ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΩΝ ΦΩΝΗΣ KAI HXOY

Υλοποιήθηκε από τους φοιτητές

Ευστάθιος Αρμαγανίδης – Π16008

Μηλαθιανάκης Κωνσταντίνος Μάριος – Π16079

Περιεχόμενα

[Γενικές Πληροφορίες 2](#_Toc45115036)

[Επεξήγηση Αλγορίθμου 2](#_Toc45115037)

[Επεξήγηση Scripts 2](#_Toc45115038)

[Digits\_rec.m 2](#_Toc45115039)

[Βιβλιογραφία 2](#_Toc45115040)

# Γενικές Πληροφορίες

Η εργασία υλοποιήθηκε με χρήση του **MatLab R2019b**.

## Επεξήγηση Αλγορίθμου

Για την υλοποίηση της αναγνώρισης ψηφίων από ηχογραφημένη ομιλία χρησιμοποιήθηκε μοντέλο SVM και Dataset εκπαίδευσης από την ακόλουθη διεύθυνση (<https://github.com/Jakobovski/free-spoken-digit-dataset>). Περιλαμβάνει δείγματα των ψηφίων 0 – 9 από τέσσερεις διαφορετικές φωνές.

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, μου έχει άμεση σχέση με την συχνότητα που αντλείται από το αρχείο
* mfcc[[1]](#_Βιβλιογραφία)
* mfccDelta[[1]](#_Βιβλιογραφία)
* mfccDeltaDelta[[1]](#_Βιβλιογραφία)
* pitch[[2]](#_Βιβλιογραφία)
* spectralCentroid[[3]](#_Βιβλιογραφία)

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.

# Βιβλιογραφία

1. <https://www.mathworks.com/help/audio/ref/mfcc.html>
2. <https://www.mathworks.com/help/audio/ref/pitch.html>
3. <https://www.mathworks.com/help/audio/ref/spectralcentroid.html>