



Εργασία 1

(Προθεσμία: Τρίτη 3/5/2022 @ 23.59)

1. Επεξεργασία Ήχου Βραχέως Χρόνου

α) Η ρουτίνα “frame_wind.m” σε περιβάλλον MATLAB διαιρεί μια κυματομορφή ήχου σε μικρότερα παράθυρα με επικάλυψη. Η ρουτίνα 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 έξοδος της ρουτίνας είναι ένας πίνακας X , κάθε στήλη του οποίου περιέχει τα παράθυρα που έχουμε εξάγει με τη σειρά εξαγωγής.

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

β) Η ρουτίνα “frame_recon.m” σε περιβάλλον MATLAB, ανακατασκευάζει μια κυματομορφή ήχου από τα μικρότερα παράθυρα με επικάλυψη από τον πίνακα X . Η ρουτίνα 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 “guit2.wav” και αναλύστε το με την `frame_wind` με $frame = 256$ δείγματα, $ovrlp = 0.5$. 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)`. Κανονικά, δεν πρέπει να ακούτε διαφορά μεταξύ του ήχου x και του ήχου y . (τέλεια ανακατασκευή).

2. Χωρισμός Φωνούμενου/Μη-φωνούμενου Τμήματος

α) Φορτώστε το αρχείο ήχου “guit1.wav” και αναλύστε το με την `frame_wind` με $frame = 256$ δείγματα και $ovrlp = 0.5$. Για κάθε $frame$ που εξάγει η `frame_wind` υπολογίστε τον FFT κάθε $frame$, όπως προηγουμένως φτιάχνοντας τον πίνακα XF . Για κάθε $frame$, υπολογίστε

i) Ενέργεια χαμηλών συχνοτήτων κάθε $frame$ k , σύμφωνα με τη σχέση

$$E_{low}(k) = \sum_{m=1}^{frame/4} |XF(m, k)|^2$$

ii) Ενέργεια υψηλών συχνοτήτων κάθε $frame$ k , σύμφωνα με τη σχέση

$$E_{high}(k) = \sum_{m=\frac{frame}{4}+1}^{frame/2} |XF(m,k)|^2$$

β) Αξιολογείστε το κάθε frame αν περιέχει voice ή unvoiced πληροφορία, σύμφωνα με το κριτήριο: «Αν $E_{high}(k) / E_{low}(k) < a$, τότε το frame k είναι voiced, αλλιώς το frame k είναι unvoiced.» για τιμές $a=1$.

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) 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 α, β και γ (Μπορείτε να χρησιμοποιήσετε κάποιο άλλο πρόγραμμα ηχογράφησης, όπως π.χ. το Audacity). Δοκιμάστε διάφορες τιμές για το a κοντά στη μονάδα και δείτε την επίδραση στο αποτέλεσμα του χωρισμού.

3. Εξαγωγή Ρυθμικού Σολφέζ από μουσικό κομμάτι

α) Φορτώστε το αρχείο ήχου “smoke.wav” και αναλύστε το με την frame_wind με frame = 1024 δείγματα και overlap = 0.5. Για κάθε frame που εξάγει η frame_wind υπολογίστε την Ενέργεια $E(k) = E_{high}(k) + E_{low}(k)$ 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. Χρησιμοποιώντας την εντολή diff, μπορείτε να υπολογίσετε την πρώτη παράγωγο της E. Απεικονίστε την μαζί με την E. Μπορείτε να βγάλετε έναν κανόνα (κριτήριο) που θα σας λέει πότε έχω την αρχή κάθε νότας?

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

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

4. Αυτόματος Ανιχνευτής Ομιλίας (Voice Activity Detector – VAD)

α) Φορτώστε το αρχείο ήχου “guit2.wav” και αναλύστε το με την frame_wind με frame = 256 δείγματα, overlap = 0.5. Για κάθε frame που εξάγει η frame_wind υπολογίστε την Ενέργεια $E(k)$ σύμφωνα με τις σχέσεις της Άσκησης 3.

β) Απεικονίστε τη μεταβλητή E. Η μεταβλητή E σας δείχνει αν κάποιο frame περιέχει ομιλία ή όχι. Μπορείτε να βγάλετε έναν κανόνα (κριτήριο) που θα σας λέει πότε έχετε ενεργή ομιλία?

γ) Δημιουργήστε έναν πίνακα Y ιδίων διαστάσεων με τον X της frame_wind. Αν έχετε ενεργή ομιλία στο συγκεκριμένο frame, τότε αντιγράψτε την αντίστοιχη στήλη του X στον Y.

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

ε) Υπολογίστε το ποσοστό των συνολικών frames που είναι ανενεργά. Σε ένα κινητό τηλέφωνο, το κινητό κλείνει τον πομπό του την ώρα των ανενεργών χρόνων, κάνοντας έτσι οικονομία στη

μπαταρία του. Άρα το ποσοστό που υπολογίσατε προηγουμένως δείχνει περίπου το ποσοστό της ενέργειας που κάνει οικονομία ένα κινητό τηλέφωνο.

Ξάνθη, 4/4/2022