωδικοποιητή/αποκωδικοποιητή ήχου κατά το πρότυπο Advanced Audio Coding (AAC). Η υλοποίηση αποτελείται από 3 επίπεδα (levels) του κωδικοποιητή/αποκωδικοποιητή. Κάθε επίπεδο περιέχει ένα μεγαλύτερο ποσοστό υλοποίησης του AAC κωδικοποιητή/αποκωδικοποιητή.

Στην παρούσα εργασία έχουν υλοποιηθεί και τα τρία επίπεδα, με τις συναρτήσεις κάθε επιπέδου να φαίνονται στον παρακάτω πίνακα.

Συναρτήσεις Level 1 (10 Αρχεία)	Σύντομη περιγραφή
SSC.m , filterbank.m , ifilterbank.m , AACoder1.m , iAACoder1.m , demoAAC1.m	Υλοποιήθηκαν όπως περιγράφονται στην εκφώνηση της εργασίας
mdct4.m , imdct4.m	Υλοποιεί τον ευθύ και τον αντιστροφο MDCT-4 όπως έχει υλοποιηθεί στο http://www.ee.columbia.edu/~marios/mdct/mdct_giraffe.html
SinWindow.m , KBDWindow.m	Υλοποιούν τα παράθυρα KBD και SIN. Το μεν πρώτο δέχεται σαν είσοδο το μέγεθος του παραθύρου και την παράμετρο alpha, ενώ το δεύτερο το μέγεθος του παραθύρου.

Συναρτήσεις Level 2 (16 Αρχεία)	Σύντομη περιγραφή
SSC.m , filterbank.m , ifilterbank.m , AACoder1.m , iAACoder1.m , mdct4.m , imdct4.m , SinWindow.m , KBDWindow.m	Οι απαραίτητες συναρτήσεις από το Level 1.
TNS.m , iTNS.m , AACoder2.m , iAACoder2.m , demoAAC2.m	Υλοποιήθηκαν όπως περιγράφονται στην εκφώνηση της εργασίας
quantizer.m	Κβαντίζει τον συντελεστή του TNS με βάση τις στάθμες κβαντισμού που δίνονται στην εκφώνηση.
TableB219.mat	MAT-file με τους πίνακες B.2.1.9a και B.2.1.9b του αρχείου w2230tfa

Συναρτήσεις Level 3 (29 Αρχεία)	Σύντομη περιγραφή
SSC.m , filterbank.m , ifilterbank.m , AACoder1.m , iAACoder1.m , mdct4.m , imdct4.m , SinWindow.m , KBDWindow.m , TNS.m , iTNS.m , AACoder2.m , iAACoder2.m , quantizer.m , TableB219.mat	Οι απαραίτητες συναρτήσεις και αρχεία από τα Level 1 και Level 2.
psycho.m , AACquantizer.m , iAACquantizer.m , encodeHuff.m , loadLUT.m , decodeHuff.m , AACoder3.m , iAACoder3.m , demoAAC3.m	Υλοποιήθηκαν όπως περιγράφονται στην εκφώνηση της εργασίας
HannWindow.m	Υλοποιεί παράθυρο Hann μήκους s.
huffCodebooks.mat , huffCodebookSF.mat	Απαραίτητα αρχεία για τη συμπίεση Huffman
spreadingfun.m	Συνάρτηση κατανομής μάσκας. Τα αποτελέσματα της έχουν προαποθηκευθεί.
spreadfuna.mat	Τα αποτελέσματα της spreadingfun.

Συστήματα Πολυμέσων και Εικονική Πραγματικότητα

1^ο Επίπεδο-

Στο πρώτο επίπεδο έχει υλοποιηθεί η βαθμίδα Sequence Segmentation Control, Filterbank και η αντίστροφη της iFilterbank, η AACoder1 και η αντίστροφη της iAACoder1 οι οποίες υλοποιούν την κωδικοποίηση και αποκωδικοποίηση για ένα αρχείο ήχου στο πρώτο επίπεδο καθώς και η συνάρτηση demoAAC1 η οποία επιδεικνύει την κωδικοποίηση του 1^{ου} επιπέδου.

Τα αποτελέσματα που προέκυψαν για το πρώτο επίπεδο για τα δύο διαφορετικά είδη παραθύρων είναι τα εξής:

1. Kaiser-Bessel-Derived (KBD) Window

```
Time elapsed for coding is 1.94684 seconds
Time elapsed for decoding is 1.46747 seconds

SNR =

306.7024  306.7256
```

2. Sinusoid (SIN) Window

```
Time elapsed for coding is 1.36021 seconds
Time elapsed for decoding is 1.28808 seconds

SNR =

307.8367  307.9088
```

Τα εξαιρετικά υψηλά SNR οφείλονται στο γεγονός πως η διαδικασία είναι πλήρως αντιστρεπτή καθώς οι μετασχηματισμοί από τα πεδία χρόνου-συχνότητας δεν επιφέρουν απώλεια πληροφορίας.

2^ο Επίπεδο-

Στο δεύτερο επίπεδο έχει υλοποιηθεί η βαθμίδα TNS (Temporal Noise Shaping) καθώς και η αντίστροφη της iTNS. Ακόμη, υλοποιήθηκαν οι AACoder2 και iAACoder2 που υλοποιούν την κωδικοποίηση και αποκωδικοποίηση ενός αρχείου ήχου για το δεύτερο επίπεδο. Τέλος, υλοποιήθηκε η συνάρτηση demoAAC2 η οποία επιδεικνύει την κωδικοποίηση του 2^{ου} επιπέδου. Για την ορθή λειτουργία του 2^{ου} επιπέδου είναι απαραίτητη και η χρήση των συναρτήσεων του 1^{ου} επιπέδου, οπότε συμπεριλαμβάνονται και αυτές στον φάκελο του 2^{ου} επιπέδου, όπως φαίνεται στον πίνακα της εισαγωγής.

Τα αποτελέσματα που προέκυψαν για το δεύτερο επίπεδο για παράθυρο KBD δίνονται στην συνέχεια:

```
Time elapsed for coding is 3.39432 seconds
Time elapsed for decoding is 1.37952 seconds

SNR =

306.6842  306.6802
```

Συγκρίνοντας τα αποτελέσματα του 2^{ου} επιπέδου με αυτά του 1^{ου} επιπέδου μπορούμε να παρατηρήσουμε ότι το μόνο που έχει αλλάξει σημαντικά είναι ο χρόνος που απαιτείται για την κωδικοποίηση του σήματος ήχου. Αυτός περίπου διπλασιάζεται στο δεύτερο επίπεδο, ενώ το SNR για κάθε κανάλι καθώς και ο χρόνος αποκωδικοποίησης κυμαίνονται σε ίδια επίπεδα. Προφανώς τα παραπάνω εξηγούνται από την αντιστρεπτότητα των διαδικασιών σε σχέση με τα παρόμοια SNR, και την περισσότερη επεξεργασία που υφίστανται τα σήματα σε σχέση με την αύξηση του χρόνου.

3^ο Επίπεδο-

Στο τρίτο επίπεδο έχει αναπτυχθεί η συνάρτηση psycho, η οποία υλοποιεί την βαθμίδα του ψυχοακουστικού μοντέλου για ένα κανάλι. Στην συνέχεια, αναπτύχθηκαν οι βαθμίδες AACquantizer και iAACquantizer οι οποίες υλοποιούν τον κβαντιστή και αποκβαντιστή του 3^{ου} επιπέδου λαμβάνοντας υπόψιν τα αποτελέσματα του ψυχοακουστικού μοντέλου. Επίσης, σε αυτό το επίπεδο χρησιμοποιήθηκαν οι έτοιμες συναρτήσεις encodeHuff και decodeHuff, οι οποίες δόθηκαν από τον διδάσκοντα, και οι οποίες υλοποιούν την κωδικοποίηση και αποκωδικοποίηση Huffman. Τέλος, υλοποιήθηκαν οι συναρτήσεις AACCoder3 και iAACCoder3 οι οποίες υλοποιούν την κωδικοποίηση και αποκωδικοποίηση ενός αρχείου ήχου για το τρίτο επίπεδο, καθώς και η συνάρτηση demoAAC3 η οποία επιδεικνύει την κωδικοποίηση του 3^{ου} επιπέδου.

Προτού προχωρήσουμε στην παρουσίαση των αποτελεσμάτων, θεωρούμε σκόπιμο να αναφέρουμε τη μεθοδολογία που ακολουθήθηκε για τον υπολογισμό του μεγέθους του τραγουδιού έπειτα από την κωδικοποίηση.

1. Οι κωδικοποιημένες ακολουθίες Huffman των sfc και των κβαντισμένων συντελεστών MDCT θεωρήθηκαν ως ακολουθίες από bits με μέγεθος όσο και ο αριθμός των bits που περιέχουν.
2. Τα frameType και winType απαιτούν 12 byte για κάθε frame.
3. Για παράθυρα ESH επιπλέον 128 bytes χρειάζονται για τους συντελεστές G, ενώ 32 για τους TNS.
4. Για όλα τα υπόλοιπα παράθυρα απαιτούνται επιπλέον 20 bytes.

Τα αποτελέσματα που προέκυψαν για το τρίτο επίπεδο για παράθυρο KBD δίνονται στην συνέχεια:

Time elapsed for coding is 149.08211 seconds
Time elapsed for decoding is 85.28417 seconds
Coding achieves a compression of 13.45439%
Bitrate of initial song is 187.500000 KB/sec
Bitrate after coding is 25.226981 KB/sec

SNR =

4.2881 3.9095

Σε ότι αφορά το 3^ο επίπεδο έχουμε αρκετά ενδιαφέροντα πράγματα να παρατηρήσουμε. Αρχικά, μπορούμε να διαπιστώσουμε ότι ο χρόνος που απαιτείται για κωδικοποίηση και αποκωδικοποίηση έχει αυξηθεί αρκετά, απαιτώντας περίπου δυόμιση λεπτά για κωδικοποίηση και ενάμιση λεπτό για αποκωδικοποίηση. Αυτά τα νούμερα είναι αρκετά αποθαρρυντικά για εφαρμογές πραγματικού χρόνου αλλά η συγκεκριμένη διαδικασία θα μπορούσε να χρησιμοποιηθεί για εφαρμογές όπου η κωδικοποίηση γίνεται μία φορά ενώ η αποκωδικοποίηση χρησιμοποιεί κάποια καλύτερη τεχνική για την ελάττωση του απαιτούμενου χρόνου. Επίσης, μπορούμε να παρατηρήσουμε ότι η συγκεκριμένη

Συστήματα Πολυμέσων και Εικονική Πραγματικότητα

διαδικασία επιτυγχάνει συμπίεση της τάξης του 13.5% εις βάρος της μείωσης του SNR κατά περισσότερο από 300 dB για κάθε κανάλι, κάτι που δεν επιβαρύνει αρκετά την ποιότητα του κωδικοποιημένου ήχου. Αυτό μπορούμε να το διαπιστώσουμε αν ακούσουμε το αρχείο εξόδου του κωδικοποιητή/αποκωδικοποιητή. Το γεγονός αυτό οφείλεται στα πολύ θετικά αποτελέσματα που επιτυγχάνονται με την ελάττωση της περίσσειας πληροφορίας του αρχικού σήματος χρησιμοποιώντας το ψυχοακουστικό μοντέλο καθώς και της κωδικοποίησης εντροπίας κατά Huffman.

Αναφορές

- [1] «Εντροπία, αποσυσχέτιση και μετασχηματισμοί στην κωδικοποίηση πολυμέσων», Χρήστος Δίου
- [2] Σημειώσεις «Συστήματα Πολυμέσων και Εικονική Πραγματικότητα, Α. Ντελόπουλος
- [3] «Συστήματα Πολυμέσων», P. Havaldar & G. Medioni
- [4] Enhancing the performance of perceptual audio coders by using temporal noise shaping (TNS)
- [5] Principles of Psychoacoustics, Yiqing Lin & Waleed H. Abdulla