



Αριστοτέλειο
Πανεπιστήμιο
Θεσσαλονίκης

Συστήματα Πολυμέσων και Εικονική Πραγματικότητα

Παραδοτέα Εργασίας Επιπέδου 3

(ΑΠΟ)ΚΩΔΙΚΟΠΟΙΗΤΗΣ ΦΩΝΗΣ

ETSI GSM 06.10

Κοκκώνης Γρηγόρης 9581
grkokkonis@ece.auth.gr

Τσουμάνης Στυλιανός 9511
styltsou@ece.auth.gr

08/03/2022

Περιεχόμενα

Περιεχόμενα	1
1. Εισαγωγή	2
2. Μέρος 1 ^ο	2
2.1. Κύριο πρόγραμμα	2
2.2. Short Term coder & decoder	2
2.3. Βοηθητικές συναρτήσεις	3
2.3.1 Preprocessing	3
2.3.2 ComputeRC	3
2.3.3 LARtrans & LARinv	3
2.3.4 LARquant & LARdequant	3
2.3.5 Residual	3
3. Μέρος 2 ^ο	4
3.1. Κύριο πρόγραμμα	4
3.2. Long Term coder & decoder	4
3.3. Βοηθητικές συναρτήσεις	5
3.3.1 LTP	5
3.3.2 LTPquant & LTPdequant	5
3.3.3 LongTerm_Prediction	5
3.3.4 Synthesis	5
4. Μέρος 3 ^ο	5
4.1. Κύριο πρόγραμμα	6
4.2. RPEencode & RPEdecode	6
4.3. params2bin & bin2params	6
4.4. APCM	6

1. Εισαγωγή

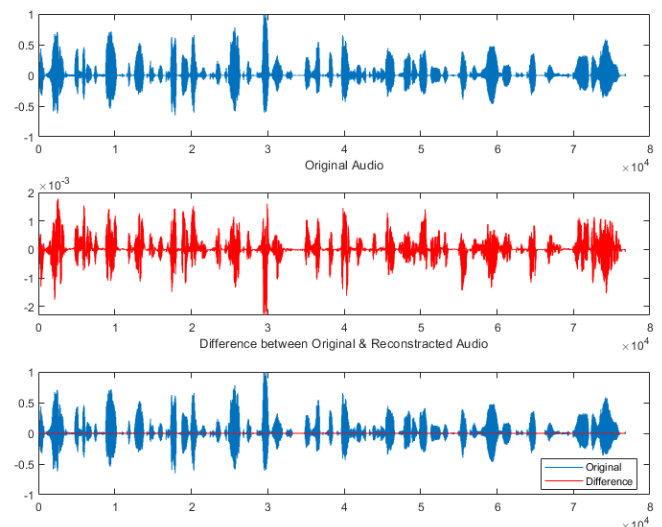
Στην παρούσα εργασία υλοποιείται ένας κωδικοποιητής / αποκωδικοποιητής φωνής με βάση το πρότυπο ETSI GSM 06.10 που χρησιμοποιεί την τεχνολογία Regular Pulse Excitation. Η σχεδίαση του έχει γίνει σε τρία ιεραρχικά μέρη, την Short Term ανάλυση, την Long Term ανάλυση και την συνολική υλοποίηση του (από)κωδικοποιητή. Για την γραφή του προγράμματος χρησιμοποιήθηκε το λογισμικό “Matlab” ενώ τα ηχητικά δείγματα λήφθηκαν από την ιστοσελίδα “[SIGNALogic](#)”.

2. Μέρος 1^ο

Στο πρώτο μέρος της εργασίας υλοποιήθηκε η βραχυπρόθεσμη ανάλυση, η οποία εκτελείται σε επίπεδο frame. Συγκεκριμένα, οι συναρτήσεις **RPE_frame_ST_coder** και **RPE_frame_ST_decoder** χρησιμοποιούνται στο αρχείο *Part1.m* όπου κωδικοποιείται και αποκωδικοποιείται ένα σύνολο δειγμάτων (αντρικής ή γυναικείας) φωνής, τα οποία είναι αποθηκευμένα σε .wav μορφή με ρυθμό δειγματοληψίας 8000 Hz (*male.wav* & *female.wav* αντίστοιχα).

2.1. Κύριο πρόγραμμα

Αρχικά, το πρόγραμμα διαβάζει το αρχείο ήχου, εξάγει τα δείγματα του σήματος και το περικόπτει κατάλληλα με σκοπό να αποθηκευτούν μόνο τα πρώτα 10 δευτερόλεπτα. Στην συνέχεια, μέσω της συνάρτησης **RPE_frame_ST_coder** εξάγονται ανά frame οι κωδικοποιημένες τιμές των συντελεστών LAR και οι τιμές της ακολουθίας που αποτελεί το “σφάλμα πρόβλεψης” του αρχικού σήματος. Έπειτα, χρησιμοποιώντας τις παραπάνω τιμές ανακατασκευάζει το αρχικό σήμα μέσω της συνάρτησης **RPE_frame_ST_decoder**. Το τελικό σήμα αναπαράγεται αυτόματα και αποθηκεύεται ως “coded1.wav”.



2.2. Short Term coder & decoder

Η συνάρτηση **RPE_frame_ST_coder**, αφού προ-επεξεργαστεί τα δείγματα του τρέχοντος frame, υπολογίζει τις τιμές των συντελεστών της συνάρτησης αυτοσυσχέτισης του σήματος. Στην συνέχεια, χρησιμοποιεί τις ισοδύναμες τιμές LAR (Log-Area Ratios) έτσι ώστε, μιμούμενη την διαδικασία αποκωδικοποίησης, να υπολογίσει τους συντελεστές του προβλέπτη. Η συνάρτηση τερματίζει επιστρέφοντας τις κωδικοποιημένες τιμές των LAR καθώς και το “σφάλμα πρόβλεψης”.

Αντιστρόφως, η συνάρτηση **RPE_frame_ST_decoder** υπολογίζει από τις κωδικοποιημένες τιμές LAR τους συντελεστές του προβλέπτη και χρησιμοποιεί την ακολουθία του “σφάλματος πρόβλεψης” για να διεγείρει το short term synthesis φίλτρο. Τέλος, επιστρέφει το ανακατασκευασμένο σήμα αφού το μετα-επεξεργαστεί.

2.3. Βοηθητικές συναρτήσεις

2.3.1 Preprocessing

Η συνάρτηση αυτή περιγράφει την διαδικασία Offset compensation και Pre-emphasis των αρχικών δειγμάτων φωνής.

2.3.2 ComputeRC

Η συνάρτηση αυτή υπολογίζει τους συντελεστές αυτοσυσχέτισης του σήματος. Χρησιμοποιώντας τις τιμές της συνάρτησης αυτοσυσχέτισης υπολογίζονται οι τιμές των συντελεστών πρόβλεψης ως λύση των κανονικών εξισώσεων $\mathbf{Rw}=\mathbf{r}$. Τέλος, μέσω της συνάρτησης poly2rc, η οποία μας δίνεται, υπολογίζονται και επιστρέφονται οι συντελεστές αυτοσυσχέτισης. Επίσης, σε μορφή σχολίων υπάρχει ο αλγόριθμος “Schur Recursion” που χρησιμοποιεί το πρότυπο για τον υπολογισμό των συντελεστών αυτοσυσχέτισης.

2.3.3 LARtrans & LARinv

Οι δύο αυτές συναρτήσεις περιγράφουν τον μετασχηματισμό των συντελεστών LAR σε συντελεστές αυτοσυσχέτισης και αντίστροφα. Συγκεκριμένα χρησιμοποιείται η προσεγγιστική μέθοδος που προτείνει το πρότυπο και όχι ο ακριβής ορισμός του Log-Area Ratio.

2.3.4 LARquant & LARdequant

Οι δύο αυτές συναρτήσεις περιγράφουν τον κβαντισμό (αποκβαντισμό) και την κωδικοποίηση (αποκωδικοποίηση) των τιμών LAR.

2.3.5 Residual

Η συνάρτηση αυτή υπολογίζει το “σφάλμα πρόβλεψης” του σήματος.

$$\begin{aligned}d(n) &= s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^8 a_k \cdot s(n-k) = \\&= s(n) + \sum_{k=1}^8 -a_k \cdot s(n-k) = \alpha \times Shat \\ \text{όπου: } \alpha &= [1 - \alpha_1 \dots - \alpha_8] \text{ και} \\ Shat &= [s(n) \ s(n-1) \dots s(n-8)]^T\end{aligned}$$

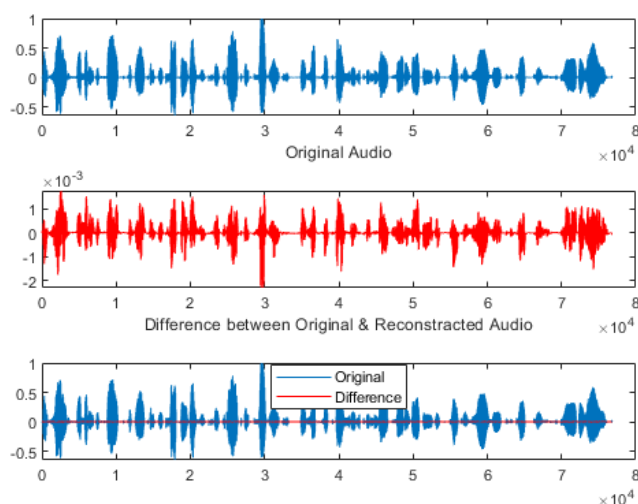
Ο πίνακας Shat κατασκευάζεται τοποθετώντας την ακολουθία τιμών του σήματος σε κάθε γραμμή και μετατοπίζοντας δεξιά (zero filling) την κάθε γραμμή ανάλογα με την τιμή k.

3. Μέρος 2^ο

Στο δεύτερο μέρος της εργασίας υλοποιήθηκε η μακροπρόθεσμη ανάλυση, η οποία εκτελείται σε επίπεδο subframe. Συγκεκριμένα, οι συναρτήσεις **RPE_frame_SLT_coder** και **RPE_frame_SLT_decoder** χρησιμοποιούνται στο αρχείο *Part2.m* όπου κωδικοποιείται και αποκωδικοποιείται ένα σύνολο δειγμάτων (αντρικής ή γυναικείας) φωνής, τα οποία είναι αποθηκευμένα σε .wav μορφή με ρυθμό δειγματοληψίας 8000 Hz (*male.wav* & *female.wav* αντίστοιχα).

3.1. Κύριο πρόγραμμα

Αρχικά, το πρόγραμμα διαβάζει το αρχείο ήχου, εξάγει τα δείγματα του σήματος και το περικόπτει κατάλληλα με σκοπό να αποθηκευτούν μόνο τα πρώτα 10 δευτερόλεπτα. Στην συνέχεια, μέσω της συνάρτησης **RPE_frame_SLT_coder** εξάγονται ανά frame οι απαραίτητοι συντελεστές για την ανακατασκευή του σήματος. Έπειτα, χρησιμοποιούνται οι παραπάνω τιμές για να κατασκευαστεί και πάλι το αρχικό σήμα μέσω της συνάρτησης **RPE_frame_SLT_decoder** και ανανεώνεται η μεταβλητή που περιέχει τις τιμές του προηγούμενου frame. Το τελικό σήμα αναπαράγεται αυτόματα και αποθηκεύεται ως “coded2.wav”.



3.2. Long Term coder & decoder

Η συνάρτηση **RPE_frame_SLT_coder**, εκτελεί αρχικά την βραχυπρόθεσμη ανάλυση που περιγράφηκε στο 1^ο μέρος για το τρέχον frame. Έπειτα, για το κάθε subframe, υπολογίζει και κωδικοποιεί τις παραμέτρους pitch period και τον παράγοντα ενίσχυσης/απόσβεσης, για να σχηματίσει την ακολουθία της Long Term πρόβλεψης. Τέλος, υπολογίζει το σφάλμα πρόβλεψης και ανασυνθέτει την ακολουθία πρόβλεψης χρησιμοποιώντας προηγούμενες τιμές της.

Η συνάρτηση **RPE_frame_SLT_decoder**, εκτελεί την ακριβώς αντίστροφη διαδικασία και αφού πραγματοποιήσει και την Short Term αποκωδικοποίηση επιστρέφει το ανακατασκευασμένο σήμα.

Και στις δύο περιπτώσεις, πριν τερματίσουν οι συναρτήσεις, ανανεώνεται η μεταβλητή όπου αποθηκεύονται οι προηγούμενες τιμές της ακολουθίας πρόβλεψης, έτσι ώστε να συμπεριλαμβάνουν και τις τιμές του τρέχοντος frame.

3.3. Βοηθητικές συναρτήσεις

3.3.1 LTP

Η συνάρτηση αυτή υπολογίζει τις Long Term παραμέτρους, pitch period και τον παράγοντα ενίσχυσης/απόσβεσης, από τις τιμές του σήματος πρόβλεψης του τρέχοντος subframe καθώς και τις ανακατασκευασμένες τιμές του από προηγούμενα subframes.

3.3.2 LTPquant & LTPdequant

Οι δύο αυτές συναρτήσεις περιγράφουν τον κβαντισμό (αποκβαντισμό) και την κωδικοποίηση (αποκωδικοποίηση) των Long Term παραμέτρων.

3.3.3 LongTerm_Prediction

Η συνάρτηση αυτή υπολογίζει το σφάλμα της Long Term πρόβλεψης χρησιμοποιώντας τις τιμές του σήματος πρόβλεψης του τρέχοντος subframe καθώς και τις ανακατασκευασμένες τιμές του από προηγούμενα subframes. Επιπλέον, οι Long Term παράμετροι πριν χρησιμοποιηθούν κωδικοποιούνται και αποκωδικοποιούνται μιμούμενοι έτσι την διαδικασία αποκωδικοποίησης.

3.3.4 Synthesis

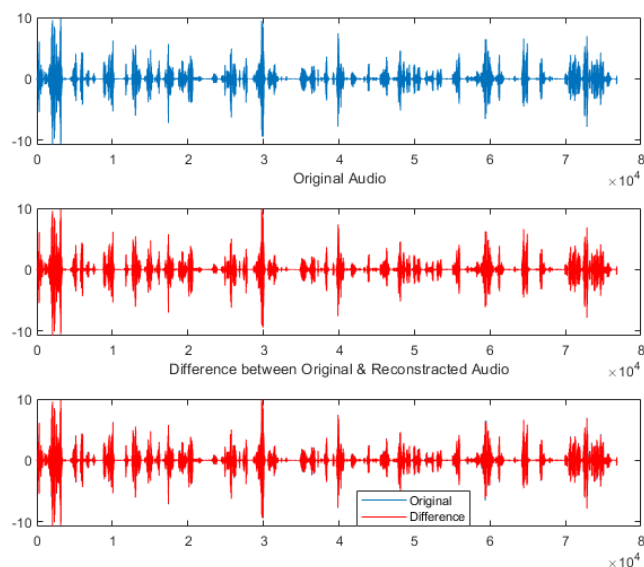
Η συνάρτηση αυτή ανασυνθέτει τις τιμές της ακολουθίας πρόβλεψης του τρέχοντος subframe χρησιμοποιώντας προηγούμενες τιμές της ίδιας ακολουθίας καθώς και του σφάλματος πρόβλεψης του subframe αυτού.

4. Μέρος 3^ο

Στο τρίτο και τελευταίο μέρος της εργασίας, συμπιέζεται περαιτέρω η ακολουθία του σήματος, υποδειγματοληπτώντας τα δείγματα των σφαλμάτων πρόβλεψης. Επιπλέον γίνεται η σύνθεση του δυαδικού block που περιέχει την απαραίτητη πληροφορία για την αποκωδικοποίηση ενός frame. Η παραπάνω διαδικασία περιγράφεται στο αρχείο *Part3.m* όπου καλούνται οι συναρτήσεις του 2^{ου} μέρους καθώς και άλλες βοηθητικές συναρτήσεις που στοχεύουν στην υποδειγματοληψία της ακολουθίας σφαλμάτων πρόβλεψης καθώς και την κωδικοποίησης αυτής.

4.1. Κύριο πρόγραμμα

Ακολουθείται η ίδια διαδικασία με αυτή του 2^{ου} μέρους με, διαφορά την υποδειγματοληψία της ακολουθίας σφαλμάτων πρόβλεψης και την κωδικοποίηση και αποθήκευση της κωδικοποιημένης πληροφορίας σε έναν και μοναδικό δυαδικό πίνακα, όπου κάθε γραμμή αντιστοιχεί σε ένα frame. Το τελικό αποτέλεσμα, έχει σαφώς πολύ θόρυβο, όπως φαίνεται και στο διπλανό διάγραμμα, καθώς υπάρχει κάποιο λάθος στην υλοποίηση της υποδειγματοληψίας ή/και της κωδικοποίησης αυτής. Επιπλέον, το δείγμα φωνής έχει περικοπεί περαιτέρω στα 2 δευτερόλεπτα καθώς το τελικό ηχητικό σήμα έχει πολύ θόρυβο.



4.2. RPEencode & RPEdecode

Οι δύο αυτές συναρτήσεις είναι υπεύθυνες, η πρώτη, για την υποδειγματοληψία της ακολουθίας σφαλμάτων πρόβλεψης και κωδικοποίηση της βέλτιστης, εκ των τεσσάρων, υποακολουθίας και η δεύτερη για την ανακατασκευή της ακολουθίας σφαλμάτων πρόβλεψης.

Κατά την κωδικοποίηση, η βέλτιστη υποακολουθία επιλέγεται με βάση την ισχύ τους, και κανονικοποιείται στο εύρος $[-1,1]$. Αφού κβαντιστούν λογαριθμικά επιστρέφονται για να γίνει η σύνθεση του δυαδικού block.

Κατά την αποκωδικοποίηση πραγματοποιείται ακριβώς η αντίθετη διαδικασία και τελικά επιστρέφεται η ανακατασκευασμένη ακολουθία των σφαλμάτων πρόβλεψης

4.3. params2bin & bin2params

Οι δύο αυτές συναρτήσεις είναι υπεύθυνες για την κωδικοποίηση και αποκωδικοποίηση των παραμέτρων και της υποακολουθίας των σφαλμάτων πρόβλεψης, σε δυαδικές κωδικο-λέξεις. Όλη η πληροφορία που είναι απαραίτητη για την αναπαραγωγή ενός frame του δείγματος φωνής, συμπιέζεται σε ένα block 260 bit, επιτυγχάνοντας έτσι ρυθμό μεταφοράς δεδομένων ίσο με 13kbps.

4.4. APCM

Η συνάρτηση αυτή είναι απλή υλοποίηση ενός λογαριθμικού κβαντιστή και αποκβαντιστή αναλόγως το mode που θα επιλεγθεί (encode & decode αντίστοιχα).