



Εργασία 1

(Προθεσμία: Τρίτη 14 Μαΐου 2024)

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 “guit1.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 “guit2.wav” και αναλύστε το με την `frame_wind` με $frame = 256$ δείγματα, $ovrlp = 0.5$. Για κάθε $frame$ που εξάγει η `frame_wind` υπολογίστε

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

$$E(k) = \sum_{m=1}^{frame} X(m,k)^2$$

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) Αξιολογείστε 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 βάζοντας μηδενικά στα unvoiced frames φτιάχνοντας την κυματομορφή $y1$. Κατόπιν, ανασυγκροτήστε τα unvoiced frames βάζοντας 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) Δοκιμάστε διάφορες τιμές για τα a και b και δείτε την επίδραση στο αποτέλεσμα του χωρισμού.

3. De-esser

Ο De-esser είναι μια εφαρμογή που χρησιμοποιείται στα προγράμματα μουσικής παραγωγής, ώστε να μειωθεί η ένταση των άφωνων φωνημάτων «ες» και «σσ» σε ηχογραφήσεις φωνητικών. Σκεφτείτε πως μπορείτε να εφαρμόσετε τον κώδικα που φτιάξατε στο προηγούμενο βήμα για να υλοποιήσετε ένα προσεγγιστικό σύστημα De-esser. Εφαρμόστε την μετατροπή του κώδικά σας στο αρχείο φωνητικών «vocals_dees.wav». Προσέξτε ότι ο ρυθμός δειγματοληψίας είναι στα 44.1 kHz κι έτσι το frame πρέπει να είναι λίγο μεγαλύτερο. Δοκιμάστε την τιμή $frame=1024$.

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, $ovrlp = 0.5$. 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 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.

5. Ανιχνευτής onset μουσικών νοτών

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

β) Υπολογίστε και απεικονίστε τη μεταβλητή:

$$\delta E(k) = |E(k) - E(k-1)|$$

- γ) Επινοείστε ένα κριτήριο για να εντοπίστε την αρχή (onset) κάθε μουσικής νότας.
- δ) Δημιουργήστε τον πίνακα Z , του οποίου οι στήλες είναι 1 αν αντιστοιχούν σε onset και 0 στην αντίθετη περίπτωση. Χρησιμοποιήστε τη `frame_recon` για να ανακατασκευάσετε την z από το Z και ακούστε το αποτέλεσμα. Το αποτέλεσμα στη μουσική θεωρία λέγεται ρυθμικό σολφέζ.

Ξάνθη, 16/4/2024