



Εργασία 1

(Προθεσμία: -----)

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. Υπολογισμός Ενέργειας ηχητικού σήματος

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 σχηματίζοντας τον πίνακα $X_f(f, t)$. Υπολογίστε την ενέργεια κάθε στήλης με τη βοήθεια της σχέσης:

$$E(t) = \sum_{f=1}^{frame} |X_f(f, t)|^2$$

Απεικονίστε την ενέργεια $E(t)$ και βρείτε τη μέγιστη ενέργεια των παραθύρων.

3. Αποθορυβοποίηση Ηχητικού Σήματος

α) Προσθέστε λευκό Γκαουσιανό θόρυβο στο αρχικό σήμα με τυπική απόκλιση $\sigma = 0.01$ με τη βοήθεια της εντολής `randn`. Ακούστε το νέο σήμα με το θόρυβο. Υπολογίστε τον X_f του ενθόρυβου σήματος, υπολογίστε την ενέργεια κάθε frame και τη μέγιστη τιμή ενέργειας.

Κατόπιν, θα προσπαθήσουμε να αφαιρέσουμε το θόρυβο αυτό με τη βοήθεια διαφόρων τεχνικών:

β) **Τεχνική 1:** Βρείτε ποιες στήλες του X_f έχουν ενέργεια μικρότερη από το 1% της μέγιστης τιμής ενέργειας που έχετε υπολογίσει παραπάνω. Μηδενίστε αυτές τις στήλες του X και ανακατασκευάστε το αποτέλεσμα. Ακούστε το και περιγράψτε τι έχει γίνει κι αν αυτό είναι αρκετό.

4. Αποθορυβοποίηση Ηχητικού Σήματος μέσω φασματικής αφαίρεσης

Η τεχνική της αποθορυβοποίησης μέσω φασματικής αφαίρεσης βασίζεται στο γεγονός ότι ο θόρυβος είναι προσθετικός στο πεδίο του χρόνου και κατ' επέκταση προσθετικός και στο πεδίο της συχνότητας. Έτσι, αν ξέραμε ένα φασματικό προφίλ του θορύβου, θα μπορούσαμε να τον αφαιρέσουμε από το σήμα στο πεδίο της συχνότητας και έτσι να έχουμε αποθορυβοποίηση.

Η εκτίμηση των προφίλ του θορύβου γίνεται από frames που δεν έχουμε πληροφορία από το σήμα και άρα εκεί υπάρχει μόνον ο θόρυβος.

Η διαδικασία της φασματικής αφαίρεσης γίνεται ως εξής

1. Χωρίζουμε το σήμα σε τμήματα (frames). (όπως στο ερώτημα 1)
2. Υπολογίζουμε τον FFT $X_f(f,t)$ κάθε τμήματος και μετά το πλάτος του $|X|$ και τη φάση του $\phi(f,t) = \text{angle}(X_f(f,t))$. Η επεξεργασία θα γίνει μόνο στο πλάτος, η φάση θα ενσωματωθεί αυτούσια μετά την επεξεργασία.
3. Υπολογίζουμε ποια τμήματα (frames) δεν περιέχουν πληροφορία σήματος (Δες Τεχνική 1). Αυτά θα μας δώσουν μια εκτίμηση του θορύβου $N(f)$.
4. Αφαιρούμε από το $|X(f,t)|$ το $N(f)$. Προσέχουμε αν η αφαίρεση δημιουργεί αρνητικούς αριθμούς πλάτους, αυτούς τους μηδενίζουμε.

$$|X_{new}(f,t)| \leftarrow \max(0, |X_f(f,t)| - N(f))$$

5. Ανασυνθέτουμε τον FFT του αποθορυβοποιημένου σήματος από το νέο πλάτος και την παλιά φάση.

$$X_{new}(f,t) = |X_{new}(f,t)| e^{j\phi(f,t)}$$

6. Γυρνάμε κάθε frame στο πεδίο του χρόνου με την εντολή `ifft` και ανασυνθέτουμε την τελική κυματομορφή. (Αν υπάρχει φανταστικό μέρος, τότε αυτό λογικά θα είναι πολύ μικρό και δεν το κρατάμε. Κρατάμε μόνο το πραγματικό μέρος)

Ανάλογα με τις διάφορες πρακτικές που μπορούμε να εκτιμήσουμε το προφίλ του θορύβου $N(f)$, έχουμε διάφορες τεχνικές και διαφορετικά αποτελέσματα.

Τεχνική 2: Το προφίλ του θορύβου $N(f)$ ισούται με το μέτρο του πρώτου frame που έχουμε χαρακτηρίσει ότι δεν περιέχει σημαντική πληροφορία με την Τεχνική 1.

Τεχνική 3: Το προφίλ του θορύβου $N(f)$ ισούται με τη μέση τιμή όλων των frame που έχουμε χαρακτηρίσει ότι δεν περιέχουν σημαντική πληροφορία με την Τεχνική 1.

Τεχνική 4: (Προσαρμοστική) Εδώ το προφίλ του θορύβου $N(f)$ αλλάζει με το χρόνο και προσαρμόζεται στο σήμα. Πιο συγκεκριμένα, η φασματική αφαίρεση ξεκινάει χρησιμοποιώντας ως το $N(f)$ της Τεχνικής 2. Καθώς, επεξεργαζόμαστε όμως frames στο χρόνο, αν συναντήσετε κάποιο frame το οποίο δεν περιέχει σημαντική πληροφορία, τότε ανανεώνετε το $N(f)$ με το συγκεκριμένο frame. Έτσι, η φασματική αφαίρεση προχωράει προσαρμοστικά, ανανεώνοντας το προφίλ του θορύβου μέχρι το τέλος.

α) Φορτώστε το αρχείο `'guit1.wav'`, προσθέστε θόρυβο, σύμφωνα με το 3.α) και εφαρμόστε τις Τεχνικές 2, 3 και 4 για την αποθορυβοποίηση με φασματική αφαίρεση. Ακούστε τα αποτελέσματα και σχολιάστε την επίδοση των διαφόρων τεχνικών (1, 2, 3, 4). Ποια θεωρείτε πιο αποτελεσματική ?

Αν $y(n)$ είναι το σήμα μετά την αποθορυβοποίηση και $x(n)$ το αρχικό, τότε υπολογίστε το Signal-to-Noise Ratio (SNR) για κάθε περίπτωση σύμφωνα με τη σχέση. Ακολουθεί τη δική σας αξιολόγηση των αποτελεσμάτων;

$$SNR_{dB} = 10 \log_{10} \frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1} [x(n) - y(n)]^2}$$

β) Ηχογραφείστε ένα δικό σας αρχείο ήχου, στο οποίο να μιλάτε σε ένα δωμάτιο. Μην προσθέσετε θόρυβο, θεωρούμε ότι ο θόρυβος υπάρχει στην ηχογράφηση σας από τον θόρυβο περιβάλλοντος. Χρησιμοποιείτε τις 4 τεχνικές για την αποθορυβοποίηση της ομιλίας σας. Δουλεύουν εδώ οι τεχνικές ? Η τεχνική που θεωρήσατε πιο αποτελεσματική πριν, εξακολουθεί να δουλεύει το ίδιο καλά και εδώ ?

Ξάνθη, 24/3/2020