



ΠΟΛΥΜΕΣΑ

Εργασία



Ευάγγελος Κωστινούδης

Νικόλαος Κατωμέρης

ΑΕΜ: 8708, 8551

✉ ekostinou@auth.gr, ngkatomer@auth.gr



9^ο ΕΞΑΜΗΝΟ
ΤΜΗΜΑ ΗΜΜΥ ΑΠΘ
2018-2019

Περιεχόμενα

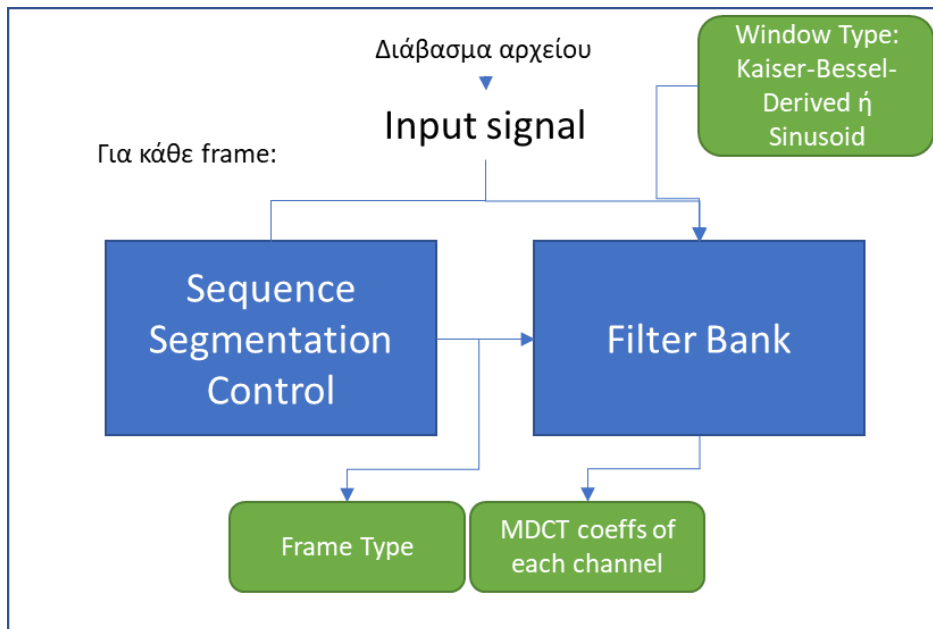
1 ^ο επίπεδο	3
Λειτουργία	3
Ενδεικτικά αποτελέσματα	4
2 ^ο επίπεδο	6
Λειτουργία	6
Ενδεικτικά αποτελέσματα	6
3 ^ο επίπεδο	7
Λειτουργία	7
Ενδεικτικά αποτελέσματα	7
Συμβάσεις	10
Κώδικας.....	11

1^ο επίπεδο

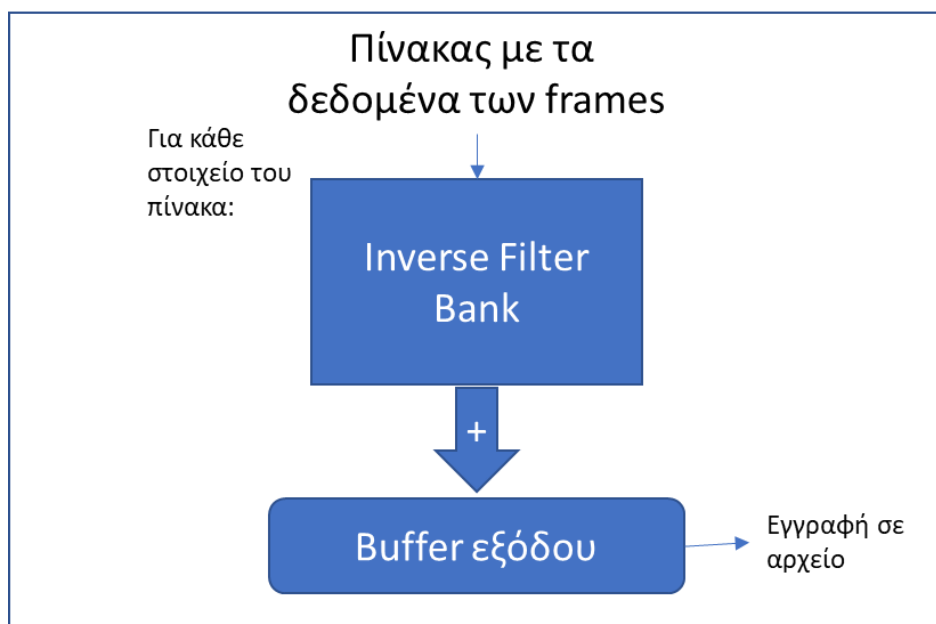
Λειτουργία

Η συνάρτηση `demoAAC1()` επιδεικνύει τη λειτουργία του 1^{ου} επιπέδου της εργασίας.

Συγκεκριμένα, πρώτα γίνεται κωδικοποίηση των δύο καναλιών του σήματος ήχου με την συνάρτηση `AACoder1()` σε ένα πίνακα που περιέχει για κάθε χωριστό frame τα στοιχεία των πράσινων πλαισίων του παρακάτω γράφου. Ο τύπος παραθύρου επιλέγεται μέσα στην συνάρτηση και στον παραδοτέο κώδικα έχει προεπιλεγεί ο τύπος Kaiser-Bessel-Derived.



Στη συνέχεια, τα δεδομένα αυτά περνούν στη συνάρτηση `iAACoder1()` η οποία αναλαμβάνει την αποκωδικοποίησή τους και γράφει το αποτέλεσμα σε ένα νέο αρχείο ήχου.



Ενδεικτικά αποτελέσματα

Η εκτέλεση της demoAAC1() με το αρχείο ήχου που δόθηκε ('LicorDeCalandracawav') πέτυχε:

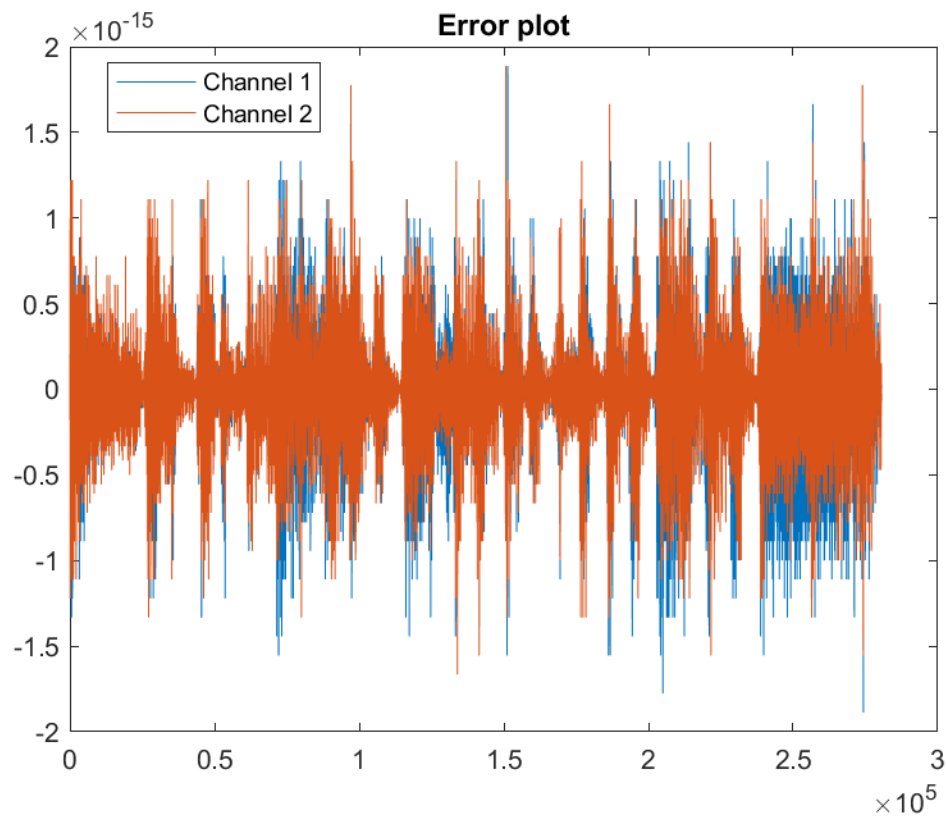
- Με παράθυρα τύπου 'KBD'

```
>> demoAAC1("LicorDeCalandracawav", "demol.wav");  
Coding: time elapsed is 0.2032 seconds  
Decoding: time elapsed is 0.1801 seconds  
Channel 1 SNR: 298.5570 dB  
Channel 2 SNR: 298.6078 dB
```

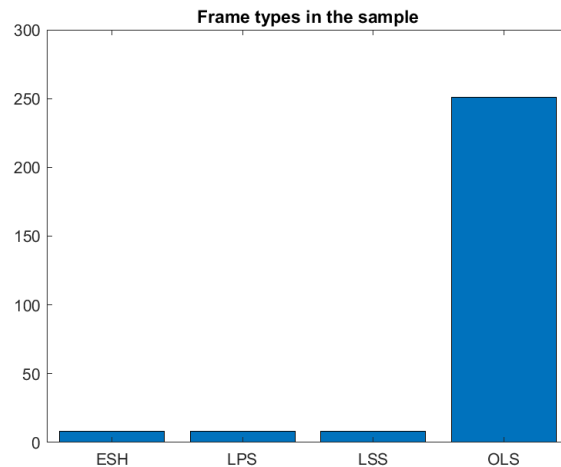
- Με παράθυρα τύπου 'SIN'

```
>> demoAAC1("LicorDeCalandracawav", "demol.wav");  
Coding: time elapsed is 0.1756 seconds  
Decoding: time elapsed is 0.1389 seconds  
Channel 1 SNR: 307.8097 dB  
Channel 2 SNR: 307.8813 dB
```

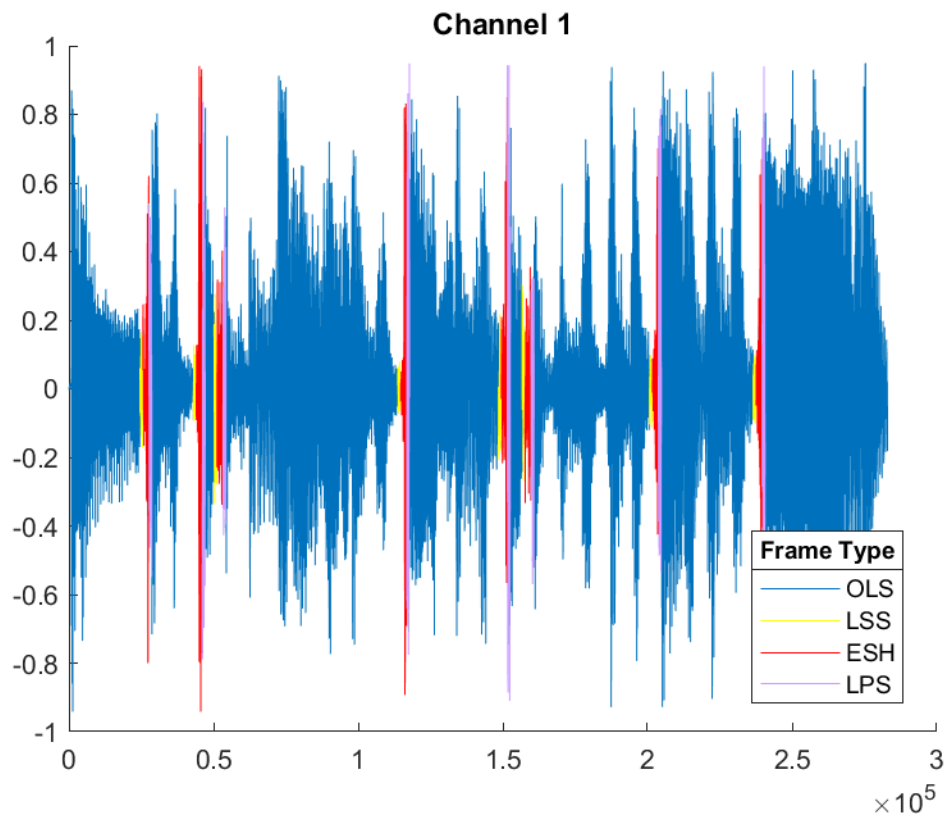
Οι τιμές θορύβου (σφάλματος) παρουσιάζονται στο επόμενο διάγραμμα και είναι της τάξης του 10^{-15} . Ο μετασχηματισμός MDCT είναι πλήρως αντιστρέψιμος, οπότε η τάξη αυτή των σφαλμάτων ήταν αναμενόμενη αφού προέρχεται από τις πράξεις.



Στο επόμενο διάγραμμα παρουσιάζεται ο αριθμός των frames όπως κατηγοριοποιήθηκαν. Όπως φαίνεται, στην συντριπτική τους πλειοψηφία δόθηκε ο τύπος: “ONLY_LONG_SEQUENCE” ενώ υπάρχουν μόλις 8 frames σε κάθε έναν από τους υπόλοιπους τύπους. Αναμενόμενα, ο αριθμός των τριών τελευταίων είναι ίδιος καθώς πριν και μετά από τα frames τύπου ‘ESH’ βρίσκονται τα ‘LPS’ και ‘LSS’.



Τα frames τύπου ‘ESH’ βρίσκονται στα σημεία με τις πολύ απότομες αλλαγές στα δείγματα του ήχου. Αυτό φαίνεται στο επόμενο διάγραμμα όπου παρουσιάζονται οι τύποι των frames πάνω στο κανάλι 1 του δείγματος.



2^ο επίπεδο

Λειτουργία

Η συνάρτηση demoAAC2() επιδεικνύει τη λειτουργία του 2^{ου} επιπέδου της εργασίας.

Στο επίπεδο αυτό, καλούνται οι συναρτήσεις AACoder2() και iAACoder2() αντίστοιχα με τις AACoder1() και iAACoder1() του 1^{ου} επιπέδου. Οι συναρτήσεις αυτές είναι όμοιες με τις αντίστοιχές τους με τη διαφορά ότι τα δεδομένα του πίνακα με τα frames περνάνε από μια νέα βαθμίδα για Temporal Noise Shaping (TNS) προτού δοθούν στην iAACoder2() που με τη σειρά της θα χρησιμοποιήσει αρχικά την iTNS() για να επαναφέρει τα TNS δεδομένα στη μορφή που τα «θέλει» η iFilterBank().

Ενδεικτικά αποτελέσματα

Η εκτέλεση της demoAAC2() με το αρχείο ήχου που δόθηκε ('LicorDeCalandracawav') πέτυχε:

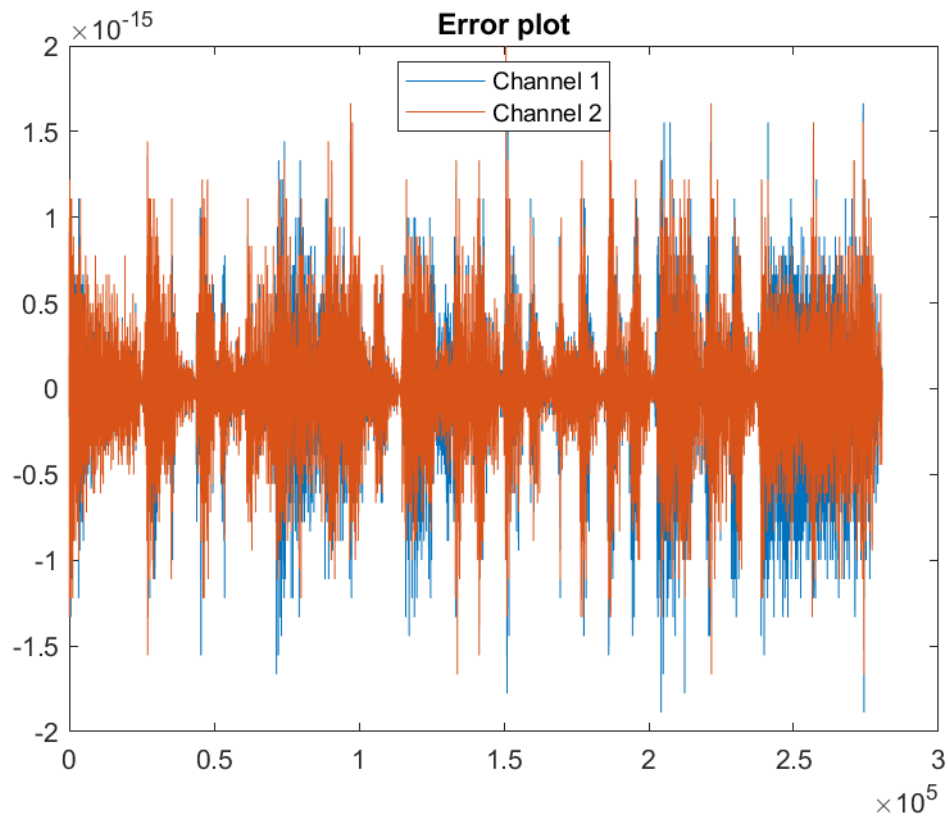
- Με παράθυρα τύπου 'KBD'

```
>> demoAAC2("LicorDeCalandracawav", "demo1.wav");
Coding: time elapsed is 0.8712 seconds
Decoding: time elapsed is 0.1888 seconds
Channel 1 SNR: 298.3997 dB
Channel 2 SNR: 298.4246 dB
```

- Με παράθυρα τύπου 'SIN'

```
>> demoAAC2("LicorDeCalandracawav", "demo2.wav");
Coding: time elapsed is 0.9782 seconds
Decoding: time elapsed is 0.1757 seconds
Channel 1 SNR: 307.0511 dB
Channel 2 SNR: 307.1191 dB
```

Οι τιμές θορύβου (σφάλματος) παρουσιάζονται στο επόμενο διάγραμμα και είναι της τάξης του 10^{-15} . Κι εδώ, η επιπλέον βαθμίδα 'TNS' είναι πλήρως αντιστρέψιμη επειδή στην ουσία εφαρμόζεται φίλτρο FIR (που υπολογίζεται ώστε να έχει ευσταθές αντίστροφο) και στη συνέχεια το ευσταθές αντίστροφό του, οπότε τα σφάλματα είναι αντίστοιχα με αυτά στο 1^ο επίπεδο.



3^ο επίπεδο

Λειτουργία

Η συνάρτηση `demoAAC3()` επιδεικνύει τη λειτουργία του 3^{ου} επιπέδου της εργασίας.

Στο επίπεδο αυτό, η ροή των δεδομένων ακολουθεί το διάγραμμα του σχήματος 1 της εκφώνησης με τον απλοποιημένο κωδικοποιητή AAC. Με εξαίρεση την βαθμίδα `bit reservoir`.

Αυτό συμβαίνει μέσα στη συνάρτηση `AACoder3()`.

Η αποκωδικοποίηση γίνεται στην συνάρτηση `iAACoder3()` όπου τα δεδομένα ακολουθούν το διάγραμμα του σχήματος 2 της εκφώνησης με τον απλοποιημένο αποκωδικοποιητή AAC.

Ενδεικτικά αποτελέσματα

Η εκτέλεση αυτής `demoAAC3()` με το αρχείο ήχου που δόθηκε ('LicorDeCalandraca.wav') πέτυχε:

- Με παράθυρα τύπου 'KBD'

```
>> demoAAC3("LicorDeCalandraca.wav", "demo3.wav", "AACSeq3.mat");  
Warning: Data clipped when writing file.  
> In audiowrite>clipInputData (line 396)  
   In audiowrite (line 176)  
   In iAACoder3 (line 90)  
   In demoAAC3 (line 28)  
Coding: time elapsed is 150.7464 seconds  
Decoding: time elapsed is 65.4369 seconds  
Uncompressed audio: 1.0795 MB (9055296 bits)  
Compressed struct: 165.2946 KB (1354093 bits)  
Compressed ratio: 14.9536 % (x 6.6874)  
Channel 1 SNR: 5.5519 dB  
Channel 2 SNR: 5.2433 dB
```

- Με παράθυρα τύπου 'SIN'

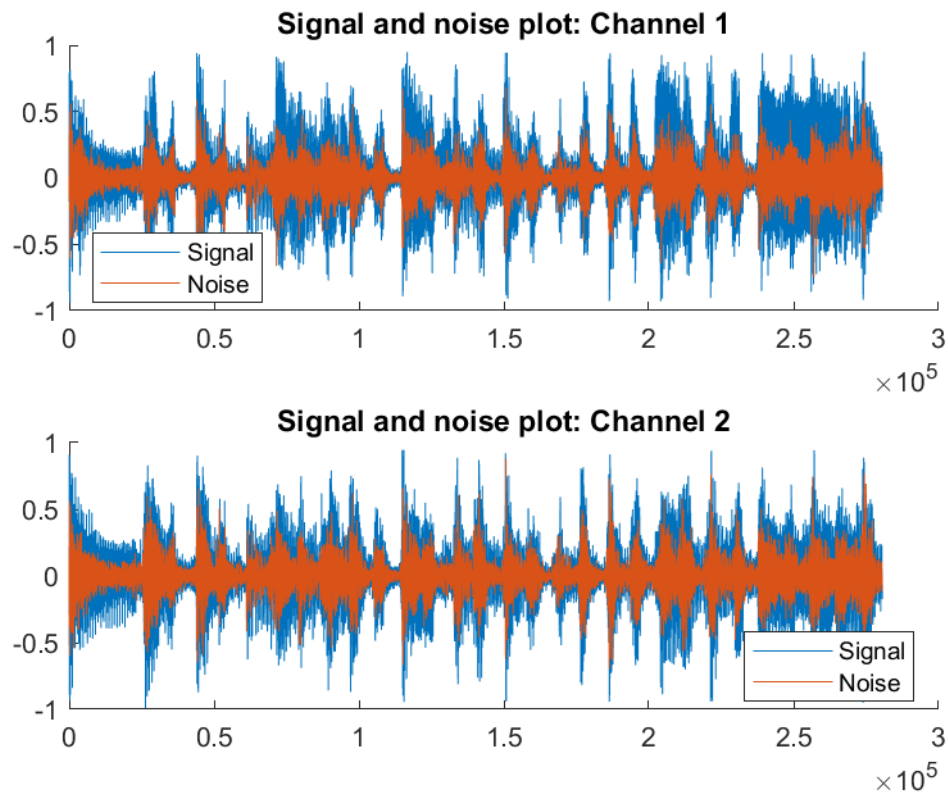
```
>> demoAAC3("LicorDeCalandraca.wav", "demo3.wav", "AACSeq3.mat");  
Warning: Data clipped when writing file.  
> In audiowrite>clipInputData (line 396)  
   In audiowrite (line 176)  
   In iAACoder3 (line 90)  
   In demoAAC3 (line 28)  
Coding: time elapsed is 151.3528 seconds  
Decoding: time elapsed is 65.4216 seconds  
Uncompressed audio: 1.0795 MB (9055296 bits)  
Compressed struct: 165.5974 KB (1356574 bits)  
Compressed ratio: 14.9810 % (x 6.6751)  
Channel 1 SNR: 5.6032 dB  
Channel 2 SNR: 5.2444 dB
```

Σημειώνεται ότι το μέγεθος αυτής συμπίεσμνης δομής που αναφέρεται είναι ενδεικτικό του θεωρητικού ελάχιστου μεγέθους της δομής αυτής και δεν ανταποκρίνεται στο πραγματικό μέγεθος που έχει στην υλοποίηση.

Το warning που εμφανίζεται αφορά αμελητέο clipping στα δεδομένα και δεν επηρεάζει την ποιότητα του αρχείου εξόδου.

Το bitrate ήταν περίπου 227.6 Kbps.

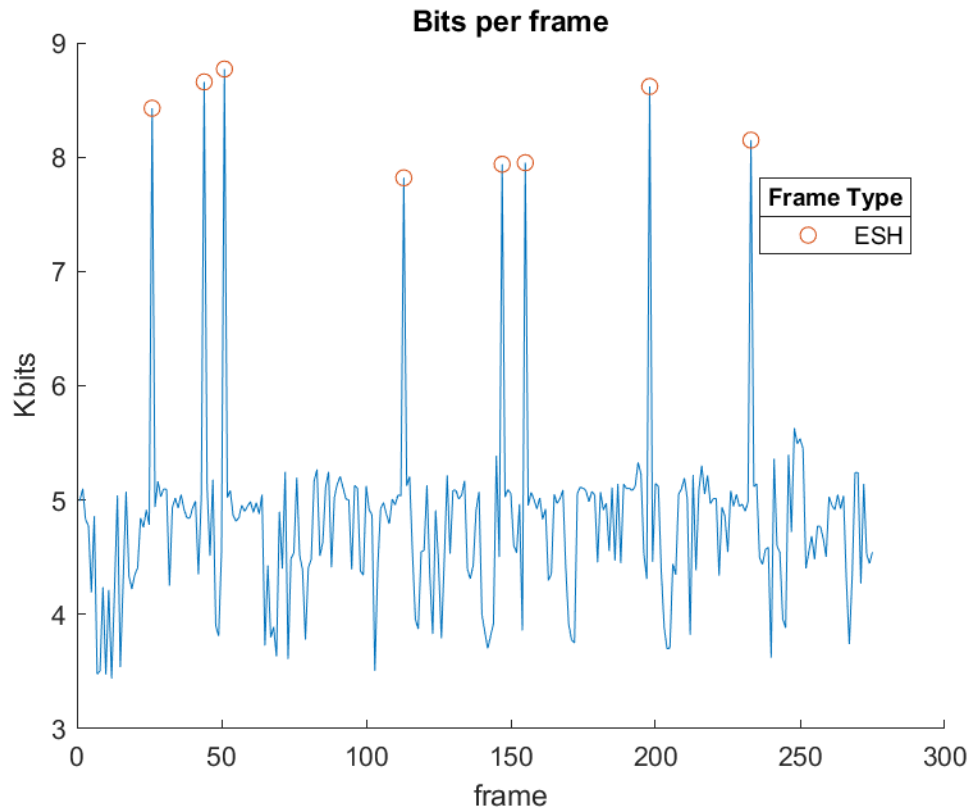
Οι τιμές του αρχικού αρχείου και του θορύβου που προέκυψε στο τελικό αρχείο παρουσιάζονται παρακάτω:



Εκ πρώτης όψεως, το σφάλμα φαίνεται αρκετά μεγάλο. Στην πράξη όμως, στο ακουστικό αποτέλεσμα δεν υπάρχει εμφανής διαφορά. Αυτό οφείλεται στη χρήση του ψυχοακουστικού μοντέλου, που επέτρεψε τον κβαντισμό του μετασχηματισμού, χωρίς απώλειες στην ουσιαστική αξία του σήματος (ακουστικό αποτέλεσμα).

Συνολικά, το αποτέλεσμα της συμπίεσης είναι ικανοποιητικό καθώς απαιτεί μόλις το 15% του αρχικού μεγέθους ενώ δεν υπάρχει ακουστικά αισθητή απώλεια στην ποιότητα του αρχείου.

Ακολουθεί διάγραμμα με τον αριθμό Kbits ανά frame.



Παρατηρήσεις:

- Όπως φαίνεται, τα bits των 'ESH' frames είναι κατά μέσο όρο διπλάσια από των υπόλοιπων τύπων. Πράγμα αναμενόμενο, καθώς τα frames αυτά περιέχουν μεγαλύτερες αλλαγές με αποτέλεσμα να είναι πιο απαιτητική η κωδικοποίησή τους.
- Ο μέσος όρος των bits ανά frame προέκυψε 4.8Kbits/frame.
- Υπάρχει μεγάλη διακύμανση στο bitrate κι αυτό οφείλεται στην κωδικοποίηση σταθερής ποιότητας κι όχι σταθερού μήκους (έλλειψη της βαθμίδας bit reservoir).

Συμβάσεις

- Το 1^ο frame το θεωρούμε τύπου 'OLS' σε όλες τις περιπτώσεις.
- Δείγματα του σήματος που 'περισσεύουν' εκτός frames (δηλαδή δεν μπορούν να συμπληρώσουν επιπλέον frame), που βρίσκονται στο τέλος των ηχητικών σημάτων, αγνοούνται.
- Στον υπολογισμό του SNR δεν χρησιμοποιούνται τα πρώτα και τελευταία 1024 δείγματα καθώς δεν υπάρχει επαρκής επικάλυψη κατά της διαδικασία κωδικοποίησης/αποκωδικοποίησης ώστε να μπορούν να συγκριθούν.
- Στον υπολογισμό του SMR στη συνάρτηση psycho(), στα αρχικά frames (τα πρώτα 2), δίνουμε την τιμή 0 στα προηγούμενα (ανύπαρκτα) frames.

Κώδικας

Ο κώδικας που αναπτύχθηκε βρίσκεται και στο [github](#).

Όλα τα προγράμματα εκτελέστηκαν με MATLAB R2017b σε σύστημα με:

- 8GB ram

-Intel i5-4670 @ 3.4GHz

- Windows 10, 64bit
