



Εργασία 2

(Προθεσμία: 17-5-2022)

Voice Morphing

α) Εξαγωγή Συντελεστών LPC

Φορτώστε το αρχείο ήχου “voice.wav” και αναλύστε το με την `frame_wind` με `frame = 256` δείγματα και `ovrlp = 0.25`. Για κάθε frame t , χρησιμοποιείτε τη ρουτίνα “`lpc_new.m`” που υπάρχει στο `eclass` για να βρείτε $p = 20$ συντελεστές γραμμικής πρόβλεψης $a_{i,t}$ και το αντίστοιχο κέρδος G_t που περιγράφουν το κάθε παράθυρο.

β) Εξαγωγή Σφάλματος Γραμμικής Πρόβλεψης

Υπολογίστε το λάθος $e[n,t]$ του γραμμικού προβλεπτή για κάθε frame t του αρχείου “voice.wav”, σύμφωνα με τη σχέση:

$$e[n,t] = (x[n,t] - \sum_{i=1}^p a_{i,t} x[n-i,t]) = x[n,t] * (1 - \sum_{i=1}^p a_i \delta[n-i])$$

γ) Τέλεια ανακατασκευή

Φιλτράρετε κάθε frame σφάλματος γραμμικής πρόβλεψης e με το φίλτρο $G/A_t(z)$ και φτιάξτε έτσι τα ανακατασκευασμένα frames στον πίνακα Y . Χρησιμοποιήστε την `frame_recon` για να ανακατασκευάσετε το τελικό αρχείο ήχου και ακούστε το. **Πως είναι ακουστικά ?** Υπολογίστε το SNR σε dB του ανακατασκευασμένου αρχείου ήχου σε σχέση με το αρχικό αρχείο ήχου. (Φροντίστε να έχουν το ίδιο μήκος για τη σύγκριση, αποκόπτοντας τα εξτρά δείγματα).

δ) Voice Morphing μεταξύ ίδιου φύλου

Φορτώστε τα αρχεία ήχου “`umale1.wav`” και “`umale2.wav`”. Επαναλάβετε τη διαδικασία του βήματος α) για να βγάλετε 20 συντελεστές γραμμικής πρόβλεψης a και το αντίστοιχο κέρδος G για frame από τα 2 αρχεία ήχου. Επαναλάβετε τη διαδικασία του βήματος β) για να βγάλετε το σφάλμα της γραμμικής πρόβλεψης για κάθε frame των αρχείων ήχου. Έτσι έχετε ένα σετ συντελεστών a_1, G_1, e_1 για το πρώτο αρχείο ήχου και ένα δεύτερο σετ συντελεστών a_2, G_2, e_2 για το δεύτερο αρχείο ήχου.

Χρησιμοποιήστε τη λογική του βήματος γ) και ανακατασκευάστε 2 αρχεία y_1 και y_2 με την εξής λογική: i) για το y_1 φιλτράρετε το σφάλμα e_1 με το φίλτρο που προκύπτει από τους

συντελεστές a_2 , G_2 , ii) για το y_2 φιλτράρετε το σφάλμα e_2 με το φίλτρο που προκύπτει από τους συντελεστές a_1 , G_1 . Μπορείτε να αναγνωρίσετε την ταυτότητα των ομιλητών στα y_1 και y_2 ;

Επαναλάβετε την ίδια διαδικασία για τα αρχεία ήχου “ufemale1.wav” και “ufemale2.wav”.

Ακούγοντας τα 2 παραδείγματα (στους άνδρες και στις γυναίκες), μπορείτε να βρείτε κάποιους λόγους για τους οποίους δεν έχει γίνει καλά η μεταφορά ταυτότητας ομιλητή ; (μπορείτε και να κοιτάξετε τις αρχικές κυματομορφές και στο πεδίο του χρόνου)

ε) *Voice Morphing* σε διαφορετικά φύλα

Επαναλάβετε την διαδικασία του βήματος δ) για τα αρχεία ήχου “umale1.wav” και “ufemale1.wav”. Ακούστε τα παραχθέντα ήχου και σχολιάστε τα αποτελέσματα. Μπορείτε να χρησιμοποιήσετε και τα αρχεία “umale2.wav” και “ufemale2.wav” για να εξάγετε τα αποτελέσματα σας. Ποια μπορεί να είναι η βασική αιτία που εδώ λειτουργεί χειρότερα ανάμεσα σε διαφορετικά φύλα ;

στ) Αναγνώριση ταυτότητας ομιλητή

Φορτώστε τα αρχεία ήχου “umale1.wav” και “umale2.wav”. Επαναλάβετε τη διαδικασία του βήματος α) για να βγάλετε 30 συντελεστές γραμμικής πρόβλεψης a και το αντίστοιχο κέρδος G για frame από τα 2 αρχεία ήχου, αλλά με $ovrlp = 0.5$.

Τα συγκεκριμένα αρχεία βγάζουν 385 διανύσματα με συντελεστές γραμμικής πρόβλεψης a για κάθε ομιλητή. Χρησιμοποιήστε τα 300 πρώτα διανύσματα και υπολογίστε το μέσο όρο τους για κάθε ομιλητή. Έτσι, θα έχουμε 2 διανύσματα: a_{1m} για τον πρώτο ομιλητή και το a_{2m} για το δεύτερο ομιλητή. Αυτά τα 2 διανύσματα θα αποτελέσουν την ταυτότητα του κάθε ομιλητή.

Θα χρησιμοποιήσουμε τα υπόλοιπα 85 διανύσματα από κάθε ομιλητή για να επιβεβαιώσουμε την ταυτότητα του κάθε ομιλητή. Ξεκινάμε με τα 85 διανύσματα του ομιλητή 1. Υπολογίζουμε την ευκλείδεια απόσταση του κάθε διανύσματος από τα a_{1m} και a_{2m} . Αν η απόσταση είναι μικρότερη από το a_{1m} θεωρούμε ότι έχει αναγνωρίσει τον ομιλητή 1, αλλιώς τον ομιλητή 2. Η διαδικασία επαναλαμβάνεται για κάθε διάνυσμα του Ομιλητή 1. Στο τέλος μετράμε το πλήθος των διανυσμάτων που έχει αναγνωριστεί ο Ομιλητής 1 και αυτών που έχει αναγνωριστεί ο Ομιλητής 2. Ο ομιλητής με το μεγαλύτερο πλήθος διανυσμάτων είναι αυτός που έχει αναγνωριστεί από το σύστημα. Το σύστημα μας είχε επιτυχία ?

Επαναλάβετε το ίδιο με τα 85 διανύσματα του Ομιλητή 2. Τι βλέπουμε εκεί ?

Επαναλάβετε το ίδιο πείραμα για τα αρχεία ήχου “ufemale1.wav” και “ufemale2.wav”.

Μπορείτε να φτιάξετε ένα σύστημα που θα αναγνωρίζει τον 1^ο ομιλητή (umale1) από τους άλλους 3 ;