



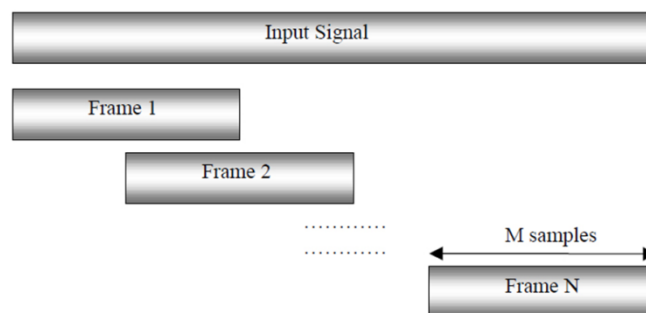
Εργασία 1

(Προθεσμία: Τετάρτη 30 Νοεμβρίου)

1. Επεξεργασία Ήχου Βραχέως Χρόνου

α) Να γραφεί μια ρουτίνα “frame_wind.m” σε περιβάλλον MATLAB, η οποία θα διαιρεί μια κυματομορφή ήχου σε μικρότερα παράθυρα με επικάλυψη, όπως φαίνεται στο ακόλουθο σχήμα. Η ρουτίνα θα παίρνει ως μεταβλητές εισόδου i) τη διακριτή κυματομορφή ήχου x , ii) το μέγεθος του παραθύρου σε πλήθος δειγμάτων $frame$, iii) το ποσοστό του παραθύρου στο οποίο θα γίνεται η επικάλυψη $ovrlp$ (από 0 ως 1, default = 0.5). Η ρουτίνα θα διαιρεί το ήχο σε παράθυρα μήκους $frame$, με επικάλυψη $ovrlp * frame$ δείγματα. Το κάθε παράθυρο θα πολλαπλασιάζεται με μια συνάρτηση παραθύρου, που ορίζεται από τη μεταβλητή $window$. Η μεταβλητή θα μπορεί να παίρνει τις τιμές "Hamming", "Hanning", "Blackman" ή "Bartlett" και τότε η συνάρτηση παραθύρου θα δίνεται από το MATLAB με την αντίστοιχη εντολή. Η έξοδος της ρουτίνας θα είναι ένας πίνακας X , κάθε στήλη του οποίου θα περιέχει τα παράθυρα που έχουμε εξάγει με τη σειρά εξαγωγής.

```
X = frame_wind (x, frame, ovrlp, window);
```



β) Να γραφεί μια ρουτίνα “frame_recon.m” σε περιβάλλον MATLAB, η οποία θα ανακατασκευάζει μια κυματομορφή ήχου από τα μικρότερα παράθυρα με επικάλυψη όπως φαίνεται στο παραπάνω σχήμα. Η ρουτίνα θα παίρνει ως μεταβλητές εισόδου i) τον πίνακα X ii) το ποσοστό του παραθύρου στο οποίο θα γίνεται η επικάλυψη $ovrlp$ (default = 0.5).

```
y = frame_recon (X, ovrlp);
```

γ) Φορτώστε το αρχείο ήχου “guit1.wav” και αναλύστε το με την `frame_wind` με $frame = 256$ δείγματα, $ovrlp = 0.5$ και παράθυρο Bartlett. Επανασυνθέστε τον ήχο με την `frame_recon`. Ελέγξτε τις υλοποιήσεις σας, ακούγοντας τον ήχο x και τον ήχο y με την εντολή `soundsc(x, 16000)`. Σε σωστές υλοποιήσεις δεν πρέπει να ακούτε διαφορά μεταξύ του ήχου x και του ήχου y . (τέλεια ανακατασκευή).

δ) Επαναλάβετε το γ) για τα 4 διαφορετικά παράθυρα και ακούστε τις τρεις ανακατασκευές. Ποιά σας φαίνεται καλύτερη και ποια χειρότερη;

ε) Χρησιμοποιείτε το αρχείο ήχου “guit1.wav” και αναλύστε το με την `frame_wind` με `frame = 256` δείγματα και `ovrlp = 0.5`. Υπολογίστε τον FFT κάθε στήλης (δηλαδή κάθε παραθύρου) χρησιμοποιώντας την εντολή ‘`fft`’ του MATLAB. Απεικονίστε το μέτρο των FFT κάθε frame ως εικόνα, χρησιμοποιώντας την εντολή “`imagesc`”. Θα δείτε ένα σχήμα που είναι συμμετρικό ως προς τον οριζόντιο άξονα. Που οφείλεται αυτή η συμμετρία; Αφαιρέστε τις μισές γραμμές του πίνακα (από 1 ως `frame/2`) και απεικονίστε το νέο πίνακα ως εικόνα. Πως ονομάζεται το διάγραμμα που έχετε απεικονίσει ; Συγκρίνετε τα αποτελέσματα με αυτά που βγάζει η εντολή “`specgram`” του MATLAB.

στ) Ηχογραφήστε ένα δικό σας σύντομο μήνυμα με την εντολή `audiorecorder`, πχ. «Πρώτη Εργασία στην Τεχνολογία Ήχου», μονοφωνικό (mono) και συχνότητα δειγματοληψίας 16 kHz και επαναλάβετε τα βήματα γ), ε). (Μπορείτε επίσης να χρησιμοποιήσετε και κάποιο άλλο πρόγραμμα ηχογράφησης, όπως π.χ. το Audacity, αρκεί να βάλετε τις ίδιες προδιαγραφές).

2. Χωρισμός Φωνούμενου/Μη-φωνούμενου Τμήματος

α) Φορτώστε το αρχείο ήχου “guit1.wav” και αναλύστε το με την `frame_wind` με `frame = 256` δείγματα, `ovrlp = 0.5` και Bartlett window. Για κάθε frame που εξάγει η `frame_wind` υπολογίστε

i) **L₁-Ενέργεια** κάθε frame k, σύμφωνα με τη σχέση

$$E(k) = \sum_{m=1}^{frame} |X(m, k)|$$

ii) **Zero-Crossing Rate** κάθε frame k, σύμφωνα με τη σχέση

$$ZCR(k) = \sum_{m=2}^{frame} (\text{sgn}(X(m-1, k)) - \text{sgn}(X(m, k)))^2$$

β) Αξιολογίστε το κάθε frame αν περιέχει voice ή unvoiced πληροφορία, σύμφωνα με το κριτήριο: «Αν $E(k) < a \cdot \max(E)$ και $ZCR(k) > b \cdot \max(ZCR)$, τότε το frame k είναι unvoiced, αλλιώς το frame k είναι voiced.» για τιμές $a=b=0.4$.

γ) Χρησιμοποιώντας τη “`frame_recon.m`” ανασυγκροτήστε τα voiced frames βαζοντας μηδενικά στα unvoiced frames φτιάχνοντας την κυματομορφή y1. Κατόπιν, ανασυγκροτήστε τα unvoiced frames βαζοντας μηδενικά στα voiced frames φτιάχνοντας την κυματομορφή y2. Ακούστε τις 2 κυματομορφές. Απεικονίστε τις 2 κυματομορφές (με την εντολή `plot.m`). Τί παρατηρείτε ? Ακούστε και απεικονίστε το άθροισμα y1+y2. Τί παρατηρείτε?

δ) Χρησιμοποιείτε το αρχείο που ηχογραφήσατε στην προηγούμενη άσκηση και επαναλάβετε τα α), β), γ).

ε) Δοκιμάστε διάφορες τιμές για τα a και b και δείτε την επίδραση στο αποτέλεσμα του χωρισμού.

3. Εξαγωγή Ρυθμικού Σολφέζ από μουσικό κομμάτι

α) Φορτώστε το αρχείο ήχου “sweetchild.wav” και αναλύστε το με την `frame_wind` με `frame = 256` δείγματα, `ovrlp = 0.5` και `window Bartlett`. Για κάθε `frame` που εξάγει η `frame_wind` υπολογίστε την Ενέργεια $E(k)$ σύμφωνα με τις σχέσεις της Άσκησης 2.

β) Απεικονίστε τη μεταβλητή E . Η μεταβλητή E σας δείχνει αν έχετε ενεργές νότες σε κάθε χρονική στιγμή. Μπορείτε να βγάλετε έναν κανόνα (κριτήριο) που θα σας λέει πότε κάθε νότα είναι ενεργή?

γ) Δημιουργήστε έναν πίνακα Y ιδίων διαστάσεων με τον X της `frame_wind`. Αν έχετε ενεργή νότα στο συγκεκριμένο `frame`, τότε γεμίστε την αντίστοιχη στήλη του X με μονάδες, αλλιώς γεμίστε τη στήλη με μηδενικά.

γ) Χρησιμοποιώντας τη “`frame_recon.m`”, ανασυγκροτήστε την κυματομορφή y από τον πίνακα Y . Ακούστε την κυματομορφή y . Τί παρατηρείτε ?

Ξάνθη, 16/11/2015