



Ευάγγελος Κωστινούδης

Νικόλαος Κατωμέρης

ΑΕΜ: 8708, 8551

📧 [ekostinou@auth.gr](mailto:ekostinou@auth.gr), [ngkatomer@auth.gr](mailto:ngkatomer@auth.gr)



πολυμεσα

Εργασία



9ο Εξάμηνο

τμημα ημμυ απθ

2018-2019

Περιεχόμενα

[1ο επίπεδο 3](#_Toc1144104)

[Λειτουργία 3](#_Toc1144105)

[Ενδεικτικά αποτελέσματα 4](#_Toc1144106)

[2ο επίπεδο 5](#_Toc1144107)

[Λειτουργία 5](#_Toc1144108)

[Ενδεικτικά αποτελέσματα 5](#_Toc1144109)

[3ο επίπεδο 6](#_Toc1144110)

[Λειτουργία 6](#_Toc1144111)

[Ενδεικτικά αποτελέσματα 6](#_Toc1144112)

[Συμβάσεις 9](#_Toc1144113)

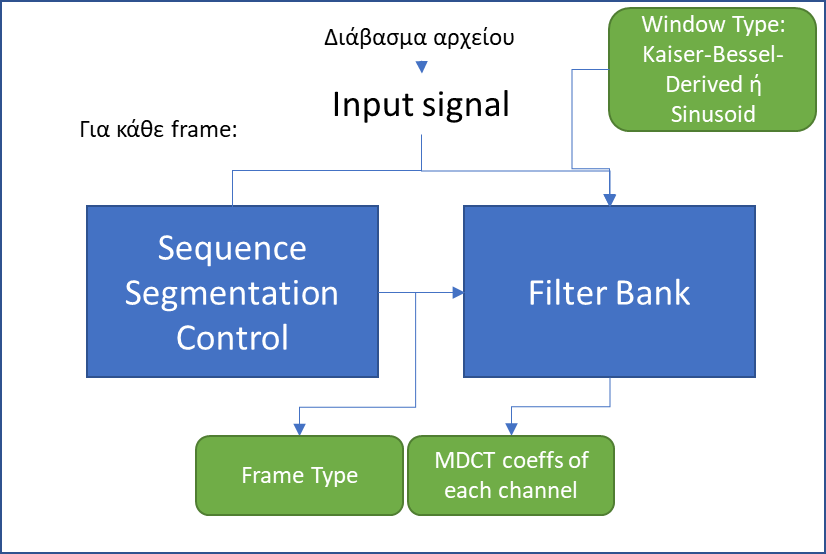
[Κώδικας 9](#_Toc1144114)

# 1ο επίπεδο

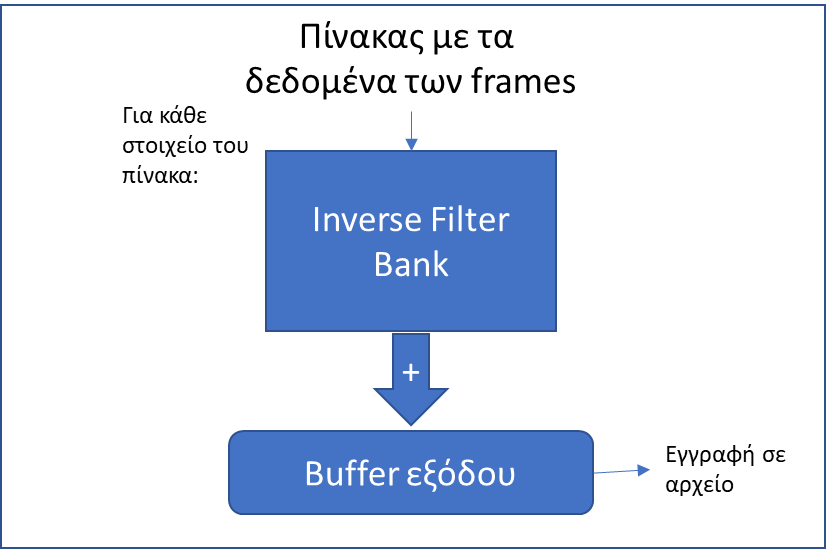
## Λειτουργία

Η συνάρτηση demoAAC1() επιδεικνύει τη λειτουργία του 1ου 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 το παραδοτέο έχει προεπιλεγμένο τον τύπο 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.



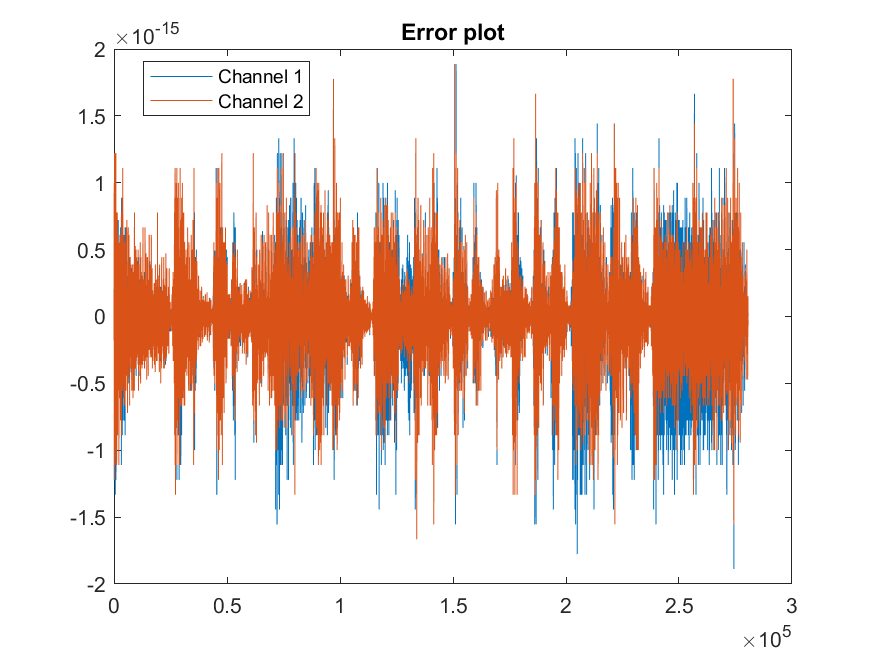
## Ενδεικτικά αποτελέσματα

H εκτέλεση της demoAAC1() με το αρχείο ήχου που δόθηκε (‘LicorDeCalandraca.wav’) πέτυχε:

Εικόνα που περιέχει τοίχος, εσωτερικό

Η περιγραφή δημιουργήθηκε με υψηλή αξιοπιστία

Οι τιμές θορύβου (σφάλματος) παρουσιάζονται στο επόμενο διάγραμμα και είναι της τάξης του . Ο μετασχηματισμός MDCT είναι πλήρως αντιστρέψιμος, οπότε η τάξη αυτή των σφαλμάτων ήταν αναμενόμενη και 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 τους είναι ίδιος καθώς πριν και μετά από τα frames τύπου ‘ESH’ βρίσκονται τα ‘LPS’ και ‘LSS’.

Εικόνα που περιέχει στιγμιότυπο οθόνης

Η περιγραφή δημιουργήθηκε με πολύ υψηλή αξιοπιστία

# 2ο επίπεδο

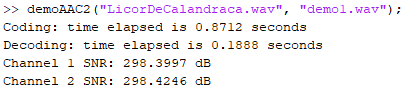
## Λειτουργία

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

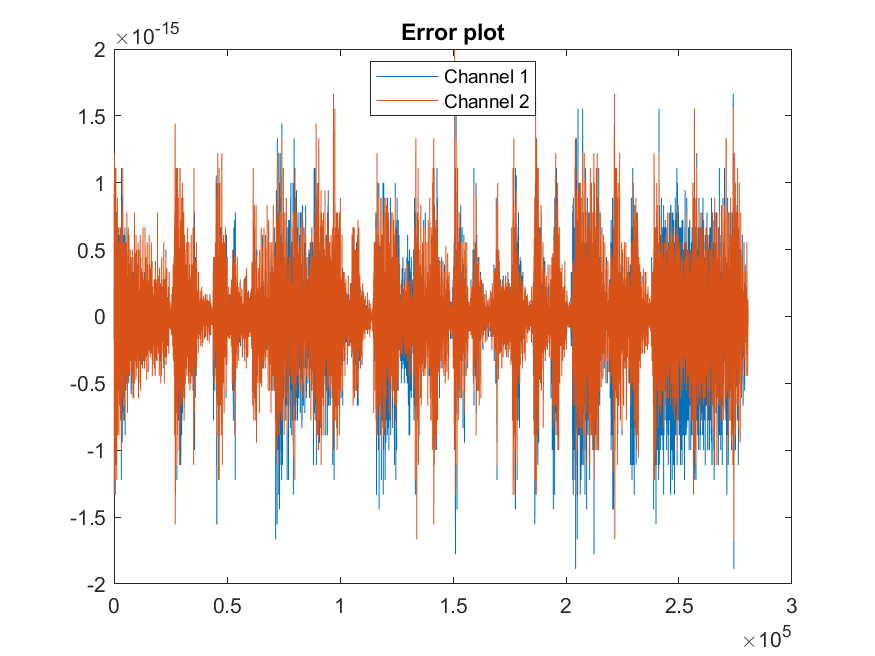
Στο επίπεδο αυτό, καλούνται οι συναρτήσεις AACoder2() και iAACoder2() αντίστοιχα με τις AACoder1() και iAACoder1() του 1ου επιπέδου. Οι συναρτήσεις αυτές είναι όμοιες με τις αντίστοιχές τους με τη διαφορά ότι τα δεδομένα του πίνακα με τα frames περνάνε από μια νέα βαθμίδα για Temporal Noise Shaping (TNS) 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

H εκτέλεση της demoAAC2() με το αρχείο ήχου που δόθηκε (‘LicorDeCalandraca.wav’) πέτυχε:



Οι τιμές θορύβου (σφάλματος) παρουσιάζονται στο επόμενο διάγραμμα και είναι της τάξης του . Κι εδώ η επιπλέον βαθμίδα ‘TNS’ είναι πλήρως αντιστρέψιμη επειδή στην ουσία εφαρμόζεται φίλτρο FIR και στη συνέχεια το ευσταθές αντίστροφό του, οπότε τα σφάλματα είναι αντίστοιχα με αυτά στο 1ο επίπεδο.



# 3ο επίπεδο

## Λειτουργία

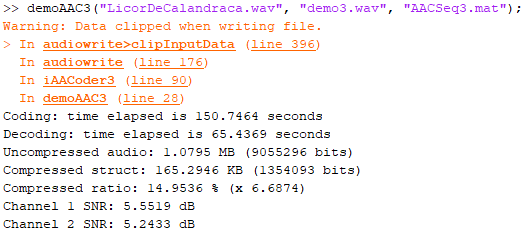
Η συνάρτηση demoAAC3() επιδεικνύει τη λειτουργία του 3ου επιπέδου της εργασίας.

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

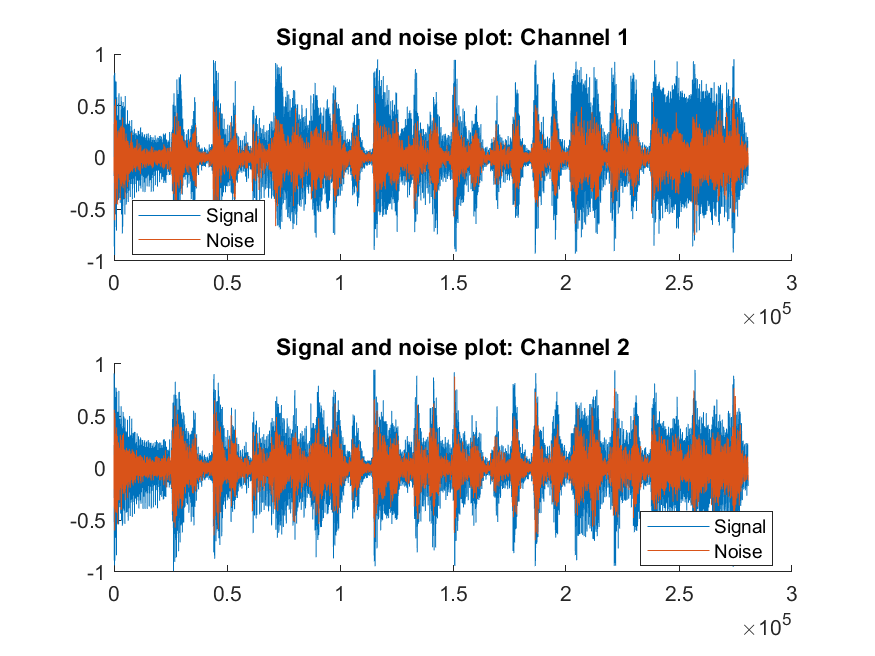
H εκτέλεση αυτής demoAAC3() με το αρχείο ήχου που δόθηκε (‘LicorDeCalandraca.wav’) πέτυχε:



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

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

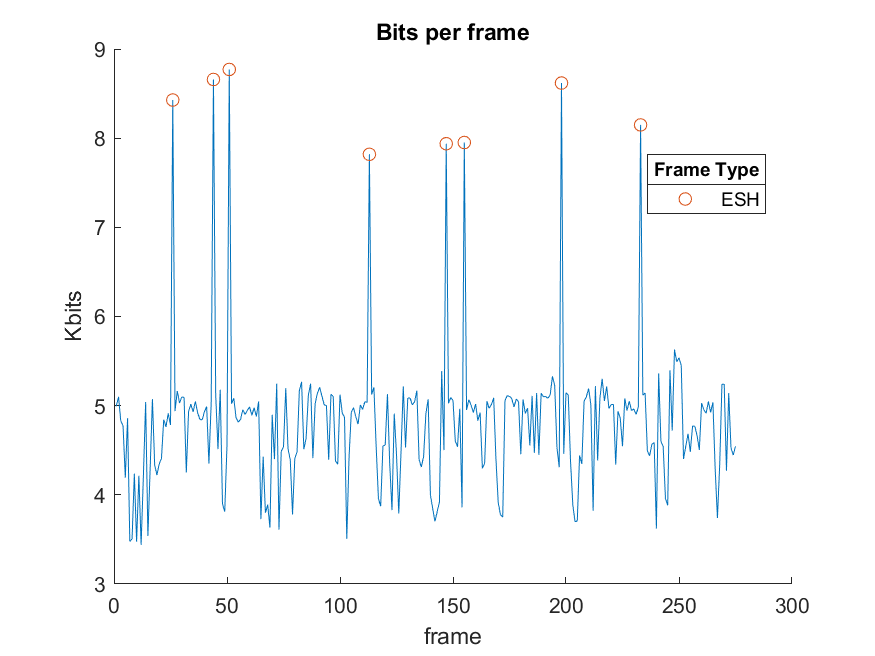
Οι τιμές του αρχικού αρχείου και του θορύβου που προέκυψε στο τελικό αρχείο παρουσιάζονται παρακάτω:



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 1ο frame το θεωρούμε τύπου ‘OLS’ σε όλες τις περιπτώσεις.
* Δείγματα του σήματος που ‘περισσεύουν’ εκτός frames (δηλαδή δεν μπορούν να συμπληρώσουν επιπλέον frame), 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](https://github.com/EKostinoudis/Simplified-ACC).

Όλα τα προγράμματα εκτελέστηκαν με MATLAB R2017b σε σύστημα με:   
- 8GB ram  
-Intel i5-4670 @ 3.4GHz  
- Windows 10, 64bit