Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Αναγνώριση Προτύπων

2016-2017

Γιώργος Δασούλας
03112010
Νικολάος Ζαρίφης
03112178

Προπαρασκευή 2ης εργαστηριακής άσκησης

Προθεσμία : 21/11/2016

Περίληψη

Σκοπός του εργαστηρίου είναι η εξαγωγή των Mel Frequency cepstral coefficients,που χρησιμοποιούνται στην εξαγώγη χαρακτηριστηκών.

Βήμα 1

Διαβάζουμε τα αρχεία μας με μια loop η οποία φτιάχνει το αντικείμενο Sound το οποίο κατά την κατασκευή του εκτελεί τα περισσότερα βήματα.

Βήμα 2

Με αντιστροφή στο **μετασχηματισμό z** ξέρουμε ότι οι σταθερές δίνουν μια συνάρτηση **Dirac** ή **Kronecker Delta** στο 0, κι ο πολλαπλασιασμός με 1/z κάνει μεταφορά όποτε η αντίστροφη θα είναι :
$F(n)= \delta(n) -\delta(n-1)*a^2$. Για να το εφαρμόσουμε απλά κάνουμε συνέλιξη με το σήμα μας, καθώς είμαστε στο πεδίο του χρόνου .
Το preemph πρακτικά είναι ένα φίλτρο το οποίο πρακτικά θα κάνει **αύξησή της ενέργειας** σε μεγαλύτερες συχνότητες.

Βήμα 3

Κάθε frame εχει 400 sample (T*fs) κάνοντας το σπάσιμο το frames μαζι με το overlap παίρνουμε τα διάφορα frames μας απο το αρχικό μας σήμα. Και στην συνέχεια εφαρμόζουμε μια συνάρτηση παραθύρου. Πραγμα που το κάνουμε γιατί θα μετατρέψουμε το σήμα μας DFT το όποιο μπορούμε να το σκεφτούμε ως Interpolation στον DTFT. Το ενδιαφέρον εδώ είναι οτι εξαιτίας του Interpolation όπος είναι λογικό το οποίο γίνεται με πολλαπλασιασμό με συνημίτονα να έχουμε συχνότητα μόνο σε συγκεκριμένα σήμεία, οπότε χρησιμοποιούμε το window για μια εξομαλυνση(δημιουργουμε spectral leakage)

Βήμα 4

Αφού θέλουμε ανώ κι κάτω φράγμα, για ανώ θα έχουμε την nyquist = 0.5fs και για κάτω 20Hz (Αρχή ακούστηκης αλλά εμείς παίρνουμε τα 100Hz γιατι τα 20 βγάζει μηδενικές τιμές). Παίρνουμε 24 ισο-αποσταση σημεία κι να γύρναμε στο αντίστροφο της mel και στην συνέχεια τα μετατρέπουμε σε κεντρικές συχνότητες σύμφωνα με τα bins του DFT μας, και στην συνέχεια με έναν πολλαπλασιασμό έχουμε το αποτελεσμά μας. (Προφάνω στο τελευταίο τρίγωνο παίρνουμε μόνο το μισο γιατί εκεί ξεκινάνε οι αρνητικές συχνότητες). Οι κεντρικές αυξάνονται λογαριθμίκα.

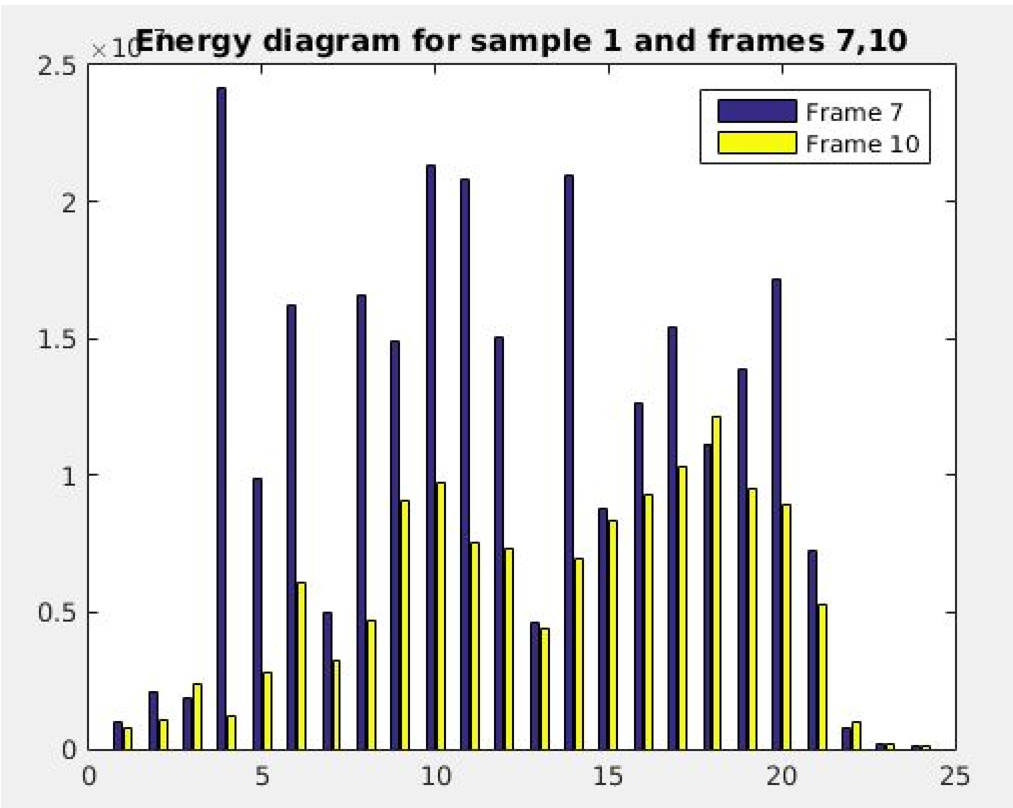
Βήμα 5

Στο βήμα αυτό υπολογίζουμε την ενέργεια απόκρισης κάθε φίλτρου . Για αυτό το λόγο , χρησιμοποιήσαμε το γεγονός ότι από **Parseval** η ενέργεια είναι το άθροισμα των **DFT** coefficients στο τετράγωνο δια το πλήθος . Πράγματι , ισχύει :

$$\sum_{n=0}^{N-1} |x[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X[k]|^2$$

where $X[k]$ is the DFT of $x[n]$, both of length N.

Προέκυψε το παρακάτω διάγραμμα :

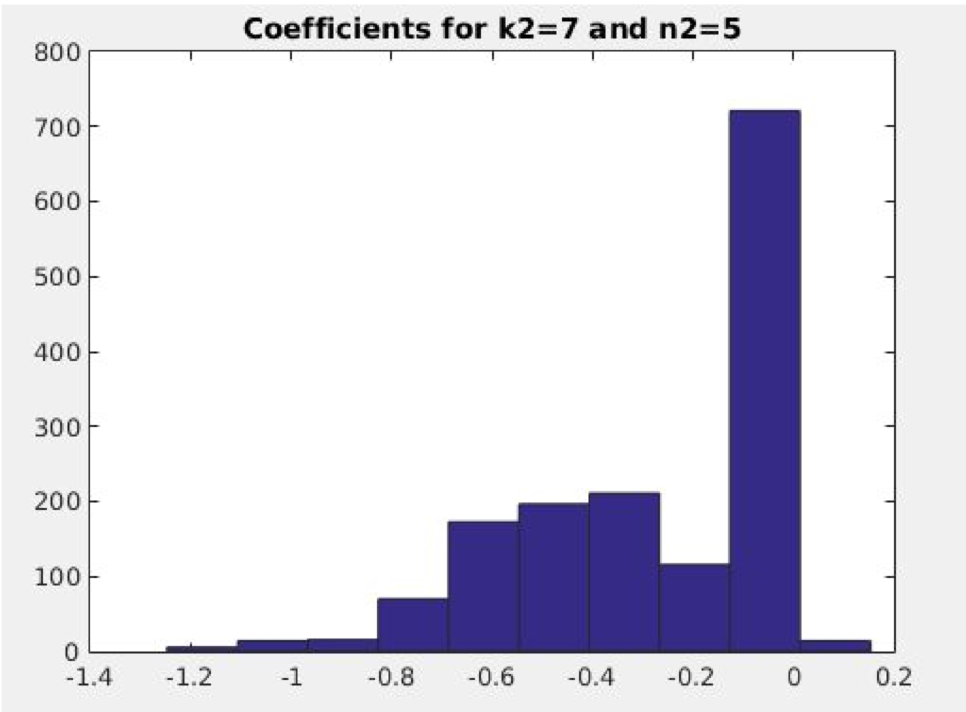


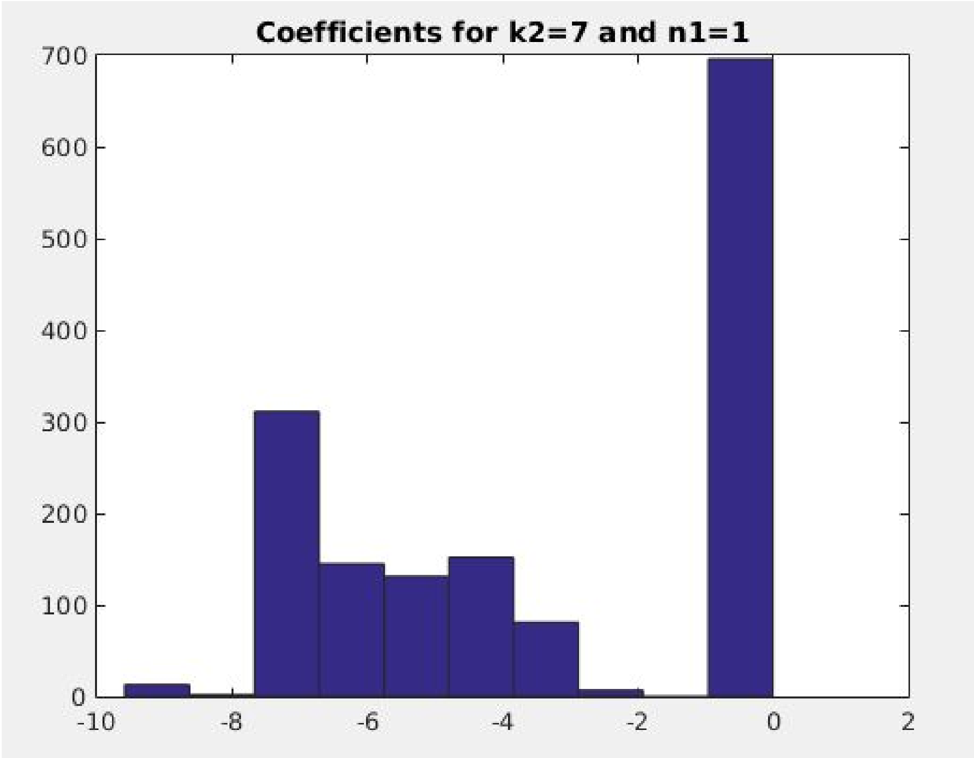
<u>Βήμα 6</u>

Στο βήμα αυτό απλώς λογαριθμήσαμε τις τιμές της ενέργειας .

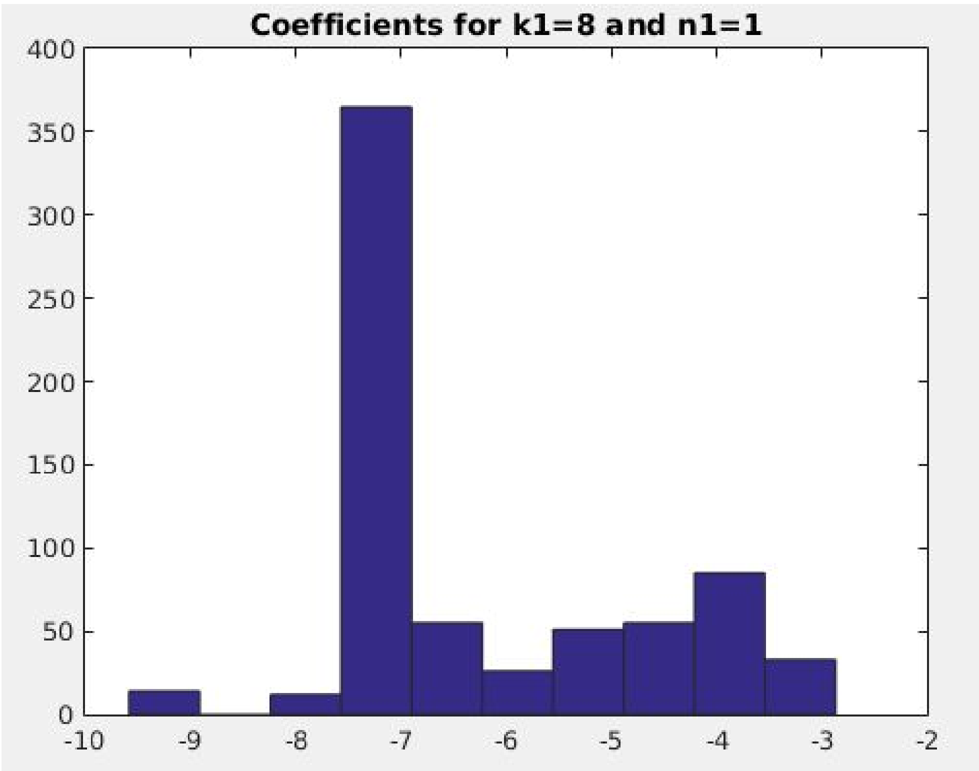
<u>Βήμα 7</u>

Για τον υπολογισμό του DCT χρησιμοποιήσαμε την αντίστοιχη συνάρτηση της MATLAB , εφαρμόσαμε την κατάλληλη κανονικοποίηση για την συμβατότητα του τύπου και από τις 24 τιμές της ενέργειας για κάθε φίλτρο, κρατήσαμε μόνο τις πρώτες 13 τιμές , καθώς το πλήθος των χαρακτηριστικών ανά frame είναι $N_c = 13$, μηδενίζοντας τις υπόλοιπες τιμές . Στη συνέχεια , για κάθε frame και κάθε sample που αντιστοιχεί στο ψηφίο $k_1 = 8$, δηλαδή τα 1-14 , βρήκαμε το ιστόγραμμα συχνοτήτων των cepstrum συντελεστών . Ομοίως , πράξαμε και για το $k_2 = 7$. Επισημαίνουμε πως σύμφωνα με τον αριθμό μητρώου είχαμε $n_1 = 1$, $n_2 = 5$. Ακολουθούν τα 4 διαγράμματα :

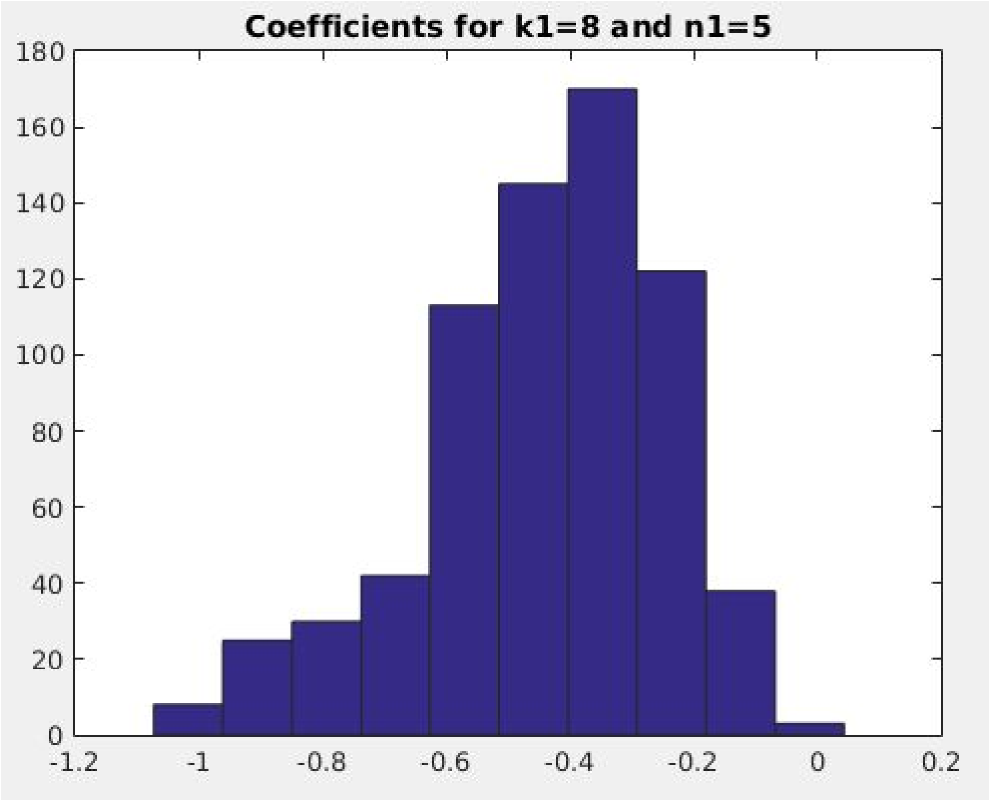




Coefficients for k2=7 and n1=1
700
600
500
400
300
200
100
0
-10
-8
-6
-4
-2
0
2



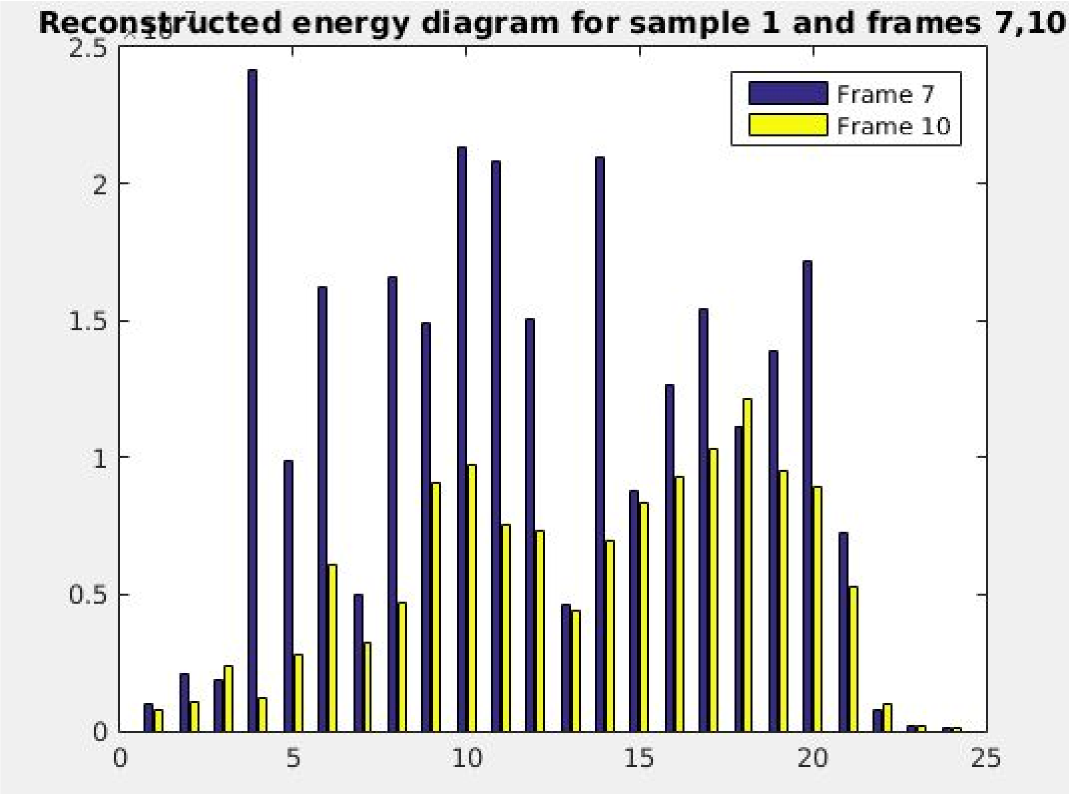
Coefficients for k1=8 and n1=1
400
350
300
250
200
150
100
50
0
-10
-9
-8
-7
-6
-5
-4
-3
-2



Βήμα 8

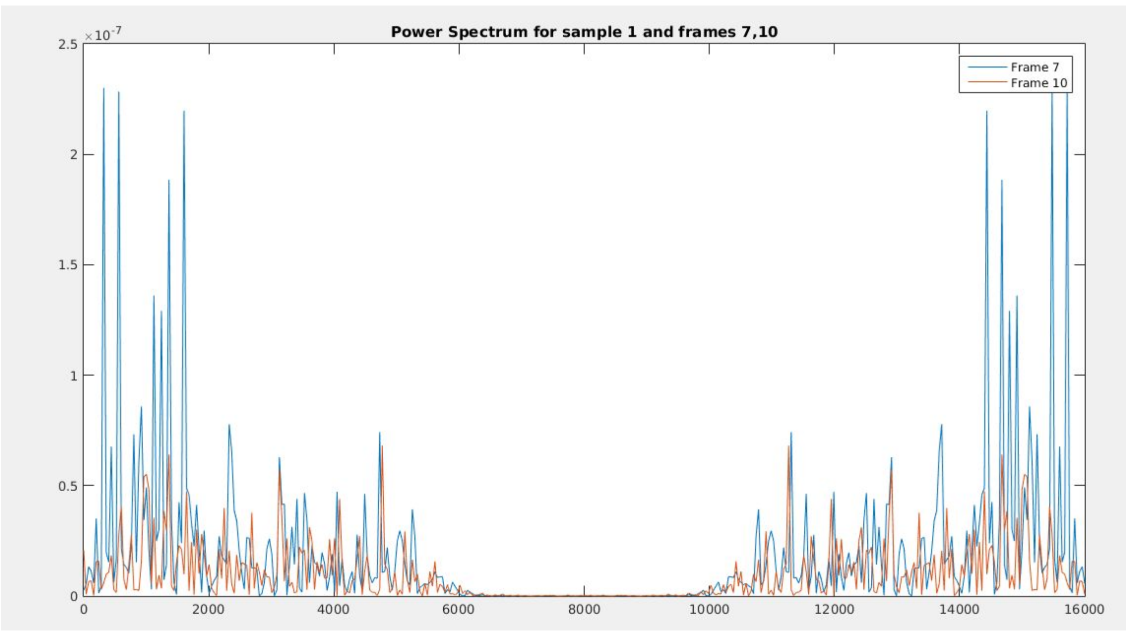
Ο μετασχηματισμός συνημιτόνου **DCT** έχει αντίστροφο τον ίδιο μετασχηματισμό με διαφορά τον συντελεστή κανονικοποίησης . Αυτό ισχύει , γιατί ο operator του **DCT** είναι ορθογώνιος (ανάγεται σε έναν ορθογώνιο πίνακα). Για την ανακατασκευή χρησιμοποιήσαμε τον αντίστροφο **DCT** και ύψωση σε δύναμη του 10 , καθώς θυμόμαστε ότι ξεκινήσαμε για το DCT με το λογάριθμο της ενέργειας .
Υλοποιώντας το σε MATLAB , είδαμε ότι είχαμε μικρή απόκλιση (της τάξης του 1-5%) ανάμεσα στο ανακατασκευασμένο και το αρχικό διάγραμμα της ενέργειας για τα συγκεκριμένα frames .

Ακολουθεί το διαγράμμα του ανακατασκευασμένου $E_i(j)$:



Χρησιμοποιήσαμε τα ίδια frames για να μπορούμε να συγκρίνουμε την ομοιότητα ανάμεσα στο αρχικό και το ανακατασκευασμένο διάγραμμα .

Επίσης , για το φάσμα ισχύος χρησιμοποιήσαμε τον τύπο : $\frac{\sum |X|^2}{N}$, αθροίζοντας πάνω στο σύνολο των συχνοτήτων , συνεπώς N=400 . Για τα ίδια frames (7 , 10) , βρήκαμε τα εξής αποτελέσματα :



Παρατηρούμε την αναμενόμενη συμμετρία με κέντρο συμμετρίας την N/2 , ανάμεσα στις αρνητικές και τις θετικές τιμές

Βήμα 9

Παρακάτω φαίνεται το διάγραμμα με τις 9 διαφορετικές κλάσεις .Παρατηρούμε αποκλίσεσι μεταξύ των συντελεστών ίδιων ψηφίων . Αυτό μπορεί να οφείλεται στην τυχαιότητα με την οποία διαλέξαμε τα frames . Παρολαυτά , φαίνεται κάποια δυνατότητα διαχωρισμού , καθώς μπορούμε να διακρίνουμε μικρές αλλά κλειστές ομάδες γειτόνων της ίδιας κλάσης .
Επίσης , παρατηρούμε πως συντελεστές διαφορετικών ψηφίων δεν πέφτουν στο ίδιο σημείο , κάτι το οποίο μας ενθαρρύνει για την δυνατότητα διαχωρισμού.