ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΩΝ ΦΩΝΗΣ KAI HXOY

Υλοποιήθηκε από τους φοιτητές

Ευστάθιος Αρμαγανίδης – Π16008 (Erasmus+)

Μηλαθιανάκης Κωνσταντίνος Μάριος – Π16079 (Erasmus+)

Περιεχόμενα

[Γενικές Πληροφορίες 2](#_Toc45282995)

[Επεξήγηση Αλγορίθμου Εξαγωγής Χαρακτηριστικών 2](#_Toc45282996)

[Επεξήγηση Scripts 2](#_Toc45282997)

[digits\_rec.m 2](#_Toc45282998)

[label\_distributer.m 2](#_Toc45282999)

[predictions.m 2](#_Toc45283000)

[Digits\_Recognition\_Start.mlapp 2](#_Toc45283001)

[Στιγμιότυπα Εκτέλεσης Εφαρμογής 3](#_Toc45283002)

[Κεντρική Σελίδα 3](#_Toc45283003)

[Προγεγραμμένο Αρχείο Ήχου 3](#_Toc45283004)

[Ηχογράφηση Μηνύματος 5](#_Toc45283005)

[Βιβλιογραφία 6](#_Toc45283006)

# Γενικές Πληροφορίες

Είμαστε περίπτωση Erasmus+, επικοινωνήσαμε μαζί σας με σχετικό email.

Η εργασία υλοποιήθηκε με χρήση του **MatLab R2019b**. Για την υλοποίηση της αναγνώρισης ψηφίων από ηχογραφημένη ομιλία χρησιμοποιήθηκε μοντέλο SVM[[4]](#_Βιβλιογραφία) και το Dataset εκπαίδευσης λήφθηκε από την ακόλουθη διεύθυνση (<https://github.com/Jakobovski/free-spoken-digit-dataset>). Περιλαμβάνει δείγματα των ψηφίων 0 – 9 από τέσσερεις διαφορετικές φωνές. Έγινε αφαίρεση μίας φωνής λόγω της αλλοιωμένης και χαμηλής ποιότητας των δειγμάτων.

## Επεξήγηση Αλγορίθμου Εξαγωγής Χαρακτηριστικών

To πρόγραμμα χωρίζει το Dataset σε αρχεία εκπαίδευσης και αρχεία δοκιμής. Από τα πρώτα αντλεί τα χαρακτηριστικά που αναφέρονται παρακάτω για την εκπαίδευση του μοντέλου.

* SampleRate, που έχει άμεση σχέση με την συχνότητα που αντλείται από το αρχείο
* Mfcc, επιστρέφει true ή false ανάλογα με την ύπαρξη ή όχι των MFCC χαρακτηριστικών[[1]](#_Βιβλιογραφία_1) [[2]](#_Βιβλιογραφία_1)
* mfccDelta, που ορίζει την διαφορά μεταξύ του τωρινού και του προηγούμενου mfcc από το δείγμα[[1]](#_Βιβλιογραφία_1) [[2]](#_Βιβλιογραφία_1)
* mfccDeltaDelta, ορίζει την διαφορά μεταξύ του τωρινού delta και του προηγούμενου από το δείγμα[[1]](#_Βιβλιογραφία_1) [[2]](#_Βιβλιογραφία_1)
* pitch[[1]](#_Βιβλιογραφία_1) [[3]](#_Βιβλιογραφία_1)
* spectralCentroid[[1]](#_Βιβλιογραφία_1) [[4]](#_Βιβλιογραφία_1)

Τα αρχεία δοκιμής συνδυάζονται σε νέα αρχεία όπου το καθένα απαρτίζεται από έξι τυχαίες ηχογραφήσεις ψηφίων, από τυχαίους ομιλητές.

# Επεξήγηση Scripts

## digits\_rec.m

Αυτό το script χρησιμοποιείται για την εκπαίδευση του μοντέλου SVM από το Digit\_Dataset. Συγκεκριμένα, τα αρχεία του dataset ανακατεύονται και διαμοιράζονται σε αρχεία εκπαίδευσης και αρχεία δοκιμής. Τα αρχεία δοκιμής αποθηκεύονται σε φάκελο με όνομα Test\_Dataset, όπου συνδυάζονται σε νέα αρχεία όπου το καθένα απαρτίζεται από έξι τυχαίες ηχογραφήσεις ψηφίων, από τυχαίους ομιλητές.

Έπειτα, εξάγονται τα προαναφερθέντα χαρακτηριστικά του ήχου από τα αρχεία εκπαίδευσης, τα οποία χρησιμοποιούνται για την εκπαίδευση του μοντέλου SVM.

## label\_distributer.m

Χρησιμοποιείται μέσω του digits\_rec.m για την σωστή κατηγοριοποίηση των αρχείων σύμφωνα με τον αριθμό που ακούγεται. Η ένδειξη αυτή βρίσκεται στο όνομα του εκάστοτε αρχείου, και με χρήση regex το πρόγραμμα αναγνωρίζει την ένδειξη.

## predictions.m

Χρησιμοποιείται από το πρόγραμμα για την ερμηνεία του αρχείου που έδωσε ο χρήστης για ανάλυση (είτε ηχογραφημένο είτε ένα από τα αρχεία δοκιμής). Το αρχείο χωρίζεται σε κομμάτια ανάλογα με το που ανιχνεύεται η ύπαρξη ομιλίας. Στην συνέχεια, γίνεται εξαγωγή χαρακτηριστικών από τα επιμέρους τμήματα του αρχείου και προβλέπεται το ψηφίο του εκάστοτε τμήματος μέσω του εκπαιδευμένου μοντέλου.

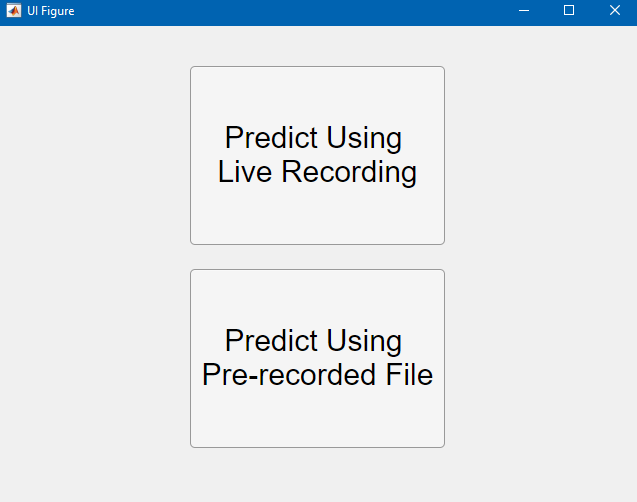
Καλείται με το πάτημα του κουμπιού “Predict” μέσω του αρχείου Digits\_Recognition\_Start.mlapp. Τα αποτελέσματα που βγάζει εμφανίζονται μέσω της ίδιας φόρμας.

## Digits\_Recognition\_Start.mlapp

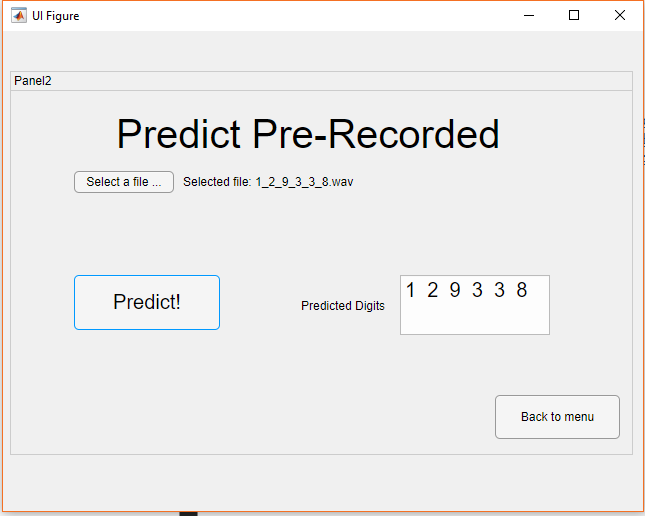
Είναι το βασικό αρχείο εκκίνησης της εφαρμογής, διαχειρίζεται τις λειτουργίες καθώς και το γραφικό περιβάλλον του. Λειτουργεί ως συντονιστής των λειτουργιών του προγράμματος καθώς δεν προσφέρει νέες λειτουργίες αλλά ομαδοποιεί τις υπάρχουσες για μεγαλύτερη ευκολία στην εκτέλεση της εφαρμογής.

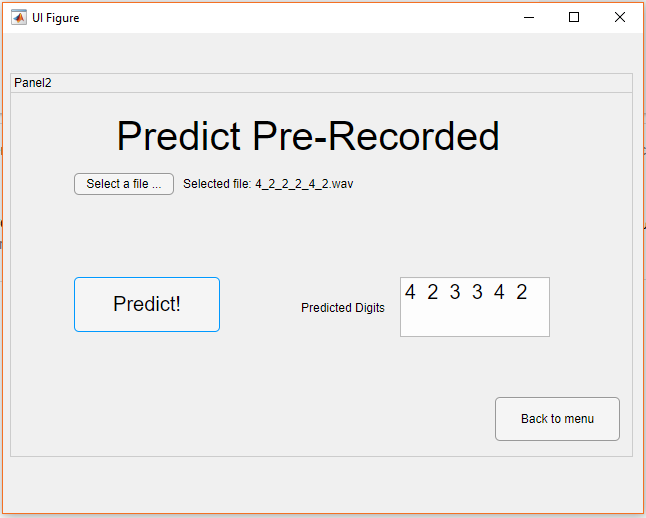
# Στιγμιότυπα Εκτέλεσης Εφαρμογής

## Κεντρική Σελίδα



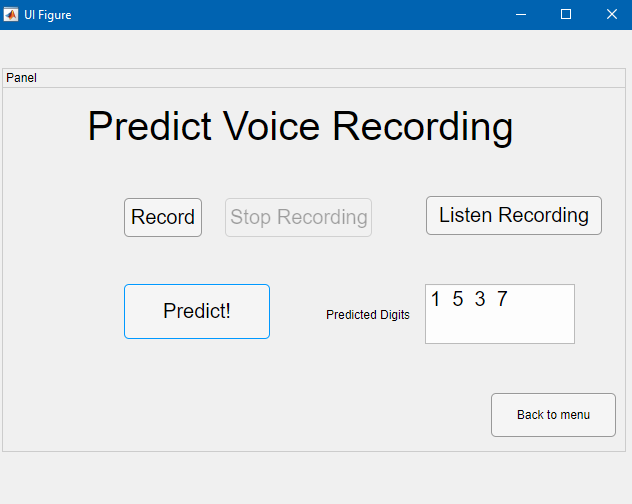
## Προγεγραμμένο Αρχείο Ήχου





## Ηχογράφηση Μηνύματος

Δείγμα ηχογράφησης αριθμών 1,5,6,7 από το μικρόφωνο του υπολογιστή. Η αναγνώριση αριθμών από το μικρόφωνο εμφανίζει μεγαλύτερο ποσοστό λάθους συγκριτικά με τα ηχογραφημένα μηνύματα του δείγματος δοκιμής που λήφθηκε από το Dataset. Παρ’ όλ’ αυτά, ακόμη και σε αυτήν την περίπτωση είναι αρκετά ακριβές.



# Βιβλιογραφία

1. <https://www.mathworks.com/help/audio/ref/audiofeatureextractor.html>
2. <https://www.mathworks.com/help/audio/ref/mfcc.html>
3. <https://www.mathworks.com/help/audio/ref/pitch.html>
4. <https://www.mathworks.com/help/audio/ref/spectralcentroid.html>
5. <https://www.mathworks.com/help/stats/fitcecoc.html>