

Συστήματα Πολυμέσων και
Εικονική Πραγματικότητα

Εργασία

Μακρυλάκη Βασιλική
makrylav@ece.auth.gr
9174

Σκοπός της εργασίας είναι η υλοποίηση ενός κωδικοποιητή / αποκωδικοποιητή ήχου κατά το πρότυπο Advanced Audio Coder (AAC). Οι προϋποθέσεις για την εκτέλεση του κώδικα είναι τα αρχεία ήχου να είναι: .wav, 2 καναλιών και συχνότητας δειγματοληψίας 48 kHz.

Level 1:

Συνάρτηση SSC:

Αρχικά ελέγχεται αν το προηγούμενο frame είναι LSS ή LPS, καθώς αν ισχύει αυτό, το τωρινό είναι OLS. Στη συνέχεια υλοποιούνται τα 4 βήματα που περιγράφονται στην ενότητα 2.1 της εκφώνησης με σκοπό τον καθορισμό του frameType για κάθε κανάλι. Το φιλτράρισμα του επόμενου frame στο βήμα 1 γίνεται με τη συνάρτηση filter του Matlab. Ο καθορισμός του τελικού frameType από τα frameTypes των δύο καναλιών γίνεται με βάση τον πίνακα απόφασης της εκφώνησης.

Συνάρτηση filterbank:

Στη συνάρτηση αυτή, αφού επιλεγθεί και εφαρμοσθεί το κατάλληλο παράθυρο, ανάλογα με τον τύπο του frame, στο σήμα εφαρμόζεται mdct.

Τα παράθυρα έχουν υλοποιηθεί στις συναρτήσεις κάτω από τη filterbank, σύμφωνα με τους τύπους που δίνονται στην εκφώνηση, με ονόματα kbd και sinW, ενώ υλοποιήθηκε και μια βοηθητική συνάρτηση με όνομα windows, για την απόφαση του αν το απαιτούμενο παράθυρο είναι τύπου long ή short. Στην kbd χρησιμοποιήθηκαν οι συναρτήσεις kaiser και cumsum του Matlab. Τέλος, η συνάρτηση mdct.m (όπως και η imdct.m) βρέθηκαν στο

https://www.ee.columbia.edu/~marios/mdct/mdct_giraffe.html .

Συνάρτηση ifilterbank:

Εδώ εφαρμόζεται στο σήμα `imdct` και το κατάλληλο παράθυρο, ακριβώς με τη λογική που περιγράφηκε παραπάνω, με σκοπό την ανάκτηση του σήματος στο χρόνο.

Συνάρτηση AACoder1:

Στη συνάρτηση αυτή υλοποιείται η κωδικοποίηση του σήματος. Αφού διαβαστεί το όνομα του αρχείου ήχου, το σήμα υφίσταται `zero-padding`, για τη μείωση του θορύβου που εμφανίζεται στα άκρα του σήματος μετά την κωδικοποίηση. Αυτό υλοποιείται με την προσθήκη από μισό `frame` μηδενικών τόσο στην αρχή όσο και στο τέλος του σήματος, ενώ στο τέλος προστίθενται επίσης όσα μηδενικά υπολείπονται μέχρι το μέγεθος του σήματος να είναι πολλαπλάσιο του 1024.

Έπειτα, αφού καθοριστεί ο αριθμός των `frames` που θα προκύψουν, επιλέγεται τυχαία το παράθυρο που θα χρησιμοποιηθεί (`KBD` ή `SIN`). Στην αρχή, το προηγούμενο `frameType` επιλέγεται να είναι `OLS`. Επίσης, κάτω απ' την `AACoder1` έχει υλοποιηθεί η βοηθητική συνάρτηση `slice`, για τον εύκολο υπολογισμό του τωρινού και του επόμενου `frame`.

Τέλος, για κάθε `frame` αποθηκεύονται στη δομή `AACSeq1` τα:

`frameType`, `winType`, `chl.frameF`, `chr.frameF` που προκύπτουν με την κλήση των συναρτήσεων `SSC` και `filterbank`.

Συνάρτηση iAACoder1:

Στη συνάρτηση αυτή υλοποιείται η αποκωδικοποίηση του κάθε `frame`, μέσω της `ifilterbank`. Μετά την αποκωδικοποίηση αφαιρούνται τα 2 ημι-Frames μηδενικών που είχαν προστεθεί κατά το `zero-padding` στην αρχή και στο τέλος, ωστόσο τα υπόλοιπα μηδενικά που είχαν προστεθεί στο τέλος παραμένουν, καθώς δε γνωρίζουμε το πλήθος τους.

Συνάρτηση demoAACoder1:

Αφού διαβαστεί το αρχείο εισόδου (.wav) και γίνει η κωδικοποίηση και η αποκωδικοποίηση, κρατιέται το κομμάτι των σημάτων με κοινό μήκος. Η διαφορά των δύο σημάτων λαμβάνεται ως θόρυβος και το SNR υπολογίζεται με τη συνάρτηση snr του Matlab. Επίσης υλοποιείται μια έξτρα συνάρτηση, η NRMSE_fun στην οποία υπολογίζεται το Normal Root Mean Square Error για κάθε κανάλι.

Αποτελέσματα

```
SNR = demoAAC1('Input.wav', 'Output.wav');
```

Με KBD παράθυρο

Chosen window: KBD

Normal Root Mean Square Error (NRMSE) for channel 1: 4.65548e-17

Normal Root Mean Square Error (NRMSE) for channel 2: 5.02628e-17

SNR: 301.658 dB

Elapsed time is 3.683346 seconds.

Με SIN παράθυρο

Chosen window: SIN

Normal Root Mean Square Error (NRMSE) for channel 1: 4.91655e-17

Normal Root Mean Square Error (NRMSE) for channel 2: 5.13683e-17

SNR: 307.824 dB

Elapsed time is 3.332012 seconds.

Level 2:

Συνάρτηση TNS:

Για την κανονικοποίηση των συντελεστών MDCT, όπως περιγράφεται στο βήμα 1 της εκφώνησης, υλοποιήθηκε η βοηθητική συνάρτηση `normaliseMDCT`. Έπειτα, για τη γραμμική πρόβλεψη των συντελεστών του `fir` φίλτρου χρησιμοποιήθηκε η συνάρτηση `lpc` του Matlab. Ο ομοιόμορφος συμμετρικός κβαντιστής βήματος 0.1 υλοποιείται στη βοηθητική συνάρτηση `uniformQuantize`. Στη συνέχεια, καλείται η συνάρτηση `assertInvertible`, η οποία φροντίζει για την ευστάθεια του αντίστροφου φίλτρου. Τέλος, το `fir` φίλτρο εφαρμόζεται μέσω της συνάρτησης `filter` του Matlab.

Στη συνάρτηση `iTNS` το μόνο που συμβαίνει είναι η εφαρμογή του αντίστροφου φίλτρου.

Συναρτήσεις `AACoder2` και `iAACoder2`:

Οι συναρτήσεις αυτές είναι παρόμοιες με τις `AACoder1` και `iAACoder1` αντίστοιχα που περιγράφηκαν στο Level 1, με τη διαφορά ότι προστίθεται κατάλληλα η κλήση των συναρτήσεων `TNS` και `iTNS`.

Τέλος, το demo επίδειξης της λειτουργίας είναι ίδιο με αυτό του Level 1.

Αποτελέσματα

```
SNR = demoAAC2('Input.wav', 'Output.wav');
```

Με KBD παράθυρο

Chosen window: KBD

Normal Root Mean Square Error (NRMSE) for channel 1: 4.67432e-17

Normal Root Mean Square Error (NRMSE) for channel 2: 5.0807e-17

SNR: 301.654 dB

Elapsed time is 5.573821 seconds.

Με SIN παράθυρο

Chosen window: SIN

Normal Root Mean Square Error (NRMSE) for channel 1: 5.0819e-17

Normal Root Mean Square Error (NRMSE) for channel 2: 5.23101e-17

SNR: 307.781 dB

Elapsed time is 7.247727 seconds.

Level 3:

Συνάρτηση psycho:

Στο πρώτο βήμα αποθηκεύεται σε έναν πίνακα το αποτέλεσμα της spreadingFunction για κάθε δυνατό ζευγάρι μπαντών. Η spreadingFunction υλοποιείται σε μια βοηθητική συνάρτηση με όνομα spreadingfun.

Στη συνέχεια, υλοποιείται το παράθυρο Hann και εφαρμόζεται στο εκάστοτε frame ή subframe (τωρινό, προηγούμενο και προ-προηγούμενο). Σε αυτές τις νέες ακολουθίες εφαρμόζεται fft (στα πρώτα 1024 δείγματα) μέσω της βοηθητικής συνάρτησης FFT η οποία επιστρέφει το μέτρο και τη φάση του fft.

Τα υπόλοιπα βήματα είναι απλές υλοποιήσεις των εξισώσεων που δίνονται στην εκφώνηση, με τη μόνη σημείωση ότι στο βήμα 7 το Tonality Index περιορίζεται στο [0.0001, 0.9999].

Συνάρτηση AACQuantizer:

Αρχικά υπολογίζονται τα κατώφλια ακουστικότητας του κωδικοποιητή μέσω της ενέργειας των συντελεστών MDCT σε κάθε scale factor band και του SMR.

Έπειτα, αφού γίνει αρχικοποίηση του scale factor gain, για κάθε μπάντα πραγματοποιείται μη ομοιόμορφη κβάντιση και αποκβάντιση των συντελεστών MDCT. Η διαδικασία επαναλαμβάνεται όσο η ισχύς του σφάλματος κβαντισμού είναι κάτω από το κατώφλι ακουστικότητας και η μέγιστη διαφορά μεταξύ 2 διαδοχικών scale factors είναι κάτω του 60, αυξάνοντας κάθε φορά το scale factor gain κατά 1.

Η συνάρτηση iAACQuantizer υλοποιεί την αντίστροφη διαδικασία εφαρμόζοντας τον αποκβαντιστή.

Συνάρτηση AACCoder3:

Η συνάρτηση αυτή είναι παρόμοια με την AACCoder2 με τις εξής παραλλαγές:

- α) Αρχικοποιείται το προηγούμενο και προ-προηγούμενο frame ως μηδενικό.
- β) Προστίθεται κατάλληλα η κλήση των συναρτήσεων psycho, threshold (βοηθητική συνάρτηση που υπολογίζει τα κατώφλια του ψυχοακουστικού μοντέλου για κάθε κανάλι), AACquantizer και encodeHuff.
- γ) Στο τέλος κάθε επανάληψης ενημερώνεται το προηγούμενο και προ-προηγούμενο frame που θα χρησιμοποιηθούν στην επόμενη επανάληψη.
- δ) Τέλος, σώζει ένα mat αρχείο με το όνομα που έχει δοθεί στην μεταβλητή fnameAACoded.

Συνάρτηση iAACCoder3:

Είναι παρόμοια με την υλοποίηση της iAACCoder2, με τη διαφορά ότι προστίθεται κατάλληλα η κλήση των συναρτήσεων decodeHuff, iAACquantizer, και η τελική ακολουθία κανονικοποιείται για να αποφευχθεί το 'clipping'.

Συνάρτηση demoAAC3:

Μέχρι και τον υπολογισμό του Normal Mean Square Error η συνάρτηση είναι ίδια με το demoAAC2. Στη συνέχεια, γίνεται υπολογισμός του bitrate τόσο της αρχικής όσο και της τελικής ακολουθίας. Αυτό γίνεται με τη βοήθεια της συνάρτησης audioinfo του Matlab από την οποία εξάγονται τα BitsPerSample τα οποία πολλαπλασιάζονται επί fs. Τέλος, το compression υπολογίζεται ως bitrate_τελικό / bitrate_αρχικό.

Αποτελέσματα

```
[SNR, bitrate, compression] = demoAAC3('Input.wav',  
'Output.wav', 'AACSeq3.mat')
```

Με KBD παράθυρο

Encoding:

Chosen window: KBD

Decoding:

Normal Root Mean Square Error (NRMSE) for channel 1: 0.00741829

Normal Root Mean Square Error (NRMSE) for channel 2: 0.0730227

SNR: 9.00308 dB

Bitrate: 768 kbps

Compression: 1

Elapsed time is 260.961459 seconds.

Με SIN παράθυρο

Encoding:

Chosen window: SIN

Decoding:

Normal Root Mean Square Error (NRMSE) for channel 1: 0.00740295

Normal Root Mean Square Error (NRMSE) for channel 2: 0.0815288

SNR: 8.45216 dB

Bitrate: 768 kbps

Compression: 1

Elapsed time is 321.333695 seconds.

Προφανώς, λόγω της κωδικοποίησης η τιμή του SNR έχει πέσει, ωστόσο στο αυτό δεν είναι αντιληπτή η διαφορά με το πρωτότυπο κομμάτι. Επίσης, το compression λογικά θα έπρεπε να έχει μικρότερη τιμή, γεγονός που αποκαλύπτει κάποιο λάθος ίσως στον υπολογισμό του.