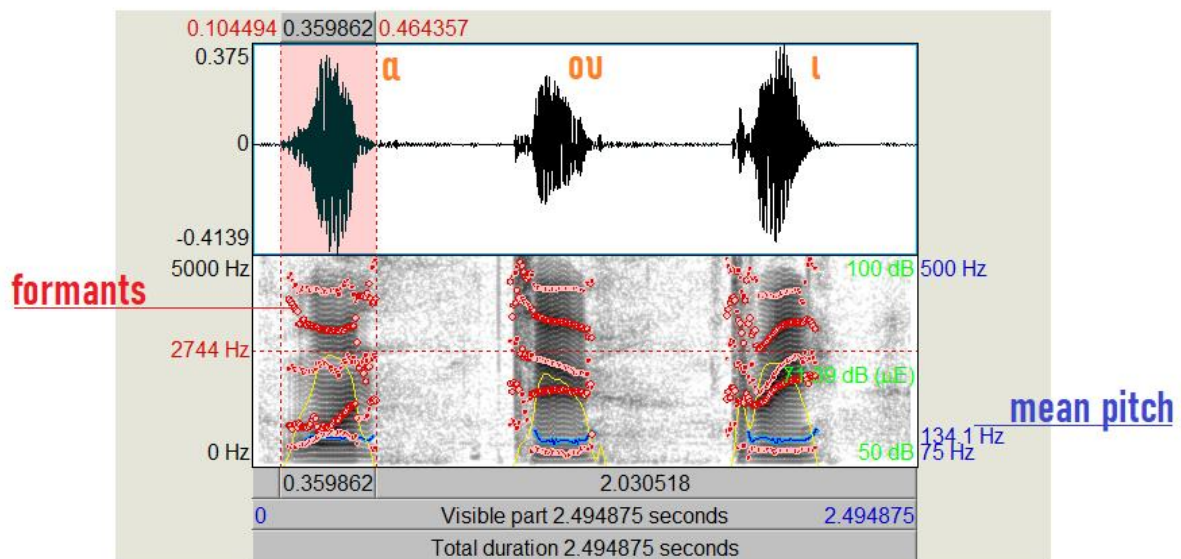


## BHMA 1:



730.93

Hz, 1404.52 Hz, 2837.89 Hz.

Speakers	Vowels	mean pitch	formants		
		p	F1	F2	F3
Άνδρας	A	137.95	741.48	1123.20	2395.21
	ΟΥ	131.03	414.33	1809.95	2430.69
	I	131.75	396.56	2014.55	2543.73
Γυναίκα	A	180.55	732.17	1409.23	2839.28
	ΟΥ	187.70	532.23	1932.50	2842.53
	I	183.1	375.41	1916.77	2583.46



Το pitch παραμένει για τον κάθε ομιλητή σχεδόν το ίδιο για τα διαφορετικά φωνήεντα και μάλλον είναι χαρακτηριστικό της εκάστοτε 'φωνής', δεν φαίνεται να μπορεί να βοηθήσει ως χαρακτηριστικό ταξινόμησης των διαφορετικών γραμμάτων. (γυναίκα μεγαλύτερο pitch από τον άνδρα ομιλητή)

Αντίθετα παρατηρούμε ότι τα formants είναι πιο χαρακτηριστικά για το κάθε γράμμα και κυρίως για τα f1 και f2 ενώ τα f3 παρατηρώ ότι είναι παραπλήσια για κάθε ομιλητή για κάθε φωνήεν.

## BHMA 2:

Με τον parser δημιουργούμε 3 λίστες:

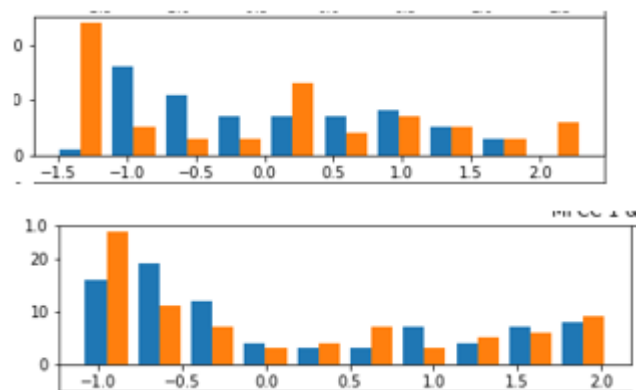
- **wav**: περιέχει το αρχείο ήχου (εκφώνηση)
- **speaker**: με τους ομιλητές
- **digit**: περιέχει τα ψηφία

## BHMA 3:

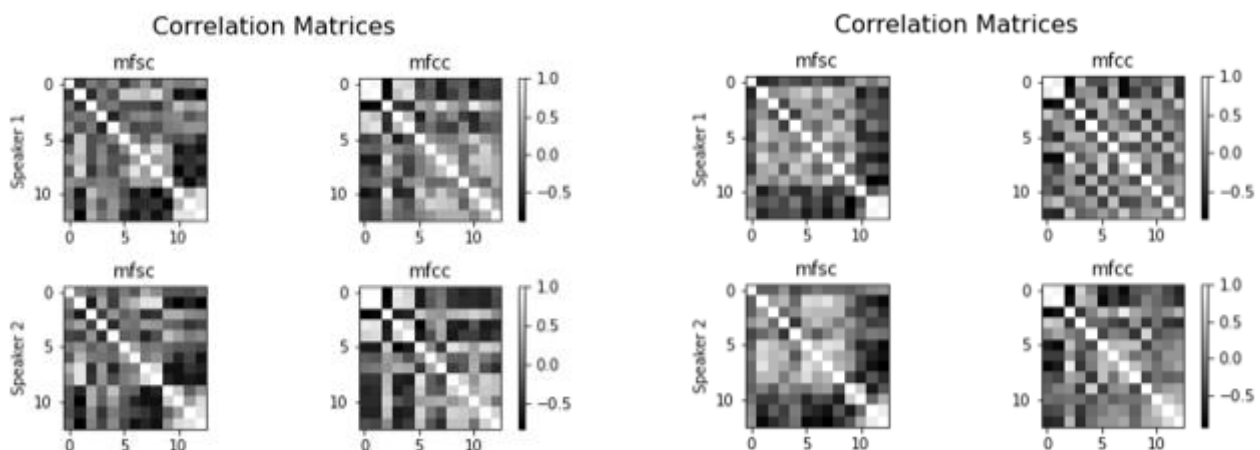
Δημιουργούμε τρεις λίστες με τα **mfccs**, **delta**, και **delta2** με τις συναρτήσεις **librosa.feature.mfcc** και **librosa.feature.delta** (order 1 & 2) αντίστοιχα. Δημιουργούμε 13 χαρακτηριστικά ανά wav με μήκος παραθύρου 0.025 sec και **sampling rate** 22050.

## BHMA 4:

Χρησιμοποιώντας τα ψηφία από τον AM (003400**n1:3n2:7**). Δημιουργήσαμε 2 λίστες με όλα τα samples για κάθε ψηφίο (#15). Για κάποιες τιμές των mfccs παρατηρούμε σημαντικές αποκλίσεις στις συχνότητες ενώ για άλλες μικρότερες.

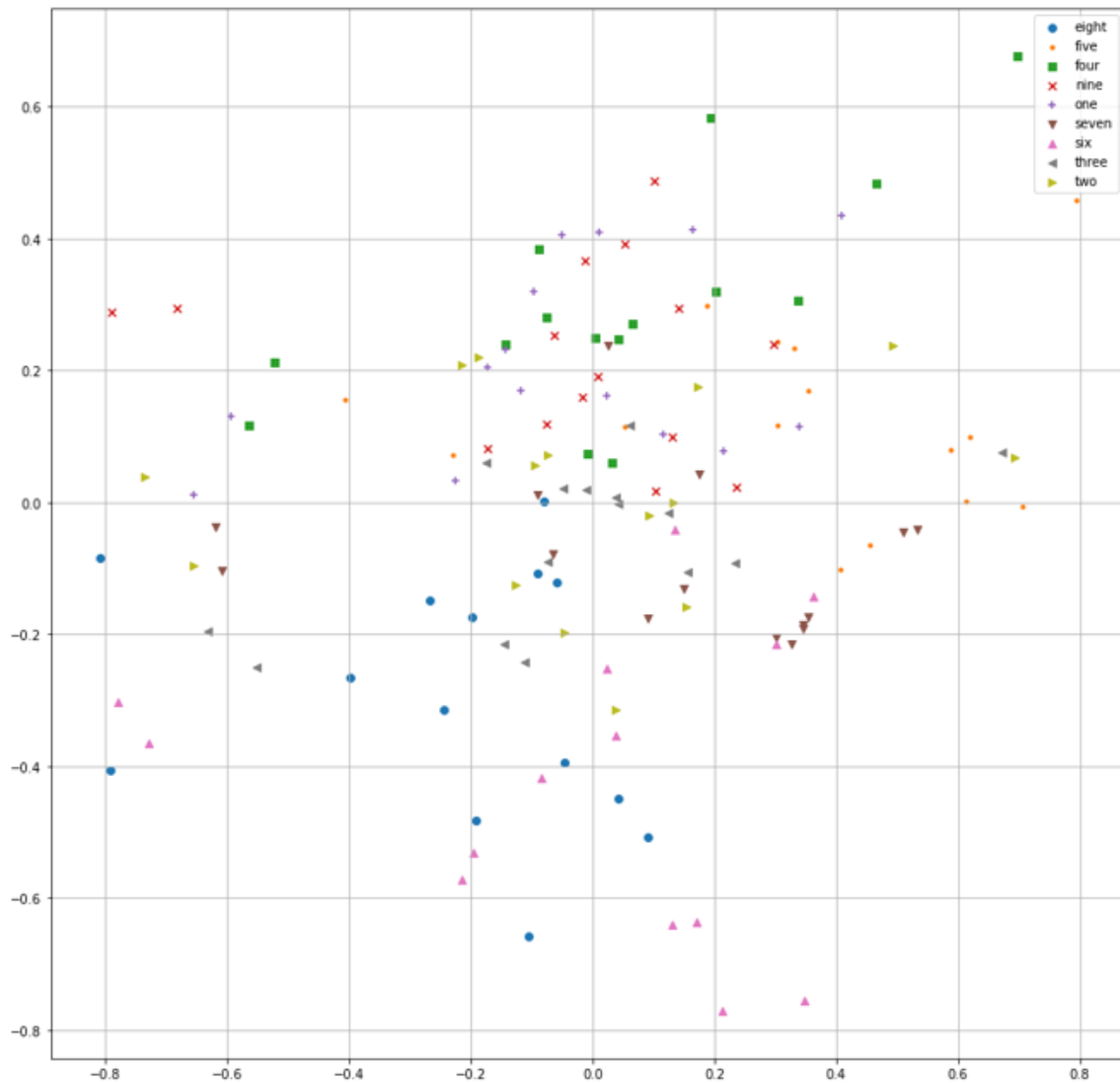


Στην συνέχεια χρησιμοποιώντας το idct δημιουργούμε τα διανύσματα με τα mfsc στα οποία και παρατηρούμε μεγαλύτερες συσχετίσεις, μέσω των correlation graphs.



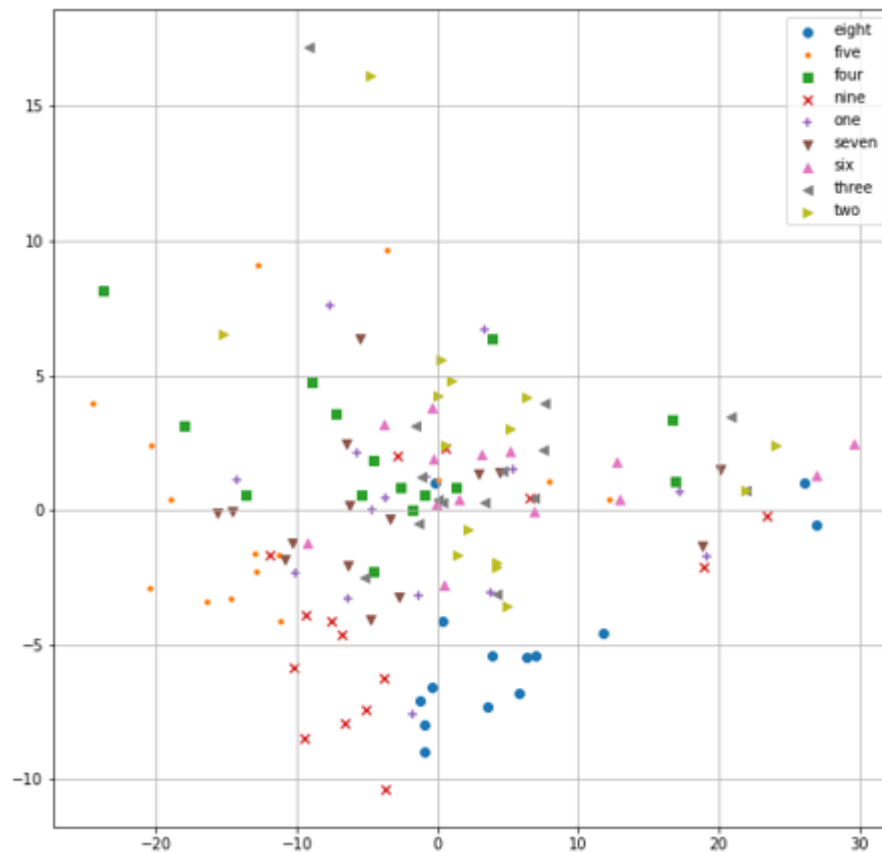
### **ΒΗΜΑ 5:**

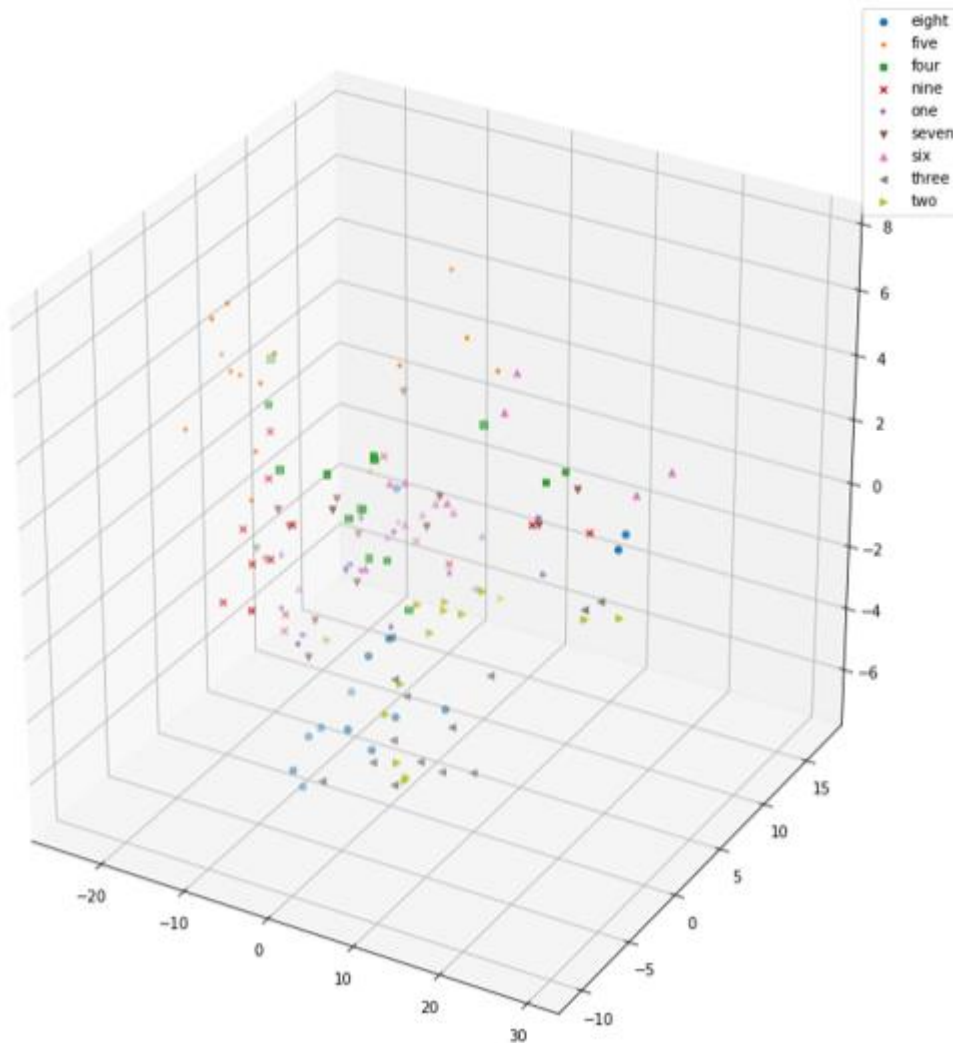
Παρατηρούμε ότι χρησιμοποιώντας μόνο 2 διαστάσεις από τα χαρακτηριστικά δεν είναι δυνατό να αποφανθούμε για τον διαχωρισμό των ψηφίων.



### **ΒΗΜΑ 6:**

Χρησιμοποιώντας 2 διαστάσεις στην PCA ακόμα δεν μπορούμε να αποφανθούμε για τον διαχωρισμό των δεδομένων αν και διακρίνονται κάποιες περιοχές. Αντίθετα χρησιμοποιώντας 3 διαστάσεις και ως εκ' τούτου τρισδιάστατη απεικόνιση παρατηρούμε σημαντική βελτίωση στην διάκριση των χαρακτηριστικών. Η χρήση της PCA ερμηνεύει την διακύμανση των δεδομένων γύρω από τον άξονα που προβάλλονται.





### **BHMA 7:**

Classification Results without ZC

```
NB: 0.870968
SVC: 0.139785
RFC: 1.000000
KNN: 0.795699
```

Classification Results with Z C

```
NB: 0.881720
SVC: 0.193548
RFC: 1.000000
KNN: 0.731183
```

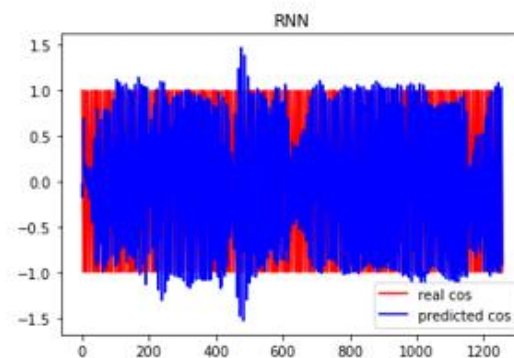
Ο Random Forest Classifier φαίνεται να αποδίδει καλύτερα όπως και ο Naive Bayes. Τέλος, προσθέτοντας το Zero Crossing Rate βλέπουμε μια περεταίρω βελτίωση σε όλους τους ταξινομητές εκτός του KNN.

### **BHMA 8:**

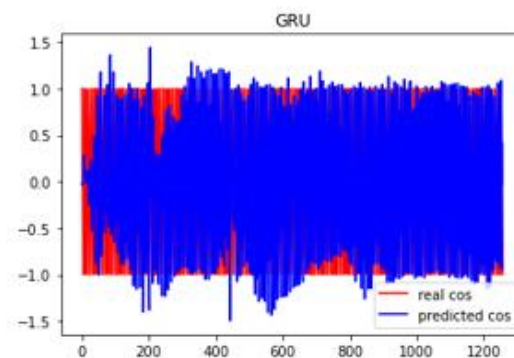
Τα RNN δεν δυσκολεύονται να συνδέσουν παρελθοντικές πληροφορίες με το παρόν λόγω του κανόνα της αλυσίδας που χρησιμοποιείται στο εν' λόγω δίκτυο. Το πρόβλημα στην εκπαίδευσή

τους προκύπτει όταν το κενό στην πληροφορία μεταξύ των πληροφοριών μεγαλώνει. Το πρόβλημα αυτό δημιουργείται λόγω του vanishing gradient problem. Όταν χρησιμοποιούμε backpropagation τα βάρη του ηη ανανεώνονται ανάλογα με το partial derivative της συνάρτησης κόστους σε κάθε επανάληψη. Το πρόβλημα προκύπτει όταν τα gradients μικραίνουν πάρα πολύ και έτσι παρόλες τις επαναλήψεις τα βάρη δεν ανανεώνονται με αποτέλεσμα το ηη να σταματήσει. Για να επιλυθεί αυτό το πρόβλημα χρησιμοποιούμε τα LSTM. Τα LSTM έχουν διαφορετική εσωτερική δομή – παραπάνω layers τα οποία ελέγχουν ποιες πληροφορίες διατηρούνται, ξεχνιούνται και μεταφέρονται στην επόμενη κατάσταση. Από την χρήση των δικτύων RNN, GRU, και LSTM, παρατηρούμε ότι το LSTM έχει την καλύτερη απόδοση.

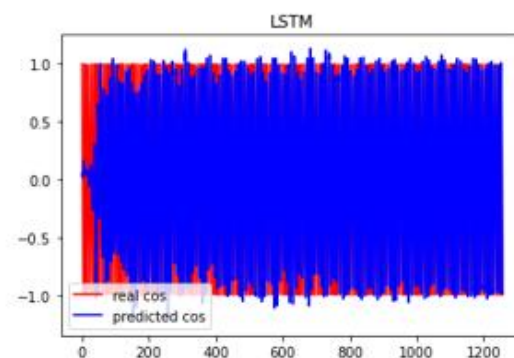
rnn



gru



lstm



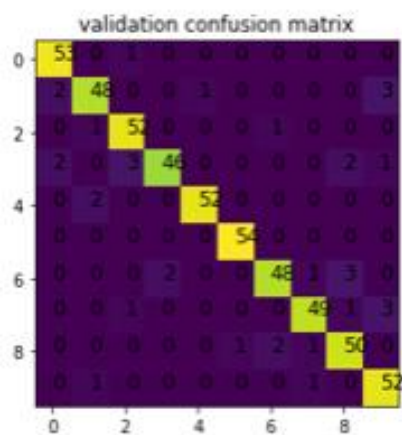
### **BHMA 9:**

Με την χρήση της βοηθητικής συνάρτησης φορτώνουμε τα δεδομένα με stratified split για την διατήρηση ιδίου αριθμού διαφορετικών ψηφίων μεταξύ training και validation.

### **BHMA 10-13:**

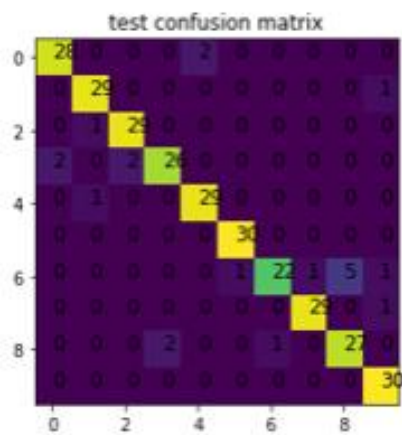
Το καλύτερο μοντέλο GMM-HMM προκύπτει για:

```
GridSearch with best results on validation set:  
Best number for states: 1  
Best number for Gaussian Mixture: 5  
Best Accuracy: 0.9333333333333333
```



Για την αποφυγή του overfitting η εκπαίδευση θα πρέπει να γίνεται όχι μόνο στο training αλλά και στο validation set. Ωστόσο η αξιολόγηση θα πρέπει να γίνει σε εντελώς καινούργια δεδομένα.

Όπως φαίνεται από τα confusion matrices η ταξινόμηση ήταν επιτυχημένη.



### **BHMA 14:**

**L2 Normalization:** κανονικοποίηση ανάλογα με το εύρος των βαρών αναλογικά (προηγούμενη τιμή) – αποφυγή overfitting μαζί με κατάλληλη αρχικοποίηση του  $\lambda_{l2}$ .

**Dropout:** σε κάθε επανάληψη και ανάλογα με τις πιθανότητες σε κάθε νευρώνα συγκεκριμένοι νευρώνες αποκόπτονται  $(1-p)$  και άλλοι διατηρούνται  $(p)$  για την αποφυγή και πάλι του overfitting.

Τα bidirectional LSTM αν και πιο αργά από τα ‘απλής κατεύθυνσης’ αποδίδουν καλύτερα μιας και διαχειρίζονται της εισόδους και προς τις 2 κατευθύνσεις (παρελθόν-παρόν & παρόν – μέλλον).