Ανέστης Δαλγκίτσης
Γεώργιος-Θεόδωρος Καλαμπόκης
Χρήστος Παλαμιώτης
January 17, 2015

Σύστημα φωνητικής αναγνώρισης στη Matlab

Αναγνώριση μεμονωμένων λέξεων

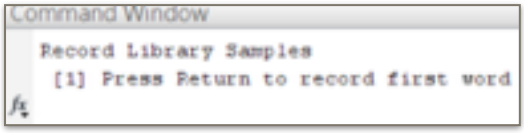
Εισαγωγή

Το πρόγραμμα αυτό αποτελεί μια μικρή προσομοίωση ενός πραγματικού προβλήματος, αυτού της φωνητικής αναγνώρισης λέξεων. Συγκεκριμένα, υλοποιείται η δημιουργία μιας μικρής βιβλιοθήκης τεσσάρων ηχογραφημένων λέξεων, καθώς και η αναγνώρισή τους μέσω μιας πέμπτης ηχογράφησης.

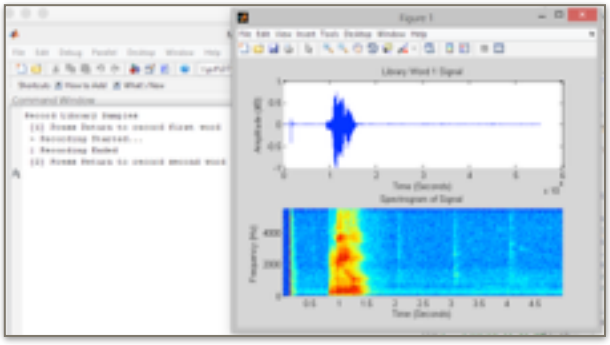
Οδηγίες χρήσης

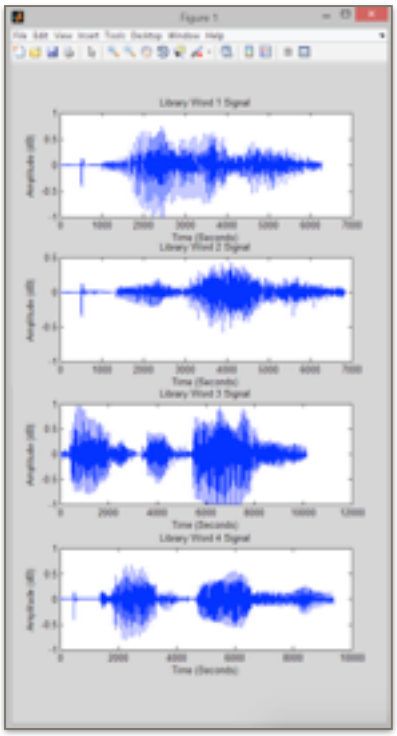
Το πρόγραμμα αυτό προορίζεται για εκτέλεση στο τερματικό της Matlab. Κατά την διάρκεια της εκτέλεσης ο χρήστης καλείται να ηχογραφήσει τέσσερις λέξεις για την δημιουργία μιας μικρής βιβλιοθήκης λέξεων και ακόμη μια για την επίδειξη της λειτουργίας φωνητικής αναγνώρισης. Αναλυτικότερα:

- Το κεντρικό πρόγραμμα βρίσκεται στο αρχείο VoiceRecognition.m, ανοίγουμε το αρχείο και το εκτελούμε.



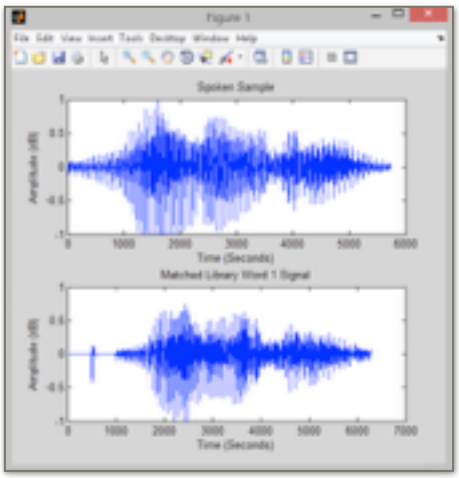
- Για τις επόμενες τέσσερις φορές ο χρήστης καλείται να ηχογραφήσει λέξεις για την δημιουργία μιας υποτυπώδης βιβλιοθήκης. Για την έναρξη της ηχογράφησης πατάμε το πλήκτρο Enter.
- Μετά την κάθε ηχογράφηση εμφανίζεται η κυματομορφή της λέξης καθώς και το γράφημα του φάσματός της.
- Όταν ολοκληρωθεί η διαδικασία της ηχογράφησης εμφανίζονται οι κυματομορφές όλων των





ηχογραφηθέντων λέξεων και ξεκινά η αναπαραγωγή τους σύμφωνα με την σειρά ηχογράφησης.

- Μετά την δημιουργία της βιβλιοθήκης, ο χρήστης καλείται να ηχογραφήσει μια τελευταία λέξη η οποία θα συγκριθεί με τις υπόλοιπες της βιβλιοθήκης ώστε να γίνει η αναγνώριση. Συνεπώς πιέζουμε ξανά το πλήκτρο Enter.
- Τέλος αναπαράγονται και εμφανίζονται οι κυματομορφές των λέξεων που ταιριάζουν περισσότερο.



Θεωρητικό υπόβαθρο

- Αναγνώριση ανθρώπινης φωνητικής δραστηριότητας (VAD)
  - Διάσπαση σήματος σε μικρότερα τμήματα (frames).
    - Η διάσπαση σε μικρότερα τμήματα γίνεται με βάση το μήκος της ηχογράφησης, για να δίνεται η δυνατότητα της αλλαγής σε περίπτωση που δεν επαρκεί ή περισσεύει ο χρόνος της ηχογράφησης.
  - Υπολογισμός του threshold που ορίζει αν υπάρχει φωνή ή θόρυβος.
- Αποκοπή ήχου (Endpoint Detection & Silence removal)
  - Απόρριψη τμήματος ήχου του δεν ξεπερνάει το threshold που ορίζει το VAD
    - Δημιουργία σήματος (threshold wave) που οριοθετεί τις περιοχές που πρόκειται να μηδενιστούν.
    - Εξομάλυνση σήματος threshold wave για την αποφυγή απότομων αλλαγών.
    - Αντιγραφή και αντικατάσταση αρχικού σήματος με νέο σήμα, χωρίς θόρυβο πριν και μετά την ομιλία, που αποτελείται μόνο από τις περιοχές που το σήμα threshold wave έχει τιμή άνω του 0.03 (Στατιστική τιμή από τις πηγές που αναγράφονται κάτω).

- Υπολογισμός MFCC
    - Διάσπαση σήματος σε μικρότερα τμήματα για επιμέρους ανάλυση. Κάθε τμήμα αποτελεί ένα απλό διάνυσμα.
    - Χρήση παραθύρων Hamming (Hamming Windowing) για την αποφυγή κενών μεταξύ των τμημάτων.
    - Εκτέλεση FFT, έναντι του DFT για γρηγορότερη απόκριση του προγράμματος, σε κάθε παράθυρο με σκοπό την εξαγωγή του πλάτους (Magnitude).
    - Χρήση Filterbank, ένα σύνολο από 26 τριγωνικά bandpass φίλτρα των οποίων οι κεντρικές συχνότητες αντιστοιχούν στις κεντρικές συχνότητες της κλίμακας mel (mel-frequency scale). Δίνεται έμφαση στις κεντρικές συχνότητες της κλίμακας mel γιατί είναι κρίσιμες στην αντίληψη της ομιλίας.
    - Λογαριθμικός μετασχηματισμός συχνότητας (mel-frequency scale) κάθε παραθύρου.
    - Υπολογισμός DCT, για την εξαγωγή των συντελεστών MFCC
- Δυναμική παραμόρφωση χρόνου & Υπολογισμός απόστασης σημάτων
    - Υλοποίηση του αλγορίθμου Dynamic Programming (DP) που με βάση την Ευκλείδεια απόσταση, υπολογίζει την ελάχιστη απόσταση μεταξύ των τιμών των σημάτων.
    - Οι αποστάσεις των τιμών των σημάτων χρησιμοποιούνται για την ταυτοποίηση της λέξης. Επιλέγεται το σήμα που έχει την μικρότερη συνολική απόσταση.

Επεξήγηση προγράμματος

- Το πρόγραμμα αποτελείται από τέσσερα αρχεία τύπου m. Το κύριο πρόγραμμα και τρεις κύριες συναρτήσεις κάθε μια από τις οποίες επιτελεί μια συγκεκριμένη λειτουργία.

Πηγές

- https://www.clear.rice.edu/elec301/Projects99/wrcocee/endpt.htm
- http://www.hindawi.com/journals/tswj/2014/146040/
- http://mirlab.org/jang/books/audiosignalprocessing/speechFeatureMfcc.asp?title=12-2%20MFCC
- http://www.ee.columbia.edu/ln/LabROSA/doc/HTKBook21/node54.html
- http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/
- http://www.diva-portal.org/smash/get/diva2:347957/FULLTEXT01.pdf

- http://www.mathcs.emory.edu/~lxiong/cs730_s13/share/slides/searching_sigkdd2012_DTW.pdf
- http://revistaie.ase.ro/content/46/s%20-%20furtuna.pdf
- Wikipedia
- MathWorks
- Διαφάνειες μαθήματος