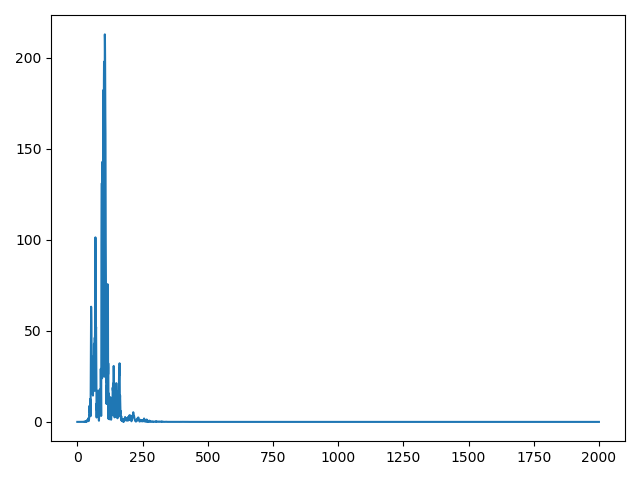
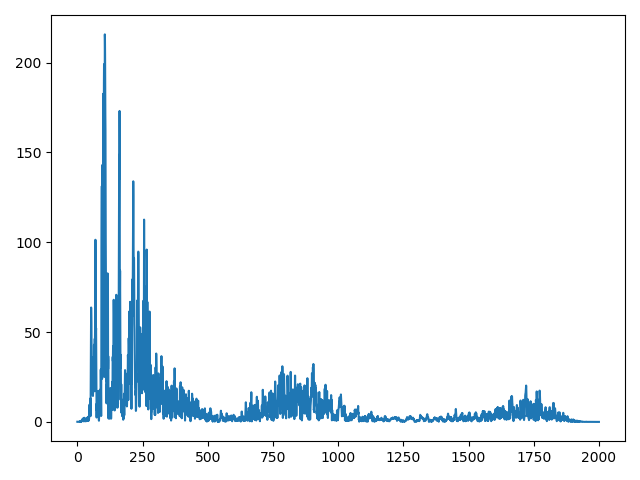
# Signal Preprocessing

## Filtering

Στα αρχεία εισόδου εφαρμόστηκε φίλτρο ώστε να μειωθούν τα επίπεδα του θορύβου και ταυτόχρονα να διατηρηθεί η πληροφορία του σήματος που μας ενδιαφέρει να αναλύσουμε. Με το φίλτρο διατηρήσαμε τα δεδομένα του σήματος που ανήκουν στην τυπική ζώνη συχνοτήτων που σχετίζονται με τον μέσο άνθρωπο. Ένας ενήλικος άνδρας παράγει ήχους από 85 έως 180 Hz ενώ μία γυναίκα από 165 έως 255 Ηz. Παρότι οι άνθρωποι όταν παράγουν αυτές τις συχνότητες παράλληλα παράγουν και ανώτερες αρμονικές τους συχνότητες δεν κρίθηκε σκόπιμο αυτές να διατηρηθούν καθώς σκοπός μας δεν είναι η πιστότερη αναπαραγωγή ήχου αλλά η αναγνώριση και κατηγοριοποίησή του.



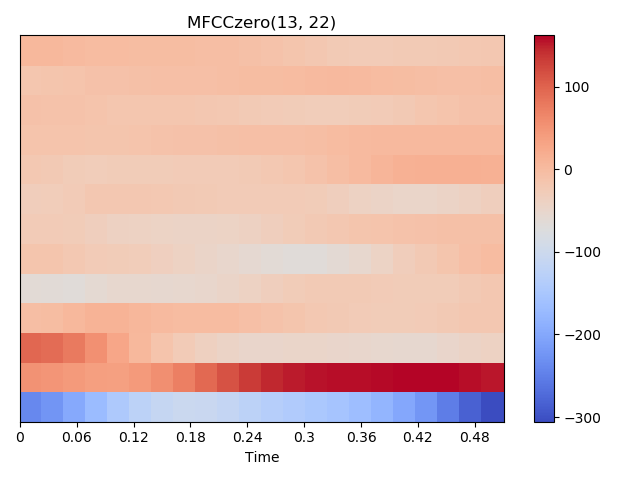
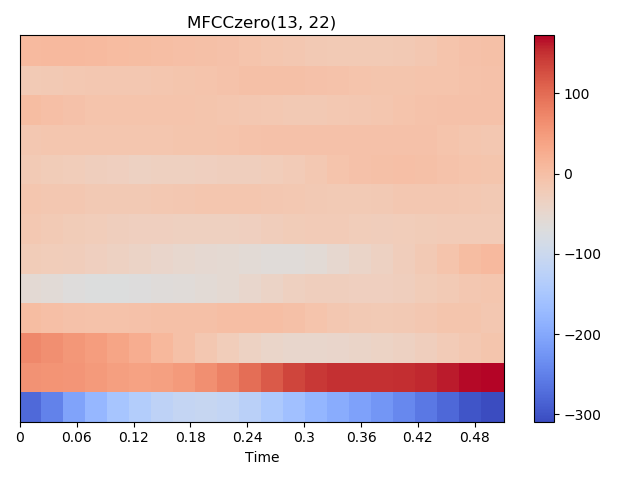
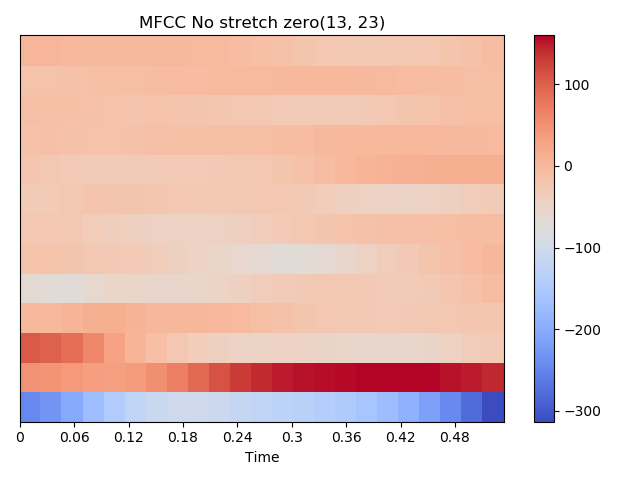
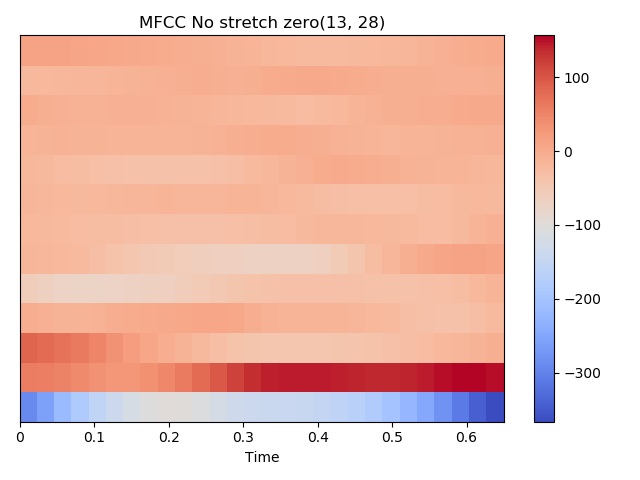
Το φίλτρο που υλοποιήθηκε είναι bandpass με σημεία αποκοπής τα 85-255 Hz και είναι τύπου IIR butterworth.

## Trimming

Με την εφαρμογή του trimming αφαιρούμε τυχόν παύσεις πριν ή μετά την εκφώνηση των ψηφίων και έτσι το σύστημα διατηρεί μόνο την χρήσιμη πληροφορία για την αναγνώριση ομιλίας. Πιο αναλυτικά, εξασφαλίζουμε ότι οι φθόγγοι που προφέρονται από τον κάθε ομιλητή θα βρεθούν σε χρονικό σημείο όσο το δυνατό πιο κοντινό μεταξύ των διαφορετικών δεδομένων εισόδου. Ουσιαστικά με αυτή την επεξεργασία εξάγουμε τις σχετικές θέσεις τόσο των αυτούσιων σημείων του σήματος (raw audio ή fourier) όσο και του spectrogram και των features.

Εάν δεν εφαρμόζονταν trimming οι θέσεις των patterns θα βρίσκονται σε διαφορετικά σημεία στον χρόνο (άξονας X) με αποτέλεσμα να προσθέσουν θόρυβο στην διαδικασία.

## Streching



Εφαρμόζοντας stretching μετά το trimming οι τιμές που επιστρέφει ο αλγόριθμος MFCC είναι πολύ πιο κοντά μεταξύ τους σε σχέση με το αρχικό σήμα. Το αποτέλεσμα της εφαρμογής του stretching είναι η χρονική διάρκεια του σήματος να είναι ίδια για όλα τα δεδομένα και με αυτό τον τρόπο επιτρέπεται στο νευρονικό δίκτυο να εντοπίσει με μεγαλύτερη ευκολία patterns. Στην περίπτωση που δεν εφαρμοστεί stretching, τα patterns στον άξονα Χ έχουν μεγαλύτερο μήκος ανάλογα με την ταχύτητα που εκφωνούνται από τον κάθε ομιλητή. Με αυτή την μέθοδο τα δεδομένα εισόδου κανονικοποιούνται χωρίς να χάνεται χρήσιμη πληροφορία για την ταξινόμηση τους.

Παρατηρώντας το σχήμα των patterns χωρίς strech προκύπτει ότι το στοιχείο στο χρονικό σημείο 0.2 σημαντική διαφορά στον χρωματισμό του, ενώ στην δεύτερη περίπτωση ο χρωματισμός είναι ο ίδιος. Τέλος όπως παρατηρούμε ο αλγόριθμός του stretch κανονικοποιεί τα δεδομένα και ως προς το magnitude.

Για να επηρεαστούν όσο το δυνατό λιγότερο τα στοιχεία των αρχείων υπολογίζεται η μέση διάρκεια ενός clip ήχου και προσαρμόζονται όλα τα clips στην διάρκεια αυτή.