

**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**  
**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ**  
**Χειμερινό Εξάμηνο 2016-17**

**Προπαρασκευή 2ης Εργαστηριακής Άσκησης:**  
**Εξαγωγή χαρακτηριστικών από φωνή για χρήση σε εφαρμογή αναγνώρισης**

**ΠΕΡΙΓΡΑΦΗ**

Σκοπός είναι η υλοποίηση ενός συστήματος επεξεργασίας και αναγνώρισης φωνής, με εφαρμογή σε αναγνώριση μεμονωμένων λέξεων. Το πρώτο μέρος αποσκοπεί στην εξαγωγή κατάλληλων ακουστικών χαρακτηριστικών από φωνητικά δεδομένα, χρησιμοποιώντας συναρτήσεις οι οποίες σας δίνονται για τους σκοπούς της εργαστηριακής άσκησης. Τα εν λόγω χαρακτηριστικά είναι στην ουσία ένας αριθμός συντελεστών cepstrum που εξάγονται μετά από ανάλυση των σημάτων με μια ειδικά σχεδιασμένη συστοιχία φίλτρων (filterbank). Η συστοιχία αυτή είναι εμπνευσμένη από ψυχοακουστικές μελέτες.

Πιο συγκεκριμένα, το σύστημα που θα αναπτύξετε αφορά σε αναγνώριση μεμονωμένων ψηφίων (isolated digits) στα Αγγλικά. Τα δεδομένα που θα χρησιμοποιήσετε περιέχουν εκφωνήσεις 9 ψηφίων από 15 διαφορετικούς ομιλητές σε ξεχωριστά .wav αρχεία. Συνολικά θα βρείτε 133 αρχεία, αφού 2 εκφωνήσεις θεωρήθηκαν προβληματικές και δεν έχουν συμπεριληφθεί. Τα ονόματα των αρχείων (π.χ. eight8.wav ) υποδηλώνουν τόσο το ψηφίο που εκφωνείται (π.χ. eight), όσο και τον ομιλητή (οι ομιλητές είναι αριθμημένοι από 1-15). Οι εκφωνήσεις έχουν ηχογραφηθεί με συχνότητα δειγματοληψίας ίση με  $F_s = 16\text{kHz}$  και η διάρκειά τους διαφέρει.

**ΕΚΤΕΛΕΣΗ**

Κατεβάστε τα δεδομένα από το παρακάτω link:

[http://cvsp.cs.ntua.gr/courses/patrec/labs\\_material2014/digits2016.zip](http://cvsp.cs.ntua.gr/courses/patrec/labs_material2014/digits2016.zip)

Τα δεδομένα είναι προστατευμένα. Χρησιμοποιήστε ως password τον κωδικό που σας έχει δοθεί.

Χρησιμοποιώντας Matlab εκτελέστε τα παρακάτω βήματα:

**Βήμα 1**

Εισάγετε τα δεδομένα στο περιβάλλον Matlab και από κάθε ένα .wav αρχείο εξάγετε ένα σήμα φωνής  $s_0(n)$  .

**Βήμα 2**

Κάθε σήμα  $s_0(n)$  πρέπει να περάσει από ένα σύστημα προέμφασης, δίνοντας στην έξοδο σήμα  $s_p(n)$  . Το σύστημα προέμφασης έχει την εξής συνάρτηση μεταφοράς:

$$H_{preemph}(z) = 1 - \tilde{a} z^{-1}$$

Η παράμετρος είναι  $\tilde{a} = 0.97$  .

**Βήμα 3**

Το κάθε σήμα  $s_p(n)$  χωρίζεται σε πλαίσια  $s_i(n) = s_p(n)w(n(i-1)N)$ ,  $i=1 \dots N_f$  , όπου  $N_f$  το πλήθος των πλαισίων. Τα πλαίσια είναι διάρκειας ίσης με  $T(ms)$  με  $T_{overlap}$  επικάλυψη. Προκειμένου να ελαχιστοποιηθούν οι ασυνέχειες στα άκρα των πλαισίων, σε κάθε πλαίσιο εφαρμόζεται παραθύρωση Hamming. Το παράθυρο Hamming στο χρόνο έχει μορφή:

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1$$

#### Βήμα 4

Ανάλυση σε συστοιχία φίλτρων σε κλίμακα Mel. Η συστοιχία φίλτρων που θα χρησιμοποιηθεί αποτελείται από έναν αριθμό  $Q$  τριγωνικών φίλτρων  $H^j, j=1, \dots, Q$ , των οποίων οι κεντρικές συχνότητες  $f_c^j$  είναι ισοκατανεμημένες με βάση την κλίμακα mel:

$$f_c^j = 2595 \cdot \log\left(1 + \frac{f^j}{700}\right), j=1 \dots Q$$

όπου  $f^j$  είναι οι κεντρικές συχνότητες στη γραμμική κλίμακα. Θεωρείστε επιπλέον ότι  $H^j(f_c^j) = 1$ .

Το εύρος ζώνης  $b^j$  του κάθε φίλτρου προσδιορίζεται έτσι ώστε οι συχνότητες αποκοπής του να ταυτίζονται με τις κεντρικές συχνότητες των δυο γειτονικών του φίλτρων στην κλίμακα mel.

#### Βήμα 5

Υπολογίστε την ενέργεια  $E_i(j)$  απόκρισης κάθε (καναλιού) φίλτρου  $j$  της συστοιχίας με είσοδο το σήμα  $s_i(n)$ . Απεικονίστε γραφικά τους συντελεστές  $E_i(j)$  για δυο διαφορετικά πλαίσια. Στην ίδια γραφική παράσταση (ξεχωριστά για κάθε πλαίσιο).

#### Βήμα 6

Στη συνέχεια εξάγετε τους συντελεστές  $G_i(j)$ , όπου:

$$G_i(j) = \log(E_i(j)), j=1, \dots, Q$$

#### Βήμα 7

Εφαρμόστε τον μετασχηματισμό συνημιτόνου (DCT) στους συντελεστές  $G_i(j)$ , οπότε και θα προκύψουν οι συντελεστές  $C_i(n)$ :

$$C_i(n) = \sum_{j=1}^Q G_i(j) \cos\left(n\left(j - \frac{1}{2}\right)\frac{\pi}{Q}\right), n=0 \dots N_c-1, N_c < Q$$

Επιλέξτε ένα ψηφίο  $k_1$ . Απεικονίστε γραφικά τα ιστογράμματα των συντελεστών  $C_i(n_1)$ ,  $C_i(n_2)$  για το σύνολο των πλαισίων που αντιστοιχούν στις εκφωνήσεις του ψηφίου αυτού. Επαναλάβετε για ένα ψηφίο  $k_2$ .

#### Βήμα 8

Ανακατασκευάστε την έξοδο της συστοιχίας φίλτρων από τους συντελεστές  $C_i(n)$  για τα πλαίσια που επιλέξατε στο Βήμα 5. Απεικονίστε γραφικά την ανακατασκευασμένη έξοδο,  $\hat{E}_i(j)$ . Απεικονίστε το φάσμα ισχύος του σήματος  $s_i(n)$  στην ίδια γραφική παράσταση.

#### Βήμα 9

Για κάθε εκφώνηση του ψηφίου  $k_1$  υπολογίστε τις μέσες τιμές κατά μήκος όλων των πλαισίων των συντελεστών  $C_i(n_1)$ ,  $C_i(n_2)$ . Επαναλάβετε για κάθε εκφώνηση και για κάθε ψηφίο. Απεικονίστε τα 133 ζεύγη τιμών που βρήκατε στο επίπεδο. Χρησιμοποιήστε διαφορετικό σύμβολο για κάθε ψηφίο (τάξη). Στο ίδιο γράφημα απεικονίστε και τα 9 ζεύγη μέσων τιμών, ένα ζεύγος για κάθε ψηφίο. Σχολιάστε ως προς τη δυνατότητα διαχωρισμού μεταξύ των ψηφίων.

## ΠΑΡΑΔΟΤΕΑ

(1) Σύντομη αναφορά (σε pdf) που θα περιγράφει τη διαδικασία που ακολουθήθηκε σε κάθε βήμα, καθώς και τα σχετικά αποτελέσματα.

(2) Κώδικας Matlab (συνοδευόμενος από σύντομα σχόλια).

Συγκεντρώστε τα (1) και (2) σε ένα .zip αρχείο το οποίο πρέπει να αποσταλεί μέσω του mycourses.ntua.gr πριν από τη διεξαγωγή του εργαστηρίου.

### ΣΗΜΕΙΩΣΕΙΣ

1. Ενδεικτικά σας δίνονται οι παρακάτω τιμές για τις διάφορες παραμέτρους που εμφανίζονται στην εκφώνηση:
  - Η διάρκεια κάθε πλαισίου είναι  $T=25\text{msec}$  και η επικάλυψη μεταξύ των πλαισίων είναι  $T_{\text{overlap}}=10\text{msec}$
  - Ο αριθμός των φίλτρων στη συστοιχία είναι  $Q=24$
  - Ο αριθμός των ζητούμενων χαρακτηριστικών ανά πλαίσιο είναι  $N_c=13$
2. Η επιλογή των δύο ψηφίων  $k_1$  ,  $k_2$  γίνεται με βάση τον αριθμό μητρώου του κάθε φοιτητή-τρια/τή.  $k_1$  είναι το τελευταίο ψηφίο ενώ  $k_2$  το προτελευταίο ψηφίο του αριθμού μητρώου. Αν κάποιο από τα δύο ψηφία είναι το 0 τότε αντικαθίσταται από το αμέσως προηγούμενο ή επόμενο του άλλου ψηφίου, το οποίο δεν είναι μηδενικό.
3. Η επιλογή των συντελεστών cepstrum  $C_i(n_1)$  ,  $C_i(n_2)$  για κάθε φοιτητή με 4 τελευταία ψηφία αριθμού μητρώου  $d_1 d_2 d_3 d_4$  προκύπτει ως  $n_1=(d_1 d_3) \text{ modulo } N_c$  και  $n_2=(d_2 d_4) \text{ modulo } N_c$  . Εάν  $n_1=n_2$  επιλέξτε  $n_2 \leftarrow n_2+1 (\text{modulo } N_c)$  .