ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Αναγνώριση Μεμονωμένων Νοημάτων της Ελληνικής Νοηματικής Γλώσσας με Κρυφά Μαρκοβιανά Μοντέλα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Σταύρου Ι. Θεοδωράκη

Επιβλέπων: Πέτρος Α. Μαραγκός
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2008

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Αναγνώριση Μεμονωμένων Νοημάτων της Ελληνικής Νοηματικής Γλώσσας με Κρυφά Μαρκοβιανά Μοντέλα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Σταύρου Ι. Θεοδωράκη

Επιβλέπων: Πέτρος Α. Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 17 Ιουλίου 2008.

.....	.....	.....
Πέτρος Μαραγκός Καθηγητής Ε.Μ.Π.	Κώστας Τζαφέστας Λέκτορας Ε.Μ.Π.	Στέφανος Κόλλιας Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2008

.....
Σταύρος Ι. Θεοδωράκης
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Σταύρος Ι. Θεοδωράκης, 2008.
Με επιφύλαξη παντός δικαιώματος.All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τη συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τη συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η παρούσα διπλωματική εργασία επικεντρώνεται στην υλοποίηση, επέκταση ενός συστήματος αναγνώρισης της νοηματικής γλώσσας. Η εφαρμογή αυτή εμπίπτει στο ερευνητικό πεδίο της Αναγνώρισης Προτύπων και ειδικότερα στον κλάδο της αλληλεπίδρασης ανθρώπου - μηχανής. Πιο συγκεκριμένα, στην εργασία εξετάζονται όλα τα στάδια για την διαδικασία της αναγνώρισης, και συγκεκριμένα τα στάδια της μοντελοποίησης της νοηματικής γλώσσας, της εκπαίδευσης των μοντέλων χρησιμοποιώντας κάθε κανάλι πληροφορίας ξεχωριστά και τέλος του συνδυασμού των διαφορετικών καναλιών πληροφορίας ώστε να παρθεί η τελική απόφαση (stream fusion). Ιδιαίτερη έμφαση δίνεται στις μεθόδους που χρησιμοποιούνται για τον συνδυασμό των καναλιών όπου εξετάζονται επεκτάσεις των HMM όπως τα Parallel και Product HMMs.

Η εργασία είναι δομημένη σε τέσσερα κεφάλαια, κάθε ένα από τα οποία ασχολείται με ένα διαφορετικό υποπρόβλημα. Επίσης παρουσιάζονται αποτελέσματα για τις μεθόδους που χρησιμοποιήθηκαν τα οποία είναι πολύ ενθαρρυντικά.

Λέξεις - Κλειδιά

αναγνώριση προτύπων, αναγνώριση νοηματικής γλώσσας, αλληλεπίδραση ανθρώπου - μηχανής, μοντελοποίηση νοηματική γλώσσας, Hidden Markov Models, fusion, Parallel HMMs, Product HMMs.

Abstract

The goal of this thesis is to test, extend, develop a sign language recognition system. This application is part of the more general research area refferred to as -human - machine interaction- which is strongly realated to the science and technology of Pattern Recognition. More specifically, this thesis focuses on all the stages that are required for recognition, sign language modeling, training each stream seperatly and streams fusion. Special effort was made to the methods we used for streams fusion where we test some exensions to HMMs like Parallel and Product HMMs.

This thesis is organized in four chapters, each of which focuses on a separate subproblem of the Sign Language Recognition. Some first and hopefull experiments are presented along with basic conclusions .

Keywords

Pattern Recognition, sign language recognition, human - machine interaction, sign language modeling, Hidden Markov Models, fusion, Parallel HMMs, Product HMMs.

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τον επιβλέποντα καθηγητή μου κύριο Π. Μαραγκό του οποίου οι διαλέξεις στα μαθήματα Όραση Υπολογιστών και Αναγνώριση Προτύπων αποτέλεσαν το βασικότερο κίνητρο για να ασχοληθώ με την ερευνητική περιοχή της Αναγνώρισης Νοηματικής Γλώσσας.

Επίσης θα ήθελα να ευχαριστήσω την Όλγα Διαμαντή με την οποία είχαμε μια εξαιρετική συνεργασία. Ακόμα, θα ήθελα να ευχαριστήσω ιδιαίτερα τον Νάσσο Κατσαμάνη, τον Δημήτρη Δημητριάδη και γενικότερα όλα τα παιδιά από το εργαστήριο για την βοήθεια που μου πρόσφεραν απλόχερα όποια στιγμή και αν τους την ζήτησα.

Επίσης θα ήθελα να ευχαριστήσω το ΙΕΛ και ιδιαίτερα την Ε. Ευθυμίου και την Ε. Φωτεινέα για την βάση δεδομένων που μας παραχώρησαν πάνω στην οποία κάναμε όλα τα πειράματα μας.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου που με στήριξαν όλα αυτά τα χρόνια.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Η επιστήμη των ηλεκτρονικών υπολογιστών πρέπει να κάνει σημαντικά βήματα ακόμα έτσι ώστε η επικοινωνία ανθρώπου - υπολογιστή να είναι αρκετά φυσική. Για τον άνθρωπο (χρήστη) ο πιο φυσικός τρόπος επικοινωνία με τον υπολογιστή είναι μέσω της φωνής και των χειρονομιών. Η αναγνώριση φωνής έχει κάνει μεγάλα βήματα προόδου τα τελευταία χρόνια [19], ενώ αντίθετα η αναγνώριση χειρονομιών βρίσκεται ακόμα αρκετά πίσω [33, 4]. Όμως οι χειρονομίες στην επικοινωνία μεταξύ ανθρώπων έχουν πολύ σημαντικό ρόλο και παρέχουν πληροφορίες που η ομιλία από μόνη της δεν μπορεί να δώσει [9]. Έτσι η συμβολή και των δύο μαζί θα δημιουργούσε ένα σύστημα επικοινωνίας ανθρώπου-υπολογιστή πιο φυσικό από το ήδη υπάρχον.

Τα πλεονεκτήματα της αναγνώρισης νοηματικής γλώσσας είναι αρκετά. Πρώτον, η επικοινωνία ανθρώπων που έχουν προβλήματα ακοής με ανθρώπους που δεν γνωρίζουν την Νοηματική Γλώσσα (ΝΓ) είναι δύσκολη έως και αδύνατη κάποιες φορές έτσι η δημιουργία ενός συστήματος αναγνώρισης ΝΓ θα γεφύρωνε αυτό το επικοινωνιακό κενό.

Δεύτερον, τα ήδη υπάρχοντα συστήματα που βασίζονται στην επικοινωνία με ομιλία και αναγνώριση φωνής (speech-based interfaces) δεν βοηθάνε καθόλου τους ανθρώπους με προβλήματα ακοής οπού ο βασικός τρόπος επικοινωνίας τους είναι μέσω της νοηματικής γλώσσας με αποτέλεσμα να απωθούνται από τους υπολογιστές. Έτσι εάν το επίπεδο της αναγνώρισης της νοηματικής γλώσσας πλησιάσει το επίπεδο της αναγνώρισης φωνής τότε θα δώθει σοβαρό κίνητρο στην χρησιμοποίηση των υπολογιστών από αυτή την ομάδα ανθρώπων.

Τρίτον, ένα αποτελεσματικό σύστημα αναγνώρισης ΝΓ όπως επίσης και σύνθεσης ΝΓ θα βοηθήσει πάρα πολλούς ανθρώπους με προβλήματα ακοής ώστε να μπορούν να παρευρεθούν σε συνέδρια, σεμινάρια, συζητήσεις οπού η επικοινωνία γίνεται μέσω της ομιλίας.

Τέταρτον ένα τέτοιο σύστημα θα δώσει μεγάλη ώθηση σε ερευνητικά πεδία

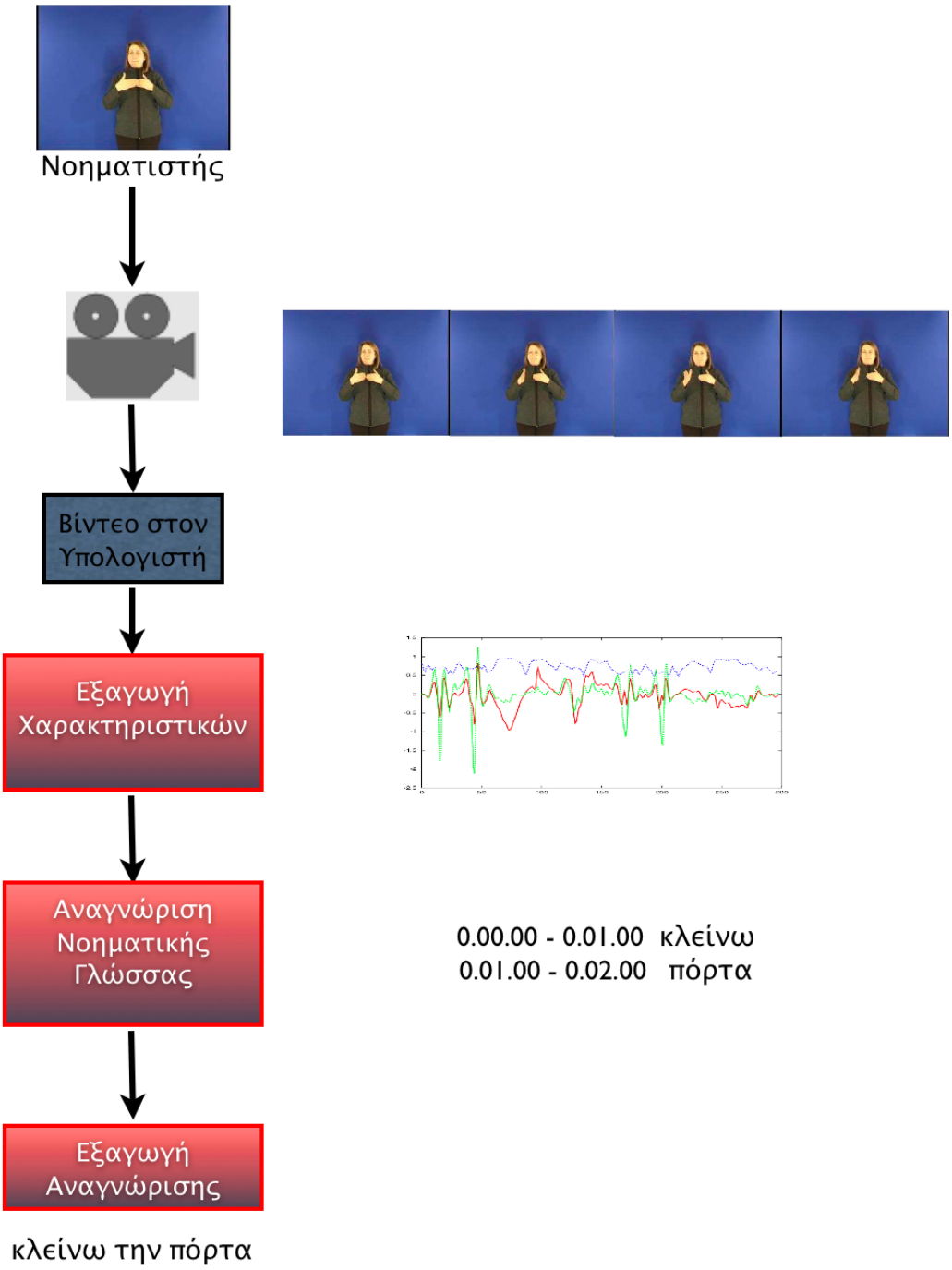
όπως της Επεξεργασίας Φυσικής Γλώσσας. Ένα από τα πρώτα βήματα τα οποία πρέπει να γίνουν όταν έχουμε μια μεγάλη βάση δεδομένων με στόχο την γλωσσική έρευνα είναι η αντιστοίχηση του τι ειπώθηκε και ποία ακριβώς χρονική στιγμή από τον νοηματιστή, των λεγόμενων μεταγραφών (transcriptions), διαδικασία αρκετά κουραστική και χρονοβόρα για τον άνθρωπο. Έτσι η αυτοματοποίηση αυτής της διαδικασίας με την βοήθεια ενός υπολογιστή θα βοηθούσε ιδιαίτερα. Με τα μέχρι τώρα συστήματα αναγνώρισης που έχουν δημιουργηθεί αυτό είναι αδύνατον να πραγματοποιηθεί.

1.1 Απεικόνιση του Συστήματος Αναγνώρισης της Νοηματικής Γλώσσας

Η αναγνώριση της νοηματικής γλώσσας όπως θα περιγραφεί σε αυτή την διπλωματική αποτελεί ένα κομμάτι ενός πολύ μεγαλύτερου συστήματος. Ο στόχος αυτού του συστήματος είναι η μετατροπή ακολουθίας νοημάτων σε κείμενο (sign-to-text) που είναι παρόμοιο με το σύστημα μετατροπής λόγου σε κείμενο (speech-to-text) που αναφέρεται για την αναγνώριση φωνής. Ουσιαστικά από την μια έχουμε έναν άνθρωπο ο οποίος λέει κάτι στην νοηματική γλώσσα και από την άλλη το σύστημα μας πρέπει να το μετατρέψει σε γραπτή γλώσσα. Αρχικά αυτό μπορεί να φαίνεται απλό όμως στην πραγματικότητα το σύστημα μπορεί να σπάσει σε τρία εξίσου δύσκολα και σημαντικά συστήματα α) εξαγωγή οπτικών χαρακτηριστικών από το βίντεο που θέλουμε να αναγνωρίσουμε β) την αναγνώριση της πρότασης που είπε ο νοηματιστής γ) την μετατροπή της πρότασης αυτής που αναγνωρίσαμε σε μορφή που μπορεί ένας άνθρωπος να διαβάσει. Ένα παράδειγμα μπορούμε να δούμε στο Σχήμα 1.1.

Στο πρώτο στάδιο εξάγουμε την απαραίτητη πληροφορία από κάθε πλαίσιο (frame) του βίντεο (όπως την θέση των δύο χεριών του νοηματιστή πάνω στην εικόνα, το σχήμα των χεριών του (χειρομορφή) και την διεύθυνση που δείχνει η χειρομορφή). Αυτή την πληροφορία την ονομάζουμε διάνυσμα χαρακτηριστικών (features vector) και αντιπροσωπεύει ένα πλαίσιο (frame) από ολόκληρο το βίντεο. Τελικά υπολογίζοντας τα διανύσματα χαρακτηριστικών για κάθε πλαίσιο του βίντεο δημιουργούμε μια ακολουθία διανυσμάτων χαρακτηριστικών η οποία περιέχει την απαραίτητη πληροφορία για όλο το βίντεο.

Στο δεύτερο στάδιο χρησιμοποιώντας την ακολουθία διανυσμάτων χαρακτηριστικών επιτυγχάνουμε την αναγνώριση των νοημάτων που εμφανίζονται στο υπό ανάλυση βίντεο. Δηλαδή ουσιαστικά προσπαθούμε να αναγνωρίζουμε ποιο νόημα ειπώθηκε και σε ποια χρονική στιγμή. Ένα πολύ σημαντικό



Σχήμα 1.1: Ένα ολοκληρωμένο σύστημα Αναγνώρισης Νοηματικής Γλώσσας. Στα πλαίσια αυτής της διπλωματικής θα ασχοληθούμε με το προτελευταίο στάδιο (Αναγνώριση Νοημάτων)

σημείο είναι ότι σε αυτό το στάδιο δεν έχει καμία σημασία το ποια νοηματική γλώσσα χρησιμοποιήθηκε.

Στο τρίτο και τελευταίο στάδιο μετατρέπουμε την πρόταση που αναγνωρίσαμε σε μορφή που μπορεί ένας άνθρωπος να διαβάσει ανάλογα με την γλώσσα που χρησιμοποίησε ο νοηματιστής.

Έτσι από τα παραπάνω διαπιστώνουμε ότι η δημιουργία ενός συστήματος αναγνώρισης νοηματικής γλώσσας απαιτεί την συμβολή τριών μεγάλων επιστημονικών πεδίων. Πρώτον της Όρασης Υπολογιστών (Computer Vision) για την εξαγωγή των οπτικών χαρακτηριστικών από το βίντεο. Δεύτερον της Αναγνώρισης Προτύπων (Pattern Recognition) για την αναγνώριση των νοημάτων που αναπαριστάται από το σήμα δεδομένων. Τρίτον της Επεξεργασία Φυσικής Γλώσσας για την εξαγωγή της πρότασης που αναγνωρίσαμε σε μορφή που μπορεί ένας άνθρωπος να διαβάσει.

Σε αυτή τη διπλωματική εργασία θα ασχοληθούμε μόνο με το στάδιο, της αναγνώριση νοημάτων.

1.2 Η Σχέση με την Αναγνώριση Φωνής

Σε αυτή την διπλωματική θα αναφέρομαι συχνά στην αναγνώριση φωνής και ο λόγος είναι ότι σχετίζεται αρκετά με την αναγνώριση νοηματικής γλώσσας. Λόγο της πολυετούς εμπειρίας που υπάρχει στην αναγνώριση φωνής μας δίνεται η δυνατότητα να μην ξεκινάμε από το μηδέν με αποτέλεσμα να μην χρειάζεται να λύσουμε κάθε νέο πρόβλημα που συναντάμε αφού τα περισσότερα έχουν ήδη αντιμετωπιστεί. Η τωρινή κατάσταση στην Αναγνώριση Νοηματικής Γλώσσας είναι αρκετά πίσω από την Αναγνώριση Φωνής. Για αυτό το λόγο θα κάνουμε μια σύνοψη πάνω στην Αναγνώριση Φωνής.

Η είσοδος και η έξοδος ενός συστήματος αναγνώρισης φωνής είναι παρόμοιες με αυτές του συστήματος αναγνώρισης νοηματικής γλώσσας. Η είσοδος είναι μια ακολουθία από διανύσματα χαρακτηριστικών που αντιπροσωπεύουν τις ακουστικές ιδιότητες της φράσης που εκφωνήθηκε. Στόχος του συστήματος είναι να αναγνωριστεί από τα διανύσματα χαρακτηριστικών τι ειπώθηκε και ποια χρονική στιγμή.

Το αρχικό φωνητικό σήμα που παίρνουμε από την καταγραφή με την βοήθεια ενός μικροφώνου δεν είναι δόκιμο για αναγνώριση. Έτσι το πρώτο βήμα που πρέπει το σύστημα μας να κάνει είναι να αναπαραστήσει το φωνητικό σήμα σε μια μορφή όπου η αναγνώριση θα είναι εφικτή. Έχει γίνει μεγάλη έρευνα για αυτή την αναπαράσταση και δύο παραδείγματα μετασχηματισμών που χρησιμοποιούνται αρκετά είναι filter bank βασισμένο στον Fast Fourier Transform [6] και linear predictive coding η οποία μειώνει το μέγεθος του κωδικοποιημένου σήματος προβλέποντας τις τιμές με την βοήθεια των

προηγούμενων [24], [25]. Οι συγκεκριμένοι τρόποι αναπαράστασης δεν είναι χρήσιμοι στην αναγνώριση νοηματικής λόγο τις διαφοράς μεταξύ φωνητικών και βίντεο σημάτων.

Το επόμενο βήμα είναι η αναγνώριση μέσω της ακολουθίας των διανυσματικών χαρακτηριστικών. Στο σημείο αυτό η σχέση με την αναγνώριση νοηματικής γλώσσας είναι πολύ μεγάλη. Το μεγάλο πρόβλημα είναι η μεταβλητότητα των χαρακτηριστικών. Ακόμα και αν το ίδιο πρόσωπο εκφέρει την ίδια λέξη δύο φορές το ακουστικό σήμα δεν θα είναι ακριβώς το ίδιο και τις δύο φορές με αποτέλεσμα να πρέπει να δημιουργηθούν μοντέλα τα οποία να είναι ανεπηρέαστα από αυτή την μεταβλητότητα.

Μια πολύ γνωστή μέθοδος που χρησιμοποιήθηκε είναι -η λεγόμενη- ¨ template matching¨. Με αυτή την μέθοδο τα διανύσματα χαρακτηριστικών συγκρίνονταν με τα προκαθορισμένα μοντέλα μας και το αποτέλεσμα ήταν η επιλογή του μοντέλου που ταίριαζε καλύτερα. Στις μέρες μας όμως χρησιμοποιείται ένα είδος στατιστικών μοντέλων που ονομάζεται Hidden Markov models (HHMs) [23]. Κάθε ακουστικό σήμα αντιπροσωπεύεται από ένα σύνολο πιθανοτικών μοντέλων, ένα HMM για κάθε λέξη και η αναγνώριση εστιάζεται στο να βρεθεί το σύνολο των μοντέλων τα θα έχουν την μεγαλύτερη πιθανότητα να αντιπροσωπεύουν το συγκεκριμένο φωνητικό σήμα. Έτσι χρησιμοποιώντας πιθανοτικά μοντέλα επιτυγχάνεται η ανεξαρτητοποίηση από την μεταβλητότητα που υπάρχει στα φωνητικά σήματα.

Επίσης ένα ακόμα μεγάλο πλεονέκτημα που έχουν τα HMMs είναι ότι μπορούν να εκπαιδευτούν αυτόματα. Η διαδικασία εκπαίδευσης των μοντέλων είναι η εξής: δίνεται στο σύστημα ένα σύνολο από δεδομένα όπου έχει καταγραφεί η αντιστοιχία τους με την γλώσσα που έχει χρησιμοποιηθεί δηλαδή κάθε δεδομένο εκπαίδευσης γνωρίζουμε τι σημαίνει. Μετά με την βοήθεια κάποιων μαθηματικών συναρτήσεων ρυθμίζουμε τις παραμέτρους των μοντέλων μας μεγιστοποιώντας την πιθανότητα σωστής αναγνώρισης των δεδομένων εκπαίδευσης [23].

Για την αξιολόγηση του συστήματος χρησιμοποιούμε ένα άλλο σύνολο δεδομένων το οποίο ονομάζουμε test data set όπου και σε αυτό γνωρίζουμε την αντιστοιχία του κάθε φωνητικού σήματος με το τι έχει ειπωθεί έτσι μπορούμε να συμπεράνουμε κατά πόσο αποτελεσματική είναι η αναγνώριση που έκανε το σύστημα μας. Βασικός κανόνας είναι να μην έχουμε χρησιμοποιήσει δεδομένα από το test data set στο σύνολο δεδομένων εκπαίδευσης γιατί τότε η αξιολόγηση δεν θα είναι ενδεικτική.

Γνωρίζοντας ότι τα HMMs μοντέλα έχουν πολύ καλά αποτελέσματα στην αναγνώριση φωνής βγάζουμε το συμπέρασμα ότι θα μπορούσαμε να τα χρησιμοποιήσουμε και στην αναγνώριση νοηματικής γλώσσας.

Υπαρχουν δύο βασικά προβλήματα στην αναγνώριση φωνής : 1)Αναγνώριση μεμονωμένων λεξεων. Κάθε λέξη έχει εκφωνηθεί ξεχωριστά χωρίς να υ-

πάρχουν συμφραζόμενα και η αρχή και το τέλος είναι καθορισμένα. 2) Αναγνώριση συνεχούς λόγου. Το πρόβλημα είναι δυσκολότερο για τους λόγους που αναφέρονται παρακάτω.

Πρώτον εάν δεν έχουν οριστεί το όρια για την κάθε λέξη τότε το σύστημα αναγνώρισης θα πρέπει από μόνο του να κάνει -όπως λέγεται- κατάτμηση (segmentation), πρόβλημα πολύ δύσκολο και για την αναγνώριση φωνής αλλά και για την νοηματική γλώσσα. Δεύτερον η αναγνώριση συνεχούς λόγου έχει να κάνει με ολόκληρες προτάσεις με αποτέλεσμα η κάθε λέξη να επηρεάζεται από τα συμφραζόμενα λόγω της γραμματικής που επιβάλλεται. Έτσι πχ η λέξη ¨διαβάζω¨ μπορεί στην συγκεκριμένη πρόταση να είναι ¨διάβαζα¨ εάν η πρόταση αναφέρεται στο παρελθόν.

Προβλήματα όπως το παραπάνω ήταν αδύνατον να λυθούν αν μοντελοποιούσαμε κάθε λέξη ξεχωριστά γιατί πρώτον ο αριθμός των μοντέλων μας θα αύξανε δραματικά με την αύξηση του λεξιλογίου και δεύτερον είναι αδύνατον να βρεθεί αρκετός αριθμός εκφωνήσεων για την κάθε λέξη έτσι ώστε να μπορεί να γίνει σωστά η εκπαίδευση. Μια ιδέα η οποία έλυσε το πρόβλημα αυτό ήταν να γίνει μοντελοποίηση των φωνημάτων που συντελούν στην δημιουργία μιας λέξης όπου ο αριθμός τους είναι περιορισμένος και ανεξάρτητος από το εύρος λεξιλογίου που θα χρησιμοποιήσουμε. Έτσι, με αυτή την ιδέα η αναγνώριση μεγάλου εύρους λεξιλογίου και συνεχούς λόγου μπορούσε πλέον να είναι εφικτή.

Βασιζόμενοι σε αυτή την ιδέα θα μπορούσαμε να εφαρμόσουμε το ίδιο και στην αναγνώριση νοηματικής γλώσσας με την διαφορά ότι αυτό το σπάσιμο είναι πολύ πιο δύσκολο μέχρι στιγμής. Αυτός θα μπορούσε να είναι ένας λόγος για τον οποίο η αναγνώριση νοηματικής γλώσσας είναι πιο πίσω από την αναγνώριση φωνής.

1.3 Σχετική Έρευνα

Δυο είναι τα βασικά πεδία επιστημονικού ενδιαφέροντος, που έχουν ως αντικείμενο την Νοηματική Γλώσσα. Το πρώτο επικεντρώνεται στη μετατροπή τμημάτων γραπτού ή προφορικού λόγου σε τμήματα λόγου μιας Νοηματικής Γλώσσας ενώ τι δεύτερο, που παρουσιάζει και τις περισσότερες δυσκολίες, ασχολείται με την αναγνώριση νοημάτων και την απόδοσή τους στην ομιλουμένη γλώσσα.

Όσον αφορά το δεύτερο επιστημονικό πεδίο τα Συστήματα Αναγνώρισης ΝΓ διακρίνονται σε εκείνα που χρησιμοποιούν ηλεκτρομηχανικές μονάδες, όπως τα γάντια δεδομένων (data-gloves) και σε συστήματα τα οποία εκμεταλλεύονται τεχνικές επεξεργασίας εικόνας. Τα συστήματα της δεύτερης κατηγορίας εστιάζουν το ενδιαφέρον τους στην αναγνώριση ακολουθιών εικόνων,

καθεμιά από τις οποίες αντιστοιχεί σε ένα ή περισσότερα νοήματα. Είναι λογικό ότι το επίπεδο δυσκολίας συστημάτων Αναγνώρισης ΝΓ με επεξεργασία εικόνας, που αποτελεί μέρος της αναγνώρισης, αυξάνει σε σχέση με συστήματα που χρησιμοποιούν data-gloves για την εξαγωγή των χαρακτηριστικών.

Έχουν γίνει πολλές διαφορετικές προσεγγίσεις και έχουν υλοποιηθεί ποικίλα συστήματα αναγνώρισης. Οι Doner και Hagen στο [3] περιγράφουν δύο τμήματα ενός συστήματος αναγνώρισης της Αμερικάνικης ΝΓ: έναν ανιχνευτή χεριού (hand tracker) και έναν ASL-parser. Ο ανιχνευτής χεριού εξάγει από ακολουθίες εικόνων πληροφορίες, που αφορούν στην χειρομορφή, τη θέση και την κίνηση του χεριού. Ο νοηματιστής φοράει ένα ζευγάρι γάντια με χρώματισμένες τις θέσεις των κλειδώσεων και των ακροδακτύλων, για να διευκολυνθεί η διαδικασία αναγνώρισης.

Οι Waldron και Kim στο [34] χρησιμοποιούν Νευρωνικά Δίκτυα Πίσω Διάδοσης (Back-propagation Neural Networks) για την αναγνώριση μεμονωμένων νοημάτων της Αμερικάνικης ΝΓ. Τα στοιχεία που δίνονται ως είσοδοι στο σύστημα αφορούν στη θέση και το σχήμα του χεριού και λαμβάνονται μέσω ενός γαντιού δεδομένων (data-glove). Στο πρώτο επίπεδο του συστήματος υπάρχουν 4 Νευρωνικά Δίκτυα, καθένα από τα οποία αναγνωρίζει μια διαφορετική κατηγορία φωνολογικών στοιχείων της ΑΝΓ. Συνολικά αναγνωρίζονται 36 χειρομορφές, 10 θέσεις χεριού, 11 προσανατολισμού χεριού και 11 κατευθύνσεις κίνησης χεριού, χρησιμοποιώντας το σύστημα του Stokoe's για τα transcriptions [30], χωρίζοντας την χειρομορφή από την γωνία περιστροφής, και την κίνηση του κάθε νοήματος. Τα στοιχεία αυτά συνδυαζόμενα μεταξύ τους προσδιορίζουν συγκεκριμένα νοήματα της ΑΝΓ.

Οι Ong και Bowden στο [21] παρουσίασαν μια πρωτότυπη προσέγγιση για την εκπαίδευση ενός ανιχνευτή χεριών ο οποίος δεν ήταν μόνο ικανός να κάνει ανίχνευση των χεριών στην εικόνα αλλά και να κάνει ταξινόμηση ως προς την χειρομορφή χρησιμοποιώντας ένα -Boosted Classifier Tree-.

Ο Holden στο [15] παρουσιάζει ένα σύστημα αναγνώρισης στατικών και δυναμικών νοημάτων της Αυστραλιανής Νοηματικής Γλώσσας. Για την εξαγωγή χαρακτηριστικών του χεριού χρησιμοποιείται ένα καταλλήλως χρωματισμένο στις αρθρώσεις των δακτύλων γάντι. Σχηματίζοντας, έτσι, ακολουθίες διανυσμάτων χαρακτηριστικών που αντιστοιχούν σε νοήματα. Η ταξινόμηση των ακολουθιών πραγματοποιείται με ένα Προσαρμοζόμενο Ασαφές Έμπειρο Σύστημα.

1.3.1 Συστήματα Αναγνώρισης Νοηματικής Γλώσσας με Μαρκοβιανά μοντέλα

Θα δώσω ιδιαίτερη έμφαση σε συστήματα αναγνώρισης ΝΓ, που βασίζονται στα Κρυφά Μαρκοβιανά Μοντέλα αφού αυτά θα χρησιμοποιηθούν για την μοντελοποίηση της ΕΝΓ στα πλαίσια αυτής της διπλωματικής. Παρακάτω θα κάνω μια ανασκόπηση σχετικών εργασιών.

Ο Bowden και οι συνεργάτες του στο [4] παρουσίασαν μια προσέγγιση για την αναγνώριση νοηματικής γλώσσας χρησιμοποιώντας ένα λεξικό που περιελάμβανε 43 διαφορετικά νοήματα με ποσοστό αναγνώρισης 97,67% χρησιμοποιώντας μικρό αριθμό δεδομένων εκπαίδευσης. Το κλειδί της προσέγγισης αυτής ήταν ότι χρησιμοποίησαν 2 στάδια για την ταξινόμηση (classification) στο πρώτο κάνανε εξαγωγή χαρακτηριστικών για την χειρομορφή και για την κίνηση των χεριών και στο δεύτερο στάδιο χρησιμοποίησαν HMM σε συνδυασμό με Independent Component Analysis για την τελική απόφαση.

Οι Ong και Ranganath στο [22] έκαναν μια γενική επισκόπηση για την εξαγωγή χαρακτηριστικών και τις μεθόδους ταξινόμησης για την νοηματική γλώσσα. Επίσης ασχολήθηκαν με την μοντελοποίηση των μεταβάσεων που εισάγονται κατά την εκτέλεση δυο διαφορετικών νοημάτων στην συνεχή νοηματική γλώσσα και την ανεξαρτησία από τον νοηματιστή κ.α. Ακόμα πρότειναν μελλοντικές προεκτάσεις της έρευνας πάνω στην αναγνώρισης της νοηματική γλώσσα.

Οι Staner και Pentland στο [27] κατέγραψαν εικόνες βίντεο και χρησιμοποίησαν ένα σύστημα HMMs με την ίδια τοπολογία 4 καταστάσεων για όλα τα νοήματα. Ο νοηματιστής φορούσε μονόχρωμα υφασμάτινα γάντια. Στο [28] οι ίδιοι ερευνητές υλοποίησαν και μια δεύτερη κατηγορία πειραμάτων κατά την οποία ο νοηματιστής δε φοράει γάντια. Τα στοιχεία, που χρησιμοποίησαν για την εκπαίδευση των μαρκοβιανών μοντέλων, σχετίζονται με την θέση του χεριού ως προς το σώμα του νοηματιστή και όχι με την ακριβή περιγραφή του σχήματος του χεριού. Το πλήθος των λέξεων που εξετάστηκαν σε όλες τις παραπάνω εργασίες είναι 40 και πέτυχαν αναγνώριση από 92% μέχρι 99%. Ωστόσο χρησιμοποίησαν μια αυστηρά καθορισμένη γραμματική και πολύ απλές συντακτικά προτάσεις.

Οι Bauer,Hienz και Kraiss στο [14] παρουσιάζουν ένα σύστημα αναγνώρισης προτάσεων της Γερμανικής Νοηματικής Γλώσσας (ΓΝΓ). Οι προτάσεις σχηματίστηκαν από ένα λεξιλόγιο 52 νοημάτων και ήταν γραμματικά σωστές,

ακολουθώντας του κανόνες της ΓΝΓ. Στο σύστημα χρησιμοποιήθηκαν HMMs διαφορετικών τοπολογιών. Χρησιμοποίησαν HMMs με μικρό αριθμό καταστάσεων για την μοντελοποίηση μικρών νοημάτων και HMMs με μεγάλο αριθμό καταστάσεων για νοήματα με μεγαλύτερη χρονική διάρκεια. Το σύστημα χρησιμοποίησε ταυτόχρονα ένα στοχαστικό γλωσσικό μοντέλο το οποίο λάμβανε υπόψη του τις πιθανότητες εμφάνισης απλών νοημάτων ή νοημάτων σε ακολουθία και πέτυχε 95% ακρίβεια αναγνώρισης.

Ο Kadous χρησιμοποιώντας Power Gloves πέτυχε αναγνώριση 95 μεμονωμένων νοημάτων με ποσοστό επιτυχίας 80% δίνοντας έμφαση στην ταχύτητα των υπολογιστικών μεθόδων [17] . Ο K. Grobel και M. Assam χρησιμοποίησαν HMMs για την αναγνώριση 262 μεμονωμένων νοημάτων με ποσοστά επιτυχίας 91,3% κάνοντας την εξαγωγή χαρακτηριστικών από νοηματιστές οι οποίοι φορούσαν χρωματιστά γάντια [13].

Οι Wang, Gao, και Ma βασιζόμενοι σε HMMs περιέγραψαν ένα σύστημα αναγνώρισης μεμονωμένων λέξεων της κινέζικης νοηματικής γλώσσας με μεγάλο λεξιλόγιο που περιελάμβανε πάνω από 5000 νοήματα [5]. Χρησιμοποίησαν τεχνικές από την αναγνώριση φωνής όπως clustering Gaussian probabilities και fast matching για να επιτύχουν σε πραγματικό χρόνο (real-time) αναγνώριση με ποσοστά επιτυχίας 95%.

Οι Vogler και Metaxas στα [32, 31, 33] ασχολήθηκαν με την συνεχή νοηματική γλώσσα βασιζόμενοι σε HMMs και χωρίζοντας τα νοήματα σε φωνήματα σε αντιστοιχία με την αναγνώριση φωνής με ένα μοντέλο Movements-Hold. Επίσης χώρισαν τα χαρακτηριστικά σε ξεχωριστά κανάλια πληροφορίας σύμφωνα με το σύστημα του Stokoe's και χρησιμοποίησαν Parallel HMMs για τον συνδιασμό τους. Τα αποτελέσματα ήταν πολύ καλά με συνέπεια τα τελευταία χρόνια τα PaHHMs να έχουν αρχίσει να χρησιμοποιούνται αρκετά.

Στους Πίνακες 1.1 και 1.2 βλέπουμε τα αποτελέσματα όπου έχουμε συνοψίσει από σχετικές έρευνες που έχουν γίνει στην αναγνώριση νοηματικής γλώσσας. Στον πίνακα 1.1 έχουν χρησιμοποιηθεί ηλεκτρομηχανικές μονάδες για την εξαγωγή των χαρακτηριστικών ενώ στον Πίνακα 1.2 έχουν χρησιμοποιηθεί μέθοδοι όρασης υπολογιστών για την επεξεργασία εικόνας.

Όπου οι συντομογραφίες που χρησιμοποιήσαμε φαίνονται παρακάτω:
I/C: isolated ή Continuous signing.
S/B:Singe hand ή Both hands
(a) Μακριά ρούχα

Work	Νοήματα	I/C	S/B	Classification methods	Rec %
Fang [10]	208 CSL	C	B	HMM, SOFM	92.1
Kadous [17]	95 ASL	I	S	Instance-based lerning Decision tree learning	80
Wang [5]	5119 CSL	I	B	HMM model sequential subunits	92.8
Wang [5]	5119 CSL	C	B	HMM model sequential subunits	86.2

Πίνακας 1.1: Αποτελέσματα προηγούμενων ερευνών Αναγνώρισης ΝΓ με εξαγωγή χαρακτηριστικών χρησιμοποιώντας ηλεκτρομηχανικές μονάδες πχ DataGloves

Work	Νοήματα	I/C	S/B	Classification methods	Restrictions Features	Rec %
Assan [1]	262 NSL	I	B	HMM	a,b,c | 2D moment-based	91.3
Assan [1]	262 NSL	C	B	HMM	a,b,c | 2D moment-based	91.3
Bauer [2]	100 GSL	I	S	HMM	a,b,c | 2D moment-based	92.5
Cui [7]	28 ASL	I	S	PCA+MDA	d,e,f,h,i | 2D segmented handposition	93.2
Starner [29]	40 ASL	C	B	HMM	a,d,e,g | 2D moment-based	92-98
Vogler [31]	22 ASL	C	B	HMM + PaHMM		95.5

Πίνακας 1.2: Αποτελέσματα προηγούμενων ερευνών Αναγνώρισης ΝΓ με εξαγωγή χαρακτηριστικών χρησιμοποιώντας μεθόδους Όρασης Υπολογιστών

(b) Χρωματιστά Γάντια

(c) Ομοιόμορφο background

(d) Complex αλλά σταθερό background

(e) Κεφάλι/πρόσωπο σταθερά ή πολύ μικρή κίνηση σε σχέση με τα χέρια

(d) Διαρκείς κίνηση των χεριών

(g) Συγκεκριμένη θέση του σώματος και στάση ή συγκεκριμένη αρχική θέση των χεριών

(h) Το αριστερό χέρι ή/και το κεφάλι βρίσκονται εκτός της εικόνας

(i) Περιορισμένο λεξιλόγιο έτσι ώστε να μην έχουμε επικάλυψη χεριών και προσώπου

Αξίζει να σημειωθεί ότι τα αποτελέσματα στους πίνακες είναι ενδεικτικά και ότι δεν έχουν νόημα οι συγκρίσεις τη στιγμή που δεν έχουν γίνει πειράματα στην ίδια βάση δεδομένων.

1.4 Ερευνητικοί Στόχοι

Η παρούσα διπλωματική εργασία ασχολείται με την κατασκευή ενός ολοκληρωμένου συστήματος αναγνώρισης μεμονωμένων νοημάτων της ελληνικής νοηματικής γλώσσας βασιζόμενο σε οπτικά χαρακτηριστικά που έχουν προκύψει από την επεξεργασία εικόνας με μεθόδους της Όρασης Υπολογιστών. Συνολικά λοιπόν η έρευνα μας επικεντρώθηκε στα παρακάτω σημεία:

- Μοντελοποίηση της Νοηματικής Γλώσσας (Κεφάλαιο 2)
- Εκπαίδευση των μοντέλων χρησιμοποιώντας κάθε κανάλι πληροφορίας (stream) ξεχωριστά (Κεφάλαιο 4)
- Τελικός συνδυασμός των καναλιών (streams fusion) (Κεφάλαιο 5)

Μοντελοποίηση της Νοηματικής Γλώσσας

Δημιουργήσαμε ένα μοντέλο για κάθε νόημα (Sign - level modeling). Κάθε μοντέλο το χωρίσαμε σε τέσσερα διαφορετικά κανάλια πληροφορίας. Το πρώτο κανάλι αναφέρεται στην θέση - κίνηση του πρωτεύων (Strong) χεριού που είναι το χέρι το οποίο εκτελεί τα νοήματα που εκτελούνται με το ένα χέρι και συνήθως είναι το δεξί. Το δεύτερο κανάλι αναφέρεται στην χειρομορφή του πρωτεύων (Strong) χεριού. Το τρίτο κανάλι αναφέρεται στην θέση - κίνηση του δευτερεύων (Weak) χεριού που είναι το αντίθετο χέρι από το strong και το τέταρτο κανάλι αναφέρεται στην χειρομορφή του δευτερεύων (Weak) χεριού.

Εκπαίδευση κάθε καναλιού (stream) ξεχωριστά

Για την εκπαίδευση των μοντέλων χρησιμοποιώντας κάθε κανάλι πληροφορίας (stream) ξεχωριστά χρησιμοποιήθηκαν απλά HMM μοντέλα. Η επιλογή των HMM βασίστηκε πρώτον στην ανάγκη τα μοντέλα μας να είναι ανεπηρέαστα από την μεταβλητότητα που εμφανίζεται στην εκτέλεση νοημάτων. Δεύτερον στην ικανότητα των HMM να αναγνωρίζουν ακολουθίες δεδομένων και τρίτον στην δυνατότητα που έχουν να παραμένουν στην ίδια κατάσταση

για περισσότερα του ενός χρονικά διαστήματα, με αποτέλεσμα να έχουν την δυνατότητα να αναγνωρίζουν διαφορετικού μήκους ακολουθίες εικόνων, οι οποίες αντιστοιχούν στην ίδια λέξη.

Για το κανάλι της θέσης - κίνησης χρησιμοποιήσαμε ένα left-right HMM 6 καταστάσεων ενώ για το κανάλι της χειρομορφής χρησιμοποιήσαμε ένα left-right HMM 5 καταστάσεων.

Τελικός συνδυασμός των καναλιών (streams fusion)

Για τον συνδυασμό των καναλιών (streams fusion) της θέσης - κίνησης και της χειρομορφής των χεριών χρησιμοποιήσαμε δύο διαφορετικές επεκτάσεις των HMM τα Parallel HMM (PaHMM) και τα Product HMM (PHMM). Τα καλύτερα αποτελέσματα τα είχαμε με τα PaHMM όμως και τα αποτελέσματα χρησιμοποιώντας τα PHMM ήταν πολύ ενθαρρυντικά για περισσότερες λεπτομέρειες βλ. Κεφάλαιο 5.

Κεφάλαιο 2

Μοντελοποίηση Ελληνικής Νοηματικής Γλώσσας

Στόχος αυτής της διπλωματικής είναι η μοντελοποίηση της ελληνικής νοηματικής γλώσσας (ΕΝΓ) με σκοπό την αναγνώριση μεμονωμένων λέξεων. Σε αυτό το κεφάλαιο θα κάνω μια σύντομη περιγραφή της Νοηματικής Γλωσσάς και θα αναφερθώ σε μερικές μοντελοποιήσεις που χρησιμοποιήθηκαν στο παρελθόν όπως το σύστημα του Stokoe [30], το μοντέλο Movement-Hold που περιγράφτηκε από τους Vogler,Metaxa [31] και το Fenomic μοντέλο που περιγράφτηκε από τους B.Bauer και K-F.Kraiss [2]. Επίσης θα αναφέρω την μοντελοποίηση που χρησιμοποίησα στα πλαίσια αυτής της διπλωματικής όπως και τα προβλήματα που συνάντησα.

2.1 Βασικά

Η ΕΝΓ χρησιμοποιείται σαν πρωταρχικός τρόπος επικοινωνίας από ένα μεγάλο αριθμό ανθρώπων, οι οποίοι έχουν προβλήματα ακοής ή ομιλίας. Τα γλωσσικά μέσα που χρησιμοποιεί η ΕΝΓ (όπως και οι άλλες νοηματικές γλώσσες) για να διατυπώσει τις έννοιες και για να δημιουργήσει μορφολογία και σύνταξη, βασίζονται στην κίνηση των χεριών, στην στάση ή στην κίνηση του σώματος, και στην έκφραση του προσώπου. Οι βασικές μονάδες του λόγου (τις οποίες η επιστήμη της γλωσσολογίας ονομάζει γλωσσικά σημεία) της ΕΝΓ ονομάζονται νοήματα. Τα νοήματα μπορούν να έχουν λεξική ή γραμματική σημασία, ακριβώς όπως τα μορφήματα και οι λέξεις στις φυσικές γλώσσες.

Τα νοήματα δεν πρέπει να συγχέονται με το δακτυλικό αλφάβητο, το οποίο είναι απλώς ένας τρόπος μεταγραφής του ελληνικού αλφαβήτου. Οι νοηματιστές, ως φυσικοί ομιλητές της ΕΝΓ, χρησιμοποιούν το δακτυλικό αλ-

φάβητο είτε για να αποδώσουν τα ακρώνυμα και τα κύρια ονόματα, είτε για να σχηματίσουν νοήματα στα οποία τα στοιχεία του δακτυλικού αλφαβήτου χρησιμοποιούνται ως χειρομορφές. Για παράδειγμα, το νόημα που σημαίνει ¨κοινωνία¨ σχηματίζεται από το ¨κ¨ του δακτυλικού αλφαβήτου σε συνδυασμό με κίνηση.

Το χαρακτηριστικότερο συστατικό ενός νοήματος ονομάζεται χειρομορφή. Η χειρομορφή είναι το σχήμα που παίρνει η παλάμη και η θέση στην οποία τοποθετούνται τα δάκτυλα τη στιγμή που αρχίζει να σχηματίζεται ένα νόημα. Η ίδια η χειρομορφή όμως από μόνη της δεν αποτελεί φορέας σημασίας.

Για να αποκτήσει σημασία, για να δημιουργηθεί δηλαδή ένα νόημα, η χειρομορφή πρέπει να συνοδεύεται και από τα παρακάτω στοιχεία [36], [37]:

- Τον ¨προσανατολισμό¨ της παλάμης, δηλαδή την κατεύθυνση προς την οποία στρέφεται η χειρομορφή κατά το σχηματισμό του νοήματος: ο δείκτης που δείχνει προς τα πάνω ή στρέφεται προς τα δεξιά αποτελεί τμήμα διαφορετικών νοημάτων.

- Τη θέση της χειρομορφής στο χώρο ή επάνω στο σώμα: τα νοήματα παράγονται σε καθορισμένο χώρο που λέγεται χώρος νοηματισμού. Ο χώρος αυτός αντιστοιχεί περίπου σε ένα τετράγωνο που ορίζεται από την κορυφή της κεφαλής ως τον άνω κορμό και εκτείνεται σε 20-30 εκατοστά δεξιά και αριστερά από τα μπράτσα. Αν χρησιμοποιήσουμε μία χειρομορφή έξω από το χώρο αυτό, π.χ. με τα μπράτσα κρεμασμένα δίπλα στο σώμα, το αποτέλεσμα δεν είναι αναγνωρίσιμο ως νόημα.

- Την κίνηση του χεριού, χωρίς την οποία δεν μπορεί να ολοκληρωθεί ένα νόημα: ο δείκτης που δείχνει προς τα πάνω ή στρέφεται προς τα δεξιά χωρίς να κινείται δεν είναι ολοκληρωμένο νόημα, δεν αντιστοιχεί δηλαδή σε ορισμένη σημασία. Εκτός από τη συμμετοχή της στο σχηματισμό του νοήματος, η κίνηση μπορεί να είναι και φορέας άλλων σημασιών, για παράδειγμα να δηλώνει τον αριθμό (ενικό ή πληθυντικό), το μέγεθος ενός αντικειμένου (μικρότερο ή μεγαλύτερο), ακόμα και τη συχνότητα μίας ενέργειας.

- Την στάση (ή κίνηση) του σώματος και/ή την έκφραση του προσώπου, που αποτελούν επίσης συστατικά του νοήματος με την έννοια ότι λειτουργούν για να μεταφέρουν πληροφορία όπως αυτή που δηλώνεται από τον τόνο της φωνής στις ομιλούμενες γλώσσες. Για παράδειγμα, η έννοια του μέλλοντος διατυπώνεται στην ΕΝΓ συνδυάζοντας το νόημα με μία ελαφρά κλίση του σώματος προς τα εμπρός.

Η πλειοψηφία των νοημάτων εκτελούνται με το ένα χέρι αλλά υπάρχουν και πολλά που χρησιμοποιούν και τα δύο. Είναι πολύ σημαντικό να κάνουμε ένα διαχωρισμό των δύο χέριων έτσι ονομάζουμε -strong hand - το χέρι το οποίο εκτελεί τα νοήματα που εκτελούνται με ένα χέρι, και παρέχει την μεγαλύτερη πληροφορία σε νοήματα που εκτελούνται με δύο χέρια και το αντίθετο χέρι το ονομάζουμε -weak hand-. Σε ένα συνηθισμένο νοηματιστή

το δεξί χέρι είναι συνήθως το -strong hand - και το αριστερό το -weak hand-.

Η βασική διαφορά μεταξύ της Νοηματικής Γλώσσας (ΝΓ) από της καθομιλουμένης είναι στον τρόπο εκφοράς των κλίσεων (διαφορετική πτώση, κλίση ρημάτων, αριθμό (ενικό ή πληθυντικό)) της κάθε λέξης. Για να εκφέρουμε στην καθομιλουμένη την λέξη πχ διάβαζα απλά πρέπει να αλλάξουμε την κατάληξη της λέξης διαβάζω, έτσι παρατηρούμε ότι οι κλίσεις της κάθε λέξης στην καθομιλουμένη εκφέρονται αρκεί να αλλάξουμε την κατάληξη ή να κάνουμε μια μικρή διαφοροποίηση (σειριακή) της πρωταρχικής λέξης. Αντίθετα στην Νοηματική Γλώσσα το επιτυγχάνουμε είτε παράλληλα με την εκτέλεση της πρωταρχικής λέξης μέσω των εκφράσεων του προσώπου, της κίνησης του υπόλοιπου σώματος, το πόσες φορές θα εκτελέσουμε το νόημα κ.α. είτε σειριακά χρησιμοποιώντας διαφορετική αρχική ή τελική θέση.

Η ύπαρξη των παράλληλων διαδικασιών στην ΕΝΓ κάνει το πρόβλημα της αναγνώρισης πολύ πιο δύσκολο από την Αναγνώριση Φωνής όπου η εκφορά μιας λέξης γίνεται σειριακά. Ο λόγος είναι ότι δεδομένου ενός σήματος που αντιπροσωπεύει μια εκφώνηση οι παράλληλες διαδικασίες είναι δυσκολότερο να μοντελοποιηθούν αναφορικά με την ξεχωριστή συνεισφορά της κάθε μιας από ότι οι σειριακές.

Στο παρελθόν έχουν γίνει πολλές προσεγγίσεις για την δημιουργία μοντέλων για κάθε νόημα ξεχωριστά -whole-sign modeling - . Όμως επειδή κάθε νόημα μπορεί να εμφανιστεί με πολλούς διαφορετικούς τρόπους (διαφορετική πτώση, κλίση, αριθμό (ενικό ή πληθυντικό)) όπως αναφέραμε και παραπάνω για κάθε ένα από αυτά θα πρέπει να δημιουργήσουμε ένα μοντέλο, με αποτέλεσμα εάν σκεφτούμε ότι έχουμε ένα λεξικό 5000 λέξεων τότε ο αριθμός των μοντέλων που πρέπει να δημιουργήσουμε γίνεται πάρα πολύ μεγάλος. Η εκπαίδευση ενός τόσο μεγάλου αριθμού μοντέλων χρειάζεται μια τεράστια βάση δεδομένων με πολλές εκφωνήσεις από την κάθε διαφορετική εμφάνιση του κάθε νοήματος το οποίο είναι πρακτικά αδύνατο. Έτσι ξεκίνησαν να γίνονται κάποιες προσπάθειες να σπάσουν τα νοήματα σε φωνήματα σε αντιστοιχία με την αναγνώριση φωνής και σε διαφορετικά κανάλια πληροφορίας των παράλληλων διαδικασιών για την πιο εύκολη μοντελοποίηση. Παρακάτω θα αναφερθώ εκτενέστερα σε αυτούς τους τρόπους.

2.2 Μοντέλο του Stokoe

Ο W. Stokoe [30] ήταν ο πρώτος που σκέφτηκε να σπάσει τα νοήματα σε μικρότερα κομμάτια και να χρησιμοποιήσει αυτή την παρατήρηση για την μοντελοποίηση της ΝΓ [30]. Θεώρησε ότι κάθε νόημα μπορεί να χαρακτηριστεί από τρεις παραμέτρους, την θέση (tabula or tab), την κίνηση (signation or sig) και την χειρομορφή (designator or dez). Μοντελοποιώντας τις παρα-

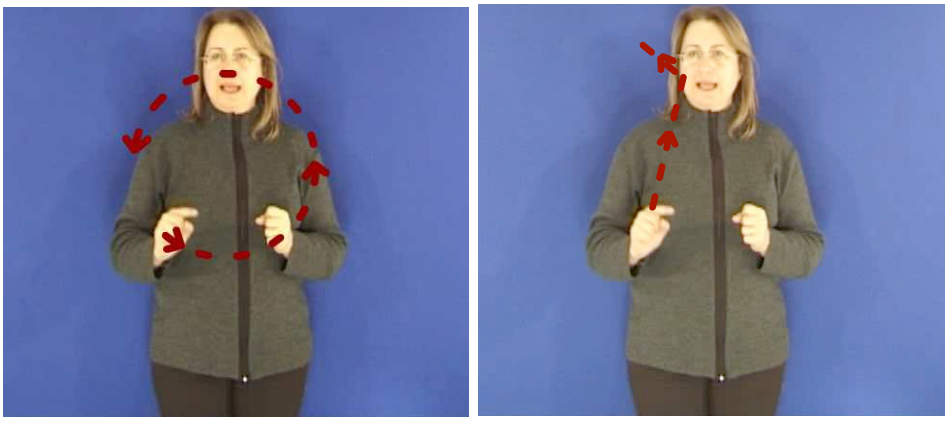
μέτρους αυτές για όλες τις πιθανές τιμές που μπορούν να πάρουν ουσιαστικά επιτυγχάνουμε την μοντελοποίηση κάθε νοήματος και κατά συνέπεια της ΝΓ. Έτσι γνωρίζοντας τις τρεις παραμέτρους μπορούμε να αναγνωρίσουμε ποιο νόημα εκτελείται κάθε φορά. Στο Σχήμα 2.1 μπορούμε να δούμε ένα παράδειγμα των παραμέτρων αυτών.

Η μοντελοποίηση αυτή παρουσιάζει ένα μεγάλο πρόβλημα. Επειδή έχει γίνει η θεώρηση ότι οι παράμετροι αυτοί παραμένουν σταθερές καθ όλη την διάρκεια εκτέλεσης ενός νοήματος τυχόν αλλαγές σε μια από τις παραμέτρους δεν λαμβάνεται υπόψιν. Έτσι ως αποτέλεσμα μπορεί να μην γίνει πχ διαχωρισμός σε νοήματα τα οποία έχουν ίδια θέση, χειρομορφή και κίνηση αλλά για το ένα εκτελείται κίνηση μια φορά ενώ για το άλλο εκτελείται η ίδια κίνηση δυο ή περισσότερες φορές. Ένα παράδειγμα φαίνεται στο Σχήμα 2.2. Το παραπάνω πρόβλημα είναι πολύ σημαντικό και πρέπει να ληφθεί υπόψη στην μοντελοποίηση αφού δημιουργεί πολλά προβλήματα στην αναγνώριση.

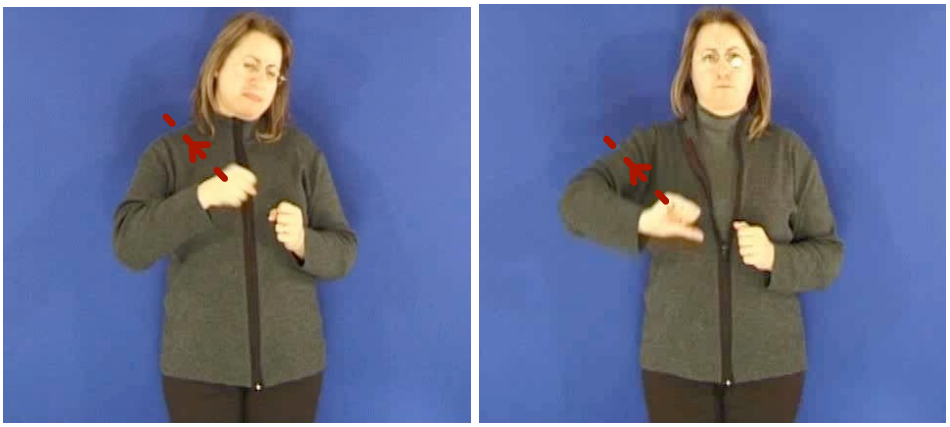
2.3 Movement Hold Μοντέλο

Οι Vogler και Metaxas [31] σε αντίθεση με τον Stokoe προσπάθησαν να δώσουν λύση στο παραπάνω πρόβλημα δημιουργώντας ένα νέο μοντέλο το Movement-Hold όπου όρισαν δυο διαφορετικές κλάσεις τις οποίες ονόμασαν movements και holds. Η κλάση movements προσδιορίζει τις χρονικές στιγμές όπου κάποιο από τα χαρακτηριστικά (στοιχεία) του νοήματος αλλάζει πχ αλλαγή της χειρομορφής, αλλαγή της κίνησης των χεριών, αλλαγή του προσανατολισμού της παλάμης. Η κλάση holds προσδιορίζει τις χρονικές στιγμές όπου όλα τα χαρακτηριστικά (στοιχεία) του νοήματος παραμένουν αμετάβλητα για σημαντικό χρονικό διάστημα.

Κάθε νόημα μπορεί να περιγραφεί από μια ακολουθία των κλάσεων αυτων. Συνηθισμένες ακολουθίες που χρησιμοποιούνται είναι ΗΜΗ (ξεκινάμε από μια στάση -hold- κάνουμε μια κίνηση -movement- και τελικά καταλήγουμε σε μια στάση -hold-) ένα παράδειγμα βλέπουμε στο Σχήμα 2.3, ΜΗ (ξεκινάμε κάνοντας μια κίνηση -movement- και καταλήγουμε σε μια στάση -hold-) ένα παράδειγμα βλέπουμε στο Σχήμα 2.4, ΜΜΜΗ (κάνουμε 3 συνεχόμενες διαφορετικές κινήσεις -movement- και καταλήγουμε σε μια στάση -hold-). Οι χρονικές στιγμές που αντιπροσωπεύονται από την κλάση -movement- περιέχουν χαρακτηριστικά που προσδιορίζουν τον τύπο της κίνησης (κυκλική,γραμμική,υπό γωνία κτλ). Επίσης και οι δυο κλάσεις παρέχουν πληροφορίες για την χειρομορφή, την θέση των χεριών και τον προσανατολισμό της παλάμης όπως μπορούμε να δούμε στο Σχήμα 2.5. Ένα μεγάλο αρνητικό της παραπάνω μοντελοποίησης είναι η δυσκολία κατασκευής των μεταγραφών (transcriptions). Δεν υπάρχει αυτοματοποιημένος τρόπος για



(α΄) Κύκλος (β΄) Έχω ιδέα



(γ΄) Διαφωνώ (δ΄) Πρώτο

Σχήμα 2.1: Στις εικόνες α,β βλέπουμε την εκτέλεση δύο διαφορετικών νοημάτων όπου έχουν κοινή θέση και χειρομορφή αλλά διαφέρουν στην κίνηση και στις εικόνες γ,δ βλέπουμε την εκτέλεση δύο διαφορετικών νοημάτων που διαφέρουν στην χειρομορφή ενώ έχουν κοινή θέση και κίνηση.

(α') Θέλω (β') Θέλω πολύ

Σχήμα 2.2: Ένα παράδειγμα για την αδυναμία του Stokoe's μοντέλου να διαχωρίσει τα νοήματα στις εικόνες α,β , ενώ η διαφορά μεταξύ μιας μόνο κίνησης (εικόνα α) και της επαναλαμβανόμενης (εικόνα β) είναι εμφανής .

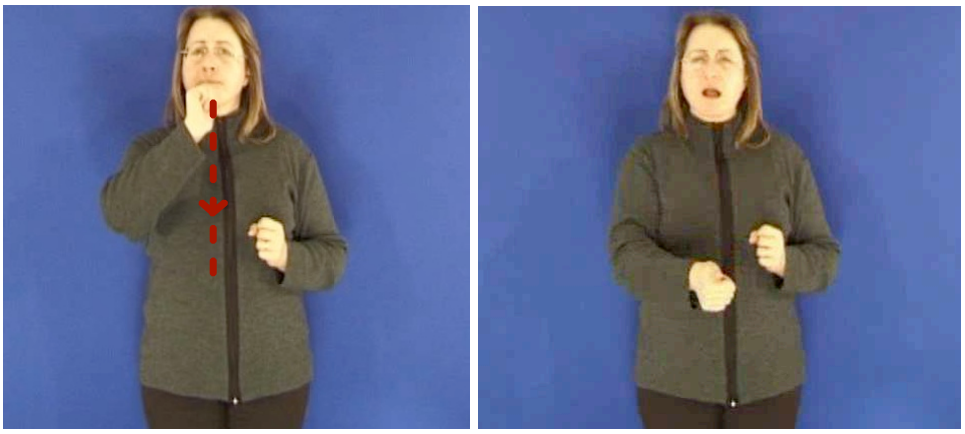
τον προσδιορισμό των -transcriptions- με αποτέλεσμα να πρέπει να γίνει με το χέρι έτσι εάν έχουμε ένα μεγάλο λεξιλόγιο τότε αυτό γίνεται αδύνατον. Επίσης δεν παρέχει πληροφορία τόσο για τις εκφράσεις του προσώπου όσο και για την κίνηση του υπόλοιπου σώματος με αποτέλεσμα ο προσδιορισμός της γραμματικής, μπορεί να επιτευχθεί μόνο από τα συμφραζόμενα.

2.4 Fenomic Μοντέλο

Το fenomic μοντέλο [2], δημιουργείται αυτόματα από τα δεδομένα που έχουμε και δεν βασίζεται σε καμία φωνητική έννοια χαρακτηριστικά που το καθιστούν ιδανικό για την μοντελοποίηση της Νοηματικής Γλώσσας. Επίσης έχει την δυνατότητα να μοντελοποιεί νέες λέξεις που δεν εμπεριείχε το λεξικό μέχρι εκείνη την στιγμή.

2.4.1 Αλγόριθμος K-means

Στόχος του είναι να χωρίσει τα υπάρχοντα διανύσματα χαρακτηριστικών σε ένα set K κλάσεων. Η λειτουργία του είναι η εξής: Υπολογίζει για κάθε διάνυσμα χαρακτηριστικών την απόσταση του από τα K centroids και το ταξινομεί στην κλάση με την μικρότερη απόσταση. Αφού ταξινομήσει όλα τα διανύσματα χαρακτηριστικών κάνει ανανέωση του centroid της κάθε κλάσης υπολογίζοντας τον μέσο όρο των διανυσμάτων χαρακτηριστικών που ανήκουν

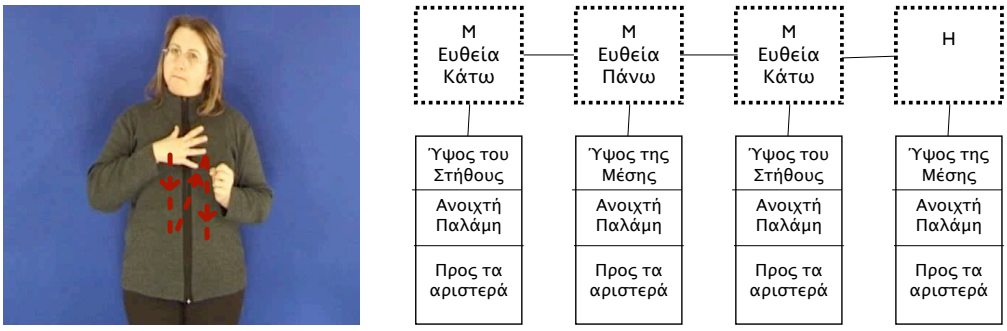


(α') Μαύρο, αρχή (β') Μαύρο, τέλος

Σχήμα 2.3: ΗΜΗ ακολουθία, το νόημα μαύρο αποτελείται από μία στάση -hold- του strong χεριού πάνω στο πιγούνι ακολουθούμενη από μια κίνηση -movement- προς τα κάτω (βλ. εικόνα α) και τέλος από μια στάση -hold- στο ύψος του στήθους (βλ. εικόνα β).

(α') Συμφωνώ, αρχή (β') Συμφωνώ, τέλος

Σχήμα 2.4: ΜΗ ακολουθία, το νόημα Συμφωνώ αποτελείται από μια κίνηση -movement- του strong χεριού προς τα κάτω (βλ. εικόνα α) και από μια στάση -hold- στο ύψος του στήθους (βλ. εικόνα β).



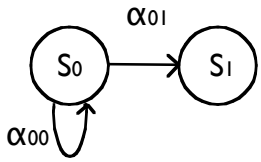
Σχήμα 2.5: Διαγραμματική περιγραφή του νοήματος Θέλω πολύ χρησιμοποιώντας το μοντέλο Movement-Hold. Αποτελείται από μα ακολουθία ΜΜΜΗ δηλαδή τρεις κινήσεις -movements- και μια στάση -hold-.

στην ίδια κλάση και επαναλαμβάνει την ίδια διαδικασία μέχρι η αλλαγή των centroids να είναι αμελητέα. Η αρχικοποίηση των centroids γίνεται τυχαία.

2.4.2 Κατασκευή του Fenomic Μοντέλου

Αρχικά παίρνουμε για όλα τα νοήματα που έχουμε στο λεξικό μας όλα τα διανύσματα χαρακτηριστικών και εκτελούμε τον αλγόριθμο K-means. Μετά από την εκτέλεση του K-means έχουμε χωρίσει τα διανύσματα χαρακτηριστικών μας σε K κλάσεις (τον αριθμό K τον έχουμε ορίσει εμείς) όπου κατά κάποιο τρόπο τα χαρακτηριστικά σε κάθε κλάση είναι όμοια. Έτσι λαμβάνοντας υπόψη τις ομοιότητες αυτές ορίζουμε κάθε κλάση σαν ένα fenomic φώνημα. Η κατασκευή του μοντέλου για κάθε λέξη που ανήκει στο λεξικό γίνεται με τα fenomic φωνήματα που ορίσαμε προηγουμένος. Μετά την εφαρμογή του αλγόριθμου K-means έχει δημιουργηθεί ένα codebook που περιέχει για κάθε διάνυσμα χαρακτηριστικών για όλα τα νοήματα σε ποία κλάση ανήκει το καθένα με αποτέλεσμα να μπορούμε να βρούμε την ακολουθία των fenomic φωνημάτων για κάθε νόημα. Στο Σχημα 2.6 φαίνεται ένα παράδειγμα fenomic μοντελοποίησης.

Σχήμα 2.6: Διαγραμματική περιγραφή της fenomic μοντελοποίησης για ένα λεξιλόγιο τριών λέξεων, Συμφωνώ, Κύκλος, Θέλω.



Σχήμα 2.7: Πιθανό μοντέλο ενός fenomic HMM

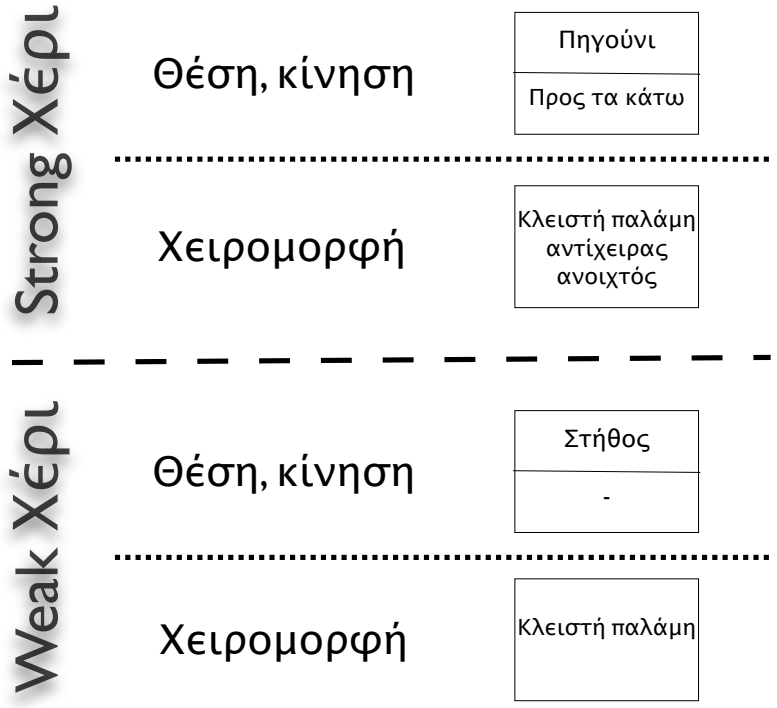
2.5 Μοντελοποίηση

Στόχος αυτής της διπλωματικής είναι βασικά η αναγνώριση μεμονωμένων νοημάτων της ελληνικής νοηματικής γλώσσας. Το λεξικό που χρησιμοποίησα περιελάμβανε 100 διαφορετικά νοήματα έτσι δημιούργησα 100 διαφορετικά μοντέλα, ένα για κάθε νόημα.

Η μοντελοποίηση αυτή όπως λέγεται -whole-word modeling- έχει το πλεονέκτημα ότι η δημιουργία των transcriptions είναι αρκετά εύκολη αφού υπάρχει αντιστοιχία ένα προς ένα (για κάθε νόημα ένα μοντέλο). Όμως έχει και σημαντικά μειονεκτήματα όπως ο αριθμός των μοντέλων για την μοντελοποίηση είναι ανάλογος του εύρους του λεξικού με αποτέλεσμα όταν έχουμε ένα πολύ μεγάλο λεξικό καθιστά αδύνατη την μοντελοποίηση αφού εκτός του ότι ο αριθμός των μοντέλων αυξάνεται δραματικά, για την εκπαίδευση των μοντέλων αυτών χρειάζεται μια τεράστια βάση δεδομένων με πολλές εκφωνήσεις από το κάθε νόημα κάτι που είναι πρακτικά δύσκολο. Στην περίπτωση μας όμως επειδή το λεξικό μας έχει μόνο 100 λέξεις και η βάση δεδομένων που έχουμε είναι αρκετά μεγάλη η μοντελοποίηση αυτή είναι εφικτή.

Πιο συγκεκριμένα κάθε μοντέλο συμπεριλαμβάνει 4 διαφορετικά κανάλια πληροφορίας. Το πρώτο κανάλι αναφέρεται στην κίνηση και θέση του Strong Hand, το δεύτερο στο είδος της χειρομορφής που χρησιμοποιήθηκε από το Strong Hand, το τρίτο στην κίνηση και θέση του Weak Hand και το τέταρτο στο είδος της χειρομορφής που χρησιμοποιήθηκε από το Weak Hand. Ένα παράδειγμα μπορούμε να δούμε στο Σχημα 2.8.

Η μοντελοποίηση κάθε καναλιού έγινε ανεξάρτητα από όλα τα υπόλοιπα βασιζόμενοι στην υπόθεση ότι τα κανάλια είναι τελείως ανεξάρτητα μεταξύ τους. Στα παρακάτω κεφάλαια θα αναφερθώ εκτενέστερα για κάθε ένα από τα παραπάνω κανάλια.



Σχήμα 2.8: Για το νόημα Μαύρο βλέπουμε την μοντελοποίηση των τεσσάρων διαφορετικών και ανεξάρτητων καναλιών

2.6 Προβλήματα κατά την μοντελοποίηση

Μερικά πολύ σημαντικά προβλήματα που πρέπει να ληφθούν υπόψη στην μοντελοποίηση είναι τα εξής:

-Όταν εκτελείται ένα νόημα, ο προσανατολισμός της παλάμης αλλάζει συνεχώς με αποτέλεσμα εάν χρησιμοποιούμε μια κάμερα για την καταγραφή του νοήματος πολλές φορές είναι αδύνατον να προσδιορίσουμε τον τύπο της χειρομορφής λόγο της δισδιάστατης προβολής πχ εάν εκτελείται ένα νόημα με χειρομορφή όπου όλα τα δάκτυλα είναι κλειστά εκτός από τον δείκτη ο οποίος είναι ανοιχτός τότε εάν ο προσανατολισμός της παλάμης είναι προς την κάμερα (δηλαδή ο δείκτης δείχνει την κάμερα) τότε είναι αδύνατον να προσδιορίσουμε την χειρομορφή αφού η δισδιάστατη προβολή της ισοδύναμη με την προβολή της χειρομορφής όπου όλα τα δάκτυλα είναι κλειστά όπως μπορούμε να δούμε και στο Σχήμα 2.9.

-Στην Νοηματική Γλώσσα έχουμε πολλούς διαφορετικούς τύπους κινήσεων. Κινήσεις που αλλάζει θέση το χέρι του νοηματιστή που η καταγραφή των οποίων είναι αρκετά εύκολη αφού η κίνηση είναι μεγάλη. Τοπικές κινήσεις όπως πχ αλλαγή του προσανατολισμού τις παλάμης ή κίνηση μόνο των δακτύλων του νοηματιστή όπου η καταγραφή τους είναι αρκετά δύσκολη αφού η κίνηση είναι πολύ μικρή. Αυτό δημιουργεί ένα ιδιαίτερο πρόβλημα αφού θέλουμε από την μια να μπορούμε να καταγράψουμε μεγάλες κινήσεις αλλά από την άλλη να μην αγνοούμε τις μικρές τοπικές κινήσεις. Ένα παράδειγμα μπορούμε να δούμε στο Σχήμα 2.10.

-Κατά την διάρκεια εκτέλεσης ενός νοήματος πολλές φορές έχουμε επικάλυψη μεταξύ των χεριών ή του προσώπου του νοηματιστή με αποτέλεσμα να είναι αδύνατον εάν χρησιμοποιούμε μια μόνο κάμερα να προσδιορίσουμε τον τύπο τις χειρομορφής. Μερικά παραδείγματα μπορούμε να δούμε στο Σχήμα 2.11.

(α΄) Κύκλος, αρχή (β΄) Κύκλος, τέλος

Σχήμα 2.9: Στις εικόνες α,β έχουμε δύο frames από την εκτέλεση του νοήματος Κύκλος όπου έχουμε και στις δύο την ίδια χειρομορφή, βλέπουμε ότι είναι αδύνατος ο σωστός προσδιορισμός της χειρομορφής στην εικόνα β.

(α΄) Βιβλίο, αρχή (β΄) Βιβλίο, ενδιάμεσο σημείο

Σχήμα 2.10: Βλέπουμε μια τοπική κίνηση με την περιστροφή του καρπού για την εκτέλεση του νοήματος βιβλίο.

(α') Επικάλυψη χεριού-κεφαλιού

(β') Επικάλυψη χεριών

(γ') Ολική Επικάλυψη

Σχήμα 2.11: Διαφόρων ειδών επικαλύψεις [35]

Κεφάλαιο 3

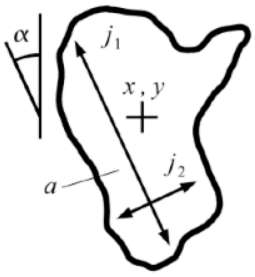
Εξαγωγή Οπτικών Χαρακτηριστικών για Αναγνώριση Νοηματικής Γλώσσας

Το σχήμα είναι το πιο σημαντικό χαρακτηριστικό για την αναγνώριση ενός αντικειμένου. Δύο αντικείμενα για να θεωρηθεί ότι έχουν το ίδιο σχήμα θα πρέπει να μπορεί το ένα να προκύψει από το άλλο με ένα κατάλληλο αριθμό μετατοπίσεων, περιστροφών και ομοιόμορφες αλλαγές κλίμακας [38]. Έτσι το σχήμα είναι όλη η γεωμετρική πληροφορία για το αντικείμενο που παραμένει αμετάβλητη από την θέση, την περιστροφή και την αλλαγή κλίμακας.

Σε αυτό το κεφάλαιο θα κάνουμε μια σύντομη αναφορά στα χαρακτηριστικά που χρησιμοποιήσαμε για την περιγραφή του σχήματος (χειρομορφή) και της θέσης των χεριών στην αναγνώριση που έγινε σε συνεργασία με την Ο. Διαμαντή [35], [8] .

Τα χαρακτηριστικά που χρησιμοποιήσαμε μπορούμε να τα χωρίσουμε στα παρακάτω υποσύνολα:

- Συντεταγμένες του centroid του κεφαλιού πάνω στην εικόνα.
- Συντεταγμένες των centroids των χεριών πάνω στην εικόνα.
- Region-Based χαρακτηριστικά για το σχήμα της χειρομορφής.
- Fourier Descriptors ως προς το σχήμα της περιβάλλουσα
- Moments χαρακτηριστικά για το σχήμα της χειρομορφής.
- Συντελεστές Cepstrum της καμπυλότητας.



$$r = \frac{j_1}{j_2}$$

$$c = \frac{4\pi\alpha}{\pi\epsilon\rho\iota\mu\epsilon\tau\rho\sigma\varsigma^2}$$

$$e = \frac{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}}{\alpha}$$

μ_{pq} : κεντρικές ροπές

Σχήμα 3.1: Τα Region-Base χαρακτηριστικά σχήματος [35].

3.1 Ανάλυση Χαρακτηριστικών

3.1.1 Region-Based Χαρακτηριστικά

Τα Region-Based χαρακτηριστικά βασίζονται στις κεντρικές ροπές και χρησιμοποιούνται συχνά για την ταξινόμηση σχημάτων, επειδή επιτυγχάνουν πολύ μεγάλη συμπίεση, αφού ένας πολύ μικρός αριθμός αυτών 2-3 αρκούν για να περιγράψουν επαρκώς το σχήμα. Οι ποσότητες αυτές παρουσιάζονται στο Σχήμα 3.1.

- Κέντρο μάζας (x, y) : Δίνει μια εκτίμηση για τη θέση του σχήματος στην εικόνα, τα x, y μετρώνται σε pixels.

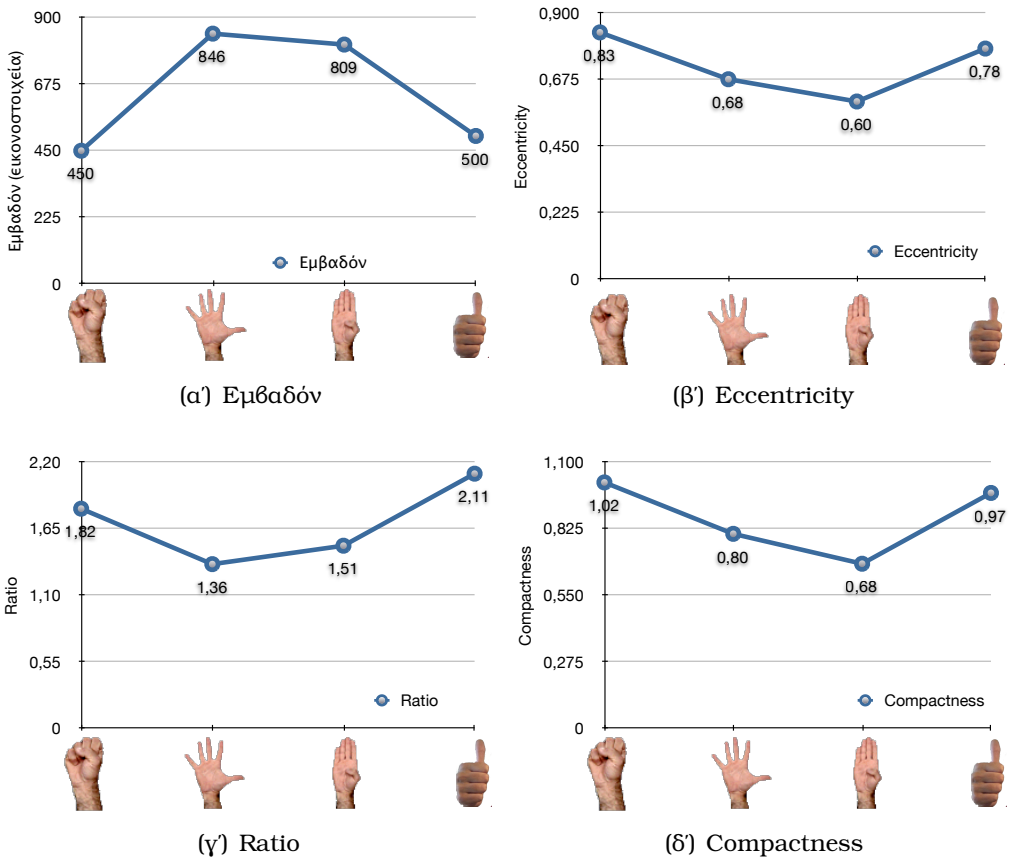
- Εμβαδόν a : Προσεγγιστικά ο αριθμός των pixels που ανήκουν στην περιοχή.

- Εκκεντρότητα e : Ο λόγος του μήκους της μακρύτερης χορδής της περιοχής προς το μήκος της μακρύτερης χορδής μεταξύ των καθέτων στην πρώτη χορδή. Πρόκειται για την εκκεντρότητα της έλλειψης που προσεγγίζει καλύτερα το σχήμα, και δίνει ένα μέτρο του πόσο -επιμηκές- είναι το σχήμα.

- Βαθμός -Συμπύκνωσης- (Compactness) c : Προσεγγιστικά ο λόγος της επιφάνειας δια το τετράγωνο της περιμέτρου του σχήματος. Δίνει ένα μέτρο του πόσο -κυκλικό- είναι το σχήμα.

- Κατεύθυνση του κύριου άξονα α : Η γωνία περιστροφής του κύριου άξονα του σχήματος από τον κάθετο.

- Λόγος αδράνειας r : Το πηλίκο του μεγαλύτερου άξονα της έλλειψης που προσεγγίζει καλύτερα το σχήμα προς το μήκος του μικρότερου άξονα αυτής.



(α′) Εμβαδόν (β′) Eccentricity

(γ′) Ratio (δ′) Compactness

Σχήμα 3.2: Τα Region-Base χαρακτηριστικά σχήματος για διαφορετικές χειρομορφές

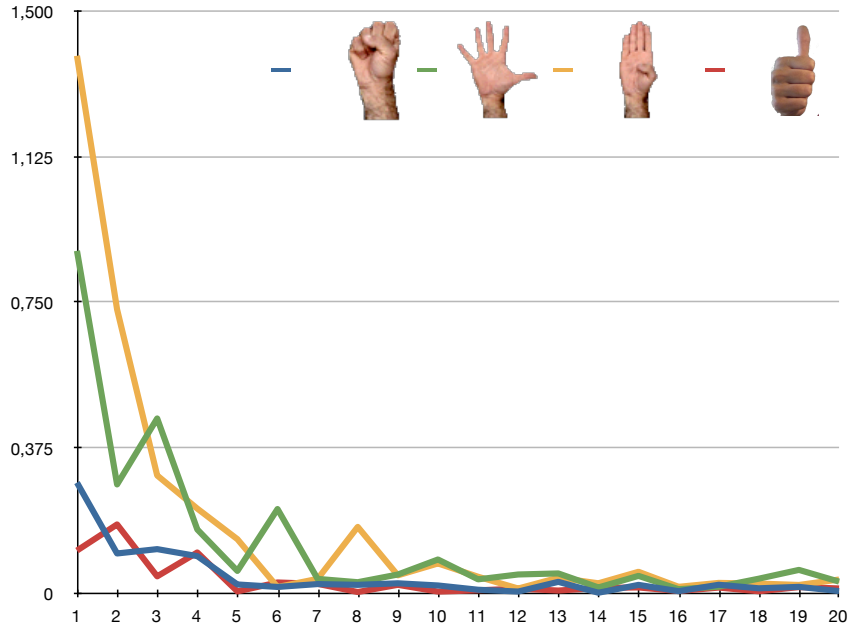
Στο Σχήμα 3.2 φαίνονται γραφικές παραστάσεις των Region Based χαρακτηριστικών για διαφορετικές χειρομορφές.

3.1.2 Περιγραφητές Fourier

Οι περιγραφητές Fourier (Fourier Descriptors) είναι ακόμα ένα ευρέως χρησιμοποιούμενο σύνολο χαρακτηριστικών σχήματος, που έχει χρησιμοποιηθεί και για εφαρμογές που σχετίζονται με τη νοηματική γλώσσα. Διαθέτουν αρκετές ιδιότητες που είναι επιθυμητό να έχουν οι περιγραφητές σχήματος:

- Ανεξαρτησία από μετασχηματισμούς μετατόπισης

- Ανεξαρτησία από περιστροφές



Σχήμα 3.3: Οι Fourier Descriptors για διαφορετικές χειρομορφές

- Ανεξαρτησία από αλλαγή κλίμακας

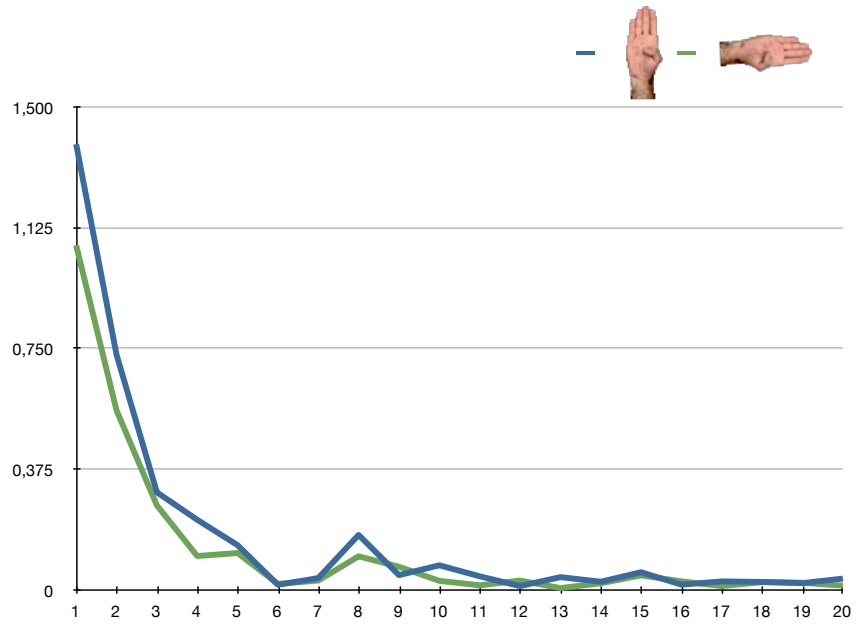
Οι ιδιότητες αυτές εξασφαλίζουν ότι η περιγραφή του αντικειμένου δεν θα εξαρτάται από την ακριβή θέση και το μέγεθος του αντικειμένου στην εικόνα, και συνεπώς παράμετροι όπως απόσταση του αντικειμένου από το σημείο λήψης της εικόνας και η θέση του αντικειμένου ως προς τον οπτικό άξονα της κάμερας λήψης δεν θα επηρεάζουν τα χαρακτηριστικά σχήματος (βλ Σχήμα 3.4). Στο Σχήμα 3.3 βλέπουμε μια γραφική παράσταση των περιγραφητές Fourier για διαφορετικές χειρομορφές.

3.1.3 Χαρακτηριστικά από Ροπές (Moments)

Ένα διαφορετικό είδος χαρακτηριστικών για την περιγραφή σχήματος όπου χρησιμοποιεί στατιστικές ιδιότητες ονομάζονται -Moments-. Ο υπολογισμός τους φαίνεται παρακάτω:

$$\mu_{ij} = \frac{\sum_{x=1}^{N} \sum_{y=1}^{N} x^i y^i f(x,y)}{\sum_{x=1}^{N} \sum_{y=1}^{N} f(x,y)} \tag{3.1}$$

Τα -Moments- είναι ανεξάρτητα από περιστροφές, μετατοπίσεις, αλλαγές κλίμακας με αποτέλεσμα λόγο των ιδιοτήτων αυτών να χρησιμοποιούνται



Σχήμα 3.4: Οι Fourier Descriptors για δύο χειρομορφές όπου η μια είναι περιστραμμένη κατά 90 μοίρες σε σχέση με την άλλη

συχνά για την περιγραφή σχήματος.

3.1.4 Συντελεστές Cepstrum της Καμπυλότητας

Το Cepstrum είναι μια μέθοδος ανάλυσης σήματος που χρησιμοποιείται ευρύτατα στην ανάλυση και αναγνώριση φωνής. Είναι ιδιαίτερα χρήσιμο για την περιγραφή σημάτων που μπορούν να θεωρηθούν ως συνελίξεις ενός σήματος από μια πηγή με ένα φίλτρο, όπως είναι η φωνή. Το Cepstrum εξυπηρετεί στο διαχωρισμό του φάσματος της πηγής από το φάσμα του φίλτρου. Επίσης είναι ικανό να εξάγει τυχόν περιοδικότητες και αρμονικές ζώνες που υπάρχουν στο φάσμα (αν αυτό θεωρηθεί ως ένα νέο σήμα), και οι οποίες μπορεί να μην είναι εμφανείς στο ίδιο το φάσμα λόγω υπέρθεσης. Προκύπτει δε από αντίστροφο Fourier μετασχηματισμό του λογαρίθμου του Fourier μετασχηματισμού του σήματος.

Η ίδια τεχνική εφαρμόστηκε για περίπτωση της εικόνας, και συγκεκριμένα στην καμπυλότητα του σχήματος. Συγκεκριμένα, ως χαρακτηριστικά σχήματος θεωρήθηκε ένα σύνολο από τους N_c πρώτους συντελεστές του πραγματικού ceprstrum της συνάρτησης καμπυλότητας.

Κεφάλαιο 4

Αναγνώριση Νοηματικής Γλώσσας

Σε αυτό το κεφάλαιο θα ασχοληθούμε με το είδος των στατιστικών μοντέλων που θα χρησιμοποιήσουμε για την μοντελοποίηση του προβλήματος μας. Έτσι θα κάνουμε μια παρουσίαση των Κρυφών Μαρκοβιανών Μοντέλων και τους λόγους που μας κάνανε να καταλήξουμε σε αυτό το είδος μοντέλων. Επίσης θα κάνουμε ανάλυση των δύο διαφορετικών καναλιών πληροφορίας της θέσης - κίνησης των χεριών και του σχήματος (χειρομορφής) των χεριών παρουσιάζοντας τα αποτελέσματα της αναγνώρισης χρησιμοποιώντας κάθε κανάλι ξεχωριστά.

4.1 Κρυφά Μαρκοβιανά Μοντέλα (HMMs)

Ένα από τα πιο βασικά σημεία τα οποία πρέπει να λάβουμε υπόψη στην μοντελοποίηση είναι ότι εάν ένας νοηματιστής εκτελέσει το ίδιο νόημα δύο φορές οι δύο αυτές εκτελέσεις δεν θα ακριβώς οι ίδιες ακόμα και αν ο νοηματιστής προσπαθήσει για αυτό. Έτσι τα μοντέλα μας δεν θα πρέπει να επηρεάζονται από αυτή την μεταβλητότητα με αποτέλεσμα να πρέπει να χρησιμοποιήσουμε κάποιου είδους στατιστικά μοντέλα.

Τα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models - HMMs) είναι ένα είδος στατιστικών μοντέλων προσαρμοσμένα σε Bayesian μοντέλα που τα κάνουν ιδανικά στο να μπορούν να αντιμετωπίσουν την μεταβλητότητα αυτή. Επίσης από την φύση τους μπορούν να περιγράψουν την μεταβολή του σήματος στο πέρασμα του χρόνου το οποίο είναι βασικό στην αναγνώριση νοηματικής γλώσσας.

Παρακάτω θα περιγράψουμε τις ιδιότητες των Κρυφών Μαρκοβιανών Μοντέλων και πως αυτές σχετίζονται με την Νοηματική Γλώσσα.

4.1.1 Ορισμός των HMMs

Ένα HMM αποτελείται από ένα σύνολο N διαφορετικών καταστάσεων

$$S = \{S_1, \ldots, S_N\}$$

όπου η μετάβαση από την μια κατάσταση σε μια άλλη γίνεται σύμφωνα με κάποια πιθανότητα. Κάθε χρονική στιγμή του σήματος αντιπροσωπεύεται από μια κατάσταση q_1 (από τις S) όπου t είναι η χρονική στιγμή στην οποία αναφερόμαστε και προσπαθεί να την μοντελοποιήσει με διάφορους τρόπους.

Κάθε μετάβαση από μια κατάσταση S_i σε μια S_j χαρακτηρίζεται από μια πιθανότητα a_{ij}. Επίσης κάθε κατάσταση έχει μια αρχική πιθανότητα π_i που προσδιορίζει την πιθανότητα το σύστημα να ξεκινήσει από αυτή. Οι δύο πιθανότητες αυτές έχουν την παρακάτω μορφή :

$$a_{ij} = P[q_i = j | q_{i-1} = i], 1 \leq i, j \leq N \tag{4.1}$$

$$\pi_i = P[q_i = i], 1 \leq i \leq N \tag{4.2}$$

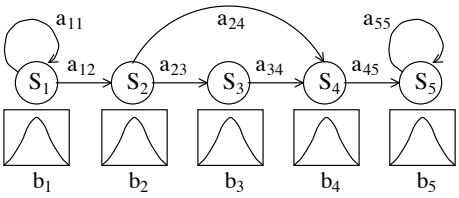
με ιδιότητες : $a_{ij} \geq 0$, $\sum_{i=1}^{n} a_{ij} = 1$ για κάθε j και $\sum_{i=1}^{n} \pi_i = 1$, για κάθε i, j. Επίσης κάθε κατάσταση i παράγει μια έξοδο k με μια πιθανότητα $b_i(k)$ η οποία ανήκει στο σύνολο όλων των πιθανών παρατηρήσεων Ω και ορίζεται ως εξής:

$$b_i(k) = P[o_t = k | q_t = i] \tag{4.3}$$

όπου k ανήκει στο σύνολο Ω. Σε συστήματα όπως αυτό της αναγνώρισης νοηματικής γλώσσας το Ω ε R^n και η $b_i(k)$ ορίζεται σαν μια μίξη συνεχών Γκαουσιανών κατανομών (Gaussian Mixture Model - GMM) όπως φαίνεται παρακάτω :

$$b_i(O) = \sum_{m=1}^{M} c_{im} G_m(O, \mu_{im}, U_{im}) \tag{4.4}$$

όπου O ε Ω =R^n, M είναι ο αριθμός των διαφορετικών Γκαουσιανών που χρησιμοποιούμε στην κατάσταση i , c_{im} είναι ο συντελεστής βάρους για την m Γκαουσιανή και κατάσταση i και G_m είναι η Γκαουσιανή κατανομή με μέση τιμή μ_{im} και απόκλιση U_{im}. Έτσι παρατηρούμε ότι με την βοήθεια των Γκαουσιανών μπορούμε να επιτύχουμε μια συνεχή αναπαράσταση του χώρου των παρατηρήσεων με αποτέλεσμα να μπορούμε να χρησιμοποιήσουμε τα HMMs και σε προβλήματα όπου το σύνολο των παρατηρήσεων είναι συνεχή σήματα και όχι διακριτά. Επίσης η χρησιμοποίηση πολλών Γκαουσιανών M μας δίνει την δυνατότητα να μπορούμε να προσεγγίσουμε οποιαδήποτε κατανομή [23], αν και στην πράξη για να το επιτύχουμε χρειαζόμαστε πολύ μεγάλη βάση δεδομένων για την εκπαίδευση του μοντέλου οπότε συνήθως ο αριθμός των Γκαουσιανών που χρησιμοποιούμε είναι από 1 έως 3. Έτσι ένα HMM μοντέλο



Σχήμα 4.1: Left-right HMM μοντέλο, επιτρέπονται μεταβάσεις μόνο από αριστερά προς τα δεξιά.

ορίζεται από 3 σύνολα πιθανοτικών μεγεθών A,B,π και για ευκολία ορίζουμε $\lambda = (A, B, \pi)$. Παρακάτω μπορούμε να δούμε ένα παράδειγμα ενός HMM μοντέλου Σχήμα 4.1 όπου λέγεται left-right λόγο του ότι επιτρέπονται μεταβάσεις από μια κατάσταση μόνο σε μια κατάσταση που βρίσκεται δεξιότερα δηλαδή $a_{ij} \geq 0$ μόνο αν $i \leq j$. Αυτό το είδος HMM χρησιμοποιείται αρκετά για την μοντελοποίηση χρονικών σημάτων.

Όπως ανέφερα παραπάνω το διάνυσμα O_t αντιπροσωπεύει την παρατήρηση την χρονική στιγμή t. Στην αναγνώριση ΕΝΓ το O_t αντιπροσωπεύει την δομή ενός νοήματος μια δεδομένη χρονική στιγμή. O_t είναι ουσιαστικά ένας πίνακας χαρακτηριστικών ο οποίος παρέχει πληροφορία για την θέση των χεριών στην εικόνα και της χειρομορφής. Έτσι ένα νόημα μπορεί να παρασταθεί από μια ακολουθία παρατηρήσεων $O = O_1, O_2, ..., O_T$ όπου Τ είναι ο αριθμός των frames του βίντεο.

Η πιθανότητα να προκύψει μια ακολουθία παρατηρήσεων $O = O_1, O_2, ..., O_T$ από μια ακολουθία καταστάσεων $Q = q_1, q_2, ..., q_T$ όπου $q_i \; \varepsilon \; S$ για ένα HMM μοντέλο λ, είναι :

$$P(O, Q|\lambda) = \pi_{q_1} b_{q_1}(O_1) \prod_{t=2}^{T} a_{q_{t-1}q_t} b_{q_t}(O_t) \tag{4.5}$$

όπου π_{q_1} είναι η πιθανότητα του συστήματος να ξεκινήσει από την κατάσταση q_1 , $b_{q_1}(O_1)$ είναι η πιθανότητα να έχουμε την παρατήρηση O_1 στην κατάσταση q_1 και $a_{q_{t-1}q_t}$ η πιθανότητα μετάβασης από την κατάσταση q_{t-1} στην κατάσταση q_t.

Δυο σημαντικά συμπεράσματα μπορούμε να βγάλουμε από την παραπάνω πιθανότητα: Πρώτον στην μοντελοποίηση HMM γίνεται η υπόθεση ότι οι παρατηρήσεις O_t , O_{t+1} είναι ανεξάρτητες. Δεύτερον γίνεται η λεγόμενη υπόθεση Markov όπου δηλαδή η πιθανότητα μετάβασης εξαρτάται μόνο από την τρέχουσα κατάσταση και όχι από τις προηγούμενες [23]. Αυτή η υπόθεση αν και δεν ισχύει γενικά σε χρονικά μεταβαλλόμενα σήματα επιτυγχάνει μεγάλη μείωση της πολυπλοκότητας.

4.1.2 Τα Τρία Βασικά Προβλήματα των HMM

Αφού αναφέραμε την μορφή των HMM στην προηγούμενη ενότητα θα πρέπει να λυθούν τα 3 βασικά προβλήματα για να μπορεί το μοντέλο αυτό να χρησιμοποιηθεί σε εφαρμογές [23]. Τα προβλήματα αυτά είναι τα παρακάτω:

Πρόβλημα 1

Δεδομένης μιας ακολουθίας παρατηρήσεων $O = O_1, O_2, ..., O_T$ και του HMM μοντέλου $\lambda = (A, B, \pi)$ πώς μπορούμε να υπολογίσουμε την πιθανότητα $P(O|\lambda)$ να προκύψει η συγκεκριμένη ακολουθία παρατηρήσεων από αυτό το μοντέλο.

Πρόβλημα 2

Δεδομένης μιας ακολουθίας παρατηρήσεων $O = O_1, O_2, ..., O_T$ και του HMM μοντέλου $\lambda = (A, B, \pi)$ πώς μπορούμε να υπολογίσουμε την πιο πιθανή ακολουθία καταστάσεων $Q = q_1, q_2, ..., q_T$.

Πρόβλημα 3

Πως αλλάζουμε τις παραμέτρους του HMM μοντέλου $\lambda = (A, B, \pi)$ έτσι ωστε να μεγιστοποιήσουμε την πιθανότητα $P(O|\lambda)$.

Η λύση του πρώτου προβλήματος -scoring problem- ουσιαστικά μας δίνει την δυνατότητα εάν έχουμε πολλά HMM μοντέλα να προσδιορίσουμε το μοντέλο το οποίο ταιριάζει καλύτερα στην συγκεκριμένη ακολουθία παρατηρήσεων επιλέγοντας αυτό με την μεγαλύτερη πιθανότητα $P(O|\lambda)$.

Ένας τρόπος για τον υπολογισμό της $P(O|\lambda)$ είναι να υπολογίσουμε την πιθανότητα για κάθε πιθανή ακολουθία $Q = q_1, q_2, ..., q_T$. Ο αριθμός των πιθανών ακολουθιών είναι N^T. Έτσι η πιθανότητα της ακολουθίας παρατηρήσεων O δεδομένης της ακολουθίας καταστάσεων Q είναι:

$$P(O|Q, \lambda) = \prod_{t=1}^{T} P(O_t|q_t, \lambda) \tag{4.6}$$

όπου υποθέτουμε ότι οι παρατηρήσεις είναι ανεξάρτητες μεταξύ τους οπότε:

$$P(O|Q, \lambda) = b_{q_1}(O_1)b_{q_2}(O_2)...b_{q_T}(O_T) \tag{4.7}$$

Η πιθανότητα για μια τέτοια ακολουθία καταστάσεων Q είναι:

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} ... a_{q_{T-1} q_T} \tag{4.8}$$

έτσι η συνολική πιθανότητα είναι το γινόμενο των δυο παραπάνω:

$$P(O, Q|\lambda) = P(O|Q, \lambda) P(Q|\lambda) \tag{4.9}$$

όμως ο υπολογισμός της παραπάνω πιθανότητας με αυτό τον τρόπο έχει πολύ μεγάλη πολυπλοκότητα αφού όπως είδαμε ο αριθμός της ακολουθίας πιθανών καταστάσεων είναι N^T έτσι εάν έχουμε ένα μεγάλο αριθμό N ή T πχ $N = 5, T = 100$ τότε έχουμε $2 * 100 * 5^{100} = 10^{72}$ υπολογισμούς. Έτσι καταφεύγουμε σε μια άλλη μέθοδο η οποία λέγεται -The Forward Algorithm-.

Ορίζουμε την πιθανότητα $\alpha_t(i) = P(O_1 O_2 ... O_t, q_t = i|\lambda)$ που είναι η πιθανότητα της εμφάνισης την ακολουθίας παρατηρήσεων $O_1, O_2, ..., O_t$ μέχρι την χρονική στιγμή t και το μοντέλο να βρίσκεται στην κατάσταση i την χρονική στιγμή αυτή. Έτσι χρησιμοποιώντας αυτή την πιθανότητα και την παρακάτω αναδρομική σχέση μπορούμε να βρούμε την πιθανότητα $P(O|\lambda)$.

Αναδρομική σχέση:

1. Αρχικοποίηση

$$\alpha_1(i) = \pi_i b_i(O_1), 1 \leq i \leq N \tag{4.10}$$

2. Αρχή αναδρομής

$$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(O_{t+1}), 1 \leq t \leq T-1 \tag{4.11}$$

3. Τερματισμός

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{4.12}$$

Μπορούμε να παρατηρήσουμε ότι με αυτή την μέθοδο η πολυπλοκότητα μειώνεται δραματικά αφού χρειαζόμαστε $N^2 T$ υπολογισμούς σε αντίθεση με την προηγούμενη μέθοδο όπου θέλουμε $2TN^T$.

Στο δεύτερο πρόβλημα στόχος είναι να βρούμε την ακολουθία $Q = q_1, q_2, ..., q_T$ που μεγιστοποιεί την πιθανότητα $P(Q, O|\lambda)$ για μια δεδομένη ακολουθία παρατηρήσεων $O = O_1, O_2, ..., O_T$. Για την λύση του προβλήματος ορίζουμε την πιθανότητα :

$$\delta_t(i) = max_{q_1, q_2, .., q_{t-1}} P[q_1, q_2, .., q_{t-1}, q_t = i, O_1, O_2, .., O_t|\lambda] \tag{4.13}$$

που είναι η μέγιστη πιθανότητα για μια διαδρομή προσμετρώντας τις πρώτες t παρατηρήσεις και βρίσκεται στην κατάσταση i στην στιγμή t. Έτσι χρησιμοποιώντας την σχέση 4.13 και τον πίνακα $\psi_t(j)$ για την καταγραφή των ορισμάτων που μεγιστοποιούν την σχέση 4.13 με την παρακάτω αναδρομική σχέση μπορούμε να υπολογίσουμε την ιδανική ακολουθία καταστάσεων.

Αναδρομική σχέση (Viterbi Algorithm):

1. Αρχικοποίηση

$$\delta_1(i) = \pi_i b_i(O_1), 1 \leq i \leq N \tag{4.14}$$

$$\psi_1(i) = 0 \tag{4.15}$$

2.Αρχή αναδρομής

$$\delta_t(j) = \max_{1 \leq i \leq N}[\delta_{t-1} a_{ij}] b_j(O_t), 2 \leq t \leq T, 1 \leq j \leq N \tag{4.16}$$

$$\psi_t(i) = arg \max_{1 \leq i \leq N}[\delta_{t-1} a_{ij}] \tag{4.17}$$

3.Τερματισμός

$$P^* = \max_{1 \leq i \leq N}[\delta_T(i)] \tag{4.18}$$

$$q_T^* = arg \max_{1 \leq i \leq N}[\delta_T(i)] \tag{4.19}$$

4.Οπισθοδρόμηση

$$q_t^* = \psi_{t+1}(q_{t+1}^*), t = T-1, T-2, ..., 1 \tag{4.20}$$

Ένα πολύ σημαντικό πλεονέκτημα του αλγόριθμου του Viterbi είναι ότι στον υπολογισμό του $\delta_t(j)$ υπάρχει η δυνατότητα να λαμβάνει υπόψη του μόνο τις καταστάσεις j όπου η πιθανότητα $\delta_{t-1}(j)$ είναι μεγαλύτερη από μια τιμή που έχουμε ορίσει εμείς με αποτέλεσμα να μειώνεται η πολυπλοκότητα πάρα πολύ. Αυτή την ιδιότητα την χρησιμοποιούμε συνήθως σε μεγάλα δίκτυα HMM όπου η πολυπλοκότητα αυξάνει κατακόρυφα.

Ένας δεύτερος εξίσου σημαντικός τρόπος επίλυσης του 2ου προβλήματος είναι ο λεγόμενος -token passing algorithm- [26]. Σε κάθε επανάληψη κάθε χρονική στιγμή $tok_t(i) = \delta_t(i)$ για κάθε i, οπότε ο token passing αλγόριθμος είναι ισοδύναμος με τον Viterbi αλγόριθμο. Η μόνη διαφορά που έχουν είναι ότι ο Viterbi κάνει ανανέωση των πιθανοτήτων βασιζόμενος στις μεταβάσεις από την κατάσταση που υπολογίζεται ενώ αντίθετα ο token passing προς την κατάσταση που υπολογίζεται. Ακόμα ο token passing αλγόριθμος έχει την δυνατότητα σε κάθε token να προσθέτει πληροφορίες όπως την διαδρομή μέσα στο δίκτυο HMM κ.α.

1: Αρχικοποίηση κάθε κατάστασης i με ένα token $tok_t(i) = \delta_t(i)$
2: $for\ t = 2\ to\ T$ do
3: for κάθε κατάσταση i do
4: $tok_t(i) = 0$
5: end for
6: for κάθε κατάσταση i do
7: for κάθε κατάσταση j που ενώνεται με i do
8: $tok_t(j) = max\{tok_t(j), tok_{t-1}(i)a_{ij}b_j(O_t)\}$
(πέρασμα του token από κατάσταση i στην κατάσταση j)
9: end for
10: end for
11: end for

Το τρίτο και δυσκολότερο πρόβλημα είναι να βρούμε μια μέθοδο για των υπολογισμό των παραμέτρων (A, B, π) του μοντέλου. Δεν υπάρχει αναλυτικός τρόπος για τον υπολογισμό των παραμέτρων που μεγιστοποιούν την πιθανότητα για μια συγκεκριμένη ακολουθία παρατηρήσεων. Μπορούμε όμως να υπολογίσουμε τις παραμέτρους $\lambda = (A, B, \pi)$ έτσι ώστε να επιτύχουμε ένα τοπικό μέγιστο χρησιμοποιώντας αλγόριθμους όπως του Baum-Welch [23]. Για την περιγραφή του υπολογισμού των παραμέτρων $\lambda = (A, B, \pi)$ ορίζουμε αρχικά την πιθανότητα την χρονική στιγμή t να βρισκόμαστε στην κατάσταση i και την χρονική στιγμή $t + 1$ να βρισκόμαστε στην κατάσταση j.

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j|O, \lambda) \tag{4.21}$$

Επίσης πρέπει να ορίσουμε την backward πιθανότητα όπου είναι η πιθανότητα να έχουμε μια ακολουθία παρατηρήσεων $O_{t+1}, O_{t+2}, .., O_T$ από την χρονική στιγμή $t + 1$ μέχρι το τέλος T δεδομένου ότι βρισκόμαστε στην κατάσταση i την χρονική στιγμή t και δεδομένου του μοντέλου λ.

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, .., O_T|q_t = i, \lambda) \tag{4.22}$$

Έτσι η backward πιθανότητα υπολογίζεται με την παρακάτω αναδρομική σχέση:

1. Αρχικοποίηση
$$\beta_T(i) = 1, 1 \leq i \leq N \tag{4.23}$$
2. Αρχή αναδρομής
$$\beta_t(i) = \sum_{j=1}^{N} a_{ij}b_j(O_{t+1})\beta_{t+1}(j), t = T - 1, T - 2, .., 1 \tag{4.24}$$

Με την βοήθεια των forward (την ορίσαμε προηγουμένως στο πρόβλημα 1 και backward πιθανοτήτων η σχέση 4.21 μπορεί να γραφεί ως εξής:

$$\begin{aligned} \xi_t(i,j) &= \frac{P(q_t = i, q_{t+1} = j, O|\lambda)}{P(O|\lambda)} \\ &= \frac{\alpha_t(i)a_{ij}b_j(Ot+1\beta_{t+1}(j)}{P(O|\lambda)} \\ = \frac{\alpha_t(i)a_{ij}b_j(Ot+1\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)} \end{aligned} \tag{4.25}$$

Επίσης η πιθανότητα να βρίσκεται στην κατάσταση i την χρονική στιγμή t δεδομένου της ακολουθίας των παρατηρήσεων και του μοντέλου είναι ίση με:

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \tag{4.26}$$

Εάν τώρα αθροίσουμε την παραπάνω πιθανότητα για όλες τις χρονικές στιγμές τότε ουσιαστικά βρίσκουμε τον αριθμό των επισκέψεων της κατάστασης i ενώ εάν κάνουμε το ίδιο ακριβώς και για την πιθανότητα $\xi_t(i,j)$ τότε παίρνουμε τον αριθμό των μεταβάσεων από την κατάσταση i στην κατάσταση j.

$\sum_{t=1}^{T-1}\gamma_t(i) =$ αριθμός μετάβασης από κατάσταση i σε οποιαδήποτε άλλη.

$\sum_{t=1}^{T-1}\xi_t(i,j) =$ αριθμός μετάβασης από κατάσταση i στην κατάσταση j.

Έτσι χρησιμοποιώντας όλα τα παραπάνω μπορούμε να προσδιορίσουμε της παραμέτρους $\lambda = (A, B, \pi)$ του μοντέλου για μια ακολουθία παρατηρήσεων:

$$\pi_j = \gamma_1(i) \tag{4.27}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{4.28}$$

$$c_{jm} = \frac{\sum_{t=1}^{T} \gamma_t(j,m)}{\sum_{t=1}^{T} \sum_{k=1}^{M} \gamma_t(j,k)} \tag{4.29}$$

$$\mu_{jm} = \frac{\sum_{t=1}^{T} \gamma_t(j,m) O_t}{\sum_{t=1}^{T} \gamma_t(j,m)} \tag{4.30}$$

$$U_{jm} = \frac{\sum_{t=1}^{T} \gamma_t(j,m)(O_t - \mu_{jm})(O_t - \mu_{jm})^T}{\sum_{t=1}^{T} \gamma_t(j,m)} \tag{4.31}$$

Η πολυπλοκότητα του αλγόριθμου Baum-Welch είναι $O(N^2T)$. Επαναλαμβάνοντας τον αλγόριθμο του Baum-Welch μέχρι την επίτευξη κάποιου από τα κριτήρια σύγκλισης [23] καταφέρνουμε να υπολογίσουμε της παραμέτρους του μοντέλου επιτυγχάνοντας ένα μέγιστο της συνολικής πιθανότητας. Επίσης όταν κάνουμε εκπαίδευση με περισσότερες της μιας ακολουθίας δεδομένων τότε υπολογίζονται οι παράμετροι π,a,c,μ,U ξεχωριστά για κάθε δεδομένο και τελικά κρατάμε το μέσο όρο αυτών.

Έχοντας πλέον ορίσει τα HMMs μοντέλα στις επόμενες ενότητες θα ασχοληθούμε με την μοντελοποίηση των δυο καναλιών πληροφορίας, της θέσης-κίνησης και του σχήματος (χειρομορφής) των χεριών χρησιμοποιώντας απλά HMMs.

4.2 Περιγραφή του Συστήματος

4.2.1 Επιλογή Συστήματος HMMs

Η δύναμη των Μαρκοβιανών Μοντέλων έγκειται στην ικανότητα τους να μοντελοποιούν ακολουθίες δεδομένων. Έτσι, μπορούν να αναγνωρίσουν ακολουθίες εικόνων, που έχουν δεσμευτεί σε κοντινά μεταξύ τους χρονικά διαστήματα κατά την διάρκεια της κίνησης των χεριών. Τα διανύσματα χαρακτηριστικών, που εξάγονται από τις ακολουθίες εικόνων εκπαίδευσης, χρησιμοποιούνται στη συνέχεια για την εκπαίδευση των HMMs, που συνιστούν το σύστημα αναγνώρισης. Η ικανότητα των HMMs, να παραμένουν στην ίδια κατάσταση για περισσότερα του ενός χρονικά διαστήματα, τους δίνει τη δυνατότητα να μπορούν να αναγνωρίζουν διαφορετικού μήκους ακολουθίες εικόνων, οι οποίες αντιστοιχούν στην ίδια λέξη. Αυτό σημαίνει ότι δυο εκδοχές της ίδιας λέξης, που σχηματίζονται με διαφορετικές ταχύτητες, μπορούν να

αποδοθούν ορθά στο ίδιο HMM. Συνεπώς, τα μαρκοβιανά μοντέλα αποτελούν μια ελκυστική επιλογή για την επεξεργασία δεδομένων, που προέρχονται από νοήματα, καθώς έχουν την ικανότητα να περιγράφουν τον τρόπο χρονικής μεταβολής ενός νοήματος. Αυτό το πετυχαίνουν αφενός αποδίδοντας διαφορετικές στιγμές του νοήματος σε διαφορετικές καταστάσεις, αφετέρου παραμένοντας στην ίδια κατάσταση για περισσότερα του ενός χρονικά διαστήματα (frames), εφόσον κάτι τέτοιο είναι απαραίτητο.

Ο αριθμός των καταστάσεων ενός HMM και η τοπολογία του παίζουν σημαντικό ρόλο στην αναγνώριση, που θα επιτευχθεί από το σύστημα. Καθώς ο σχηματισμός των νοημάτων στην νοηματική γλώσσα είναι μια διαδικασία, που μεταβάλλεται στο χρόνο, κάθε νόημα θα ήταν καλό να αντιστοιχιστεί σε μια από-αριστερά προς-τα-δεξιά τοπολογία (left-right topology). Ο βέλτιστος αριθμός καταστάσεων σε μια τοπολογία HMM εξαρτάται από τη συχνότητα των frames καθώς και από την πολυπλοκότητα των νοημάτων και η εύρεση του στην περίπτωση μας γίνεται εμπειρικά.

Στα πειράματα που έγιναν χρησιμοποιήθηκαν HMMs ίδιας τοπολογίας και ίδιου αριθμού καταστάσεων για όλα τα νοήματα καθορίζοντας των αριθμό των καταστάσεων πειραματικά, καθώς και HMMs διαφορετικού αριθμού καταστάσεων για νοήματα διαφορετικής χρονικής διάρκειας. Ο καθορισμός του αριθμού καταστάσεων για κάθε νόημα έγινε χρησιμοποιώντας τον μέσο όρο του πλήθους των frames που συνιστούν τα παραδείγματα εκπαίδευσης για το συγκεκριμένο νόημα, καταλήγοντας σε 5 διαφορετικές τοπολογίες HMMs των 3,4,5,6 και 7 καταστάσεων.

4.2.2 Λεξιλόγιο

Στα πειράματα που έγιναν στα πλαίσια αυτής της διπλωματικής χρησιμοποιήθηκε ένα λεξιλόγιο το οποίο αποτελείτο από 100 διαφορετικές λέξεις.

4.2.3 Σύνολα Δεδομένων

Τα δεδομένα μας τα χωρίζουμε σε δυο υποσύνολα. Το ένα το χρησιμοποιούμε για την εκπαίδευση των μοντέλων μας και το άλλο για τον έλεγχο της απόδοσης του συστήματος. Τα δυο αυτά υποσύνολα είναι ανεξάρτητα μεταξύ τους δηλαδή δεδομένα από το υποσύνολο ελέγχου δεν έχουν χρησιμοποιηθεί ποτέ κατά την εκπαίδευση των μοντέλων.

Για κάθε νόημα χρησιμοποιούμε από 3 μέχρι 12 παραδείγματα εκπαίδευσης και από 1 μέχρι 4 παραδείγματα ελέγχου, διατηρώντας την αναλογία 2 προς 1 ή 3 προς 1, που συνηθίζεται να τηρούν άλλοι ερευνητές [31, 4, 2]. Συνεπώς για το σύνολο των 100 λέξεων, χρησιμοποίησα τελικά 600 παραδείγματα εκπαίδευσης και 200 παραδείγματα ελέγχου.

4.2.4 Κριτήρια Απόδοσης

Για τα κριτήρια απόδοσης στην αναγνώριση μεμονωμένων λέξεων-νοημάτων ορίζεται μόνο ένας ρυθμός αναγνώρισης που δηλώνει το ποσοστό των ορθώς αναγνωρισμένων λέξεων-νοημάτων δηλαδή W/N που W είναι ο αριθμός των λέξεων που αναγνωρίστηκαν σωστά και N είναι το σύνολο των λέξεων που χρησιμοποιήσαμε για την αξιολόγηση του συστήματος. Τα λάθη που λαμβάνουν χώρα σε αυτή την περίπτωση είναι αποκλειστικά αντικαταστάσεις λέξεων. Εφόσον δεν είναι δυνατόν να παρατηρηθούν εισαγωγές λέξεων, παραλείπουμε την ακρίβεια λέξεων ως μέτρο απόδοσης του συστήματος η οποία σε αυτή την περίπτωση ταυτίζεται με το ρυθμό αναγνώρισης.

Για τον έλεγχο της απόδοσης του συστήματος χρησιμοποιήσαμε την μέθοδο 10-fold cross-validation [18]. Δηλαδή χωρίζουμε τυχαία το σύνολο των δεδομένων μας σε δυο υποσύνολα ένα για την εκπαίδευση των μοντέλων μας και ένα για τον έλεγχο της απόδοσης με αναλογία 2 προς 1 ή 3 προς 1, και παίρνουμε το ποσοστό των ορθώς αναγνωρισμένων λέξεων-νοημάτων επαναλαμβάνουμε την παραπάνω διαδικασία 10 φορές και το τελικό ποσοστό απόδοσης του συστήματος είναι ο μέσος όρος των ποσοστών των ορθώς αναγνωρισμένων λέξεων-νοημάτων για τις 10 επαναλήψεις.

4.3 Κανάλι για την Θέση-Κίνηση των χεριών

Η πληροφορία για τη Θέση και Κίνηση των χεριών παίζει πολύ σημαντικό ρόλο στην Νοηματική Γλώσσα. Κάθε νόημα ορίζεται από μια αρχική και μια τελική θέση των χεριών όπως επίσης και από την κίνηση που εκτελούν τα χέρια ξεκινώντας από την αρχική τους θέση για να καταλήξουν στην τελική τους θέση. Μερικά παραδείγματα διαφορετικής κίνησης και θέσης των χεριών μπορούμε να δούμε στην ενότητα 2.2.

Επειδή η εξαγωγή της πληροφορίας της θέσης και κίνησης των χεριών από μια ακολουθία εικόνων (βίντεο) είναι σχετικά εύκολη διαδικασία και πολύ αξιόπιστη θα της δώσουμε πολύ μεγάλο βάρος στην αναγνώριση. Στις παρακάτω ενότητες θα αναφερθούμε σε πειραματικά αποτελέσματα της αναγνώρισης χρησιμοποιώντας μόνο την πληροφορία της θέσης και κίνησης των χεριών.

4.3.1 Μοντελοποίηση της Θέσης-Κίνησης των χεριών

Για την μοντελοποίηση της νοηματικής γλώσσα όπως είδαμε και σε προηγούμενο κεφάλαιο (βλ. ενότητα 2.5) χρησιμοποιήσαμε ένα κανάλι πληροφορίας για την θέση, κίνηση για κάθε χέρι ξεχωριστά. Πιο συγκεκριμένα τα κανάλια αυτά αποτελούνται από τα εξής διανύσματα χαρακτηριστικών:

- Την θέση του χεριού πάνω στην εικόνα

- Την κίνηση του χεριου

- Την απόσταση των δύο χεριών μεταξύ τους

Για τον υπολογισμό της θέσης του χεριού χρησιμοποιήσαμε τις συντεταγμένες του κέντρου μάζας της καμπύλης που το περικλείει, για περισσότερες λεπτομέρειες (βλ. 3.1.1 και [35]). Έχοντας την θέση του χεριού κάθε χρονική στιγμή ο υπολογισμός της κίνησης γίνεται βρίσκοντας την πρώτη παράγωγο της θέσης χρησιμοποιώντας τις διαφορές της θέσης από ένα frame με το αμέσως προηγούμενο. Η απόσταση των δύο χεριών υπολογίζεται με την City - Block απόσταση αφαιρώντας τις συντεταγμένες των δύο χεριών μεταξύ τους.

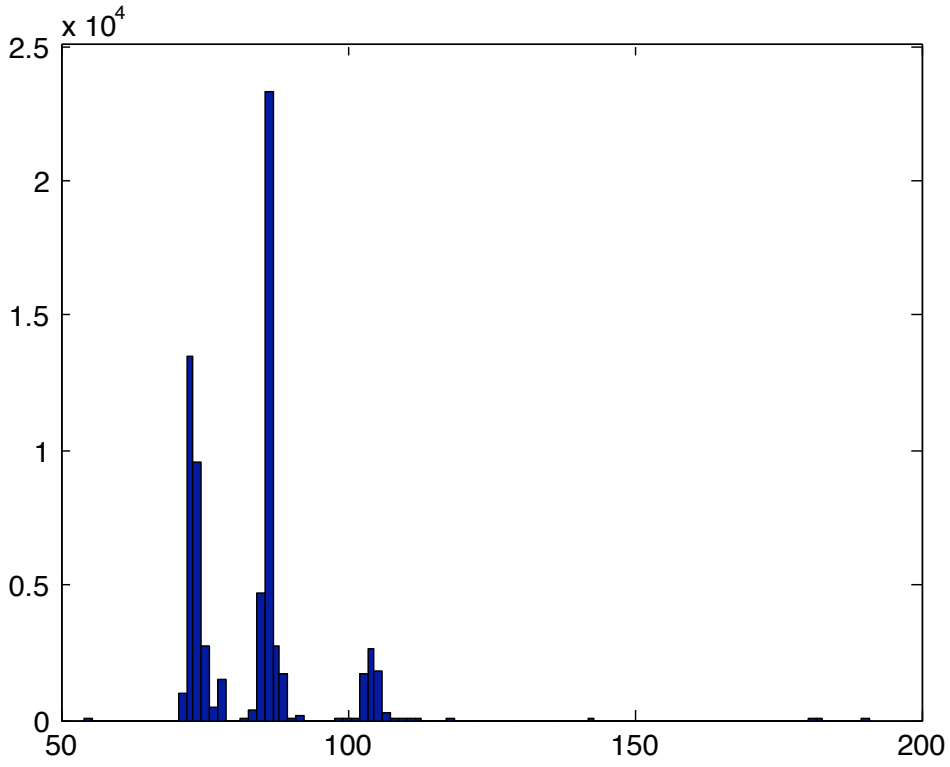
4.3.2 Κανονικοποίηση της Θέσης

Η τιμές για τις συντεταγμένες της θέσης των χεριών πρέπει να κανονικοποιηθούν έτσι ώστε να είναι ανεξάρτητες από το ύψος του νοηματιστή και γενικότερα από την θέση που βρίσκεται σε σχέση με την κάμερα. Την ανάγκη για normalization μπορούμε να την δούμε στο Σχήμα 4.2 όπου έχουμε απεικονίσει τις θέσεις των κεφαλιών σε κάποια νοήματα για τους 3 νοηματιστές από την βάση δεδομένων που χρησιμοποιήσαμε. Αυτός ο προφανής διαχωρισμός των τριών νοηματιστών μεταξύ τους δημιουργείται λόγο της διαφοράς ύψους που έχουν.

Έτσι ένα απλός και αποτελεσματικός τρόπος για να κάνουμε κανονικοποίηση είναι να θεωρήσουμε σημείο αναφοράς των μετρήσεων των συντεταγμένων των χεριών το κεφάλι του νοηματιστή δηλαδή κάθε χρονική στιγμή αφαιρούμε τις συντεταγμένες των χεριών από τις συντεταγμένες του κεφαλιού. Έτσι αποκτάμε ανεξαρτησία των μετρήσεων από το πού βρίσκεται ο νοηματιστής σε σχέση με την κάμερα αλλά και από το ύψος του.

4.4 Πειραματικά αποτελέσματα ως προς τα διανύσματα χαρακτηριστικών

Σε αυτή την ενότητα θα ασχοληθούμε με τα διανύσματα χαρακτηριστικών που θα χρησιμοποιήσουμε στο κανάλι πληροφορίας για την θέση. Θα εξετάσουμε ποια μας παρέχουν σημαντική πληροφορία και θα καταλήξουμε σε αυτά που θα χρησιμοποιήσουμε.



Σχήμα 4.2: Απεικόνιση του προφανή διαχωρισμού των τριών νοηματιστών λόγο της διαφοράς ύψους που έχουν μεταξύ τους.

4.4.1 Πειραματικά αποτελέσματα χρησιμοποιώντας μόνο την θέση

Στον Πίνακα 4.1 μπορούμε να δούμε τα αποτελέσματα αναγνώρισης χρησιμοποιώντας μόνο την πληροφορία της θέσης του strong χεριού δηλαδή τις συντεταγμένες (x,y) του κέντρου βάρους κάθε χρονική στιγμή πάνω στην εικόνα, με και χωρίς να κάνουμε normalization.

	Ποσοστό Αναγνώρισης %	H	S	N
Χωρίς Normalization	39.15	77	120	197
Με Normalization	61.12	120	77	197

Πίνακας 4.1: Αποτελέσματα με και χωρίς normalization της θέσης του strong χεριού.

Από τα αποτελέσματα του Πίνακα 4.1 επαληθεύεται η ανάγκη για normalization που αναφέραμε στην ενότητα 4.3.2. Σε όλα τα πειράματα που θα ακολουθήσουν θα χρησιμοποιήσουμε την πληροφορία της θέσης normalized.

4.4.2 Πειραματικά αποτελέσματα χρησιμοποιώντας την θέση και κίνηση

Χρησιμοποιώντας την πληροφορία της θέσης του strong χεριού (x,y) και της ταχύτητας (κίνησης) του δηλαδή της πρώτης χρονικής παραγώγου της θέσης του παίρνουμε τα αποτελέσματα που φαίνονται στον Πίνακα 4.2.

Διανύσματα Χαρακτηριστικών	Ποσοστό Αναγνώρισης %	H	S	N
Θέση,Κίνηση $x, y, \dot{x}, \dot{y}$	74.09	146	51	197

Πίνακας 4.2: Αποτελέσματα της θέσης και κίνησης του strong χεριού.

Από τα αποτελέσματα του Πίνακα 4.2 παρατηρούμε μια σημαντική αύξηση του ποσοστού αναγνώρισης που οφείλεται στο ότι χρησιμοποιώντας τις παραγώγους της θέσης δίνουμε μεγάλη έμφαση στην κίνηση των χεριών.Έτσι διαφορετικές εκτελέσεις του ίδιου νοήματος που σίγουρα θα έχουν την ίδια κίνηση στο χώρο (πχ κυκλική) αλλά μπορεί να να μην εκτελούνται ακριβώς στην ίδια θέση αναγνωρίζονται σωστά.

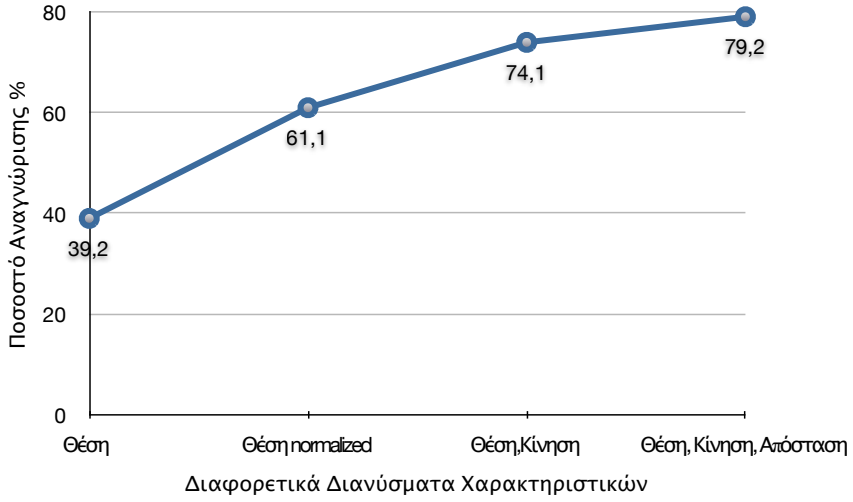
4.4.3 Πειραματικά αποτελέσματα χρησιμοποιώντας την θέση, κίνηση και απόσταση των χεριών

Χρησιμοποιώντας την πληροφορία της θέσης,κίνησης του strong χεριού και της απόστασης των χεριών μεταξύ τους παίρνουμε τα αποτελέσματα που φαίνονται στον Πίνακα 4.3.

Διανύσματα Χαρακτηριστικών	Ποσοστό Αναγνώρισης %	H	S	N
Θέση,Κίνηση,Απόσταση Χεριών $x, y, \dot{x}, \dot{y}, x_s - x_w, y_s - y_w$	79.18	156	41	197

Πίνακας 4.3: Αποτελέσματα της θέσης, κίνησης του strong χεριού και της απόστασης των χεριών

Όπου x_s, y_s είναι οι συντεταγμένες του strong χεριού και x_w, y_w είναι οι συντεταγμένες του weak χεριού. Η αύξηση του ποσοστού αναγνώρισης είναι λογική αφού συμπεριλάβαμε την πληροφορία της θέσης του weak χεριού ή οποία είναι πολύ σημαντική στις περιπτώσεις που εκτελούνται νοήματα στα οποία χρησιμοποιούνται και τα δύο χέρια.



Σχήμα 4.3: Συνολικά αποτελέσματα για τα διαφορετικά διανύσματα χαρακτηριστικών.

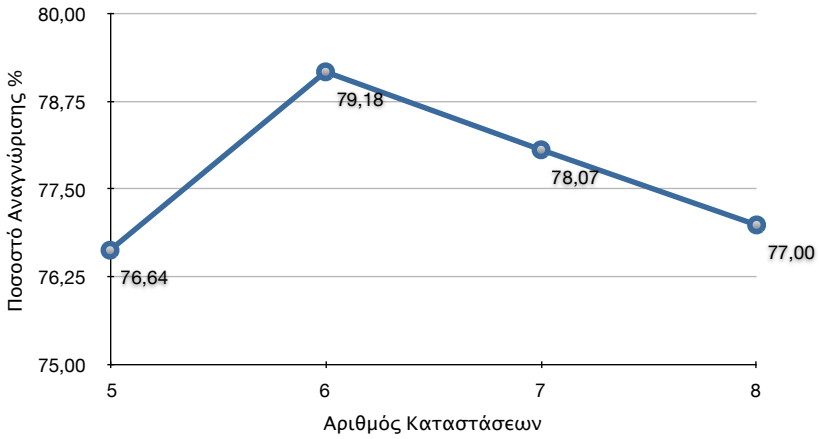
Έτσι καταλήγουμε ότι τα διανύσματα χαρακτηριστικών που θα χρησιμοποιήσουμε θα συμπεριλαμβάνει τη θέση την κίνηση και την απόσταση των χεριών. Στις επόμενες ενότητες θα κάνουμε πειράματα πάνω στην τοπολογία των HMM μοντέλων.

4.5 Πειραματικά αποτελέσματα για την τοπολογία των HMM μοντέλων

Σε αυτή την ενότητα θα ασχοληθούμε με την τοπολογία των HMM μοντέλων. Θα παρουσιάσουμε πειραματικά αποτελέσματα για μια left-right τοπολογία με κοινό αριθμό καταστάσεων για όλα τα μοντέλα όπως επίσης και για διαφορετικό αριθμό καταστάσεων για νοήματα διαφορετικής χρονική διάρκειας. Επίσης θα αναφερθούμε και σε μια διαφορετική τοπολογία η οποία ονόμαζεται X-Model.

4.5.1 Πειράματα με κοινό αριθμό καταστάσεων

Στην παρούσα παράγραφο θα παρουσιάσουμε πειραματικά αποτελέσματα χρησιμοποιώντας κοινό αριθμό καταστάσεων για όλα τα HMM μοντέλα. Στο Σχήμα 4.4 μπορούμε να δούμε τα αποτελέσματα για κοινό αριθμό καταστάσεων από 5 έως 8.



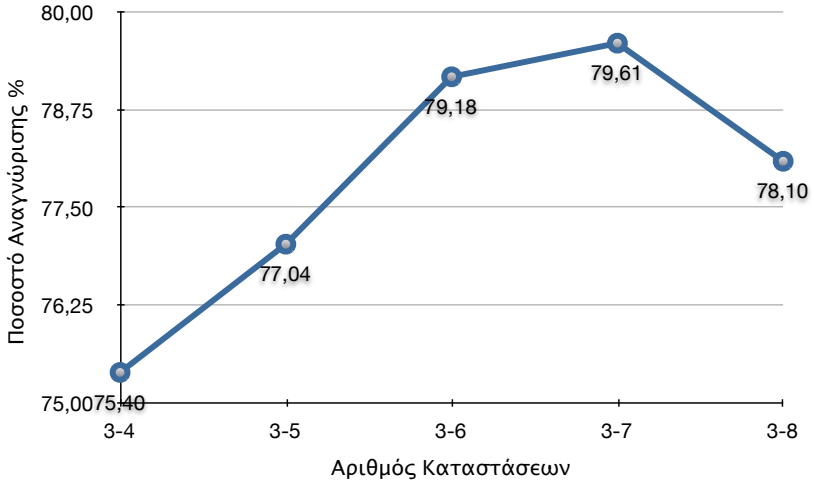
Σχήμα 4.4: Αποτελέσματα με κοινό αριθμό καταστάσεων όλων των HMM μοντέλων.

Από το Σχήμα 4.4 βγάζουμε το συμπέρασμα ότι έχουμε τα καλύτερα αποτελέσματα όταν χρησιμοποιούμε 6 καταστάσεις.

4.5.2 Πειράματα με διαφορετικό αριθμό καταστάσεων

Στην παρούσα παράγραφο θα παρουσιάσουμε πειραματικά αποτελέσματα χρησιμοποιώντας διαφορετικό αριθμό καταστάσεων για νοήματα διαφορετι-

κής χρονικής δίαρκειας. Τα νοήματα θα τα χωρίσουμε σε HMMs ανάλογα με τον μέσο όρο του αριθμού των frames από τα δεδομένα εκπαίδευσης. Χρησιμοποιώντας διαφορετικό αριθμό καταστάσεων θα προσπαθήσουμε να εκμεταλλευτούμε την διαφορετική χρονική διάρκεια που έχουν τα νοήματα με σκοπό να βοηθηθούμε στην αναγνώριση. Στο Σχήμα 4.5 μπορούμε να δούμε τα αποτελέσματα για διάφορες τοπολογίες.



Σχήμα 4.5: Αποτελέσματα με διαφορετικό αριθμό καταστάσεων για νοήματα διαφορετικής χρονικής δίαρκειας.

Μπορούμε να παρατηρήσουμε ότι τα αποτελέσματα είναι παρόμοια με αυτά που είχαμε όταν χρησιμοποιήσαμε κοινό αριθμό καταστάσεων.

4.5.3 Πειράματα με το X-Model

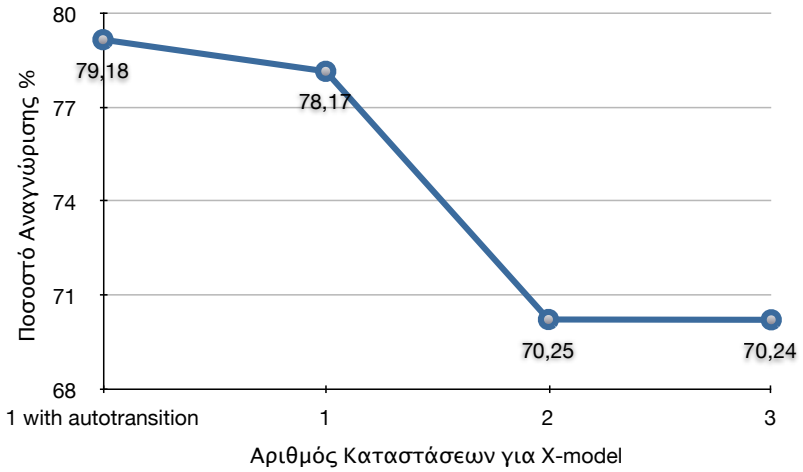
Μια άλλη τοπολογία που εφαρμόσαμε ήταν αυτή του μοντέλου X-Model. Η τοπολογία του X-Model είναι η εξής: Προσθέτουμε στην αρχή και στο τέλος του αρχικού μοντέλου μας άλλο ένα μοντέλο HMM που το ονομάζουμε X, με αριθμό καταστάσεων που τον ορίζουμε εμείς πειραματικά ένα παράδειγμα φαίνεται στο Σχήμα 4.6. Στόχος του μοντέλου X είναι να επιτύχουμε ανεξαρτησία από την αρχική και τελική θέση του νοήματος στην αναγνώριση.

Ο λόγος που το κάνουμε αυτό είναι επειδή είναι αρκετά συνηθισμένο για το ίδιο νόημα κάποιος νοηματιστής να ξεκινάει από διαφορετική αρχική θέση ή να καταλήγει σε διαφορετική τελική θέση ή να κάνει μια παραπάνω κίνηση είτε στην αρχή είτε στο τέλος του νοήματος με αποτέλεσμα να δημιουργείται πρόβλημα στην αναγνώριση. Η εμφάνιση του προβλήματος αυτού είναι πολύ

Σχήμα 4.6: Το Μοντέλο X-Model.

έντονη στην συνεχή νοηματική γλώσσα πού λόγω της εκτέλεσης συνεχόμενων νοημάτων, στο χρονικό διάστημα μεταξύ το τέλος ενός νοήματος και την αρχή του επόμενου έχουμε υποχρεωτικά την εισαγωγή μιας παραπάνω κίνησης των χεριών από την τελική θέση του πρώτου νοήματος στην αρχική θέση του επόμενου. Το πρόβλημα αυτό έχει ονομαστεί - Movement Epenthesis Problem- [33, 22].

Έτσι με το νέο αυτό μοντέλο επιτυγχάνουμε μια αρκετά ικανοποιητική επίλυση του προβλήματος αφού η αναγνώριση δεν επηρεάζεται τόσο πολύ από την αρχική και τελική θέση του νοήματος λόγο των μοντέλων X που έχουμε βάλει στην αρχή και στο τέλος του αρχικού μοντέλου. Στο Σχήμα 4.7 μπορούμε να δούμε πειραματικά αποτελέσματα σε σχέση με τον αριθμό των καταστάσεων για το X μοντέλο. Σε όλες τις περιπτώσεις εκτός από την πρώτη δεν επιτρέπεται η μετάβαση από μια κατάσταση στον εαυτό της έτσι ώστε να μην παραμένουμε στο X μοντέλο για πολύ χρόνο.

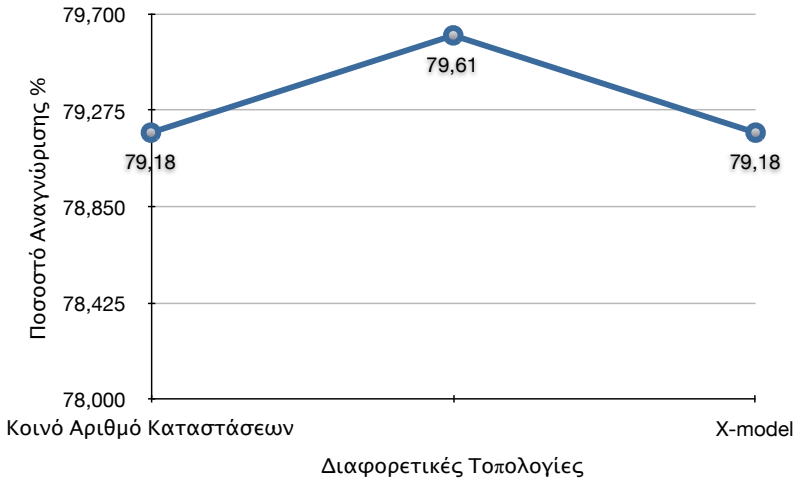


Σχήμα 4.7: Αποτελέσματα με X-Model για διαφορετικό αριθμό καταστάσεων για το X μοντέλο

Μπορούμε να παρατηρήσουμε η ότι η τοπολογία που δίνει τα καλύτερα αποτελέσματα για το X-Model είναι όταν χρησιμοποιούμε ένα HMM με

μια κατάσταση και επιτρέπουμε μεταβάσεις στον εαυτό της. Αυτό είναι πολύ λογικό γιατί με αυτό τον τρόπο δεν αναγκάζουμε το μοντέλο μας να μείνει συγκεκριμένο αριθμό frames στο μοντέλο X αλλά να κάνει μόνο του την επιλογή ανάλογα με τα δεδομένα. Όμως βλέπουμε ότι παρόλα αυτά τα αποτελέσματα είναι παρόμοια με τα προηγούμενα αυτό συμβαίνει επειδή τα βίντεο που είχαμε στην βάση δεδομένων ήταν κομμένα ακριβώς την στιγμή που άρχιζε και τελείωνε το νόημα με αποτέλεσμα να μην εμφανίζεται το πρόβλημα στο οποίο αναφερθήκαμε παραπάνω. Για την αναγνώριση συνεχούς λόγου τα αποτελέσματα είναι πιθανό να είναι καλύτερα αφού η εμφάνιση του -Movement Epenthesis Problem- είναι πολύ συχνή.

Στο Σχήμα 4.8 μπορούμε να δούμε τα συνολικά αποτελέσματα αναγνώρισης χρησιμοποιώντας μόνο το κανάλι για την θέση-κίνηση των χεριών για τις 3 διαφορετικές τοπολογίες HMM που χρησιμοποιήσαμε.



Σχήμα 4.8: Συνολικά αποτελέσματα για τις 3 τοπολογίες που εφαρμόσαμε για το κανάλι της θέσης-κίνησης των χεριών

4.6 Κανάλι για την Χειρομορφή των χεριών

Το είδος της χειρομορφής παίζει πολύ σημαντικό ρόλο στην νοηματική γλώσσα. Πολλές φορές είναι το μοναδικό χαρακτηριστικό που αλλάζει με αποτέλεσμα να είναι απαραίτητη η πληροφορία αυτή για το διαχωρισμό νοημάτων.

Μέχρι στιγμής έχει γίνει πολύ μικρή έρευνα για την αναγνώριση του είδους της χειρομορφής και ο λόγος είναι επειδή η εξαγωγή των χαρακτηριστικών για τον προσδιορισμό της είναι διαδικασία αρκετά δύσκολη. Οι

περισσότεροι αλγόριθμοι αναγνώρισης έχουν χρησιμοποιήσει 2D δεδομένα όπου η εξαγωγή των κατάλληλων χαρακτηριστικών γίνεται πολύ δύσκολη. Όμως ακόμα και αλγόριθμοι που χρησιμοποιούν 3D δεδομένα δεν έχουν καταφέρει να προσδιορίσουν την χειρομορφή με ακρίβεια απαραίτητη για την αναγνώριση της νοηματικής γλώσσας, εκτός και αν χρησιμοποίησαν Data Gloves όπου το πρόβλημα ξεφεύγει από την όραση υπολογιστών αφού η εξαγωγή χαρακτηριστικών δεν γίνεται μέσω της επεξεργασία εικόνας.

Στην βάση δεδομένων που χρησιμοποιήσαμε στα πλαίσια αυτής της διπλωματικής χρησιμοποιήθηκαν 10 διαφορετικές χειρομορφές, παρά τον μικρό αριθμό διαφορετικών κλάσεων η εξαγωγή χαρακτηριστικών για την σωστή αναγνώριση ήταν πολύ δύσκολη λόγο του ότι χρησιμοποιήσαμε μια μόνο κάμερα για την καταγραφή τους, περισσότερες λεπτομέρειες θα αναφέρουμε παρακάτω. Στο Σχήμα 4.9 μπορούμε να δούμε τις 10 διαφορετικές χειρομορφές.

Σχήμα 4.9: Οι 10 διαφορετικές χειρομορφές που χρησιμοποιήθηκαν στα πλαίσια αυτής της διπλωματικής

Όπως μπορούμε να δούμε και στην ενότητα 2.5 για την μοντελοποίηση της νοηματικής γλώσσας χρησιμοποιήσαμε ένα κανάλι πληροφορίας για την χειρομορφή για κάθε χέρι ξεχωριστά. Πιο συγκεκριμένα τα κανάλια αυτά αποτελούνται από τα εξής διανύσματα χαρακτηριστικών:

- Region Based
- Fourier Descriptors
- Moments

- Συντελεστές Cepstrum

Για περισσότερες λεπτομέρειες για τα διανύσματα χαρακτηριστικών (βλ. Κεφάλαιο 3).

4.6.1 Μοντελοποίηση της χειρομορφής των χεριών

Για την μοντελοποίηση του καναλιού χειρομορφής εφαρμόσαμε δυο διαφορετικές μοντελοποιήσεις. Στην πρώτη φτιάξαμε 10 μοντέλα ένα μοντέλο για κάθε διαφορετική χειρομορφή ενώ στην δεύτερη μοντελοποιήσαμε το κάθε νόημα ξεχωριστά χρησιμοποιώντας την πληροφορία για την χειρομορφή. Στις επόμενες παραγράφους θα αναλύσουμε τους δυο παραπάνω τρόπους μοντελοποίησης.

Μοντελοποίηση διαφορετικών χειρομορφών

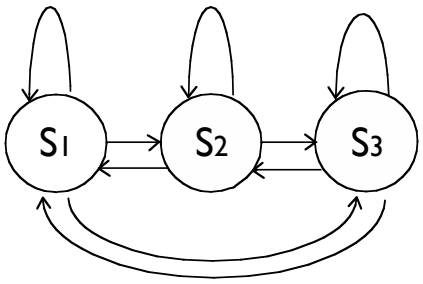
Στόχος της μοντελοποίησης αυτής ήταν η δημιουργία ενός μοντέλου για κάθε διαφορετική χειρομορφή ώστε γνωρίζοντας το είδος της χειρομορφής και σε συνδυασμό με την πληροφορία από το κανάλι της θέσης να μπορούμε να επιτύχουμε μια σωστή αναγνώριση. Συναντήσαμε ένα πολύ σημαντικό πρόβλημα. Λόγω του ότι τα δεδομένα μας είχαν προκύψει από μόνο μια κάμερα λόγο της δισδιάστατης προβολής κάθε χειρομορφή περιγράφεται από περισσότερα το ενός σχήματα ανάλογα με τον προσανατολισμό της παλάμης (βλ. ενότητα 2.6).

Έτσι για την μοντελοποίηση χρησιμοποιήσαμε ένα μη εργοδικό HMM (όπου δηλαδή επιτρέπονται μεταβάσεις από όλες τις καταστάσεις σε όλες βλ. Σχήμα 4.10) με αριθμό καταστάσεων τον αριθμό όλων των πιθανών δυσδιάστατων προβολών του χεριού κάθε χειρομορφής. Έτσι κάθε κατάσταση του HMM αντιπροσώπευε μια δισδίαστατη προβολή του χεριού.

Μοντελοποίηση νοημάτων με την πληροφορία για την χειρομορφή

Με την μοντελοποίηση αυτή δημιουργούμε ένα μοντέλο για κάθε διαφορετικό νόημα χρησιμοποιώντας την πληροφορία για την χειρομορφή. Βασιζόμενοι στην ικανότητα των HMM να αναγνωρίζουν ακολουθίες δεδομένων και την ιδιότητα τους να παραμένουν στην ίδια κατάσταση για περισσότερα του ενός χρονικά διαστήματα, χρησιμοποιήσαμε ένα εργοδικό μοντέλο HMM με αριθμό καταστάσεων και τοπολογία που καθορίσαμε πειραματικά.

Όπως αναφέραμε παραπάνω λόγω της αλλαγής του προσανατολισμού της παλάμης εξαιτίας της κίνησης των χεριών σε ένα νόημα εμφανίζεται μια χρονική μεταβολή της δισδίαστατης προβολής της χειρομορφής (ακόμα και εάν η



Σχήμα 4.10: Μη εργοδικό μοντέλο HMM 3 καταστάσεων

χειρομορφή σε ένα νόημα παραμένει σταθερή) την οποία εκμεταλλευτήκαμε για την μοντελοποίηση του κάθε νοήματος. Ουσιαστικά με αυτό τον τρόπο μοντελοποίησης αποφύγαμε το πρόβλημα που αναφέραμε προηγούμενως και ταυτόχρονα το εκμεταλλευτήκαμε για την μοντελοποίηση.

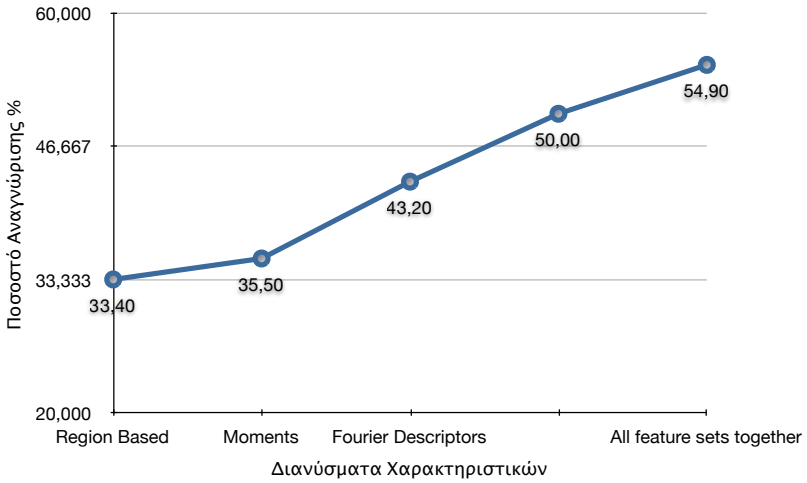
4.7 Πειραματικά αποτελέσματα για το κανάλι της χειρομορφής

Για όλα τα πειραματικά αποτελέσματα που θα ακολουθήσουν θα χρησιμοποιήσουμε λεξιλόγιο 100 λέξεων, ο χωρισμός σε δεδομένα εκπαίδευσης και ελέγχου γίνεται με τον τρόπο που υποδείξαμε στην ενότητα 4.2.3 όπως και για τα κριτήρια απόδοσης του συστήματος που χρησιμοποιήσαμε μπορούμε να ανατρέξουμε στην ενότητα 4.2.4.

4.7.1 Μοντελοποίηση κάθε χειρομορφής

Στην παρούσα παράγραφο θα παρουσιάσουμε τα αποτελέσματα αναγνώρισης της χειρομορφής μοντελοποιώντας κάθε διαφορετική χειρομορφή ξεχωριστά με ένα μη εργοδικό HMM όπως αναφέραμε παραπάνω. Τα αποτελέσματα φαίνονται στο Σχήμα 4.11.

Από τα αποτελέσματα του σχήματος 4.11 παρατηρούμε ότι δεν έχουμε καθόλου καλή αναγνώριση αν και οι διαφορετικές κλάσεις (χειρομορφές) είναι μόνο 10. Αυτό οφείλεται στο ότι είναι πολύ δύσκολη η εκπαίδευση του μη εργοδικού HMM μοντέλου έτσι ώστε κάθε κατάσταση να αντιπροσωπεύει μια δισδιάστατη προβολή της χειρομορφής γιατί λόγο της συνεχούς αλλαγής του προσανατολισμού της παλάμης εμφανίζονται πάρα πολλές διαφορετικές δισδιάστατες προβολές. Έτσι εγκαταλείψαμε αυτή την μοντελοποίηση.



Σχήμα 4.11: Αποτελέσματα αναγνώρισης του είδους της χειρομορφής για τα 4 διανύσματα χαρακτηριστικών και τον συνδυασμό τους.

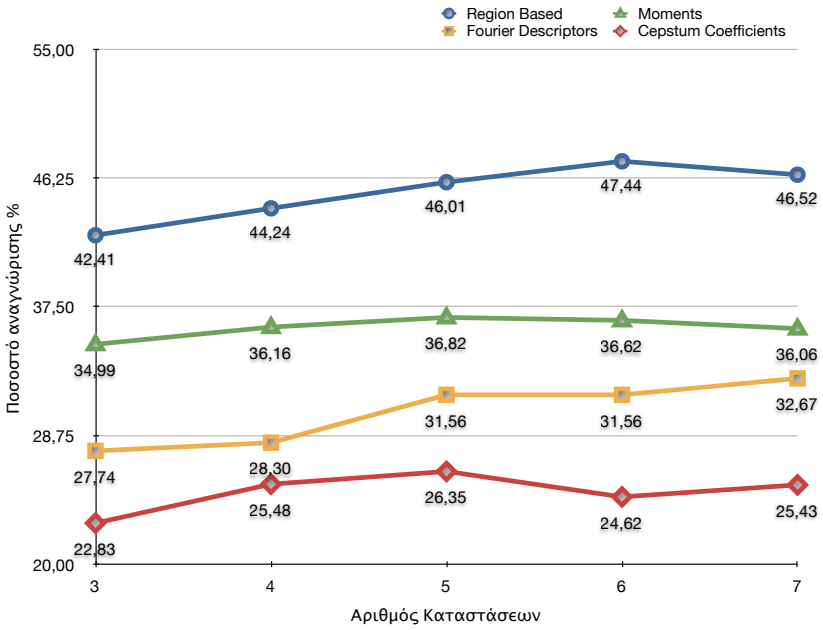
4.7.2 Μοντελοποίηση κάθε νοήματος

Στην παράγραφο αυτή θα εξετάσουμε ποία διανύσματα χαρακτηριστικών για την χειρομορφή μας δίνουν μεγάλη πληροφορία, μοντελοποιώντας κάθε νόημα ξεχωριστά με βάση τα διανύσματα αυτά.

Στο Σχήμα 4.12 μπορούμε να δούμε τα αποτελέσματα αναγνώρισης χρησιμοποιώντας ξεχωριστά τα 4 διαφορετικά διανύσματα χαρακτηριστικών (Region Based, Fourier Descriptors, Moments, Cepstrum Coefficients) χρησιμοποιώντας κοινό αριθμό καταστάσεων για όλα τα HMM μοντέλα.

Όπως βλέπουμε τα καλύτερα αποτελέσματα τα έχουμε με τα Region Based χαρακτηριστικά στην δεύτερη θέση έρχονται τα Moments μετά τα Fourier Descriptors και τέλος οι συντελεστές Cepstrum. Τα αποτελέσματα μπορεί να μην φαίνονται ικανοποιητικά αλλά αν σκεφτούμε ότι έχουμε 10 διαφορετικές χειρομορφές και προσπαθούμε να προσδιορίσουμε ποίο νόημα από τα 100 του λεξιλογίου εκτελείται χρησιμοποιώντας μόνο την πληροφορία για την χειρομορφή διαπιστώνουμε ότι έχουμε επιτύχει σε αρκετά ικανοποιητικό βαθμό την μοντελοποίηση της μεταβολής της δισδιάστατης προβολής της χειρομορφής.

Στην συνέχεια θα χρησιμοποιήσουμε διαφορετικό αριθμό καταστάσεων για νοήματα διαφορετικής χρονικής διάρκειας. Θα χωρίσουμε τα νοήματα σε HMM ανάλογα με τον μέσο όρο του αριθμού των frames βασιζόμενοι στα δεδομένα εκπαίδευσης. Έτσι χρησιμοποιώντας διαφορετικό αριθμό καταστάσεων θα προσπαθήσουμε να επιτύχουμε καλύτερη αναγνώριση εκμεταλλευόμενοι τη διαφορετική χρονική διάρκεια που έχουν τα νοήματα. Όμως όπως μπορο-



Σχήμα 4.12: Αποτελέσματα αναγνώρισης για τα 4 διανύσματα χαρακτηριστικών χρησιμοποιώντας διαφορετικό αριθμό κοινών καταστάσεων σε όλα τα μοντέλα.

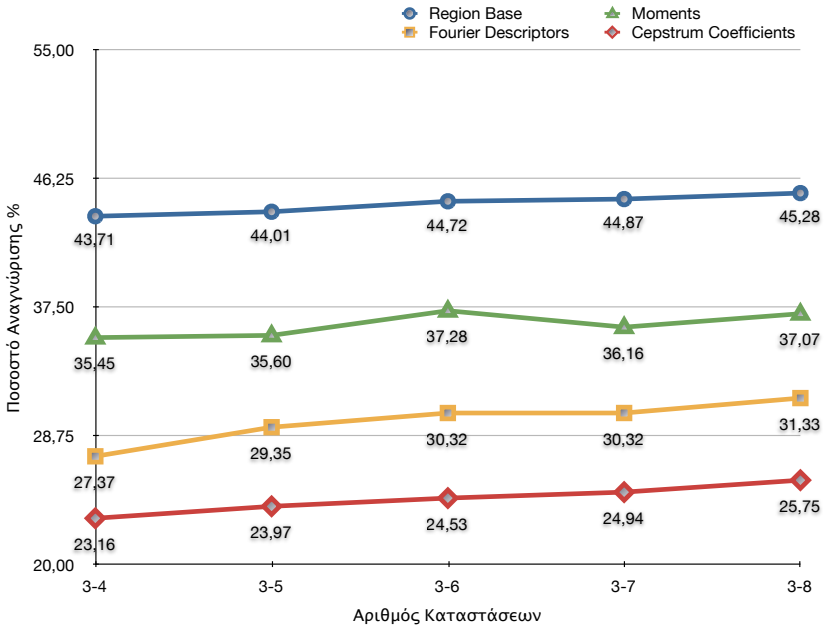
ύμε να διαπιστώσουμε από τα πειραματικά αποτελέσματα που φαίνονται στο Σχήμα 4.13 δεν βοηθηθήκαμε καθόλου έτσι στο κανάλι για την χειρομορφή θα χρησιμοποιήσουμε κοινό αριθμό καταστάσεων για όλα τα μοντέλα.

Χρησιμοποιώντας και τα 4 διανύσματα χαρακτηριστικών και με κοινό αριθμό καταστάσεων για όλα τα μοντέλα HMM παίρνουμε τα αποτελέσματα που φαίνονται στο Σχήμα 4.14.

4.8 Συμπεράσματα

Στο κεφάλαιο αυτό ορίσαμε τα απλά HMM και εξηγήσαμε τους λόγους που επιλέξαμε αυτά τα στατιστικά μοντέλα για την μοντελοποίηση. Επίσης αναλύσαμε τα δύο κανάλια πληροφορίας το πρώτο για την θέση-κίνηση και το δεύτερο για την χειρομορφή των χεριών.

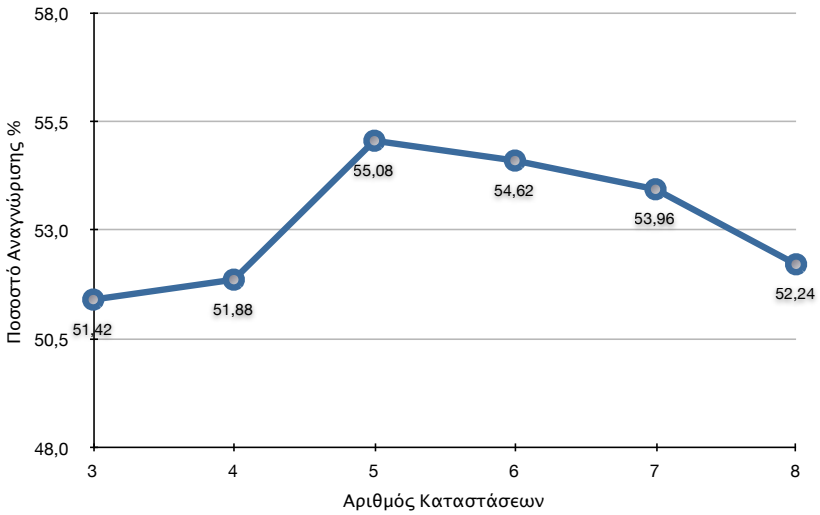
Ασχοληθήκαμε με την τοπολογία των HMM μοντέλων δοκιμάζοντας ποικίλες τοπολογίες. Πειραματιστήκαμε με ένα κοινό τύπο HMM για όλα τα μοντέλα μας με κοινό αριθμό καταστάσεων και καταλήξαμε ότι σε αυτή την περίπτωση ο ιδανικότερος αριθμός καταστάσεων που πρέπει να χρησιμοποιήσουμε είναι 6 καταστάσεις για το κανάλι της θέσης-κίνησης επιτυγχάνοντας



Σχήμα 4.13: Αποτελέσματα αναγνώρισης για τα 4 διανύσματα χαρακτηριστικών χρησιμοποιώντας διαφορετικό αριθμό καταστάσεων για νοήματα διαφορετικής χρονικής διάρκειας.

ποσοστό αναγνώρισης 79.2% και 5 καταστάσεις για το κανάλι της χειρομορφής επιτυγχάνοντας ποσοστό αναγνώρισης 55,1%. Επίσης πειραματιστήκαμε χωρίζοντας τα νοήματα μας σε HMMs με διαφορετικό αριθμό καταστάσεων ανάλογα με την χρονική τους διάρκεια, καταλήγοντας για το κανάλι της θέσης-κίνησης σε 5 διαφορετικούς τύπους HMMs, των τριών μέχρι επτά καταστάσεων όπου και είχαμε τα καλύτερα αποτελέσματα 79.7% ενώ για το κανάλι της χειρομορφής παρατηρήσαμε ότι είχαμε μείωση της αναγνώρισης οπότε αποφασίσαμε την χρησιμοποίηση 5 καταστάσεων για όλα τα μοντέλα που είχαμε και τα καλύτερα αποτελέσματα.

Τέλος για το κανάλι της θέσης-κίνησης πειραματιστήκαμε με το X-Model και καταλήξαμε ότι η ιδανική τοπολογία που πρέπει να χρησιμοποιήσουμε για το X μοντέλο είναι ένα HMM μιας κατάστασης όπου θα επιτρέπεται η μετάβαση στον εαυτό της επιτυγχάνοντας ποσοστό επιτυχίας 79.18%.



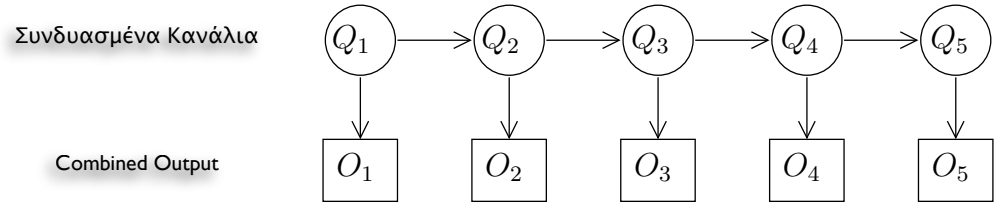
Σχήμα 4.14: Αποτελέσματα αναγνώρισης χρησιμοποιώντας και τα 4 διανύσματα χαρακτηριστικών μαζί με κοινό αριθμό καταστάσεων για όλα τα μοντέλα.

Κεφάλαιο 5

Fusion καναλιών θέσης - κίνησης και χειρομορφής των χεριών

Στο προηγούμενο κεφάλαιο αναλύσαμε τα δύο διαφορετικά κανάλια πληροφορίας της θέσης - κίνησης και της χειρομορφής των χεριών. Για την τελική αναγνώριση θα πρέπει να συνδυάσουμε τα δύο αυτά κανάλια. Με τα απλά HMM που έχουμε περιγράψει μέχρι στιγμής αυτό που θα μπορούσαμε να κάνουμε είναι να συγχωνεύσουμε τα δύο αυτά κανάλια σε ένα κοινό HMM (βλ Σχήμα 5.1). Όμως με την συγχώνευση αυτή υποθέτουμε ότι τα δυο κανάλια αυτά είναι συγχρονισμένα δηλαδή εξελίσσονται ταυτόχρονα περνώντας από την ίδια κατάσταση την ίδια χρονική στιγμή. Η υπόθεση αυτή όμως δεν ισχύει γενικά αφού είναι πολύ πιθανό την στιγμή που αλλάζει θέση το χέρι η δισδίαστατη προβολή της χειρομορφής να παραμένει σταθερή ή και το ανάποδο. Ένα παράδειγμα όπου θα είχαμε πρόβλημα στην αναγνώριση είναι το εξής: Έστω ότι εκτελείται ένα νόημα μόνο με το Strong χέρι όπου το χέρι εκτελεί μια κίνηση έτσι ώστε να μην αλλάζει ο προσανατολισμός της παλάμης με αποτέλεσμα η δισδίαστατη προβολή της χειρομορφής να παραμένει σταθερή και μετά την εκτέλεση της κίνησης του χεριού να γίνεται μια περιστροφή της παλάμης μεταβάλλοντας την προβολή της χειρομορφής αλλά χωρίς να αλλάζει θέση το χέρι, η περιστροφή του χεριού από το υπάρχων σύστημα είναι αδύνατον να προσδιοριστεί σωστά. Έτσι παρατηρούμε ότι θα χρειαστεί ένας διαφορετικός τρόπος για να κάνουμε fusion.

Σε αυτό το κεφάλαιο θα αναφέρουμε μερικές επεκτάσεις των HMM και στο τέλος θα παρουσιάσουμε τα πειραματικά αποτελέσματα της αναγνώρισης μετά το fusion.



Σχήμα 5.1: Ένα παράδειγμα ενός απλού HMM που αναγκάζει την συγχώνευσή των διαφορετικών καναλιών. Q_t είναι η κατάσταση όλων των καναλιών την χρονική στιγμή t, και O_t είναι η έξοδος όλων των καναλιών την χρονική στιγμή t.

5.1 Επέκταση των HMM

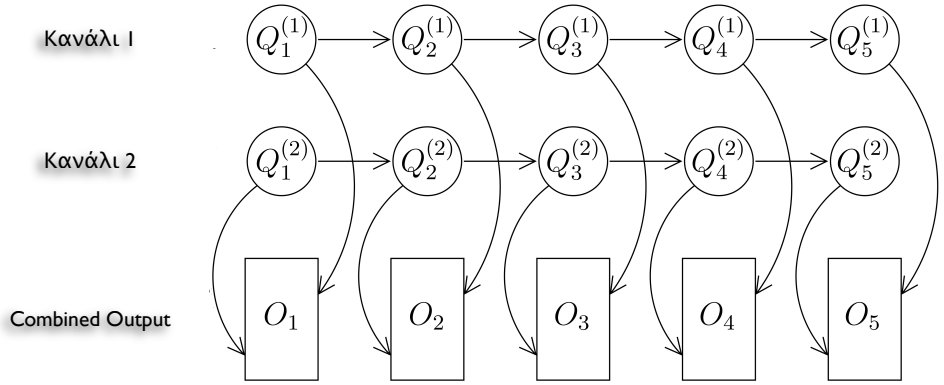
5.1.1 Factorial Hidden Markov Models (FHMMs)

Τα FHMMs [12] είναι μια γενίκευση των HMMs κατάλληλη για πολλές εφαρμογές όπου εμπλέκονται από 2 ή περισσότερα κανάλια πληροφορίας. Για κάθε κανάλι c όπου $1 \leq c \leq C$ η πιθανότητα μετάβασης από μια κατάσταση σε μια άλλη $a_{ij}^{(c)}$ και η πιθανότητα να έχουμε την παρατήρηση O στην κατάσταση i, $b_i^{(c)}(O)$ δεν εξαρτώνται από κανένα από τα υπόλοιπα κανάλια d, όπου $d \neq c$. Οι πιθανότητες παρατηρήσεων $b_i^{(c)}(O)$ για το κάθε κανάλι συνδυάζονται στο λεγόμενο -meta-state - και παράγεται μια συνολική πιθανότητα παρατήρησης. Το πώς ακριβώς συνδυάζονται οι πιθανότητες αυτές εξαρτάται από την εφαρμογή. Οι πιο συνηθισμένοι τρόποι είναι η πρόσθεση ή ο πολλαπλασιασμός τους.Ένα παράδειγμα μπορούμε να δούμε στο Σχήμα 5.2.

Επειδή η συνολική πιθανότητα παρατηρήσεων εξαρτάται από το meta-state η μέθοδος εκπαίδευσης που βασίζεται στον αλγόριθμο expectation maximization, ο οποίος υπολογίζει τις κατάλληλες παραμέτρους για ένα HMM θα είχε εκθετική πολυπλοκότητα σε σχέση με τον αριθμό των καναλιών. Για τον λόγο αυτό οι Ghahramani και Jordan περιέγραψαν μια μέθοδο η οποία χρειαζόταν πολυωνυμικό χρόνο και βασιζόταν στην mean-field theory [12].

5.1.2 Coupled Hidden Markov Models (CHMM)

Τα CHMMs [20] επιτρέπουν την αλληλεπίδραση διαφορετικών καναλιών στις πιθανότητες μετάβασης αλλά ταυτόχρονα αφήνουν κάθε κανάλι να έχει την δική του παρατήρηση. Για κάθε κανάλι c όπου $1 \leq c \leq C$ η πιθανότητα μετάβασης από την κατάσταση $i^{(c)}$ στην κατάσταση $j^{(c)}$, $a_{ij}^{(c)}$ δεν εξαρτάται μόνο από την κατάσταση $i^{(c)}$ αλλά από όλες τις καταστάσεις $i^{(d)}$ όπου $1 \leq d \leq C$.



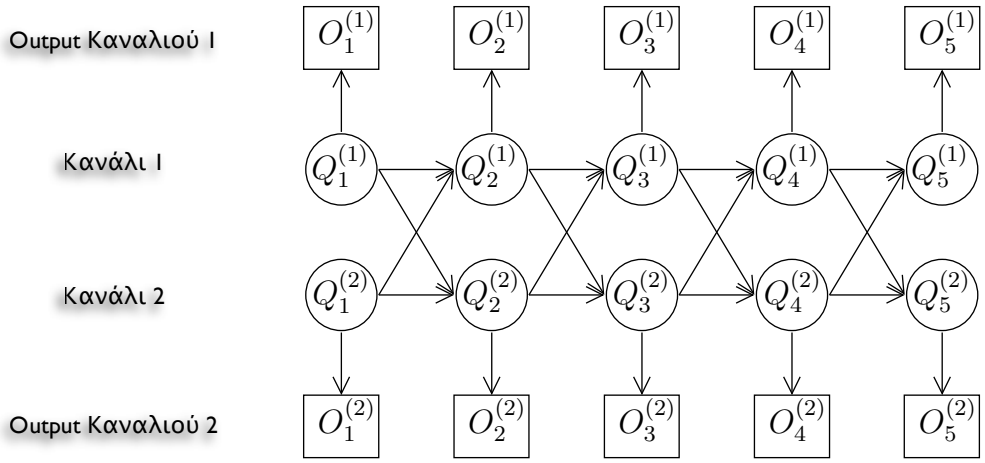
Σχήμα 5.2: Ένα παράδειγμα ενός FHMM όπου οι πιθανότητες μετάβασης από μια κατάσταση σε μια άλλη στο ίδιο κανάλι είναι ανεξάρτητες από όλα τα άλλα κανάλια, ενώ οι πιθανότητες παρατηρήσεων για κάθε κανάλι συνδυάζονται. $Q_t^{(c)}$ είναι η κατάσταση για το κανάλι c την χρονική στιγμή t και O_t είναι η συνδυασμένη έξοδος την χρονική στιγμή t για όλα τα κανάλια [32].

Όμως κάθε κανάλι έχει την δικιά του παρατήρηση οι οποίες δεν συνδυάζονται όπως γινόταν στα FHMM ένα παράδειγμα μπορούμε να δούμε στο Σχήμα 5.3.

Ουσιαστικά τα CHMM είναι το αντίθετο από τα FHMM. Στα FHMM οι πιθανότητες μετάβασης καταστάσεων είναι ανεξάρτητες για κάθε κανάλι ενώ οι πιθανότητες παρατηρήσεων συνδυάζονται για να προκύψει μια συνολική. Ενώ στα CHMM γίνεται το ακριβώς αντίθετο, δηλαδή η συνολική πιθανότητα παρατηρήσεων είναι οι πιθανότητες παρατηρήσεων για κάθε κανάλι ξεχωριστά ενώ οι πιθανότητες μετάβασης καταστάσεων εξαρτώνται από όλα τα κανάλια. Οι M. Brand, N. Oliver και A. Pentland περιέγραψαν μια μέθοδο εκπαίδευσης του μοντέλου σε πολυωνυμικό χρόνο και περιέγραψαν τα πλεονεκτήματα των CHMMs σε σχέση με τα απλά HMMs[20].

5.1.3 Product Hidden Markov Models (PHMM)

Το PHMM [16, 11] είναι ένας συνδυασμός των δύο παραπάνω δομών HMM (FHMM, CHMM). Οι πιθανότητες μετάβασης και οι πιθανότητες των παρατηρήσεων εξαρτώνται από όλα τα κανάλια πληροφορίας (streams) και καμία από τις δύο δεν είναι ανεξάρτητες όπως ισχύει για τα FHMM, CHMM. Κάθε κατάσταση του αποτελείται από τον συνδυασμό των καταστάσεων των streams που έχουμε. Όπως μπορούμε να δούμε στο Σχήμα 5.4 οι καταστάσεις που βρίσκονται στην ίδια γραμμή έχουν την ίδια κατάσταση στο stream της θέσης - κίνησης ενώ οι καταστάσεις που βρίσκονται στην ίδια στήλη έχουν την ίδια



Σχήμα 5.3: Ένα παράδειγμα για ένα CHMM όπου οι πιθανότητες παρατηρήσεων είναι ανεξάρτητες για κάθε κανάλι ενώ οι πιθανότητες μετάβασης από μια κατάσταση σε μια άλλη εξαρτάται από όλα τα κανάλια. $Q_t^{(c)}$ είναι η κατάσταση για το κανάλι c την χρονική στιγμή t και $O_t^{(c)}$ είναι η έξοδος του καναλιού c την χρονική στιγμή t [32].

κατάσταση στο stream της χειρομορφής. Έτσι βλέπουμε ότι το stream θέσης - κίνησης έχει γίνει tied για τις καταστάσεις που βρίσκονται στην ίδια γραμμή ενώ το ίδιο έχει γίνει για το stream της χειρομορφής για τις καταστάσεις που βρίσκονται στην ίδια στήλη. Οι παράμετροι ενός PHMM είναι οι εξής:

$$\pi(i) = P(q_1 = i) \tag{5.1}$$

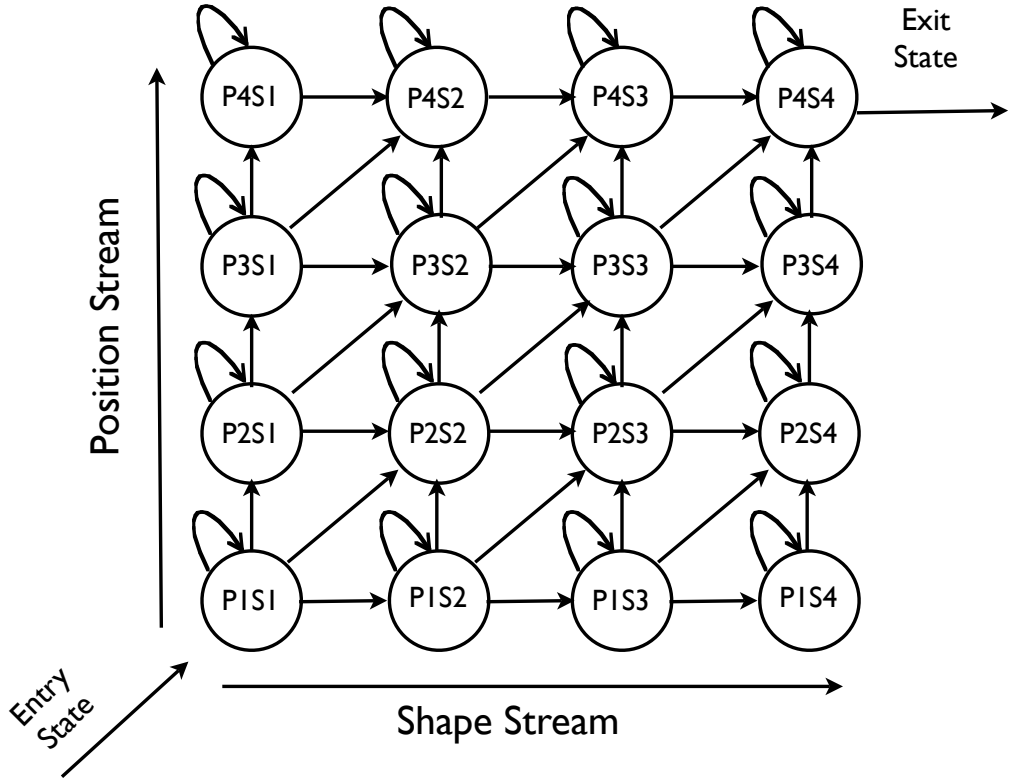
$$b_t(i) = P(O_t|q_t = i) \tag{5.2}$$

$$\alpha_{ji} = P(q_t = i|q_{t-1} = j) \tag{5.3}$$

Όπου O_t είναι οι παρατηρήσεις για όλα τα streams :

$$O_t = [(O_t^1)^T, ..., (O_t^S)^T]^T \tag{5.4}$$

Το μεγάλο πλεονέκτημα που αποκτούμε με τα PHMM είναι ότι δεν επιβάλλεται στα κανάλια να είναι συγχρονισμένα αφού στο πλέγμα αυτό, των HMM καταστάσεων υπάρχουν όλοι οι δυνατοί συνδυασμοί των καταστάσεων των δύο διαφορετικών καναλιών (streams), γεγονός πολύ βασικό για την αναγνώριση της νοηματικής γλώσσας όπως έχουμε αναφέρει και προηγουμένως.

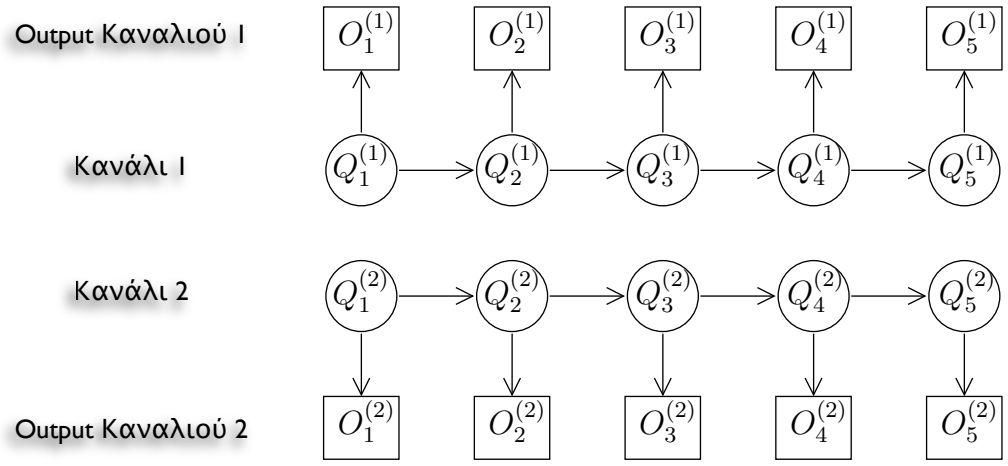


Σχήμα 5.4: Ένα παράδειγμα για ένα PHMM όπου οι πιθανότητες παρατηρήσεων και οι πιθανότητες μετάβασης από μια κατάσταση σε μια άλλη εξαρτώνται και από τα δύο κανάλια. Το P αναφέρεται στο stream της θέσης - κίνησης (Position) και το S αναφέρεται στο stream του σχήματος της χειρομορφής (Shape).

Επιπλέον με τα PHMM έχουμε την δυνατότητα να ρυθμίσουμε το πόσο ασύγχρονα θα μπορούν να είναι τα κανάλια μας κόβοντας καταστάσεις από το πλέγμα π.χ εάν κρατήσουμε μόνο την διαγώνιο του πλέγματος τότε υποχρεώνουμε τα κανάλια μας να είναι συγχρονισμένα αφού στην διαγώνιο σε όποια κατάσταση βρίσκεται το ένα κανάλι βρίσκεται και το άλλο.

5.1.4 Parallel Hidden Markov Models (PaHMMs)

Τα PaHMMs [32] μοντελοποιούν C κανάλια πληροφορίας χρησιμοποιώντας C ανεξάρτητα HMMs με ξεχωριστές πιθανότητες παρατηρήσεων για κάθε κανάλι. Σε αντίθεση με τα CHMMs η πιθανότητες μετάβασης καταστάσεων εξαρτώνται μόνο από καταστάσεις που ανήκουν στο ίδιο κανάλι πληροφορίας.



Σχήμα 5.5: Ένα παράδειγμα για ένα PaHMM όπου οι πιθανότητες παρατηρήσεων και μετάβασης από μια κατάσταση σε μια άλλη είναι ανεξάρτητες για κάθε κανάλι. $Q_t^{(c)}$ είναι η κατάσταση για το κανάλι c την χρονική στιγμή t και $O_t^{(c)}$ είναι η έξοδος του καναλιού c την χρονική στιγμή t [32].

Ένα παράδειγμα μπορούμε να δούμε στο Σχήμα 5.5.

Στα PaHMM κάνουμε την υπόθεση ότι κάθε κανάλι πληροφορίας εξελίσσεται τελείως ανεξάρτητα από τα άλλα έτσι δεν χρειάζεται τα κανάλια να είναι συγχρονισμένα πράγμα απαραίτητο για την αναγνώριση της νοηματικής γλώσσας. Βασιζόμενοι σε αυτή την υπόθεση η εκπαίδευση του HMM μοντέλου γίνεται για κάθε κανάλι ανεξάρτητα από τα υπόλοιπα και γίνεται συνδυασμός όλων των καναλιών την στιγμή της αναγνώρισης.

Υπάρχουν κάποια σημεία που πρέπει να λάβουμε υπόψη όταν ένα σύστημα χρησιμοποιεί PaHMMs για την Αναγνώριση ΝΓ. Πρώτον είναι το πότε και πώς θα γίνεται ο συνδυασμός των διαφορετικών καναλιών πληροφορίας. Δεύτερον τι κάνουμε όταν ένα κανάλι δεν περιέχει σημαντική πληροφορία πχ η πληροφορία που μας παρέχει το κανάλι του weak χεριού όταν εκτελείται ένα νόημα που χρησιμοποιεί μόνο το ένα χέρι (strong χέρι). Τέλος πως θα δεσμεύσουμε τον αλγόριθμο αναγνώρισης ώστε το αποτέλεσμα της αναγνώρισης για όλα τα κανάλια να είναι κοινό. Για παράδειγμα δεν θα είχε νόημα αν το κανάλι του weak χεριού και του strong χεριού αναγνώριζαν το κάθε ένα διαφορετικό νόημα. Παρακάτω θα αναφερθώ σε αυτά τα σημεία και θα αναφέρω τρόπους επίλυσης τους.

Συνδυασμός Καναλιών

Κάποια στιγμή κατά την διάρκεια της αναγνώρισης θα πρέπει να συνδυάσουμε την πληροφορία από τα διαφορετικά κανάλια πληροφορίας. Από εδώ και στο εξής θα γράφω τις πιθανότητες σε λογαριθμική μορφή. Έτσι το σύστημα αναγνώρισης θα πρέπει να βρει την μέγιστη πιθανότητα από τον συνδυασμό όλων των καναλιών πληροφορίας η οποία είναι:

$$\max_{Q^{(1)},..,Q^{(C)}} \{logP(Q^{(1)},...,Q^{(C)},O^{(1)},...,O^{(C)}|\lambda^{(1)},...,\lambda^{(C)})\} \tag{5.5}$$

όπου $Q^{(c)}$ είναι η ακολουθία καταστάσεων για το κανάλι c, $1 \leq c \leq C$, και $O^{(c)}$ η ακολουθία παρατηρήσεων για το κανάλι c με HMM $\lambda^{(c)}$. Επειδή τα κανάλια στα PaHMMs είναι ανεξάρτητα, ο συνδυασμός τους γίνεται με τον πολλαπλασιασμό των πιθανοτήτων για κάθε κανάλι ξεχωριστά (δηλαδή άθροισμα των log πιθανοτήτων) έτσι η εξίσωση 5.5 γίνεται :

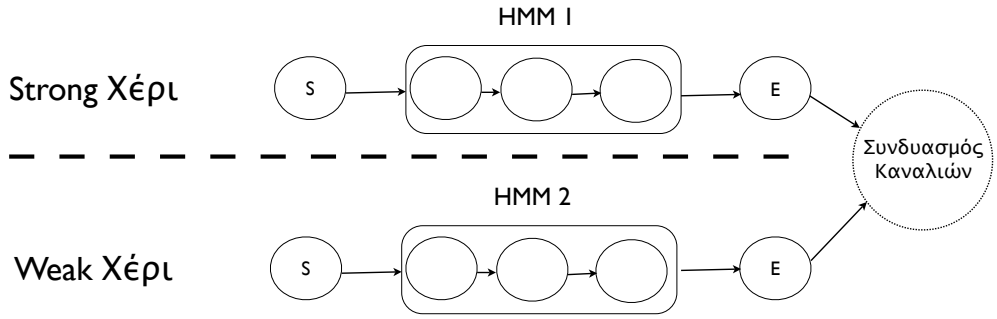
$$\begin{aligned}\max_{Q^{(1)},..,Q^{(C)}} \{logP(Q^{(1)},...,Q^{(C)},O^{(1)},...,O^{(C)}|\lambda^{(1)},...,\lambda^{(C)})\} = \\ \max_{Q^{(1)},..,Q^{(C)}} \{\sum_{c=1}^{C} logP(Q^{(c)},O^{(c)}|\lambda^{(c)})\}\end{aligned} \tag{5.6}$$

Στο Σχήμα 5.6 μπορούμε να δούμε ένα παράδειγμα συνδυασμού δύο διαφορετικών καναλιών πληροφορίας (το ένα κανάλι αναφέρεται στο strong χέρι και το άλλο στο weak χέρι). Όμως πρέπει να κάνουμε κάποιες αλλαγές για να λάβουμε υπόψη τα κανάλια που μας παρέχουν μικρή ή καθόλου πληροφορία.

Κανάλια με μικρή ή καθόλου πληροφορία

Πολλές φορές στην Ελληνική Νοηματική Γλώσσα υπάρχουν κανάλια που έχουν μικρή ή καθόλου πληροφορία. Ένα προφανές παράδειγμα είναι ή διαφορά μεταξύ νοημάτων που εκτελούνται με το ένα χέρι (strong χέρι) και αυτών που εκτελούνται και με τα δύο. Όταν εκτελείται ένα νόημα μόνο από το strong χέρι το weak χέρι δεν παίζει κανένα ρόλο με αποτέλεσμα η πληροφορία του καναλιού για το weak χέρι δεν μας είναι χρήσιμη. Επίσης το κανάλι που παρέχει πληροφορία για την θέση - κίνηση των χεριών είναι πιο αξιόπιστο από ότι το κανάλι που παρέχει πληροφορία για την χειρομορφή.

Έτσι μπορούμε να διαπιστώσουμε ότι σε κάθε κανάλι πληροφορίας πρέπει να δίνουμε βάρος ανάλογα με το πόσο σημαντική πληροφορία μας παρέχει. Έτσι ορίζουμε $\omega^{(c)}$ το βάρος για το κανάλι c με αποτέλεσμα η εξίσωση 5.6 μετατρέπεται στην:



Σχήμα 5.6: Ένα παράδειγμα συνδυασμού δυο διαφορετικών καναλιών. Τα δυο κανάλια λειτουργούν τελείως ανεξάρτητα και συνδυάζονται στο τέλος του νοήματος. S ορίζει την αρχή του νοήματος -word start- και E το τέλος - word end-.

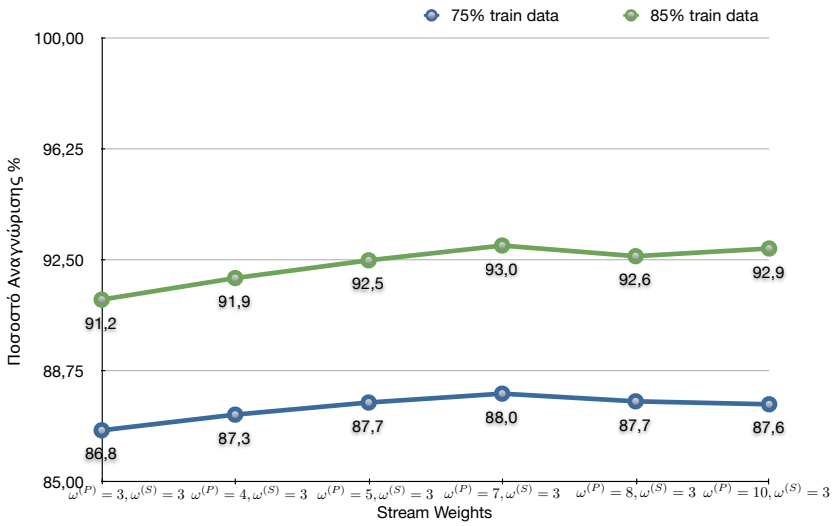
$$max_{Q^{(1)},...,Q^{(C)}}\{\sum_{c=1}^{C}\omega^{(c)}logP(Q^{(c)},O^{(c)}|\lambda^{(c)})\} \tag{5.7}$$

5.2 Πειραματικά αποτελέσματα χρησιμοποιώντας PaHMM για το Fusion

Σε αυτή την ενότητα θα παρουσιάσουμε τα πειραματικά αποτελέσματα χρησιμοποιώντας PaHMM για το Fusion των καναλιών θέσης - κίνησης και χειρομορφής των χεριών. Στα πειράματα που θα παρουσιάσουμε έχουμε κάνει χρήση δύο διαφορετικών set δεδομένων εκπαίδευσης 100 διαφορετικών νοημάτων. Στο πρώτο έχουμε χρησιμοποιήσει το 75% των συνολικών δεδομένων για εκπαίδευση και 25% για αξιολόγηση (600 δεδομένα εκπαίδευσης και 200 αξιολόγησης) και στο δεύτερο το 85% για εκπαίδευση και 15% για αξιολόγηση (680 δεδομένα εκπαίδευσης και 120 αξιολόγησης). Για τα κριτήρια αξιολόγησης βλ. ενότητα 4.2.4.

Αρχικά θα πειραματιστούμε με την επιλογή των συντελεστών βάρους για το κάθε κανάλι $\omega^{(c)}$.

Από τα αποτελέσματα του Σχήματος 5.7 βγάζουμε το συμπέρασμα ότι η καλύτερη επιλογή των stream weights είναι $\omega^{(P)} = 7$ και $\omega^{(S)} = 3$. Από εδώ και στο εξής αυτά τα stream weights θα χρησιμοποιούμε. Επίσης παρατηρούμε μια αρκετά μεγάλη αύξηση του ποσοστού αναγνώρισης όταν αυξάνουμε τα δεδομένα εκπαίδευσης γεγονός που υποδεικνύει ότι στο πρώτο set δεδομένων δεν έχουν εκπαιδευτεί αρκετά καλά τα μοντέλα μας πράγμα φυσιολογικό αφού σε αρκετά μοντέλα χρησιμοποιούμε μόνο 3 εκτελέσεις για



Σχήμα 5.7: Αποτελέσματα fusion για τα 2 διαφορετικά train sets χρησιμοποιώντας διαφορετικά stream weights για κάθε κανάλι

την εκπαίδευση τους.

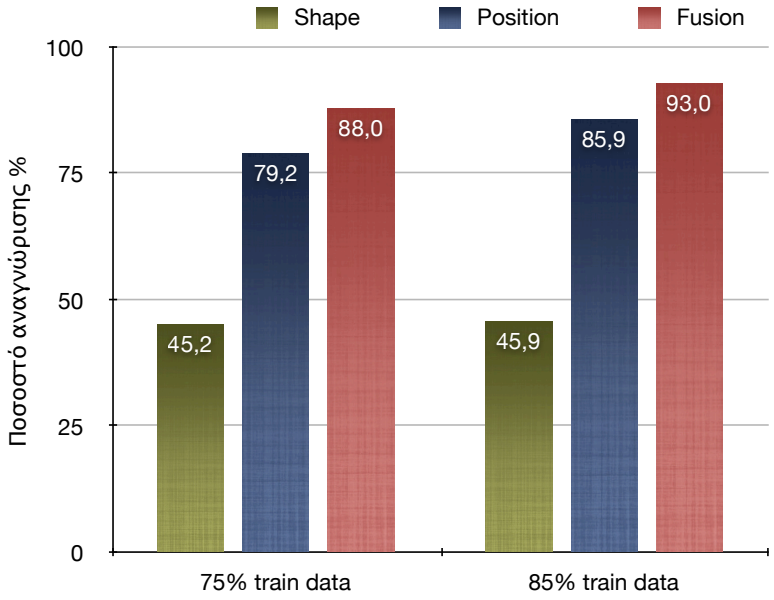
Στην συνέχεια παρουσιάζουμε στο Σχήμα 5.8 την αύξηση του ποσοστού αναγνώρισης λόγο του fusion των διαφορετικών καναλιών.

Παρατηρούμε μια σημαντική αύξηση της τάξεως του 7-9% κοινή και για τα δύο set δεδομένων που χρησιμοποιήσαμε που δείχνει ότι τα PaHMM είναι κατάλληλα για τον συνδυασμό των καναλιών της θέσης - κίνησης και της χειρομορφής των χεριών.

5.3 Πειραματικά αποτελέσματα χρησιμοποιώντας Product HMM για το Fusion

Στην ενότητα αυτή θα παρουσιάσουμε τα πειραματικά αποτελέσματα της αναγνώρισης κάνοντας fusion με Product HMM. Για την εκπαίδευση και αξιολόγηση του συστήματος θα χρησιμοποιήσουμε τα δύο διαφορετικά set δεδομένων που χρησιμοποιήσαμε και στην προηγούμενη ενότητα. Επίσης για stream weights θα χρησιμοποιούμε $\omega^{(P)} = 7$ και $\omega^{(S)} = 3$.

Ένα από τα πλεονεκτήματα των PHMM είναι ότι έχουν την δυνατότητα να περιορίζουν την ελευθερία των καναλιών ενώ αντίθετα με τα PaHMM τα κανάλια είναι τελείως ελεύθερα. Τον περιορισμό αυτό μπορούμε να τον επιτύχουμε αφαιρώντας καταστάσεις από το πλέγμα των HMM καταστάσεων πχ αφαιρώντας τις κατάλληλες καταστάσεις μπορούμε να αναγκάσουμε το



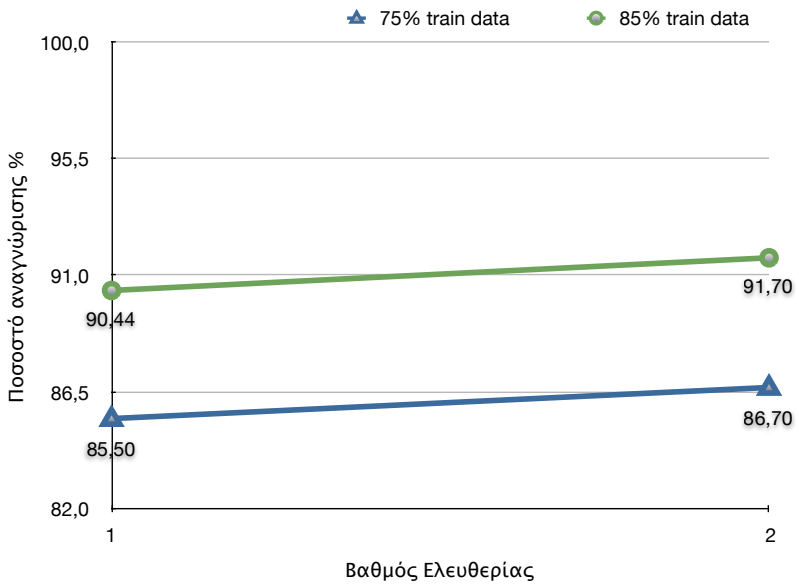
Σχήμα 5.8: Αύξηση του ποσοστού αναγνώρισης λόγο του fusion με PaHMM για τα 2 διαφορετικά train sets

κανάλι της θέσης - κίνησης να έχει την δυνατότητα να βρίσκεται το πολύ μια κατάσταση πιο μπροστά ή πιο πίσω από το κανάλι της χειρομορφής.

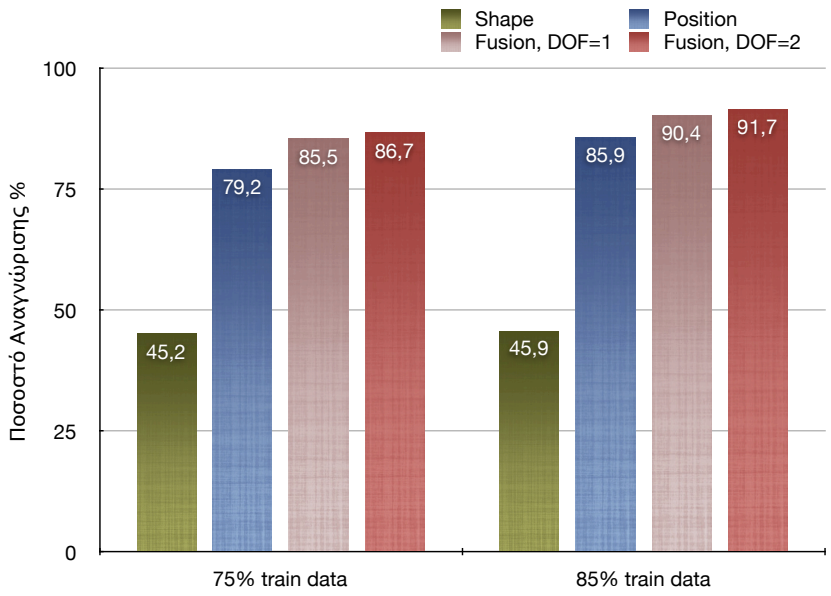
Όμως ένα από τα μειονεκτήματα των PHMM είναι ότι όσο αυξάνουμε τον βαθμό ελευθερίας των καναλιών χρειαζόμαστε όλο και μεγαλύτερο αριθμό δεδομένων για την σωστή εκπαίδευση των μοντέλων μας. Όπως θα δούμε και παρακάτω λόγω της μικρής βάσης δεδομένων που είχαμε καταφέραμε να εκπαιδεύσουμε τα μοντέλα μας με βαθμό ελευθερίας μέχρι 2 δηλαδή το ένα κανάλι να έχει την δυνατότητα να βρίσκεται το πολύ 2 καταστάσεις πιο μπροστά ή πιο πίσω από το άλλο.

Από τα αποτελέσματα του Σχήματος 5.9 παρατηρούμε μια αρκετά μεγάλη αύξηση του ποσοστού αναγνώρισης με την αύξηση των δεδομένων εκπαίδευση γεγονός που υποδεικνύει ότι η εκπαίδευση των μοντέλων μας χρησιμοποιώντας το πρώτο set δεδομένων δεν έχει γίνει αρκετά καλά πράγμα που παρατηρήσαμε και με το PaHMM.

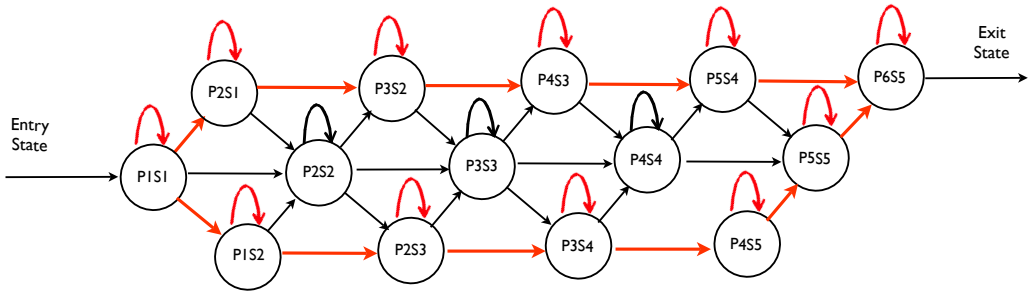
Στα Σχήματα 5.10 και 5.9 παρατηρούμε μια αύξηση ποσοστού αναγνώρισης της τάξεως του 1-2% λόγο της αύξησης του βαθμού ελευθερίας των PHMM. Λόγο της περιορισμένης βάσης δεδομένων δεν μπορέσαμε να εκπαιδεύσουμε τα μοντέλα μας με βαθμό ελευθερίας μεγαλύτερο από 2 όμως πιστεύουμε ότι με την αύξηση του βαθμού ελευθερίας θα επιτύχουμε ακόμα καλύτερα αποτελέσματα.



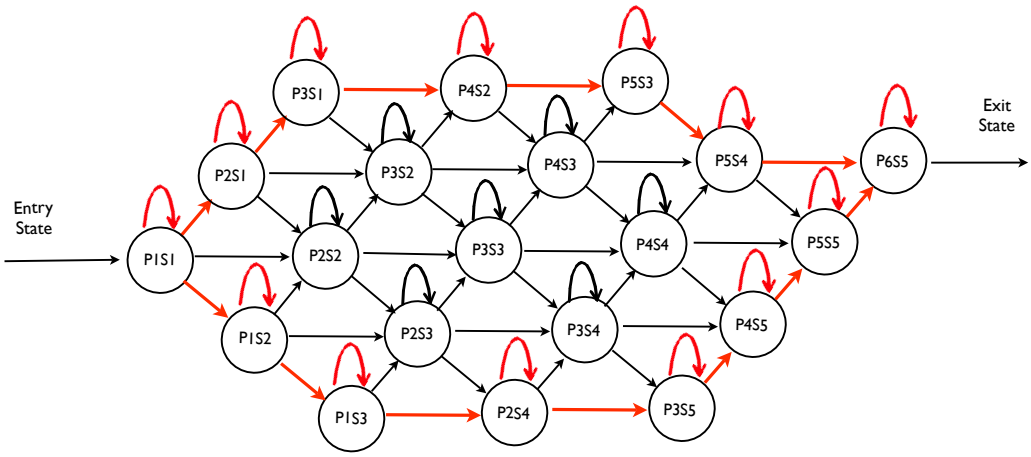
Σχήμα 5.9: Αποτελέσματα fusion με PHMM για τα 2 διαφορετικά train sets χρησιμοποιώντας βαθμό ελευθερίας 1 και 2



Σχήμα 5.10: Αύξηση του ποσοστού αναγνώρισης λόγο του fusion με PHMM με διαφορετικούς βαθμούς ελευθερίας (DOF) για τα 2 διαφορετικά train sets



Σχήμα 5.11: Αναπαράσταση με κόκκινη γραμμή των διαδρομών που ακολουθούνται περισσότερο από το σύνολο των δεδομένων αξιολόγησης πάνω στα PHMM με βαθμό ελευθερίας (DOF) 1



Σχήμα 5.12: Αναπαράσταση με κόκκινη γραμμή των διαδρομών που ακολουθούνται περισσότερο από το σύνολο των δεδομένων αξιολόγησης πάνω στα PHMM με βαθμό ελευθερίας (DOF) 2

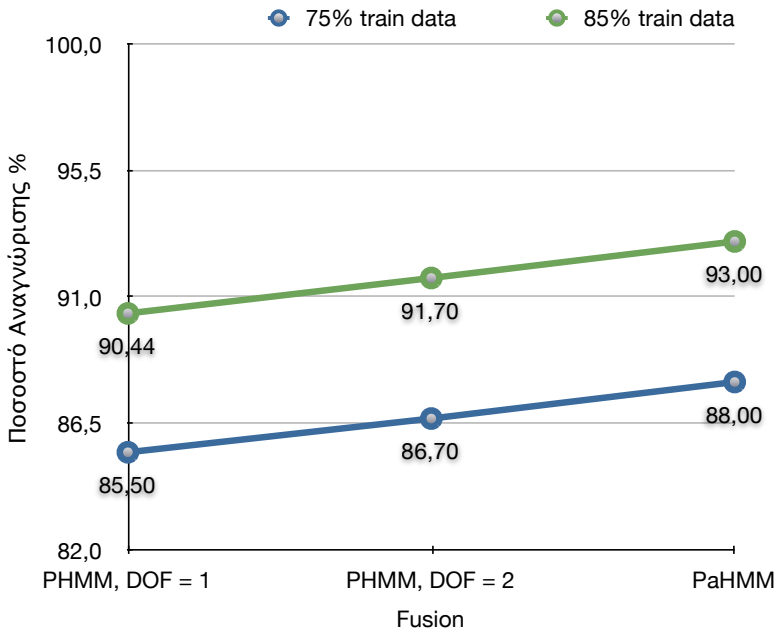
Στα Σχήματα 5.11, 5.12 έχουμε αναπαραστήσει με κόκκινη γραμμή τις διαδρομές πάνω στα PHMM με βαθμό ελευθερίας 1 και 2 αντίστοιχα που ακολουθούνται περισσότερο από το σύνολο των δεδομένων αξιολόγησης. Μπορούμε να παρατηρήσουμε ότι επιλέγονται κυρίως διαδρομές προς τα άκρα δηλαδή εξαντλείται ο βαθμός ελευθερίας που έχουμε επιλέξει για το μοντέλο μας. Έτσι από τα παραπάνω μπορούμε να βγάλουμε το συμπέρασμα ότι είναι πολύ πιθανό με ακόμα μεγαλύτερη αύξηση της ελευθερίας του μοντέλου μας να επιτύχουμε ακόμα μεγαλύτερη αύξηση του ποσοστού αναγνώρισης. Βέβαια εάν αυξήσουμε τον βαθμό ελευθερίας πάρα πολύ ουσιαστικά το PHMM μετατρέπεται σε PaHMM αφού τα κανάλια θα είναι τελείως ελεύθερα να κινούνται όπου θέλουν. Έτσι έχει μεγάλο ενδιαφέρον να διαπιστώσουμε εάν μπορούμε να βρούμε βαθμό ελευθερίας έτσι ώστε να επιτύχουμε καλύτερα αποτελέσματα από τα PaHMM.

5.4 Συγκριτικά αποτελέσματα κάνοντας Fusion με PaHMM και PHMM

Σε αυτή την ενότητα θα παρουσιάσουμε τα τελικά αποτελέσματα κάνοντας fusion με τα PHMM και τα PaHMM και θα κάνουμε μια σύγκριση των δυο αυτών διαφορετικών τεχνικών.

Στο Σχήμα 5.13 παρατηρούμε ότι τα καλύτερα αποτελέσματα τα είχαμε με τα PaHMM κατά 1,3% μεγαλύτερο ποσοστό αναγνώρισης από τα PHMM και για τα δύο sets. Όμως όπως τονίσαμε και στην προηγούμενη ενότητα πιστεύουμε ότι με κατάλληλη αύξηση του βαθμού ελευθερίας των PHMM θα επιτύχουμε καλύτερα αποτελέσματα από τα PaHMM.

Ένας ακόμη λόγος εκτός από αυτούς που αναφέραμε στην προηγούμενη ενότητα που πιστεύουμε ότι με τα PHMM θα έχουμε καλύτερα αποτελέσματα είναι ο εξής: Αν σκεφτούμε ότι το κανάλι θέσης - κίνησης των χεριών μοντελοποιεί την μεταβολή της θέσης των χεριών στην εικόνα σε σχέση με τον χρόνο και το κανάλι της χειρομορφής μοντελοποιεί την μεταβολή του σχήματος της δισδιάστατης προβολής της χειρομορφής σε σχέση με το χρόνο αντιλαμβανόμαστε ότι τα δύο αυτά κανάλια δεν είναι τελείως ανεξάρτητα μεταξύ τους. Βέβαια μπορούμε να έχουμε μεταβολή της θέσης των χεριών ενώ ταυτόχρονα το σχήμα της δισδιάστατης προβολής της χειρομορφής να παραμένει σταθερό ή και το ανάποδο όπως έχουμε αναφέρει σε παραδείγματα στα προηγούμενα κεφάλαια αλλά υπάρχουν και πάρα πολλές περιπτώσεις όπου υπάρχει εξάρτηση των δυο αυτών μεταβολών όπου η χρησιμοποίηση των PHMM για την μοντελοποίηση θα ήταν καλύτερη.



Σχήμα 5.13: Συγκριτικά αποτελέσματα για fusion με PaHMM και PHMM με διαφορετικούς βαθμούς ελευθερίας (DOF) για τα 2 διαφορετικά train sets

Κεφάλαιο 6

Συνεισφορές - Συμπεράσματα της Εργασίας και Κατευθύνσεις για Μελλοντική Έρευνα

Το κεφάλαιο αυτό αποτελεί την κατακλείδα της όλης διπλώματικής. Ανακεφαλαιώνονται τα βασικά σημεία της παρουσίασης που προηγήθηκε, παρουσιάζονται τα συμπεράσματα που προκύπτουν και γίνεται διερεύνηση ποικίλλων θεμάτων σχετικών με τις προοπτικές που αναδεικνύονται και τις κατευθύνσεις προς τις οποίες θα μπορούσε να κινηθεί η μελλοντική σχετική έρευνα.

6.1 Ανακεφαλαίωση

Στα πλαίσια της διπλωματικής αυτής έγινε μια προσέγγιση για την ανάπτυξη ενός συστήματος για την αναγνώριση μεμονωμένων νοημάτων της ελληνικής νοηματικής γλώσσας. Το σύστημα αυτό αποτελείται από τρία στάδια, το στάδιο της μοντελοποίησης, το στάδιο της εκπαίδευσης κάθε καναλιού πληροφορίας ξεχωριστά και το στάδιο του συνδυασμού των καναλιών (stream fusion) για τα τελικά αποτελέσματα της αναγνώρισης.

- Για την μοντελοποίηση δημιουργήσαμε ένα μοντέλο για κάθε νόημα. Πιο συγκεκριμένα κάθε μοντέλο το χώρισα σε 4 διαφορετικά κανάλια πληροφορίας. Το πρώτο κανάλι αναφέρεται στην θέση - κίνηση του Strong Hand, το δεύτερο στο είδος της χειρομορφής που χρησιμοποιήθηκε από το Strong Hand, το τρίτο στην θέση - κίνηση του Weak Hand και το τέταρτο στο είδος της χειρομορφής που χρησιμοποιήθηκε από το Weak Hand. Ένα παράδειγμα μπορούμε να δούμε στο Σχημα 2.8.

- Για την εκπαίδευση κάθε καναλιού (stream) ξεχωριστά χρησιμοποιήθηκαν απλά HMM μοντέλα. Η επιλογή των HMM βασίστηκε πρώτον στο ότι έπρεπε να ληφθεί υπόψη ότι εάν ένας νοηματιστής εκτελέσει το ίδιο νόημα δύο φορές οι εκτελέσεις δεν θα είναι ακριβώς οι ίδιες ακόμα και αν ο νοηματιστής προσπαθήσει γι αυτό. Έτσι τα μοντέλα μας δεν θα έπρεπε να επηρεάζονται από αυτή την μεταβλητότητα με αποτέλεσμα να έπρεπε να χρησιμοποιήσουμε κάποιου είδους στατιστικά μοντέλα. Δεύτερον στην ικανότητα τους να αναγνωρίζουν ακολουθίες δεδομένων με αποτέλεσμα να μπορούν να αναγνωρίσουν ακολουθίες εικόνων, που έχουν δεσμευτεί σε κοντινά μεταξύ τους χρονικά διαστήματα κατά την διάρκεια της κίνησης των χεριών. Τέλος στην ικανότητα τους να παραμένουν στην ίδια κατάσταση για περισσότερα του ενός χρονικά διαστήματα, με αποτέλεσμα να έχουν την δυνατότητα να αναγνωρίζουν διαφορετικού μήκους ακολουθίες εικόνων, οι οποίες αντιστοιχούν στην ίδια λέξη.

  Για το κανάλι θέσης - κίνησης χρησιμοποιήσαμε ένα left-right HMM 6 καταστάσεων επιτυγχάνοντας ποσοστά αναγνώρισης για δύο set δεδομένων εκπαίδευσης 75% και 85% των συνολικών δεδομένων που είχαμε στην διάθεση μας 79,2% και 85,9% αντίστοιχα.

  Για το κανάλι της χειρομορφής χρησιμοποιήσαμε ένα left-right HMM 5 καταστάσεων επιτυγχάνοντας ποσοστά αναγνώρισης για τα δύο παραπάνω set δεδομένων εκπαίδευσης 45,2% και 45,9% αντίστοιχα.

- Για τον συνδυασμό των καναλιών (streams fusion) χρησιμοποιήσαμε δύο διαφορετικές επεκτάσεις των HMM τα Parallel HMM και τα Product HMM. Με τα PaHMM επιτύχαμε ποσοστό αναγνώρισης για τα δύο παραπάνω set δεδομένων εκπαίδευσης 88% και 93% αντίστοιχα ενώ με τα PHMM 86,7% και 90,4%αντίστοιχα. Όμως με τα PHMM κάναμε πειράματα με βαθμό ελευθερίας μέχρι 2 λόγο της μικρής βάσης δεδομένων που είχαμε. Για τους λόγους που έχουμε αναφέρει στο προηγούμενο κεφάλαιο ελπίζουμε σε καλύτερα αποτελέσματα με τα PHMM με την κατάλληλη αύξηση του βαθμού ελευθερίας των καναλιών.

6.2 Μελλοντική Έρευνα

Η μελλοντική έρευνα θα μπορούσε να κινηθεί σε τουλάχιστον τρείς αξονες. Ο πρώτος αφορά τη μοντελοποίηση της ΝΓ χρησιμοποιώντας φωνήματα, ο δεύτερος στην ολοκληρωμένη αξιολόγηση των PHMM για το fusion διαφορετικών καναλιών πληροφορίας και ο τρίτος στην αναγνώριση συνεχής νοηματικής γλώσσας.

6.2.1 Μοντελοποίηση

Η μοντελοποίηση που χρησιμοποιήσαμε σε αυτή τη διπλωματική είναι εφικτή για μικρό εύρος λεξιλογίου. Με την αύξηση του εύρους του λεξιλογίων έχουμε ανάλογη αύξηση των μοντέλων μας αφού για κάθε λέξη έχουμε ένα μοντέλο με αποτέλεσμα όταν έχουμε ένα πολύ μεγάλο λεξιλόγιο ο αριθμός των μοντέλων μας θα είναι πολύ μεγάλος έτσι θα χρειαζόμαστε μια τεράστια βάση δεδομένων για την εκπαίδευση των μοντέλων μας πράγμα ανέφικτο. Έτσι βασιζόμενοι στην εμπειρία μας από την αναγνώριση φωνής μια μελλοντική έρευνα θα είναι η χρησιμοποίηση φωνημάτων για την μοντελοποίηση κάθε νοήματος. Έτσι θα μειώσουμε δραματικά τον αριθμό των μοντέλων και το μέγεθος της βάσης δεδομένων που θα χρειαζόμαστε για την εκπαίδευση αυτών με αποτέλεσμα η μοντελοποίηση και αναγνώριση ΝΓ μεγάλου εύρους λεξιλογίου να είναι εφικτή.

6.2.2 Fusion χρησιμοποιώντας τα PHMM

Ενδιαφέρον επίσης παρουσιάζει μια ολοκληρωμένη αξιολόγηση του fusion των διαφορετικών καναλιών πληροφορίας με την χρήση των PHMM. Όπως δείξαμε και σε αυτή την διπλωματική το fusion με την χρήση των PHMM ήταν αρκετά ικανοποιητικό αν και δεν καταφέραμε να έχουμε μια ολοκληρωμένη αξιολόγηση του, λόγο της μικρής βάσης δεδομένων που είχαμε στην διάθεση μας.

6.2.3 Αναγνώριση συνεχής νοηματικής γλώσσας

Ένα τελευταίο θέμα που παρουσιάζει ιδιαίτερο ενδιαφέρων είναι η αξιολόγηση του συστήματος σε συνεχή νοηματική γλώσσα. Το πρόβλημα αυτό είναι αρκετά πιο δύσκολο αφού δεν έχουν οριστεί τα όρια για την κάθε λέξη με αποτέλεσμα εκτός από την αναγνώριση το σύστημα θα πρέπει να κάνει -όπως λέγεται- segmentation. Επίσης η αναγνώριση συνεχούς λόγου έχει να κάνει με ολόκληρες προτάσεις με αποτέλεσμα η κάθε λέξη να σχετίζεται με τα συμφραζόμενα.

Βιβλιογραφία

[1] Marcell Assan και Kirsti Grobel. Video-based sign language recognition using hidden markov models. Στο *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*, pages 97–109, London, UK, 1998. Springer-Verlag.

[2] B.Bauer και K-F.Kraiss. Towards an automatic sign language recognition system using subunits. *International Gesture Workshop*, 2298:64–75, 2001.

[3] V.Hagen B.Doner. Towars an american sign language interface. *Artificial Intelligence Review*, 1994.

[4] Richard Bowden, David Windridge, Timor Kadir, Andrew Zisserman και Michael Brady. A linguistic feature vector for the visual interpretation of sign language. *CVSSP*, 2004.

[5] W. Gao C. Wang και J. Ma. A real-time large vocabulary recognition system for chinese sign language. *In I. Wachsmuth and T. Sowa, editors, Lecture Notes in Artificial Intelligence*, 2298:86 – 95, 2002.

[6] J. Cooley. How the fft gained acceptance. *IEEE Speech Processing Magazine*, σελίδες 10 –13, 1992.

[7] Yuntao Cui και Juyang Weng. Appearance-based hand sign recognition from intensity image sequences. *Computer Vision and Image Understanding: CVIU*, 78(2):157–176, 2000.

[8] O. Diamanti και P. Maragos. Geodesic active regions for segmentation and tracking of human gestures in sign language videos. *Proceedings of the International Conference on Image Processing*, 2008.

[9] D.McNeill. Hand and mind: what gestures reveal about thought. *University of Chicago Press*, 1992.

[10] Gaolin Fang, Wen Gao και Jiyong Ma. Signer-independent sign language recognition based on sofm/hmm. *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001. Proceedings. IEEE ICCV Workshop on*, pages 90–95, 2001.

[11] C. Neti G. Gravier, G. Potamianos. Asynchrony modeling for audio-visual speech recognition. *Proceedings of the second international conference on Human Language Technology Research*, 2002.

[12] Z. Ghahramani και M. I. Jordan. Factorial hidden markov models. *Machine Learning*, 29:245 – 275, 1997.

[13] K. Grobel και M. Assan. Isolated sign language recognition using hidden markov models. *Systems, Man, and Cybernetics, 1997. 'Computational Cybernetics and Simulation'., 1997 IEEE International Conference on*, 1:162–167 ολ.1, 1997.

[14] B. Bauer H. Hienz και K. F Kraiss. Hmm-based continuous sign language recognition using stochastic grammars. *In Gesture Workshop*, σελίδες 185–196, 1999.

[15] R. Owens J. Holden και G.Roy. Hand movement classification using an addaptive fuzzy expert system. *International Journal of Expert Systems*, 1996.

[16] G. Potamianos J. Luettin και C. Neti. Asynchromous stream modeling for large vocabulary audio-visual speech recognition. *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference*, 1:169–172, 2001.

[17] M. W. Kadous. Machine recognition of auslan signs using powergloves: Towards large-lexicon recognition of sign language. *In Proceedings of the Workshop on the Integration of Gesture in Language and Speech*, σελίδες 165 – 174, 1996.

[18] R. Kohavi. A study of cross - validation and bootstrap for accuracy estimation and model selection. *Proceedings of the Fourteenth International Joint Conference*, 1995.

[19] RP. Lippmann. Speech recognition by machines and humans. *Speech Communication*, 1997.

[20] N. Oliver M. Brand και A. Pentland. Coupled hidden markov models for complex action recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1997.

[21] Eng Jon Ong και R. Bowden. A boosted classifier tree for hand shape detection. *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, σελίδες 889–894, 2004.

[22] S.C.W. Ong και S. Ranganath. Automatic sign language analysis: a survey and the future beyond lexical meaning. *Transactions on Pattern Analysis and Machine Intelligence*, 27(6):873–891, 2005.

[23] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.

[24] L. R. Rabiner και B. H. Juang. Fundamentals of speech recognition. *Prentice - Hall*, 1993.

[25] C. Rowden. Analysis. *In C. Rowden, editor, Speech Processing*, σελίδες 35 – 96, 1992.

[26] N. Russell S. Young και J. Thornton. Token passing: a conceptual model for con- nected speech recognition systems. *Technical report, F - INFENG/TR38 Cambridge University*, 1989.

[27] T. Starner και A. Pentland. Visual recognition of american sign language using hidden marcov models. *In International Workshop on Automatic Face and Gesture Recognition*, 1995.

[28] T. Starner και A. Pentland. Real-time american sign language recognition from video using hidden marcov models. *Technical Report 375 M.I.T Media Laboratory Perceptual Computing Section*, 1996.

[29] T. Starner, J. Weaver και A. Pentland. Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375, 1998.

[30] W. C. Stokoe. Sign language structure: An outline of the visual communication system of the american deaf. *Studies in Linguistics: Occasional Papers 8. Linstok Press, Silver Spring, MD*, 1978.

[31] Christian Vogler. *American Sign Language Recognition: Reducing the Complexity of the Task with Phoneme-Based Modeling and Parallel Hidden Markov Models*. Διδακτορική Διατριβή , Department of Computer and Information Science, University of Pennsylvania, 2002.

[32] Christian Vogler και Dimitris Metaxas. Parallel hidden markov models for american sign language recognition. *International Conference on Computer Vision*, 01:116, 1999.

[33] Christian Vogler και Dimitris N. Metaxas. Handshapes and movements: Multiple-channel american sign language recognition. Στο *Gesture Workshop*, σελίδες 247–258, 2003.

[34] M. B. Waldron και S. Kim. Isolated asl sign recognition system for deaf persons. *IEEE Transactions on Rehabilitation Engineering*, 1995.

[35] Ο. Διαμαντή. Οπτική Ανάλυση Βίντεο Νοηματικής Γλώσσας : Κατάτμηση, Παρακολούθηση και Εξαγωγή Χαρακτηριστικών. Διπλωματική Εργασία, Σχολή ΗΜΜΥ, ΕΜΠ, 2007.

[36] Μ. Κατσογιάννου Ευθυμίου Ε. Από την έρευνα για την ελληνική νοηματική γλώσσα (ΕΝΓ): μελέτη του λεξιλογίου και δημιουργία λεξικού. *22η ετήσια συνάντηση του Τομέα γλωσσολογίας του Πανεπιστημίου Θεσσαλονίκης*, 2001.

[37] Μ. Κατσουάννου. Η Ελληνική Νοηματική Γλώσσα Ένα Άλλο Μέσο Επικοινωνίας.

[38] Π. Μαραγκός. *Ανάλυση Εικόνων και Όρασης Υπολογιστών*. Εκδόσεις ΕΜΠ, 2005.