Παν. Θεσσαλίας, ΤΜΗΥΤΔ ΕΠΕΞΕΡΓΑΣΙΑ ΦΩΝΗΣ & ΗΧΟΥ Ακαδ. Έτος 2012-13
Ημερομηνία: 12-13-2012 **Project Matlab – Ομάδα 5** Παραδοτέο: 07-02-2013

Η εργασία θα εκτελεστεί σε ομάδες 3-4 ατόμων. Η παράδοσή της θα γίνει ως γραπτή αναφορά και θα συνοδεύεται από CD το οποίο θα περιλαμβάνει τα ηχογραφημένα δείγματα και τον κώδικα Matlab. Η παράδοσή της θα συνοδευτεί από προφορική της εξέταση.

ΣΤΟΧΟΙ ΕΡΓΑΣΙΑΣ

Σκοπός της εργασίας αυτής είναι η εξοικείωση με τα βασικά στάδια ενός συστήματος επεξεργασίας και αναγνώρισης σήματος φωνής. Η εργασία περιλαμβάνει τα ακόλουθα μέρη και βήματα.

ΜΕΡΟΣ Α. Βασικοί Αλγόριθμοι Εκτίμησης Παραμέτρων Φωνής

Α-1. **Ηχογράφηση:** Ηχογραφήσετε το ονοματεπώνυμο ενός από τα άτομα της ομάδας σας με συχνότητα δειγματοληψίας περί τα 16–22 kHz.

Α-2. **Φασματόγραμμα Σήματος Φωνής:** Σχεδιάστε το φασματόγραμμα (spectrogram) του σήματος φωνής που ηχογραφήσατε, χρησιμοποιώντας παράθυρο Hamming μήκους περί τα 10 msec και παράθυρο Hamming μήκους περί τα 100 msec, με μετατόπιση (και στις 2 περιπτώσεις) περί τα 5 msec. Σχολιάστε τις διαφορές στα δύο φασματογράμματα.

Α-3. **Εκτίμηση Έμφωνων / Άφωνων Τμημάτων Σήματος:** Σχεδιάστε έναν αλγόριθμο που διαχωρίζει τα τμήματα του σήματος που περιέχουν έμφωνους ήχους (voiced) από αυτά που περιέχουν άφωνους ήχους (unvoiced), όπως και σιωπή (silence), που πιθανώς να υπάρχει στην ηχογράφηση σε διάφορα σημεία της, π.χ. ίσως μεταξύ του μικρού ονόματος και επιθέτου, όπως και στην αρχή και στο τέλος της. Χρησιμοποιήστε για τον σκοπό αυτόν την ενέργεια (energy) και ρυθμό διαβάσεων μηδενικής τιμής (zero-crossing rate), φυσικά υπολογισμένα σε παράθυρα του σήματος (κατάλληλου μήκους και επικάλυψης). Σχεδιάστε τις τιμές των δύο αυτών μεγεθών σε ένα παράλληλο σχήμα με αυτό της κυματομορφής του σήματος φωνής. Σχεδιάστε επίσης την έξοδο (απόφαση) του αλγορίθμου σας σε τμήματα (voiced, unvoiced, silence).

Α-4. **Εκτίμηση Θεμελιώδους Συχνότητας Διέγερσης:** Σχεδιάστε έναν αλγόριθμο που να βρίσκει την θεμελιώδη συχνότητα διέγερσης (pitch) έμφωνων ήχων, και εφαρμόστε τον στα έμφωνα τμήματα του ηχογραφημένου σήματος, όπως αυτά έχουν ανιχνευτεί από το προηγούμενο βήμα (Α-3). Σχεδιάστε την συχνότητα σε παράλληλο σχήμα με την ηχογραφημένη κυματομορφή, θεωρώντας ότι λαμβάνει μηδενική τιμή στα μη έμφωνα τμήματα.

Α-5. **Γραμμική Πρόβλεψη Σήματος:** Απομονώστε τμήμα διάρκειας 30 msec που αντιστοιχεί σε κάποιο έμφωνο και σε κάποιο άφωνο τμήμα του ηχογραφημένου σήματος. Υπολογίστε το διάνυσμα χαρακτηριστικών LPC για τα δύο κομμάτια για τάξη φίλτρου $p = 8, 12, 16$. Υπολογίστε το λάθος εκτίμησης κάθε φορά, και σχεδιάστε στο ίδιο διάγραμμα το διακριτό μετασχηματισμό Fourier (DFT) των τμημάτων του σήματος που επιλέξατε, όπως και το μέτρο της all – pole συνάρτησης μεταφοράς με βάση το μοντέλο LPC.

ΜΕΡΟΣ Β. Απλό Σύστημα Ταξινόμησης Φωνημάτων

Β-1. **Ηχογράφηση:** Ηχογραφήστε χρησιμοποιώντας συχνότητα δειγματοληψίας 22 kHz αρκετές ακολουθίες βασικών ήχων φωνηέντων, ώστε να έχετε 15 απομονωμένες εμφανίσεις για καθένα από τα 3 φωνήματα /a/, /e/, /i/. Εντοπίστε τα χρονικά όρια των φωνημάτων αυτών (αρχή και τέλος) είτε χειρονακτικά είτε με κάποιον αυτοματοποιημένο τρόπο [π.χ. με την μέθοδο που αναπτύξατε στο τμήμα Α-3] και μετέπειτα οπτική επαλήθευση των ορίων. Απομονώστε τα τμήματα του σήματος μεταξύ των ορίων αυτών και προχωρήστε με αποθήκευσή τους σε καινούργια αρχεία, ένα για κάθε ηχογραφημένο φώνημα. Έτσι θα έχετε 45 αρχεία. Προτείνεται όλα τα αρχεία να αντιστοιχούν στην φωνή ενός ομιλητή. Τα δεδομένα να ηχογραφηθούν σε «καθαρό» ηχητικό περιβάλλον.

Β-2. **Εξαγωγή Χαρακτηριστικών Σήματος:** Για κάθε ένα από τα παραπάνω αρχεία, υλοποιείστε την ανάλυση LPC ώστε να εξαχθούν ακολουθίες διανυσμάτων χαρακτηριστικών του σήματος. Αναλυτικότερα, προτείνονται οι ακόλουθες παράμετροι:

- Φίλτρο προέμφασης με συντελεστή 0.95.
- Παράθυρο Hamming με μήκος περίπου 20 ή 25 msec.
- Χρονική ολίσθηση μεταξύ διαδοχικών παραθύρων περί τα 10 msec (οπότε υπάρχει μερική επικάλυψη μεταξύ διαδοχικών παραθύρων).
- Τάξη φίλτρου $p = 12$.
- Για κάθε ένα από τα αρχεία ήχου, η ανάλυση θα πραγματοποιηθεί για όσα παράθυρα χωρούν στη διάρκεια του φωνήματος. Κάθε παράθυρο θα παράγει ένα διάνυσμα χαρακτηριστικών LPC, οπότε για κάθε αρχείο αντιστοιχεί μία ακολουθία τέτοιων διανυσμάτων.

Β-3. **Εκπαίδευση Στατιστικών Μοντέλων Φωνημάτων:** Αφήστε 5 αρχεία κάθε φωνήματος για το σύνολο δοκιμής (test set) και χρησιμοποιείστε τα υπόλοιπα 10 για το σύνολο εκπαίδευσης (training set). Εκπαιδεύσετε μία πολυδιάστατη κατανομή για κάθε φώνημα χρησιμοποιώντας το σύνολο των ηχογραφημένων αρχείων του συγκεκριμένου φωνήματος που βρίσκονται στο training set (δηλαδή το σύνολο των ακολουθιών διανυσμάτων χαρακτηριστικών του συγκεκριμένου set). Εκτιμήστε κατά συνέπεια τις παραμέτρους της κατανομής:

$$Pr(\mathbf{x}|c) = \frac{1}{(2\pi)^{p/2}(\det \mathbf{\Sigma}_c)^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_c)^T \mathbf{\Sigma}_c{}^{-1}(\mathbf{x}-\mathbf{m}_c)\}$$

όπου c είναι η κλάση (φώνημα), $\mathbf{x}$ είναι το διάνυσμα χαρακτηριστικών με διάσταση p, και τέλος $\mathbf{m}_c$ και $\mathbf{\Sigma}_c$ είναι οι παράμετροι προς εκτίμηση (διάνυσμα μέσης τιμής και πίνακας συνδιασποράς). Εκπαιδεύστε μία κατανομή για κάθε φώνημα (συνολικά 3 κατανομές).

Β-4. **Ταξινόμηση Δειγμάτων Δοκιμής:** Βρείτε την ακρίβεια ταξινόμησης των φωνημάτων του συνόλου δοκιμής (% σωστά), με βάση την τριάδα μοντέλων φωνημάτων (χρησιμοποιείστε την μέγιστη πιθανοφάνεια). Δοκιμάστε αν υπάρχει διαφοροποίηση αν χρησιμοποιηθούν διαγώνιοι πίνακες $\mathbf{\Sigma}_c$.

Β-5. **Ανθεκτικότητα σε Θόρυβο:** Επαναλάβατε το παραπάνω πείραμα (βήμα Β-4 μόνο) με θορυβώδη δεδομένα δοκιμής. Αυτά μπορούν να δημιουργηθούν προσθέτοντας τυχαίο θόρυβο (με αρκετό πλάτος!) στα ήδη ηχογραφημένα δείγματα δοκιμής. Το βήμα Β-2 πρέπει φυσικά να επαναληφθεί για την εξαγωγή των διανυσμάτων δοκιμής.