



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

**Αναγνώριση και Μοντελοποίηση Συνεχούς Νοηματικής
Γλώσσας με την Χρήση Οπτικής Επεξεργασίας και
Στατιστικών Μεθόδων**

ΕΚΘΕΣΗ ΠΡΟΟΔΟΥ ΚΑΙ ΠΡΟΤΑΣΗΣ ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

ΘΕΟΔΩΡΑΚΗΣ Ι. ΣΤΑΥΡΟΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός & Μηχανικός Υπολογιστών Ε.Μ.Π.

Αθήνα, Ιούνιος 2012



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Αναγνώριση και Μοντελοποίηση Συνεχούς Νοηματικής Γλώσσας με την Χρήση Οπτικής Επεξεργασίας και Στατιστικών Μεθόδων

ΕΚΘΕΣΗ ΠΡΟΟΔΟΥ ΚΑΙ ΠΡΟΤΑΣΗΣ ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

ΘΕΟΔΩΡΑΚΗΣ Ι. ΣΤΑΥΡΟΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός & Μηχανικός Υπολογιστών Ε.Μ.Π.

Συμβουλευτική Επιτροπή:

Καθ. Πέτρος Μαραγκός (Επιβλέπων)
Επίκ. Καθ. Κωνσταντίνος Τζαφέστας
Καθ. Γεώργιος Καραγιάννης

Εγκρίθηκε από την πενταμελή επιτροπή στις 2012:

.....
Π. Μαραγκός
Καθηγητής Ε.Μ.Π.

.....
Κ. Τζαφέστας
Επίκ. Καθηγητής Ε.Μ.Π.

.....
Γ. Καραγιάννης
Καθηγητής Ε.Μ.Π.

.....
Τ. Σελλής
Καθηγητής Ε.Μ.Π.

.....
Α. Ποταμιάνος
Αν. Καθηγητής Πολυτεχνείο Κρήτης

Αθήνα, Ιούνιος 2012

...

ΘΕΟΔΩΡΑΚΗΣ Ι. ΣΤΑΥΡΟΣ

Υποψήφιος Διδάκτορας Ηλεκτρολόγος Μηχανικός & Μηχανικός Ηλεκτρονικών Υπολογιστών
Ε.Μ.Π.

Copyright © Θεοδωράκης Ι. Σταυρος, 2012.

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 στη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

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

Περίληψη

Η παρούσα έρευνα επικεντρώνεται στην αυτόματη επεξεργασία βίντεο συνεχούς νοηματικής γλώσσας, στην εξαγωγή χαρακτηριστικών, στην μοντελοποίηση και τελικά στην αναγνώριση συνεχούς νοηματικής γλώσσας (Continuous Sign Language Recognition). Η έρευνα αυτή συνδυάζει τα ερευνητικά πεδία της Αναγνώρισης Προτύπων και Όρασης Υπολογιστών. Σε αυτά τα πλαίσια αναπτύσσονται μέθοδοι για την οπτική επεξεργασία και εξαγωγή χαρακτηριστικών από βίντεο συνεχούς νοηματικής γλώσσας σχετιζόμενα με τους αρθρωτές όπως τα χέρια και το κεφάλι. Επιπλέον αναπτύσσονται στατιστικές μέθοδοι για την μοντελοποίηση και αναγνώριση της νοηματικής γλώσσας. Πιο συγκεκριμένα αναπτύσσονται μέθοδοι για την αυτόματη οδηγούμενη από τα δεδομένα- (data-driven) μοντελοποίηση σε επίπεδο υπο-μονάδων (φωνημάτων) τα διαφορετικά είδη κίνησης, θέσης και χειρομορφών των χεριών του νοηματιστή κατά την εκτέλεσης ενός νοήματος. Επιπλέον η μοντελοποίηση των παραπάνω υπο-μονάδων εμπλουτίζεται από την αξιοποίησή γλωσσικής-φωνητικής πληροφορίας. Ακόμα λόγω των πολλαπλών παράλληλων ροών πληροφορίας (π.χ. θέση, κίνηση των χεριών, είδος χειρομορφής, εκφράσεις του προσώπου κ.τ.λ) αναπτύσσονται μέθοδοι για τον συνδυασμό των διαφορετικών καναλιών πληροφορίας με απώτερο σκοπό την αναγνώριση συνεχούς νοηματικού λόγου. Οι προταθέντες μέθοδοι συνδυάζονται σε ένα συνολικό σύστημα επεξεργασίας, εκπαίδευσης και αναγνώρισης και αξιολογούνται σε προτυποποιημένα σώματα βίντεο νοηματικών γλωσσών. Συγκρίσεις με σύγχρονες μεθόδους από την βιβλιογραφία οδηγούν σε βελτιώσεις σε σχέση με βασικές υπάρχοντες μεθόδους. Οι επιδράσεις της έρευνας αυτής και των εφαρμογών της αναμένετε να έχουν διεπιστημονικό χαρακτήρα, όπως για παράδειγμα στη γλωσσολογία και ανάλυση των νοηματικών γλωσσών καθώς και την αυτόματη επεξεργασία και επισημείωση μεγάλων σωμάτων βίντεο νοηματικών γλωσσών.

Περιεχόμενα

1	Εισαγωγή	11
1.1	Αυτόματη Αναγνώριση Νοηματικής Γλώσσας	11
1.2	Ερευνητικές Συνεισφορές	12
1.3	Επισκόπηση Σχετικής Έρευνας	15
1.3.1	Εξαγωγή χαρακτηριστικών	15
1.3.2	Αναγνώριση Νοηματικής Γλώσσας	16
2	Εξαγωγή χαρακτηριστικών για την αναγνώριση Νοηματικής Γλώσσας	21
2.1	Εισαγωγή	21
2.2	Ανίχνευση κεφαλιού και χεριών του Νοηματιστή	21
2.2.1	Πιθανοτικό Μοντέλο Χρώματος Δέρματος	21
2.2.2	Μορφολογική Επεξεργασία των εξαγομένων Μασκών Δέρματος	22
2.2.3	Μορφολογική Κατάτμηση των Μασκών Δέρματος	23
2.3	Ανίχνευση και Παρακολούθηση των Χεριών και του Κεφαλιού	24
2.4	Μοντελοποίηση και εξαγωγή χαρακτηριστικών για τις χειρομορφές	25
3	Μοντελοποίηση της Νοηματικής Γλώσσας με δεδομενοκεντρικές υπομονάδες	29
3.1	Εισαγωγή στην έννοια των υπομονάδων	29
3.2	Αυτόματη κατάτμηση σε Στατικά και Δυναμικά τμήματα	31
3.3	Ομαδοποίηση και Μοντελοποίηση των υπομονάδων	34
3.3.1	Κανονικοποίηση Χαρακτηριστικών	34
3.3.2	Ομαδοποίηση των δυναμικών τμημάτων	36
3.3.3	Δυναμικές υπομονάδες ανά χαρακτηριστικό	36
3.3.4	Δυναμικές υπομονάδες πολλαπλών ροών πληροφορίας	38
3.3.5	Ομαδοποίηση στατικών υπομονάδων	38
3.4	Φωνητικό Λεξικό με Δυναμικές/Στατικές Υπομονάδες	39
3.4.1	Συνδυασμός των Δυναμικών και Στατικών Υπομονάδων	39
3.4.2	Πολλαπλές Προφορές Νοημάτων	40
3.4.3	Αντιστοίχιση μιας ακολουθίας υπομονάδων σε πολλαπλά νοήματα	41
3.4.4	Κοινές Προφορές και Ομοιότητα Μεταξύ Νοημάτων	44
4	Μοντελοποίηση της Νοηματικής Γλώσσας με φωνητικές υπομονάδες	47
4.1	Μετατροπή των HamNoSys σε PDTS φωνητικά σύμβολα	48
4.2	Ανίχνευση και μοντελοποίηση και εκπαίδευση των φωνητικών υπομονάδων	49
4.2.1	Εκπαίδευση των Posture-Detention-Transition-Steady Shift (PDTS) φωνητικών υπομονάδων	50
4.2.2	Αντιστοίχιση των PDTS συμβόλων με τα πραγματικά δεδομένα	51
4.2.3	Αποτελέσματα Φωνητικών Υπομονάδων	51

5 Αναγνώριση Νοηματικής Γλώσσας	55
5.1 Αναγνώριση Χειρομορφών	55
5.1.1 Σώμα Δεδομένων και Επισημείωση των Χειρομορφών	55
5.1.2 Πειραματικά Αποτελέσματα	56
5.1.3 Πειράματα Ταξινόμησης	60
5.2 Αναγνώριση Μεμονωμένων Νοημάτων	63
5.2.1 Δεδομενοκεντρικές Υπομονάδες	63
5.2.2 Φωνητικές Υπομονάδες	70
6 Σύνοψη Προόδου και Μελλοντική Έρευνα	73
6.1 Ερευνητική Συνεισφορά	73
6.2 Κατευθύνσεις Μελλοντικής Έρευνας	74
Βιβλιογραφία	74
Α΄ Κατάλογος Δημοσιεύσεων του Συγγραφέα	82

Κατάλογος Σχημάτων

1.1	Σύνοψη συστήματος: ανίχνευσης, παρακολούθησης και εξαγωγής χαρακτηριστικών για τους αρθρωτές (χέρια, κεφάλι).	12
1.2	Σύνοψη συστήματος: Μοντελοποίηση της Νοηματική Γλώσσα (ΝΓ) με υπομονάδες χρησιμοποιώντας δεδομενοκεντρικές μεθόδους.	13
1.3	Σύνοψη συστήματος: Μοντελοποίηση της ΝΓ με υπομονάδες χρησιμοποιώντας Γλωσσική-Φωνητική Πληροφορία.	14
2.1	Μοντελοποίηση χρώματος δέρματος. (α, β) Παραδείγματα από επισημειωμένες περιοχές δέρματος (τετράγωνα) οι οποίες χρησιμοποιούνται για την εκπαίδευση του χρωματικού μοντέλου. (γ) Δεδομένα εκπαίδευσης στο χρωματικό χώρο a^*-b^* μαζί με την κανονική κατανομή $p_s(a^*, b^*)$. Κάθε εικονοστοιχείο αντιστοιχίζεται σε δέρμα βασιζόμενη σε ένα κατώφλι της $p_s(a^*(\vec{x}), b^*(\vec{x}))$. Η ευθεία γραμμή αντιστοιχεί στην πρώτη ιδιοκατεύθυνση μετά από PCA στα δεδομένα εκπαίδευσης.	22
2.2	Ενδιάμεσα και τελικά αποτελέσματα για την εξαγωγή της μάσκας δέρματος και της μορφολογικής κατάτμησης εφαρμοσμένα σε ένα πλαίσιο από δύο βάσεις δεδομένων. (α) Αρχικό πλαίσιο (β) Αρχική εκτίμηση της μάσκα δέρματος S_0 , η οποία έχει προκύψει από κατωφλιοποίηση της πιθανότητας δέρματος σε κάθε εικονοστοιχείο. (γ) Τελική εκτίμηση της μάσκα δέρματος S_2 , η οποία έχει προκύψει μετά από μορφολογική βελτίωση της S_0 . (δ) $Erosion\ S_2 \ominus B_c$ της S_2 με ένα μικρό δίσκο όπου εφαρμόζεται στην S_2 με σκοπό να βοηθάει στον διαχωρισμό επικαλυπτομένων αρθρωτών. (ε) Κατάτμηση της μάσκας δέρματος S_2 , σε συνεκτικές συνιστώσες μαζί με την εφαρμογή του αλγορίθμου competitive reconstruction opening (ενότητα 2.2.3).	23
2.3	Εναλλαγή πλαισίων με ύπαρξη ή μη επικαλύψεων: Σχηματική αναπαράσταση της εμπρόσθια-οπίσθια γραμμικής εκτίμησης.	25
2.4	Αποτέλεσμα της ανίχνευσης και παρακολούθησης των χεριών και του κεφαλιού του νοηματιστή και σε περιπτώσεις όπου έχουμε επικαλύψεις από την βάση δεδομένων BU400. Παράδειγμα πλαισίων με την μάσκα δέρματος και τους αρθρωτές που περιλαμβάνει κάθε κατατμημένη περιοχή. Το H αντιστοιχεί στο κεφάλι, το L στο αριστερό χέρι και το R στο δεξί χέρι.	25
2.5	Αποτέλεσμα της ανίχνευσης και παρακολούθησης των χεριών και κεφαλιού του νοηματιστή και σε περιπτώσεις όπου έχουμε επικαλύψεις από την βάση δεδομένων Dicta-Sign Corpus.	25
2.6	Αποτέλεσμα της ανίχνευσης και παρακολούθησης των χεριών και κεφαλιού του νοηματιστή και σε περιπτώσεις όπου έχουμε επικαλύψεις από την βάση δεδομένων Dicta-Sign Corpus.	26
2.7	Αποτέλεσμα της ανίχνευσης και παρακολούθησης των χεριών και κεφαλιού του νοηματιστή και σε περιπτώσεις όπου έχουμε επικαλύψεις μαζί με τις ελλείψεις που έχουν ταιριάζει σε κάθε αρθρωτή στις περιπτώσεις που έχουμε επικαλύψεις στην βάση δεδομένων BU400.	26

2.8	Ομαλοποιημένο Ταίριασμα του Μοντέλο Σχήματος-Εμφάνισης (ΜΣΕ) για δύο διαφορετικούς νοηματιστές (Ο11Α,Ο12Β) πρώτη και δεύτερη σειρά αντίστοιχα από την βάση δεδομένων Dicta-Sign Συνεχής ΕΝΓ. Σε κάθε αρχική εικόνα πλαισίου υπερθέτουμε την ανακατασκευή βασισμένοι στο μοντέλο, $A_0(W_p^{-1}(x)) + \sum \lambda_i A_i(W_p^{-1}(x))$. Στην πάνω δεξιά γωνία δείχνουμε κάθε φορά την ανακατασκευή, αλλά στο χώρο του μοντέλου Σχήματος-Εμφάνισης (ΣΕ) $A_0(x) + \sum \lambda_i A_i(x)$, η οποία καθορίζει τα βέλτιστα βάρη.	28
3.1	Παραδείγματα 3 νοημάτων της Αμερικάνικης Νοηματικής Γλώσσας (ΑΝΓ) τα οποία εμφανίζονται στην βάση δεδομένων BU400 Συνεχής ΑΝΓ. Τα παραδείγματα αυτά είναι ενδεικτικές προφορές ανάμεσα σε πολλές διαφορετικές προφορές που εμφανίζονται στην συγκεκριμένη βάση. Οι πολλαπλές αυτές προφορές οφείλονται στο συνεχή νοηματικό λόγο όπου φαινόμενα όπως συνάρθρωση εμφανίζονται, με αποτέλεσμα την διαφοροποίηση της εκτέλεσης ενός νοήματος και την εξάρτηση της από το συγκεκριμένο στο οποίο βρίσκεται.	30
3.2	Διάσπαση του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ στις PDTS υπομονάδες από τις οποίες αποτελείται.	30
3.3	Διάγραμμα το οποίο δείχνει την διαδικασία της αυτόματης κατάτμησης και ταξινόμησης σε στατικά και δυναμικά τμήματα. Στο πρώτο επίπεδο εκπαιδεύουμε ένα μοντέλο για κάθε νόημα το οποίο χρησιμοποιούμε για την κατάτμηση σε τμήματα με βάση την ταχύτητα (υψηλή -Y- και χαμηλή -X-). Στο δεύτερο επίπεδο η ίδια διαδικασία επαναλαμβάνεται χρησιμοποιώντας αυτή την φορά ένα μοντέλο το οποίο εκπαιδεύεται χρησιμοποιώντας μόνο τα στατικά τμήματα που έχουν εξαχθεί από το πρώτο επίπεδο. Στην συνέχεια επαναταξινομούμε όλα τα στατικά τμήματα με βάση την ταχύτητα (υψηλή -Y- και χαμηλή -X-).	32
3.4	Το εργοδικό κρυφό μαρκοβιανό μοντέλο δύο καταστάσεων (2S-ERG) το οποίο χρησιμοποιήθηκε για την κατάτμηση και ταξινόμηση σε δυναμικά (Δ) και στατικά (Σ) τμήματα.	32
3.5	(α) Κατανομή της ταχύτητας (Ιστόγραμμα) μαζί με τις δύο γκαουσιανές κατανομές που έχουν ταιριάζει οι οποίες αντιστοιχούν στις δύο διαφορετικές καταστάσεις του εργοδικού μοντέλου για το νόημα ΠΑΡΑΔΕΧΟΜΑΙ. (β) Σημεία κατάτμησης πάνω στο προφίλ της ταχύτητας για μια επανάληψη του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ παράλληλα με την αντιστοιχία των ετικετών Δυναμικό/Στατικό τμήμα.	33
3.6	Κατάτμηση σε δυναμικά και στατικά τμήματα για μια επανάληψη του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ της Αμερικάνικης Νοηματικής Γλώσσας (ΑΝΓ). Κάθε γραμμή αντιστοιχεί σε διαφορετικό τμήμα.	34
3.7	Αυτόματη vs. Ανθρώπινη κατάτμηση και ταξινόμηση των τμημάτων για μια επανάληψη του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ.	34
3.8	Οι τροχιές δυναμικών τμημάτων πάνω στο διδιάστατο χώρο νοηματισμού: (α) Χωρίς κανονικοποίηση (Position (P)) (β) Κανονικοποίηση με βάση την αρχική θέση (Initial Position Normalized (Pn)) (γ) Κανονικοποίηση με βάση το μέγεθος της κίνησης (Scale Normalized (Sn)) (δ) Κανονικοποίηση με βάση και την αρχική θέση και το μέγεθος της κίνησης (Scale Position Normalized (SPn)).	35

3.9	Οι τροχιές διαφορετικών δυναμικών τμημάτων πάνω στο δισδιάστατο χώρο νοηματισμού μετά από κανονικοποίηση με βάση και την αρχική θέση. Το χρώμα των τροχιών ομαδοποιεί τα δυναμικά τμήματα ανάλογα με την υπομονάδα/συστάδα (subunit/cluster) που αντιστοιχούν. (α) Τροχιές των δυναμικών τμημάτων που ομαδοποιήθηκαν στις αντίστοιχες υπομονάδες/συστάδες χρησιμοποιώντας ως χαρακτηριστικό τις x, y συντεταγμένες χωρίς κανονικοποίηση (P). (β) Τροχιές των δυναμικών τμημάτων που ομαδοποιήθηκαν στις αντίστοιχες υπομονάδες/συστάδες χρησιμοποιώντας ως χαρακτηριστικό τις κανονικοποιημένες x, y συντεταγμένες με βάση την αρχική θέση και το μέγεθος της κίνησης (SPn). (γ) Τροχιές των δυναμικών τμημάτων που ομαδοποιήθηκαν στις αντίστοιχες υπομονάδες/συστάδες χρησιμοποιώντας ως χαρακτηριστικό διάνυσμα την κατεύθυνση των τροχιών (Direction (D)). (δ) Τροχιές των δυναμικών τμημάτων που ομαδοποιήθηκαν σε δύο υπομονάδες/συστάδες χρησιμοποιώντας ως χαρακτηριστικό διάνυσμα το μέγεθός της κίνησης (Scale (S)).	37
3.10	Οι τροχιές των δυναμικών τμημάτων που αντιστοιχούν σε διαφορετικές υπομονάδες με βάση το χρώμα. Οι υπομονάδες βασίζονται ταυτόχρονα και στην κατεύθυνση αλλά και στο μέγεθος της κίνησης.	38
3.11	Παραδείγματα υπομονάδων κατεύθυνσης-κλίμακας στο επίπεδο των πραγματικών πλαισίων υπερθέτοντας το αρχικό στο τελικό πλαίσιο και τοποθετώντας ένα βέλος το οποίο υποδεικνύει το είδος της κίνησης. Οι τρεις δυναμικές υπομονάδες αντιστοιχούν σε κίνησης που εμφανίζονται στα νοήματα ΚΟΥΦΟΣ, ΑΠΟΦΑΣΙΖΩ, ΜΕΣΑ αντίστοιχα.	39
3.12	Χρήση του αλγορίθμου K-means για την ομαδοποίηση των στατικών τμημάτων και κβάντιση του δισδιάστατου νοηματικού χώρου. Κάθε στατική υπομονάδα απεικονίζεται με την χρήση διαφορετικού χρώματος.	40
3.13	Ακολουθίες υπομονάδων αντιστοιχίζονται σε δύο επίλεγμένα νοήματα τις ΑΝΓ (ΒΛΕ-ΠΩ και ΕΝΑ).	42
3.14	Δενδρογράμμο με βάση την διακριτικότητα των νοημάτων (d_s).	45
4.1	Παραδείγματα εκτέλεσης για τρία νοήματα της Ελληνική Νοηματική Γλώσσα (ΕΝΓ) : ΣΩΡΟΣ, ΑΜΕΣΩΣ και ΕΥΡΩΠΗ.	48
4.2	Νόημα ΣΩΡΟΣ : Κατάτμηση σε χρονικά τμήματα με την ενσωμάτωση των PDTS φωνητικών συμβόλων. Έχουμε κάνει υπέρθεση του πρώτου και τελευταίου πλαισίου κάθε τμήματος. Τα PDTS φωνητικών συμβόλων που αντιστοιχούν σε transitions ή epenthesis συνοδεύονται από ένα βέλος που υποδεικνύει το είδος της κίνησης. Κάθε χρονικό τμήμα αντιστοιχίζεται σε ένα PDTS σύμβολο. Τα είδη των PDTS συμβόλων είναι Epenthesis (E), Posture (P), Transition (T).	49
4.3	Αντιστοιχία των PDTS συμβόλων με τα HMM μοντέλα για το νόημα ΑΜΕΣΩΣ.	50
4.4	Φωνητικές υπομονάδες που αντιστοιχούν σε Ts και Es. (α) Η E-to-head υπομονάδα αντιστοιχεί σε επένθεση κίνηση E, (β-δ) Οι T-circular, T-down-right, T-in-left υπομονάδες αντιστοιχούν σε μεταβάσεις (Ts). Απεικονίζουμε τις τροχιές στο πραγματικό νοηματικό χώρο εφαρμόζοντας κανονικοποίηση με βάση την αρχική τους θέση. Το κόκκινο σημάδι υποδεικνύει την αρχή της τροχιάς. Στο Πίνακα 4.2 απεικονίζουμε τα αντίστοιχα PDTS σύμβολα.	52
4.5	Φωνητικές υπομονάδες που αντιστοιχούν σε Ps. Απεικονίζουμε στον νοηματικό χώρο με διαφορετικό χρώμα τα δεδομένα εκπαίδευσης τεσσάρων υπομονάδων που αντιστοιχούν σε 4 διαφορετικά P (P-forehead, P-stomach, P-shoulder, P-head-top).	53

5.1	Παράμετροι Τρισδιάστατης πόζας : (α-γ) Κατεύθυνση εκτεταμένων δακτύλων: (α) Πρόσοψη, (β) Πλάγια όψη, (γ) Κάτοψη και (δ) Προσανατολισμός της Παλάμης. Παρατηρούμε ότι έχουμε τροποποιήσει τα αντίστοιχα σχήματα από του άρθρου [34] ορίζοντας αριθμητικές παραμέτρους για κάθε διαφορετικό προσανατολισμό.	56
5.2	Χώρος χαρακτηριστικών του πειράματος D-HFSBP όπου τα μοντέλα εξαρτώνται από όλες τις παραμέτρους επισημείωσης χρησιμοποιώντας την μέθοδο Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος- Εμφάνισης (Aff-SAM). Για λόγους οπτικοποίησης προβάλλουμε τα διανύσματα χαρακτηριστικών των εκπαιδευμένων μοντέλων στο επίπεδο $\lambda_1 - \lambda_2$ όπου λ_1 και λ_2 αντιστοιχούν στις δύο πρώτες ιδιοκατευθύνσεις. (α) Τα κεντροειδή των μοντέλων μαζί με τις αντίστοιχες ετικέτες που υποδεικνύουν το είδος της χειρομορφής και του προσανατολισμού της ($[HSId]_{FSBP}$). (β) Τα κεντροειδή των μοντέλων και οι ελλείψεις που δείχνουν την διασπορά τους. (γ) Πραγματικές κομμένες εικόνες χειρομορφών για κάθε κεντροειδή.	60
5.3	Davies-Bouldin Index (DBI) σε λογαριθμική κλίμακα (άξονας y) για όλες τις μεθόδους εξαγωγής χαρακτηριστικών μεταβάλλοντας την εξάρτηση των κλάσεων στις παραμέτρους επισημείωσης.	61
5.4	Πειράματα ταξινόμησης για περιπτώσεις χωρίς επικαλύψεις στην βάση δεδομένων DS-1. Πειραματικά αποτελέσματα μεταβάλλοντας την εξάρτηση των κλάσεων στις παραμέτρους επισημείωσης του προσανατολισμού [H,F,B,S,P] (άξονας x) και της μεθόδου εξαγωγής χαρακτηριστικών (legend). Ο αριθμός των κλάσεων για κάθε πείραμα απεικονίζεται στον πίνακα 5.5.	63
5.5	Πειράματα ταξινόμησης με και χωρίς επικαλύψεις στην βάση δεδομένων DS-2. Πειραματικά αποτελέσματα μεταβάλλοντας την εξάρτηση των κλάσεων στις παραμέτρους επισημείωσης του προσανατολισμού [H,F,B,S,P] (άξονας x) και της μεθόδου εξαγωγής χαρακτηριστικών (legend). Ο αριθμός των κλάσεων για κάθε πείραμα απεικονίζεται στον πίνακα 5.6.	64
5.6	Πειραματικά αποτελέσματα χρησιμοποιώντας 94 διαφορετικά νοήματα μεταβάλλοντας τον αριθμό των υπομονάδων που χρησιμοποιήθηκαν	67
5.7	Πειραματικά αποτελέσματα χρησιμοποιώντας 94 διαφορετικά νοήματα μεταβάλλοντας τις μεθόδους και τα διανύσματα χαρακτηριστικών που χρησιμοποιήθηκαν. . . .	68
5.8	Πειραματικά αποτελέσματα μεταβάλλοντας των αριθμό νοημάτων στο λεξικό.	69
5.9	Σύγκριση των μεθόδων Data-driven Subunits (SU-DD) και Phonetic Subunits (SU-P). (α) Αναγνώριση Νοημάτων (β) Αριθμός υπομονάδων. Στα σχήματα (α,β) ο άξονας x αντιστοιχεί στην μεταβολή του αριθμού των νοημάτων που χρησιμοποιήθηκαν. Στο σχήμα (γ) ο αριθμός των νοημάτων είναι 961. Για την SU-DD μέθοδο απεικονίζουμε την επίδραση της μεταβολής του αριθμού των δυναμικών υπομονάδων (x άξονας). Ο αριθμός των στατικών υπομονάδων είναι σταθερός και ίσος με 500. Σε αντίθεση ο αριθμός των υπομονάδων για την SU-P μέθοδο είναι προκαθορισμένος και σταθερός.	71

Κατάλογος Πινάκων

3.1	Παράδειγμα δύο φωνητικών λεξικών [†]	41
3.2	Μια εκτέλεση των νοημάτων (ΕΝΑ και ΒΛΕΠΩ της ΑΝΓ) που αντιστοιχούν στην ακο- λουθία υπομονάδων 'MD12 HP16'	43
3.3	Μια εκτέλεση των νοημάτων (ΕΝΑ και ΒΛΕΠΩ της ΑΝΓ) που αντιστοιχούν σε διαφο- ρετικές ακολουθίες υπομονάδων.	43
4.1	PDTS φωνητικά σύμβολα για το νόημα ΣΩΡΟΣ της ΕΝΓ	50
4.2	Παραδείγματα των φωνητικών υπομονάδων και των νοημάτων στα οποία εμφανίζον- ται. '*' αντιστοιχεί σε πολλαπλά νοήματα.	53
5.1	Είδος Χειρομορφής -Handshape identity (HSId)- αντιστοιχεί στο είδος της χειρομορ- φής ανεξαρτήτως πόζας : 5 ενδεικτικά παραδείγματα.	56
5.2	Παραδείγματα από διαφορετικά είδη χειρομορφών και οι αντίστοιχες παραμέτρους επισημείωσης. '# insts.' αντιστοιχεί στον αριθμό των εμφανίσεων στην βάση δεδο- μένων. Σε κάθε περίπτωση ένα ενδεικτικό πραγματικό παράδειγμα απεικονίζεται το οποίο αντιστοιχεί στο συγκεκριμένο είδος χειρομορφής με τον συγκεκριμένο τρισ- διάστατο προσανατολισμό.	57
5.3	Είδος εξάρτησης των κλάσεων σε σχέση με τις παραμέτρους επισημείωσης της χει- ρομορφής. Κάθε γραμμή αντιστοιχεί σε διαφορετική εξάρτηση. Εξάρτηση ή μη- εξάρτηση σε μια παράμετρο σημειώνεται με 'Ε' ή '*' αντίστοιχα. Για παράδειγμα στην τρίτη σειρά (D-HBP) τα μοντέλα εξαρτώνται από τις παραμέτρους [HSId,B,P].	58
5.4	Σύνοψη των πειραματικών αποτελεσμάτων, μεταβάλλοντας το σώμα δεδομένων, την εξάρτηση των κλάσεων και την μέθοδο εξαγωγής των χαρακτηριστικών. Όπου η Εξ. Κλάσεων αντιστοιχεί στην εξάρτηση των κλάσεων σε σχέση με τις παραμέτρους επι- σημείωσης του προσανατολισμού της χειρομορφής. Επικ. υποδεικνύει την ύπαρξη ή όχι επικαλύψεων στις χειρομορφές που περιέχονται στο σώμα δεδομένων. # HSIds αντιστοιχεί στον αριθμό του είδους των διαφορετικών χειρομορφών ποθ περιέχονται στο σώμα δεδομένων. Avg. Acc. είναι ο μέσος όρος αναγνώρισης και Std είναι η τυπική απόκλιση.	61
5.5	Περιπτώσεις χειρομορφών χωρίς επικαλύψεις. Αριθμός κλάσεων για κάθε πείραμα στην βάση δεδομένων DS-1.	63
5.6	Περιπτώσεις χειρομορφών με και χωρίς επικαλύψεις. Αριθμός κλάσεων για κάθε πείραμα στην βάση δεδομένων DS-2.	63
5.7	Τριάντα ενδεικτικά νοήματα της ΑΝΓ μαζί με τον αριθμό επαναλήψεων τους.	65
5.8	Σύνοψη των πειραμάτων χρησιμοποιώντας διαφορετικές μεθόδους και παραμέτρους η οποίες περιγράφονται αναλυτικά στην ενότητα 5.2.1. Οι ορισμοί των ακρωνύμιων [†] κάθε μεθόδου περιγράφονται στην ενότητα 5.2.1.	65
5.9	Αντιστοίχιση του διανύσματος χαρακτηριστικών που χρησιμοποιήθηκε με τον αριθμό υπομονάδων που επιλέχθηκε.	68

Κεφάλαιο 1

Εισαγωγή

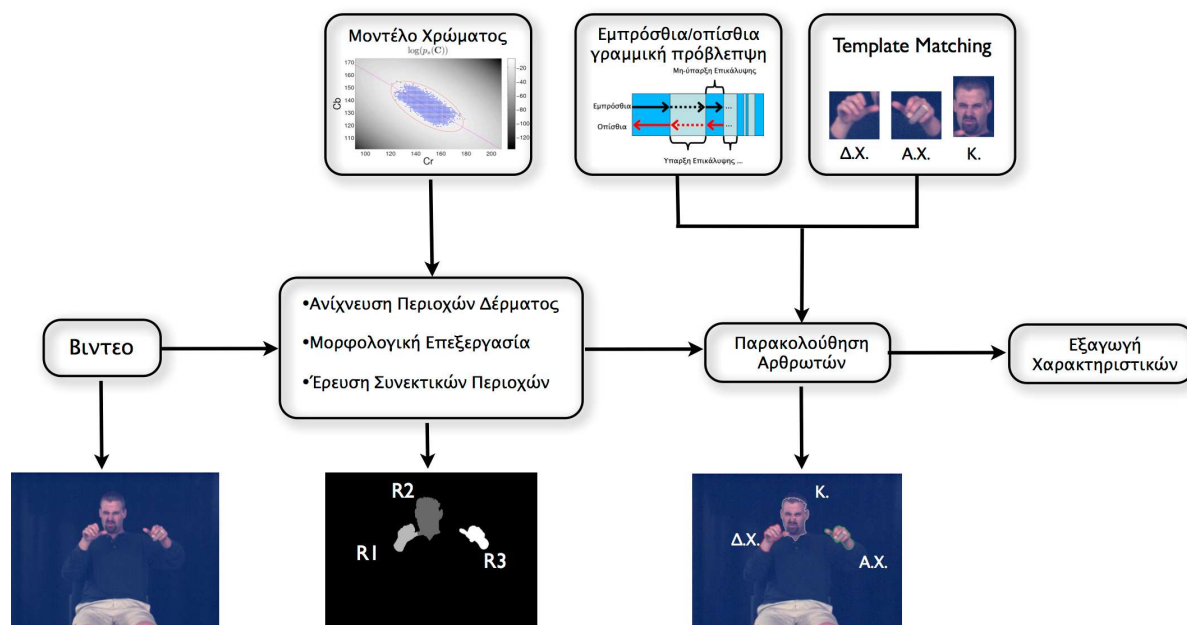
Η επιστήμη των ηλεκτρονικών υπολογιστών πρέπει να κάνει σημαντικά βήματα ακόμα έτσι ώστε η επικοινωνία ανθρώπου - υπολογιστή να είναι αρκετά φυσική. Για τον άνθρωπο (χρήστη) ο πιο φυσικός τρόπος επικοινωνία με τον υπολογιστή είναι μέσω της φωνής και των χειρονομιών. Η αναγνώριση φωνής έχει κάνει μεγάλα βήματα προόδου τα τελευταία χρόνια [50], ενώ αντίθετα η αναγνώριση χειρονομιών βρίσκεται ακόμα αρκετά πίσω. Όμως οι χειρονομίες στην επικοινωνία μεταξύ ανθρώπων έχουν πολύ σημαντικό ρόλο και παρέχουν πληροφορίες που η ομιλία από μόνη της δεν μπορεί να δώσει [53]. Έτσι η συμβολή και των δύο μαζί θα δημιουργούσε ένα σύστημα επικοινωνίας ανθρώπου-υπολογιστή πιο φυσικό από το ήδη υπάρχον.

Η δημιουργία ενός συστήματος αυτόματης αναγνώρισης νοηματικής γλώσσας θα βοηθούσε στην επίλυση αρκετών προβλημάτων. 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 όπου 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 δοθεί 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 όπου 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 αντιστοίχιση του τι ειπώθηκε και ποία ακριβώς χρονική στιγμή από τον νοηματιστή, των λεγόμενων μεταγραφών (transcriptions), διαδικασία αρκετά κουραστική και χρονοβόρα για τον άνθρωπο.

1.1 Αυτόματη Αναγνώριση Νοηματικής Γλώσσας

Στις νοηματικές γλώσσες η μετάδοση πληροφορίας γίνεται μέσω οπτικών μοτίβων και αποτελεί τον μοναδικό τρόπο επικοινωνίας μεταξύ των κωφών [72]. Επιπλέον οι χειρονομίες αποτελούν ένα εναλλακτικό και συμπληρωματικό τρόπος επικοινωνίας μεταξύ των ανθρώπων. Τα οπτικά μοτίβα σχηματίζονται μέσω των χειρών, του σώματος και του προσώπου του ανθρώπου. Ένα σύστημα αυτόματης αναγνώρισης νοηματικής γλώσσας μπορεί να ιδωθεί ως μια ενδιαμέση τεχνολογική μονάδα η οποία μπορεί να καλύψει το επικοινωνιακό κενό μεταξύ των κωφών και ακούων ανθρώπων.

Ένα σύστημα αναγνώρισης νοηματικής γλώσσας αποτελείται από τρία υποσυστήματα: 1) Εξαγωγή χαρακτηριστικών από ένα βίντεο νοηματικού λόγου, 2) Αναγνώριση νοηματικής γλώσσας



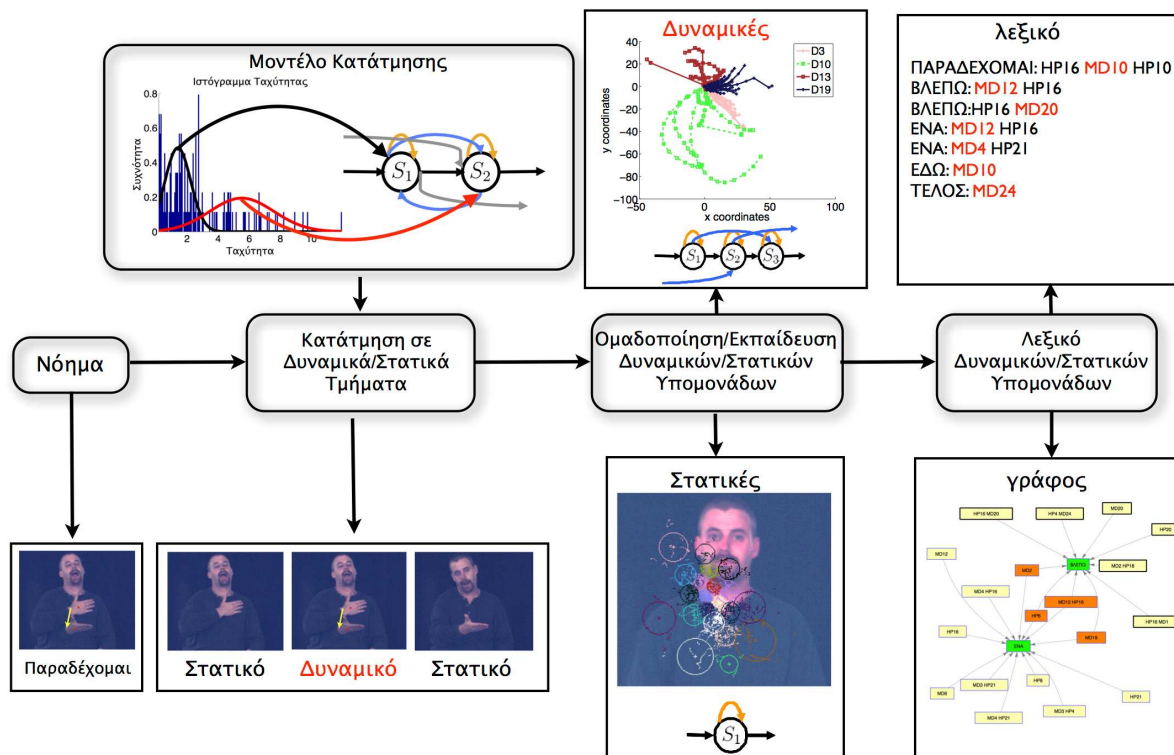
Σχήμα 1.1: Σύνοψη συστήματος: ανίχνευσης, παρακολούθησης και εξαγωγής χαρακτηριστικών για τους αρθρωτές (χέρια, κεφάλι).

και 3) Μετατροπή της πρότασης που αναγνωρίστηκε σε κείμενο. Στο πρώτο υποσύστημα 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 -χειρομορφή- κ.α.). Την πληροφορία αυτή την ονομάζουμε διάνυσμα χαρακτηριστικών και αντιπροσωπεύει ένα πλαίσιο από ολόκληρο το βίντεο. Υπολογίζοντας τα διανύσματα χαρακτηριστικών για κάθε πλαίσιο του βίντεο δημιουργούμε μια ακολουθία διανυσμάτων χαρακτηριστικών. Στο δεύτερο υποσύστημα 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. Στο τρίτο και τελευταίο υποσύστημα μετατρέπουμε την πρόταση που αναγνωρίστηκε σε μορφή που μπορεί ένας άνθρωπος να διαβάσει ανάλογα με την γλώσσα που χρησιμοποίησε ο νοηματιστής. 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. Στα πλαίσια της παρούσας διδακτορικής έρευνας ασχοληθήκαμε και αναπτύξαμε τα δύο πρώτα υποσυστήματα.

1.2 Ερευνητικές Συνεισφορές

Σε αυτή την ενότητα θα παρουσιάσουμε περιληπτικά τις κύριες ερευνητικές συνεισφορές της μέχρι τώρα διδακτορικής έρευνας. Αυτές μπορούν να συνοψισθούν ως εξής:

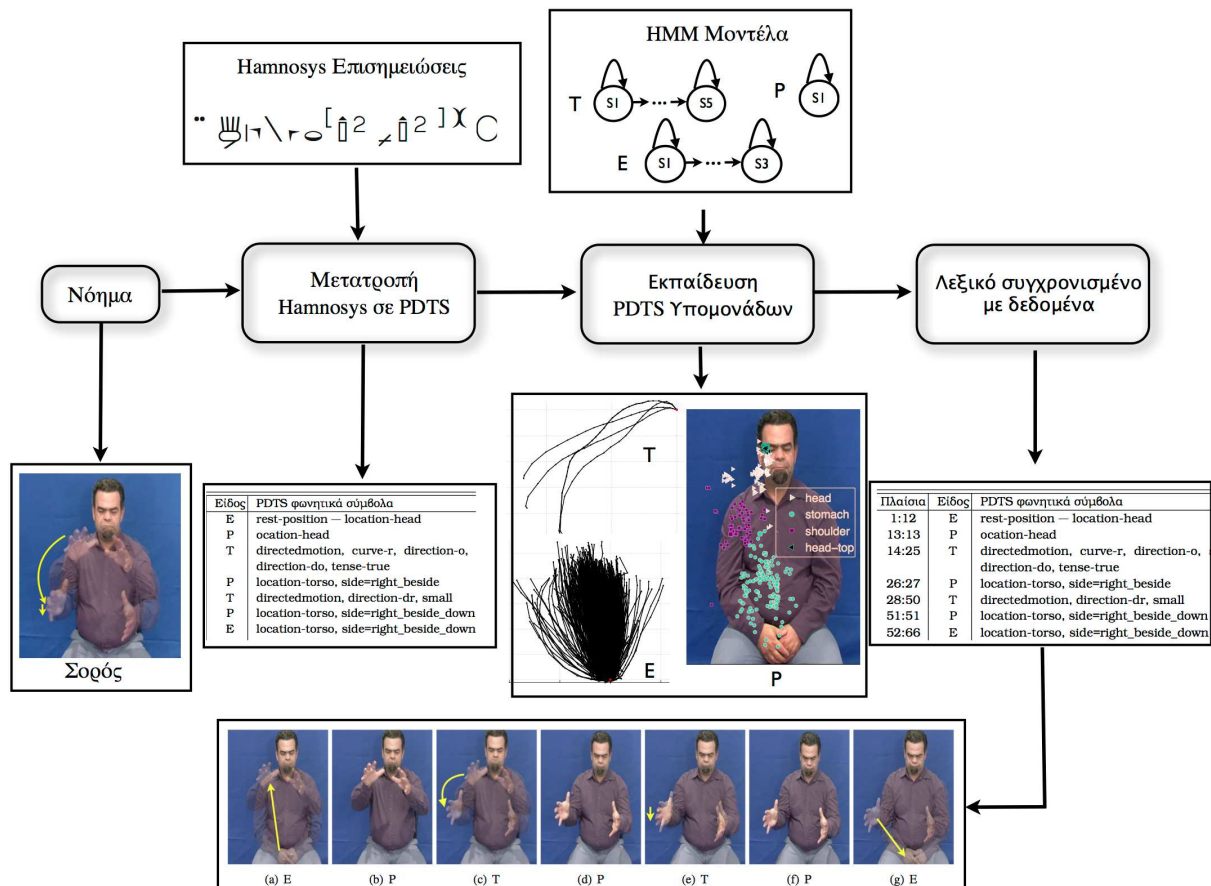
- Εξαγωγή χαρακτηριστικών για την αναγνώριση Νοηματική Γλώσσα (ΝΓ)
- Μοντελοποίηση της ΝΓ με δεδομενοκεντρικές υπομονάδες
- Μοντελοποίηση της ΝΓ με φωνητικές υπομονάδες



Σχήμα 1.2: Σύνοψη συστήματος: Μοντελοποίηση της ΝΓ με υπομονάδες χρησιμοποιώντας δεδο-μενοκεντρικές μεθόδους.

Εξαγωγή χαρακτηριστικών για την αναγνώριση ΝΓ

Αναπτύξαμε ένα σύστημα εξαγωγής χαρακτηριστικών σχετιζόμενα με τους αρθρωτές που συμ-μετέχουν κατά την διάρκεια άρθρωσης της νοηματικής γλώσσας. Μια σύνοψη του συστήματος απεικονίζεται στο Σχήμα 1.1. Πιο συγκεκριμένα το σύστημα απαρτίζεται από τα εξής: 1) Ανίχνευση των περιοχών όπου έχουμε την εμφάνιση χρώματος δέρματος κάνοντας χρήση ενός πιθανοτικού μοντέλου χρώματος. Βασίζομενοι στην παρατήρηση ότι οι αρθρωτές προς ανίχνευση/παρακολού-θηση αντιστοιχούν στο χρώμα του δέρματος χρησιμοποιούμε την πληροφορία του χρώματος για την ανίχνευση τους. Για την μοντελοποίηση του χρώματος δέρματος εκπαιδεύουμε μια γκαου-σιανή κατανομή στον χρωματικό χώρο *CIE-Lab*. 2) Μορφολογική επεξεργασία των εξαγόμενων μασκών δέρματος. Η πρώτη εκτίμηση της μάσκας δέρματος ενδέχεται να περιλαμβάνει λάθη, τρύπες κ.α. Έτσι για την ομαλοποίηση της μάσκας, κάλυψη τρυπών και αφαίρεση πολύ μικρών περιοχών εφαρμόζουμε μορφολογική επεξεργασία. 3) Μορφολογική κατάτμηση των συνεκτικών περιοχών στις περιπτώσεις όπου έχουμε επικαλύψεις οι οποίες είναι πολύ μικρές. Εφαρμόζουμε την μέθοδο *competitive reconstruction opening* η οποία βασίζεται σε αλληπάλληλα μορφολογικά φιλτραρίσματα. 4) Αντιστοίχιση των συνεκτικών περιοχών με τους αρθρωτές (χέρια, κεφάλι). Αρχικά εκτιμούμε την θέση των αρθρωτών συνδυάζοντας εμπρόσθια/οπίσθια γραμμική πρόβλεψη από τα τρία προηγούμενα πλαίσια και *template matching* μεταξύ διαδοχικών πλαισίων. Στην συνέχεια αντιστοιχίζουμε κάθε αρθρωτή στην συνεκτική περιοχή η οποία βρίσκεται πιο κοντά στην εκτιμώμενη θέση του. 5) Εξαγωγή χαρακτηριστικών που σχετίζονται με την θέση και το σχήμα των αρθρωτών. Στο κεφάλαιο 2 που ακολουθεί περιγράφουμε με περισσότερη λεπτομέρεια το σύστημα.



Σχήμα 1.3: Σύνοψη συστήματος: Μοντελοποίηση της ΝΓ με υπομονάδες χρησιμοποιώντας Γλωσσική-Φωνητική Πληροφορία.

Μοντελοποίηση της ΝΓ με δεδομενοκεντρικές υπομονάδες

Αναπτύξαμε ένα σύστημα για την μοντελοποίηση και αναγνώριση ΝΓ με δεδομενοκεντρικές υπομονάδες. Στο Σχήμα 1.2 παρουσιάζουμε μια σύνοψη του συστήματος. Πιο συγκεκριμένα το σύστημα αποτελείται από τα εξής: 1) Κατάτμηση και ταξινόμηση κάθε τμήματος σε δυναμικό ή στατικό. Βασίζομαστε στην μοντελοποίηση του προφίλ της ταχύτητα για κάθε νόημα κάνοντας χρήση ενός εργοδικού HMM 2 καταστάσεων. 2) Επανεξέταση της ταξινόμησης στα στατικά τμήματα εκπαιδεύοντας ένα εργοδικό HMM μοντέλο για όλα τα στατικά τμήματα ανεξαρτήτως νοήματος. 3) Κατασκευή των δυναμικών υπομονάδων εφαρμόζοντας τον αλγόριθμο DTW και agglomerative ιεραρχική συσταδοποίηση στα δυναμικά τμήματα. 4) Κατασκευή των στατικών υπομονάδων εφαρμόζοντας τον αλγόριθμο συσταδοποίησης K-means στα στατικά τμήματα. 5) Συνδυασμός των στατικών και δυναμικών υπομονάδων για την κατασκευή λεξικού σε επίπεδο υπομονάδας. 6) Αναγνώριση ΝΓ με βάση τις δυναμικές/στατικές υπομονάδες και του λεξικού. Στο κεφάλαιο 3 παρουσιάζεται αναλυτικά το παραπάνω σύστημα.

Μοντελοποίηση της ΝΓ με φωνητικές υπομονάδες

Αναπτύξαμε ένα σύστημα για την μοντελοποίηση και αναγνώριση ΝΓ κάνοντας χρήση Γλωσσικής-Φωνητικής Πληροφορίας. Στο Σχήμα 1.3 παρουσιάζουμε μια σύνοψη του συστήματος. Πιο συγκεκριμένα το σύστημα αποτελείται από τα εξής: 1) Μετατροπή των επισημειώσεων επιπέδου HamNoSys (HamNoSys) σε δομημένα Posture-Detention-Transition-Steady Shift (PDTS) φωνητικά

σύμβολα χρησιμοποιώντας την εργασία του C.Vogler [78]. 2) Εκπαίδευση υπολογιστικών φωνητικών υπομονάδων κάνοντας χρήση Κρυφών Μαρκοβιανών Μοντέλων. 3) Αντιστοίχιση των φωνητικών συμβόλων με τα διανύσματα χαρακτηριστικών. 4) Κατασκευή PDTs λεξικού με την χρονική αντιστοιχία των PDTs φωνητικών συμβόλων και των διανυσμάτων χαρακτηριστικών. 5) Εξαγωγή PDTs φωνητικών συμβόλων κατά την διάρκεια της αναγνώρισης. 6) Αντιστοίχιση των PDTs φωνητικών ακολουθιών με τα νοήματα βάση του PDTs λεξικού για την αναγνώριση σε επίπεδο νοήματος. Στο κεφάλαιο 4 παρουσιάζεται αναλυτικά το παραπάνω σύστημα.

1.3 Επισκόπηση Σχετικής Έρευνας

1.3.1 Εξαγωγή χαρακτηριστικών

Το πρώτο βήμα σε ένα σύστημα Αναγνώρισης Νοηματικής Γλώσσας είναι η εξαγωγή χαρακτηριστικών από ένα βίντεο νοηματικού λόγου. Η διαδικασία εξαγωγής χαρακτηριστικών αποτελείται από 3 στάδια: 1) Ανίχνευση των αρθρωτών του νοηματικού λόγου (κεφάλι, χέρια, κορμός) 2) Παρακολούθηση των αρθρωτών κατά την διάρκεια του νοηματικού λόγου και 3) Εξαγωγή διαφόρων ειδών χαρακτηριστικών από τους αρθρωτές όπως σχήμα, θέση, κίνηση κ.α.

Η ανίχνευση των αρθρωτών επιτυγχάνεται χρησιμοποιώντας διαφόρων ειδών χαρακτηριστικών της εικόνας. Τέτοια χαρακτηριστικά μπορεί να είναι το χρώμα δέρματος, οι ακμές της εικόνας, το σχήμα, η κίνηση κ.α. Η κατάτμηση με βάση το χρώμα του δέρματος, με σκοπό την ανίχνευση των χεριών και κεφαλιού του νοηματιστή έχει εφαρμοστεί σε αρκετές ερευνητικές δουλειές [2, 70, 87, 95, 67]. Η χρήση της χρωματικής συνιστώσας ενδείκνυται λόγω της ιδιαιτερότητας/διαφορετικότητας του χρώματος του ανθρώπινου δέρματος. Η μεγάλη ποικιλία στις συνθήκες φωτισμού συνιστά την χρήση χρωματικών χώρων οι οποίοι διαχωρίζουν την χρωματική συνιστώσα από την συνιστώσα του φωτισμού όπως *HSV*, *CIE-Lab*, *YCbCr* [76, 44]. Η πληροφορία της κίνησης χρησιμοποιήθηκε στα άρθρα [19, 40] κάνοντας την υπόθεση ότι τα χέρια του νοηματιστή είναι το μοναδικό αντικείμενο το οποίο κινείται σε ένα στατικό φόντο και ότι το σώμα και το κεφάλι του νοηματιστή παραμένουν σχετικά σταθερά. Μια διαφορετική προσέγγιση προτάθηκε από τους Ong και Bowden στο άρθρο [57], οι οποίοι για την ανίχνευση των χεριών βασίστηκαν στην πληροφορία του σχήματος των χεριών χρησιμοποιώντας *boosted classifier tree*. Σε άλλες ερευνητικές προσεγγίσεις όπως [6, 5, 36] χρησιμοποιήθηκε η πληροφορία του χρώματος σε συνδυασμό με τους περιορισμούς που θέτει η φυσιολογία του ανθρώπινου σώματος για την ανίχνευση του κεφαλιού, του κορμού του σώματος, των βραχιόνων, των χεριών και την συσχέτιση της θέσης και της κίνησης των αρθρωτών με το υπόλοιπο σώμα.

Το δεύτερο βήμα προς την εξαγωγή χαρακτηριστικών είναι η παρακολούθηση των αρθρωτών κατά την διάρκεια του νοηματικού λόγου. Για την παρακολούθηση των χεριών έχουν χρησιμοποιηθεί ποικίλες μέθοδοι: 1) *blob-based* [70, 75, 2], 2) *model-based* [40, 12] 3) *hand contour / boundary models* [14, 19]. Η αυξημένη συχνότητα των επικαλύψεων μεταξύ των αρθρωτών κατά την διάρκεια του αυθόρμητου νοηματικού λόγου ανάγει την παρακολούθηση των αρθρωτών σε ένα πολύπλοκο πρόβλημα. Για την επιτυχή παρακολούθηση των αρθρωτών και την αντιμετώπιση των επικαλύψεων οι μέθοδοι που περιγράφονται στα άρθρα [95, 67] βασίστηκαν σε ένα πιθανοτικό πλαίσιο για την ανάθεση ετικετών (κεφάλι, χέρια κ.τ.λ) στις περιοχές όπου έχουν ανιχνευθεί αρθρωτές. Οι Downton και Drouet στο άρθρο [24] χρησιμοποίησαν ένα τρισδιάστατο ιεραρχικό κυλινδρικό μοντέλο του άνω τμήματος του ανθρώπινου σώματος το οποίο το ταίριαζαν στις ακμές της εικόνας που είχαν ανιχνεύσει. Με αυτό τον τρόπο υπολόγιζαν τις τρισδιάστατες παραμέτρους του κινηματικού μοντέλου με αποτέλεσμα τη συνεχή παρακολούθηση των αρθρωτών. Εναλλακτικές προσεγγίσεις βασίστηκαν στην τρισδιάστατη αναπαράσταση με την χρήση πολλαπλών καμερών [79, 53]. Άλλος ένας παράγοντας ο οποίος αυξάνει την δυσκολία της παρακολούθησης των αρθρωτών έγκειται στην πολυπλοκότητα του φόντου του βίντεο. Στις περισσότερες ερευνητικές μελέ-

τες χρησιμοποιήθηκαν βίντεο με ομοιόμορφο φόντο [6, 87, 95]. Ωστόσο υπήρξαν και προσεγγίσεις [14, 1] όπου εφαρμόζοντας τεχνικές αφαίρεσης φόντου (background subtraction) κατάφεραν να είναι ανεξάρτητες από το φόντο του βίντεο.

Η εξαγωγή χαρακτηριστικών τα οποία περιγράφουν τους αρθρωτές σε κάθε πλαίσιο αποτελεί κρίσιμο στοιχείο για τα συστήματα Αναγνώρισης Νοηματικής Γλώσσας. Όσο αφορά τα χέρια, απαραίτητο χαρακτηριστικό είναι οι δισδιάστατες ή τρισδιάστατες συντεταγμένες των κεντροειδών των χεριών του νοηματιστή [70, 6, 75, 19, 79]. Χαρακτηριστικά που σχετίζονται με την κίνηση των χεριών είναι επίσης απαραίτητα: όπως η τροχιά των χεριών ή η οπτική ροή [14, 87]. Πιο περίπλοκα χαρακτηριστικά σχετίζονται με το σχήμα ή την εμφάνιση (appearance) των χεριών. Μετά την κατάτμηση των χεριών οι κομμένες εικόνες υφίστανται κανονικοποίηση σε σχέση με το μέγεθος, την φωτεινότητα και τον προσανατολισμό [19], και ακολούθως εφαρμόζονται μέθοδοι όπως ανάλυση στις κύριες συνιστώσες (Principal Component Analysis -PCA-) για την μείωση των διαστάσεων [22, 85]. Σε αρκετές ερευνητικές προσεγγίσεις χρησιμοποιήθηκαν γεωμετρικά χαρακτηριστικά βασισμένα στο σχήμα των χεριών όπως ροπές σχήματος [70, 75], το μέγεθος και οι αποστάσεις μεταξύ των δακτύλων και της παλάμης του χεριού [6, 5]. Άλλες προσεγγίσεις βασίστηκαν στο περίγραμμα του χεριού για την εξαγωγή χαρακτηριστικών ανεξαρτήτως μετατόπισης, κλίμακας και περιστροφής του χεριού όπως περιγραφίτες Fourier (Fourier descriptors) [14, 74]. Οι Huang και Jeng [40] σε αντίθεση χρησιμοποίησαν ενεργά μοντέλα σχήματος (Active Shape Models) [17] για την αναπαράσταση και μοντελοποίηση του περιγράμματος του χεριού. Οι Bowden και Sahardi [9] εφάρμοσαν PCA πάνω στα περιγράμματα των χεριών και κατασκεύασαν μη γραμμικά Point Distribution μοντέλα. Μια ακόμη δημοφιλής μέθοδος εξαγωγής χαρακτηριστικών σχήματος προκύπτει από την χρήση του Ιστογράμματος Κατευθυνόμενων Κλίσεων (Histogram of Oriented Gradients - HOG-) [20]. Αποτελεί ένα από κοινού ιστόγραμμα κβαντισμένων κατευθύνσεων της κλίσης της εικόνας και της μετατοπισμένης θέσης, στην γειτονιά κάθε εικονοστοιχείου. Στα άρθρα [11, 51], συναντώνται τέτοιου είδους περιγραφητές για την εξαγωγή χαρακτηριστικών για τα χέρια των νοηματιστών και συνδυάζουν τόσο την εμφάνιση όσο και το σχήμα των χεριών.

Εκτός από μεθόδους οι οποίες επεξεργάζονται ή μοντελοποιούν τις δισδιάστατες εικόνες των χεριών υπήρξαν και άλλες οι οποίες βασίστηκαν στην τρισδιάστατη μοντελοποίηση, με στόχο την προσέγγιση των γωνιών που σχηματίζονται μεταξύ των δακτύλων ενός χεριού και της τρισδιάστατης πόζας [37, 30, 1, 23]. Αυτές οι μέθοδοι έχουν το πλεονέκτημα ότι είναι ανεξάρτητες της οπτικής γωνίας της κάμερας και της πόζας του χεριού. Ωστόσο η προσαρμογή του μοντέλου (model fitting) δεν είναι αρκετά ακριβής και η διαδικασία είναι υπολογιστικά χρονοβόρα.

1.3.2 Αναγνώριση Νοηματικής Γλώσσας

Τα τελευταία χρόνια το ενδιαφέρον των ερευνητών έχει στραφεί αρκετά στον τομέα της αυτόματης Αναγνώρισης Νοηματικής Γλώσσας [58, 1]. Το αντικείμενο του ερευνητικού τομέα αυτού αποτελεί ένα σύνθετο πρόβλημα το οποίο θέτει σημαντικές προκλήσεις στην οπτική επεξεργασία και εξαγωγή χαρακτηριστικών από βίντεο νοηματισμού (Όραση Υπολογιστών) καθώς και στην μοντελοποίηση πολλαπλών ροών πληροφορίας με στόχο την αναγνώριση (Αναγνώριση Προτύπων). Στην προηγούμενη ενότητα αναφέρθηκαν ερευνητικές μελέτες σχετικές με την εξαγωγή χαρακτηριστικών από βίντεο νοηματισμού. Στην παρούσα ενότητα θα αναφερθούμε σε ερευνητικές εργασίες που σχετίζονται με την στατιστική μοντελοποίηση πολλαπλών ροών πληροφορίας με στόχο την αυτόματη αναγνώριση της νοηματικής γλώσσας. Για την αυτόματη αναγνώριση νοημάτων είναι απαραίτητη η δημιουργία ενός μοντέλου για κάθε νόημα που έχουμε στο λεξικό του υπομελέτη συστήματος. Ανάλογα με το επίπεδο μοντελοποίησης το μοντέλο αυτό μπορεί να είναι είτε αυτόνομο (sign/word-level) είτε να αποτελεί σύνθεση μικρότερων μοντέλων (subunit-level) αντίστοιχα με τα μοντέλα των φωνημάτων της αναγνώρισης φωνής. Η επιλογή του επιπέδου μοντελοποίησης για ένα σύστημα αναγνώριση νοηματικής γλώσσας εξαρτάται από το πλήθος των διαθέσιμων δε-

δομένων εκπαίδευσης για την εκπαίδευση των μοντέλων . Τα συστήματα αναγνώριση νοημάτων τα οποία βασίζονται σε μοντέλα επιπέδου νοήματος (sign/word-level) περιορίζονται συνήθως σε μικρό μέγεθος λεξιλογίου. Αντίθετα τα συστήματα τα οποία βασίζονται σε μοντέλα επιπέδου υπομονάδων (subunit-level) έχουν την δυνατότητα να αντιμετωπίζουν λεξικά μεγαλύτερου μεγέθους. Ο παραπάνω περιορισμός σχετίζεται με τον όγκο των διαθέσιμων δεδομένων εκπαίδευσης. Για την εκπαίδευση μοντέλων σε επίπεδο νοήματος το σώμα εκπαίδευση θα πρέπει να περιέχει πολλαπλές εκτελέσεις κάθε νοήματος που έχουμε στο λεξικό και σε διαφορετικό συγκείμενο. Έτσι για μεγάλα λεξιλόγια η παραπάνω προϋπόθεση δημιουργεί την ανάγκη ύπαρξης απαγορευτικά μεγάλων σωμάτων δεδομένων εκπαίδευσης.

Μοντελοποίηση σε Επίπεδο Νοήματος

Οι περισσότερες ερευνητικές προσεγγίσεις στον τομέα της αυτόματης αναγνώρισης νοηματικής γλώσσας έχουν βασιστεί σε μοντελοποίηση σε επιπέδου νοήματος. Στην συνέχεια αναφέρουμε ενδεικτικές ερευνητικές μελέτες που έχουν ασχοληθεί με τα εξής προβλήματα: 1) Αναγνώριση μεμονωμένων νοημάτων, 2) Εύρεση νοημάτων σε συνεχή νοηματικό λόγο 3) Αναγνώριση συνεχούς νοηματικού λόγου.

Αρκετές ερευνητικές προσπάθειες έχουν εμφανιστεί στην διεθνή βιβλιογραφία για την επίλυση του προβλήματος αναγνώρισης μεμονωμένων νοημάτων. Οι Grobel και Assan στο άρθρο [32] παρουσίασαν ένα σύστημα αναγνώρισης 262 μεμονωμένων νοημάτων χρησιμοποιώντας Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Model -HMM-) για την μοντελοποίηση κάθε νοήματος. Ο Zhang et al. στο άρθρο [93] ασχολήθηκαν με την αναγνώριση 439 μεμονωμένων νοημάτων κάνοντας χρήση HMM. Ο Zieren και Kraiss στο άρθρο [94] ασχολήθηκαν με την αναγνώριση μεμονωμένων νοημάτων ανεξαρτήτως νοηματιστή σε ένα λεξιλόγιο 221 νοημάτων χρησιμοποιώντας HMM. Οι Shanableh et al. [66] εστίασαν στην εξαγωγή χωροχρονικών χαρακτηριστικών και στην δημιουργία ενός συστήματος αναγνώριση μεμονωμένων νοημάτων. Πιο πρόσφατα οι Athitsos et al. [3] παρουσίασαν ένα σύστημα αυτόματης εύρεσης νοημάτων από ένα λεξικό το οποίο δίνει την δυνατότητα στον χρήστη να εκτελεί ένα νόημα από τα 921 νοήματα τα οποία υπάρχουν στο λεξικό και το σύστημα να παρουσιάζει τα νοήματα τα οποία ταιριάζουν καλύτερα με το νόημα το οποίο εκτελέστηκε.

Το πρόβλημα εύρεσης νοημάτων (sign spotting) μέσα σε συνεχή νοηματικό λόγο αποτελεί δυσκολότερο πρόβλημα από την αναγνώριση μεμονωμένων νοηματικών λόγω της ύπαρξης του φαινομένου της συνάρθρωσης. Πιο συγκεκριμένα κατά την διάρκεια του συνεχής νοηματικού λόγου η άρθρωση ενός νοήματος μπορεί να διαφοροποιηθεί λόγω του συγκείμενου στο οποίο βρίσκεται. Οι Nayak et al. [54, 55] στοχεύοντας στο να αντιμετωπίσουν το φαινόμενο της συνάρθρωσης στην εύρεση νοημάτων πρότειναν μια μέθοδο για την εξαγωγή και εκμάθηση του πυρήνα των νοημάτων χρησιμοποιώντας Iterated Conditional Modes (ICM) [7]. Ένα ακόμα σημαντικό πρόβλημα κατά την διάρκεια του συνεχής νοηματικού λόγου είναι η εμφάνιση της επένθεσης κίνησης μεταξύ δύο διαδοχικών νοημάτων [48]. Οι Fang et al. [31, 27] επέκτειναν την προηγούμενη εργασία τους [28] χρησιμοποιώντας transition-movement μοντέλα με στόχο την αντιμετώπιση της επένθεσης κίνησης. Για την επίλυση του προβλήματος της εύρεσης νοημάτων σε συνεχή νοηματικό λόγο, οι Yang et al. [86] πρότειναν μια καινοτόμο μέθοδο για την δημιουργία threshold μοντέλων σε conditional random field (CRF) μοντέλα. τα οποία εφαρμόζαν μια προσαρμοστική καταφλιωποίηση με στόχο την διάκριση νοημάτων και μη-νοημάτων μέσα σε ένα συνεχή νοηματικό λόγο. Στην συνέχεια οι ίδιοι εγγραφείσες στα άρθρα [89, 88] πρότειναν μια μέθοδο για την αναγνώριση της Αμερικάνικης Νοηματικής Γλώσσας με έναν αλγόριθμο Level Building ο οποίος βασιζόταν στον δυναμικό προγραμματισμό. Πρόσφατα οι Kelly et al [45] για την αυτόματη εξαγωγή και εύρεση νοημάτων από συνεχή νοηματικό λόγο πρότειναν έναν αλγόριθμο τον οποίο ονόμασαν multiple instance learning density matrix.

Το πρόβλημα της αναγνώρισης συνεχούς νοηματικού λόγου αποτελεί το δυσκολότερο πρόβλημα στην περιοχή της αναγνώρισης νοηματικής. Παρακάτω αναφέρουμε ενδεικτικές ερευνητικές προσπάθειες. Οι Starnier et al. στα άρθρα [69, 71] παρουσίασαν ένα αυτόματο σύστημα αναγνώρισης συνεχούς νοηματικής γλώσσας βασισμένο σε HMM χρησιμοποιώντας χαρακτηριστικά που σχετίζονται με την κίνηση των χεριών. Η βάση δεδομένων που χρησιμοποίησαν αποτελείτο από 40 διαφορετικά νοήματα. Οι Liang και Ouhyoung στο άρθρο [47] ασχολήθηκαν με το πρόβλημα της κατάτμησης και αναγνώρισης συνεχούς νοηματικού λόγου σε ένα λεξιλόγιο 250 διαφορετικών νοημάτων κάνοντας χρήση HMM. Το είδος των χαρακτηριστικών που χρησιμοποίησαν σχετιζόταν με την θέση, κίνηση και προσανατολισμό των χεριών. Οι Bauer και Hienz στο άρθρο [4] επέκτειναν την ερευνητική μελέτη των Grobel και Assan [32] για την αναγνώριση συνεχούς νοηματικού λόγου σε λεξιλόγιο 97 διαφορετικών νοημάτων χρησιμοποιώντας χαρακτηριστικά σχετιζόμενα με την θέση και την κίνηση των χεριών καθώς και το είδος της χειρομορφής. Οι Holden et al. [38] πρότειναν ένα σύστημα αναγνώρισης Αυστραλιανής Νοηματικής Γλώσσας βασισμένοι σε HMM. Η βάση δεδομένων που χρησιμοποιήθηκε αποτελούνταν από 163 νοήματα.

Εναλλακτικές ερευνητικές μέθοδοι είχαν στόχο την εκμετάλλευση των υποτίτλων που προϋπήρχαν σε σώματα δεδομένων νοηματικού λόγου προερχόμενα κυρίως από ειδήσεις. Οι Farhadi et al. [29] παρουσίασαν μια μέθοδο για την εύρεση νοημάτων μέσα σε συνεχή νοηματικό λόγο με στόχο την αντιστοίχιση των υποτίτλων χρησιμοποιώντας voting και discriminative μεθόδους. Οι Buehler et al. [11, 12] πρότειναν μια μέθοδο για την εκμάθηση ενός μεγάλου αριθμού νοημάτων της Αγγλικής Νοηματικής Γλώσσας από τηλεοπτικές μεταδόσεις με χρήση των υποτίτλων, εφαρμόζοντας μια συνάρτηση απόστασης.

Μοντελοποίηση σε Επίπεδο Υπομονάδων

Τα συστήματα αναγνώρισης που βασίζονται στην μοντελοποίηση επίπεδου νοήματος περιορίζονται συνήθως σε μικρού μέγεθους λεξιλόγια. Για την αντιμετώπιση μεγαλύτερων λεξιλογίων άρχισαν να εμφανίζονται συστήματα όπου η μοντελοποίηση βασιζόταν σε δομικά στοιχεία μικρότερα του νοήματος (υπομονάδες). Στη δεκαετία του 60 ο Stokoe στο άρθρο [72] αναφέρει για πρώτη φορά ότι κάθε νόημα αποτελείται από 3 βασικές παραμέτρους: την θέση και κίνηση των χεριών, και το είδος της χειρομορφής. Βασιζόμενες στην δουλειά του Stokoe αρκετές ερευνητικές εργασίες εμφανίστηκαν στην βιβλιογραφία οι οποίες μοντελοποιούσαν τις 3 παραπάνω παραμέτρους ξεχωριστά. Οι Waldron και Kim στο άρθρο [84] παρουσίασαν ένα σύστημα αναγνώρισης μεμονωμένων νοημάτων χρησιμοποιώντας αυτο-οργανωμένους χάρτες (SOM) βασισμένοι στις παραμέτρους του Stokoe για τον διαχωρισμό της κίνησης από τον προσανατολισμό και την χειρομορφή των χεριών. Οι Bowden et al. [10] πρότεινε μια μέθοδο μοντελοποίησης αυτών των γλωσσικών παραμέτρων για την ταξινόμηση νοημάτων. Ως τελικό ταξινομητή χρησιμοποίησε μαρκοβιανές αλυσίδες σε συνδυασμό με Independent Component Analysis. Οι Cooper et al. [16] επέκτειναν την προηγούμενη δουλειά τους [10] προτείνοντας μια μέθοδο για την εκπαίδευση ταξινομητών με στόχο την μοντελοποίηση των παραπάνω 3 παραμέτρων. Για τον συνδυασμό των αποκρίσεων τους ως ταξινομητής χρησιμοποιήθηκε ένα μαρκοβιανό μοντέλο για την αναγνώριση μεμονωμένων νοημάτων. Οι Ding και Martinez μοντελοποίησαν τις 3 βασικές παραμέτρους που συνθέτουν ένα νόημα. Για τον συνδυασμό τους χρησιμοποίησαν μια δενδρική δομή για την αναγνώριση μεμονωμένων νοημάτων. Οι Ong et al. [59, 60] πρότειναν μια μέθοδο για την ανάλυση της κλίσης των νοημάτων μοντελοποιώντας τις συστηματικές διαφορές στον τρόπο εκτέλεσης των νοημάτων χρησιμοποιώντας Δυναμικά Bayesian Δίκτυα με ανεξάρτητα σύνολα χαρακτηριστικών τα οποία τα ονόμασαν Multichannel Hierarchical Hidden Markov Model (MH-HMM).

Το μοντέλο που προτάθηκε από τον Stokoe δεν έχει την δυνατότητα να μοντελοποιήσει την αλληλουχία/διαδοχή φωνητικών υπομονάδων [73]. Οι Liddell και Johnson με στόχο την μοντελοποίηση της αλληλουχίας φωνητικών υπομονάδων [49] στην Αμερικάνικη Νοηματική Γλώσσα (ANF)

πρότειναν το Movement-Hold μοντέλο [48]. Το Movement-Hold μοντέλο περιγράφει την ΑΝΓ βασισμένο σε δύο ειδών φωνητικές υπομονάδες 1) Τμήματα στα οποία υπάρχει συνεχής αλλαγή μιας από τις 3 παραμέτρους του Stokoe (movement segment) 2) Τμήματα στα οποία και οι 3 παράμετροι παραμένουν αμετάβλητοι (hold segments). Αρκετές ερευνητικές προσεγγίσεις έλαβαν υπόψη τους την αλληλουχία φωνητικών υπομονάδων. Οι Vogler και Metaxas στα άρθρα [81, 83], εμπνευσμένοι από το Movement-Hold μοντέλο [48] έσπασαν τα νοήματα στις φωνητικές υπομονάδες που αποτελούνταν βασισμένοι σε ήδη επισημειωμένα νοήματα. Για την από κοινού μοντελοποίηση των παράλληλων ροών πληροφορίας χρησιμοποίησαν Παράλληλα Κρυφά Μαρκοβιανά Μοντέλα (Parallel Hidden Markov models). Οι Bauer et al. [6] ασχολήθηκαν επίσης με την μοντελοποίηση σε επίπεδο υπομονάδων προτείνοντας μια δεδομενοκεντρική προσέγγιση. Πιο συγκεκριμένα εφάρμοσαν τον αλγόριθμο K-means για την ομαδοποίηση ανεξάρτητων πλαισίων του βίντεο σε συστάδες οι οποίες αντιστοιχούσαν στις προς μοντελοποίηση υπομονάδες. Οι Fang et al. στο άρθρό [28] και οι Han et al. στο άρθρο [33] πρότειναν μια δεδομενοκεντρική μέθοδο για την μοντελοποίηση σε επίπεδο υπομονάδων. Η διαφορά τους από την δουλειά του Bauer ήταν ότι εφάρμοσαν αλγόριθμους ομαδοποίησης πάνω σε χρονικά τμήματα και όχι σε ανεξάρτητα πλαίσια λαμβάνοντας υπόψη την δυναμική η οποία αποτελεί κρίσιμο συστατικό της νοηματικής γλώσσας. Οι Yin et al. στο άρθρο [91] πρότειναν μια δεδομενοκεντρική μέθοδο για την αυτόματη εύρεση του συνόλου των υπομονάδων προς μοντελοποίηση. Εφάρμοσαν discriminative feature selection ομαδοποιώντας τα πιο όμοια χρονικά τμήματα. Πιο συγκεκριμένα τα πιο δυσδιάκριτα χρονικά τμήματα των νοημάτων αποτέλεσαν τις βασικές υπομονάδες (fenemes). Οι Kong και Ranganath στο άρθρο [46] πρότειναν επίσης μια δεδομενοκεντρική μέθοδο για την εξαγωγή των βασικών υπομονάδων της ΑΝΓ. Εφάρμοσαν ένα αλγόριθμο κατάτμησης βασισμένοι σε κανόνες (rule-based) για εύρεση/κατάτμηση των περιοχών όπου τα χέρια του νοηματιστή βρίσκονται σε κίνηση. Πιο συγκεκριμένα χρησιμοποίησαν ανάλυση κύριων συνιστωσών για την εξαγωγή χαρακτηριστικών και χρησιμοποίησαν τον αλγόριθμο K-means για την ομαδοποίηση τους.

Κεφάλαιο 2

Εξαγωγή χαρακτηριστικών για την αναγνώριση Νοηματικής Γλώσσας

2.1 Εισαγωγή

Το πρώτο βήμα προς την αυτόματη αναγνώριση της νοηματικής γλώσσας είναι η εξαγωγή χαρακτηριστικών από βίντεο νοηματικού λόγου. Πιο συγκεκριμένα είναι απαραίτητη η εξαγωγή χαρακτηριστικών σχετιζόμενα με τους αρθρωτές που συμμετέχουν στην παραγωγή της νοηματικής γλώσσας. Οι κύριοι αρθρωτές είναι τα χέρια και το κεφάλι του νοηματιστή. Σε αυτό το κεφάλαιο θα παρουσιάσουμε ένα σύστημα για την ανίχνευση και παρακολούθηση των αρθρωτών καθώς και την εξαγωγή χαρακτηριστικών σχετιζόμενων με αυτούς.

Για την εξαγωγή χαρακτηριστικών από βίντεο με συνεχή νοηματική γλώσσα χρησιμοποιούμε το σύστημα που έχει παρουσιαστεί στο άρθρο [5]. Το παραπάνω σύστημα αποτελείται από τα εξής: 1) Ανίχνευση των περιοχών όπου εμφανίζεται χρώμα δέρματος χρησιμοποιώντας ένα πιθανοτικό μοντέλο χρώματος 2) Μορφολογική επεξεργασία των εξαγόμενων μασκών δέρματος 3) Μορφολογική κατάτμηση των εξαγόμενων μασκών δέρματος 4) Παρακολούθηση και την αντιστοίχιση των εξαγόμενων μασκών δέρματος με τα μέρη του σώματος του νοηματιστή 5) Επίλυση/αποσαφήνιση των επικαλύψεων μεταξύ χεριών και κεφαλιού.

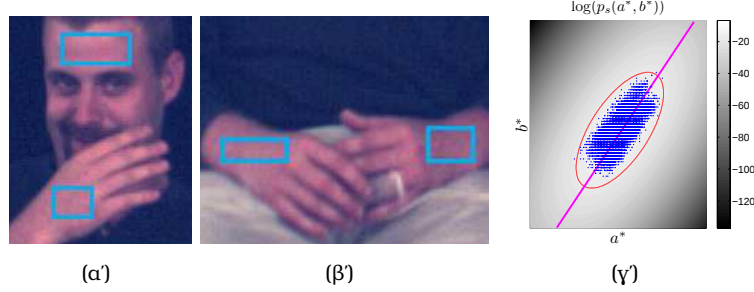
2.2 Ανίχνευση κεφαλιού και χεριών του Νοηματιστή

2.2.1 Πιθανοτικό Μοντέλο Χρώματος Δέρματος

Για την επεξεργασία των βίντεο νοηματικής βασιζόμαστε στη πληροφορία του χρώματος. Εμπνευσμένοι από υπάρχουσες προσεγγίσεις όπως [10, 2, 11]. Βασιζόμενοι στην παρατήρηση ότι το χρώμα του δέρματος του ανθρώπου έχει κάποια συγκεκριμένα χαρακτηριστικά, χρησιμοποιούμε την πληροφορία χρώματος για την ανίχνευση των αρθρωτών που μας ενδιαφέρουν (χέρια και κεφάλι). Ωστόσο απαραίτητη προϋπόθεση είναι ο νοηματιστής να φοράει μπλούζα με μακριά μανίκια. Επιπλέον το χρώμα του φόντου και των ρούχων του νοηματιστή θα πρέπει να διαφέρουν από το χρώμα του δέρματος. Το παραπάνω δεν είναι τόσο περιοριστικό καθώς το χρώμα δέρματος κάθε νοηματιστή έχει αρκετά μικρή διακύμανση [13].

Για την μοντελοποίηση του χρώματος δέρματος χρησιμοποιούμε μια γκαουσιανή κατανομή στον χρωματικό χώρο *CIE-Lab* κρατώντας μόνο τις δύο χρωματικές συνιστώσες a^*, b^* . Με αυτό τον τρόπο αποκτάμε σταθερότητα σε τυχόν μεταβολές στην φωτεινότητα [13]. Υποθέτουμε ότι οι (a^*, b^*) τιμές στα pixel όπου έχουμε δέρμα ακολουθούν μια διμεταβλητή γκαουσιανή κατανομή πυκνότητας πιθανότητας $p_s(a^*, b^*)$, η οποία εκπαιδεύεται χρησιμοποιώντας περιοχές δέρματος. Ενδεικτικά παραδείγματα από επισημειωμένες περιοχές δέρματος απεικονίζονται στα Σχήματα.2.1(α,β). Για

την ανίχνευση των περιοχών οι οποίες αντιστοιχούν σε περιοχές δέρματος υπολογίζουμε την πιθανότητα κάθε εικονοστοιχείου της εικόνας να ανήκει στην παραπάνω γκαουσιανή κατανομή. Καταυτών των τρόπο υπολογίζεται η μάσκα δέρματος S_0 . Η χρησιμοποίηση του χρωματικού μοντέλου στο σύστημα μας διευκολύνει την προσαρμογή σε διαφορετικούς νοηματιστές όπως και σε διαφορετικά βίντεο τα οποία έχουν γυριστεί σε διαφορετικές συνθήκες καθώς μόνο οι παράμετροι τις γκαουσιανής $p_s(a^*, b^*)$ χρειάζεται να προσαρμοστούν.



Σχήμα 2.1: Μοντελοποίηση χρώματος δέρματος. **(α, β)** Παραδείγματα από επισημειωμένες περιοχές δέρματος (τετράγωνα) οι οποίες χρησιμοποιούνται για την εκπαίδευση του χρωματικού μοντέλου. **(γ)** Δεδομένα εκπαίδευσης στο χρωματικό χώρο a^*-b^* μαζί με την κανονική κατανομή $p_s(a^*, b^*)$. Κάθε εικονοστοιχείο αντιστοιχίζεται σε δέρμα βασιζόμενη σε ένα κατώφλι της $p_s(a^*(\vec{x}), b^*(\vec{x}))$. Η ευθεία γραμμή αντιστοιχεί στην πρώτη ιδιοκατεύθυνση μετά από PCA στα δεδομένα εκπαίδευσης.

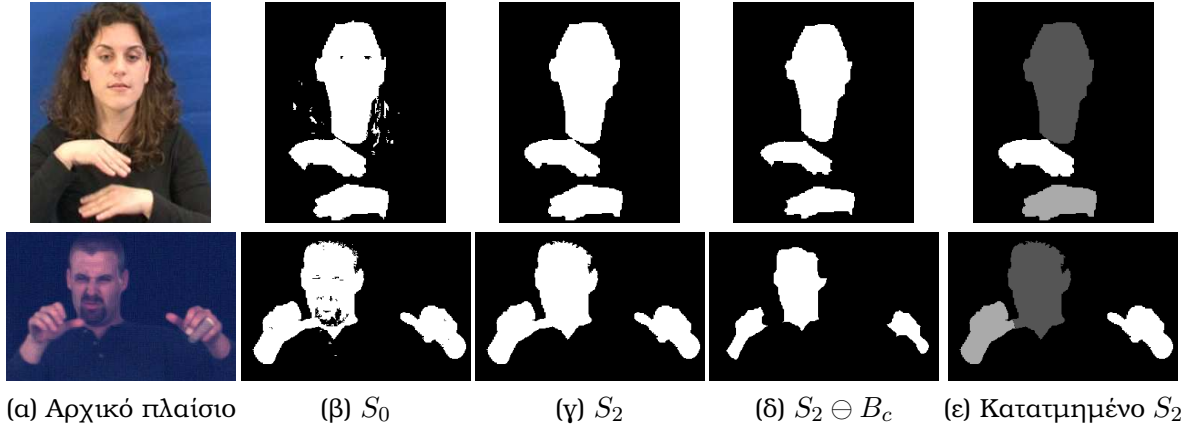
2.2.2 Μορφολογική Επεξεργασία των εξαγομένων Μασκών Δέρματος

Σε κάθε πλαίσιο (frame) γίνεται μια πρώτη εκτίμηση της μάσκας δέρματος S_0 εφαρμόζοντας κατωφλιοποίηση της πιθανότητας $p_s(a^*(\vec{x}), b^*(\vec{x}))$ για κάθε εικονοστοιχείο $\vec{x}$ της εικόνας, όπου p_s είναι η εκπαιδευμένη κατανομή του χρώματος δέρματος Σχ. 2.1(γ) (βλέπε Σχ.2.2(β)). Η τιμή της σταθεράς για την κατωφλιοποίηση υπολογίζεται έτσι ώστε ένα ποσοστό των δεδομένων εκπαίδευσης να ταξινομούνται σε χρώμα δέρματος. Αυτό το ποσοστό ορίζεται έτσι ώστε να είναι ελάχιστα μικρότερο από 100% με στόχο να μην συμπεριληφθούν πιθανά δεδομένα σε ακραίες τιμές. Στα πειράματά μας χρησιμοποιήσαμε ποσοστό 99%.

Η παραπάνω μάσκα δέρματος S_0 ενδέχεται να περιλαμβάνει λάθη, όπως τρύπες εσωτερικά του κεφαλιού εξαιτίας περιοχών με διαφορετικό χρώμα από το δέρμα όπως π.χ. μάτια, στόμα κ.τ.λ. Για αυτό το λόγο προτείνουμε ένα αλγόριθμο για την διόρθωση αυτών των λαθών ο οποίος χρησιμοποιεί εργαλεία από την μαθηματική μορφολογία [35, 52]. Χρησιμοποιούμε την έννοια των *τρυπών* (holes) $\mathcal{H}(S)$ σε μια δυαδική εικόνα S [68]. Για να γεμίσουμε κάποιες περιοχές του φόντου (hole filling) η οποίες δεν είναι τρύπες με την στενή έννοια αλλά ενώνονται με το φόντο μέσω ενός μικρού καναλιού εφαρμόζουμε το παρακάτω γενικευμένο hole filling από το οποίο προκύπτει μια καλύτερη μάσκα δέρματος S_1 :

$$S_1 = S_0 \cup \mathcal{H}(S_0) \cup \{\mathcal{H}(S_0 \bullet B) \oplus B\} \quad (2.1)$$

όπου B είναι ένα δομικό στοιχείο μικρού μεγέθους, $\oplus$ και $\bullet$ ορίζουν το Minkowski dilation και closing αντίστοιχα. Στην θεωρία η καλύτερη επιλογή για το δομικό στοιχείο B είναι ένας δίσκος, έτσι ώστε να υπάρχει ισάζια αντιμετώπιση προς όλες τις διευθύνσεις. Όμως για υπολογιστικούς λόγους στην υλοποίηση μας χρησιμοποιήσαμε ένα τετράγωνο 5×5 εικονοστοιχείων για το B αφού ο υπολογισμός του erosion/dilation με δομικό στοιχείο σχήματος τετράγωνου είναι αρκετά πιο γρήγορος.



Σχήμα 2.2: Ενδιάμεσα και τελικά αποτελέσματα για την εξαγωγή της μάσκας δέρματος και της μορφολογικής κατάτμησης εφαρμοσμένα σε ένα πλαίσιο από δύο βάσεις δεδομένων. (α) Αρχικό πλαίσιο (β) Αρχική εκτίμηση της μάσκα δέρματος S_0 , η οποία έχει προκύψει από κατωφλιοποίηση της πιθανότητας δέρματος σε κάθε εικονοστοιχείο. (γ) Τελική εκτίμηση της μάσκα δέρματος S_2 , η οποία έχει προκύψει μετά από μορφολογική βελτίωση της S_0 . (δ) Erosion $S_2 \ominus B_c$ της S_2 με ένα μικρό δίσκο όπου εφαρμόζεται στην S_2 με σκοπό να βοηθάει στον διαχωρισμό επικαλυπτομένων αρθρωτών. (ε) Κατάτμηση της μάσκας δέρματος S_2 , σε συνεκτικές συνιστώσες μαζί με την εφαρμογή του αλγορίθμου competitive reconstruction opening (ενότητα 2.2.3).

Στην συνέχεια για να διορθώσουμε πιθανά λάθη εκμεταλλευτήκαμε πρότερη γνώση : Πιο συγκεκριμένα οι συνεκτικές συνιστώσες που είναι χρώμα δέρματος πρέπει να είναι το πολύ τρεις (κεφάλι, αριστερό χέρι, δεξί χέρι). Επιπλέον θεωρούμε πως δεν μπορούμε να έχουμε περιοχή με εμβαδόν μικρότερο από ένα κατώφλι A_{min} το οποίο αντιστοιχεί στο μικρότερο δυνατό εμβαδόν του χεριού του υπό μελέτη νοηματιστή και βίντεο.¹

Μετέπειτα εφαρμόζουμε ένα area opening, βρίσκουμε τις συνεκτικές συνιστώσες/περιοχές του S_1 , υπολογίζουμε τα εμβαδά τους και τελικά απορρίπτουμε όσες περιοχές έχουν εμβαδόν μικρότερο από A_{min} . Τελικά κρατάμε τις τρεις περιοχές με το μεγαλύτερο εμβαδόν καταλήγοντας στην τελική εκτίμηση της μάσκας δέρματος S_2 Σχήμα. 2.2(γ).

2.2.3 Μορφολογική Κατάτμηση των Μασκών Δέρματος

Η μάσκα δέρματος S_2 που υπολογίσαμε στην προηγούμενη ενότητα περιλαμβάνει τους τρεις αρθρωτές (κεφάλι και χέρια). Για να τους διαχωρίσουμε ακολουθούμε την παρακάτω μέθοδο κατάτμησης βασισμένοι στην μαθηματική μορφολογία. Όταν η μάσκα S_2 περιέχει τρεις συνεκτικές περιοχές δηλαδή δεν έχουμε επικάλυψη μεταξύ των αρθρωτών, η κατάτμηση είναι άμεση εντοπίζοντας τις συνεκτικές περιοχές. Στην περίπτωση που έχουμε επικαλύψεις η οποίες δεν είναι αρκετά μεγάλες, δηλαδή όταν δυο/τρεις διαφορετικές περιοχές ενώνονται με μια μικρή γέφυρα Σχήμα. 2.2(γ) προτείνουμε την μέθοδο *competitive reconstruction opening* για να τις διαχωρίσουμε. Η μέθοδος αυτή βασίζεται σε μορφολογικά αλληπάλληλα φιλτραρίσματα.

Πιο συγκεκριμένα αν η μάσκα S_2 περιέχει N_{cc} συνεκτικές περιοχές/συνιστώσες όπου $N_{cc} < 3$ βρίσκουμε τις συνεκτικές περιοχές/συνιστώσες του $S_2 \ominus B_c$ χρησιμοποιώντας δομικό στοιχείο B_c ένα δίσκο με ακτίνα 3 εικονοστοιχεία (βλέπε Σχ. 2.2(δ) και απορρίπτουμε της περιοχές η οποίες έχουν εμβαδόν μικρότερο από A_{min} . Εάν ο αριθμός των συνεκτικών περιοχών παραμείνει N_{cc} σημαίνει πως οι αρθρωτές έχουν μεγάλη επικάλυψη μεταξύ τους με αποτέλεσμα ο διαχωρισμός

¹Ο περιορισμός αυτός προκύπτει από την λογική υπόθεση ότι το εμβαδόν του κεφαλιού είναι πάντα μεγαλύτερο από το εμβαδόν των χεριών ενός ανθρώπου [87]

τους να μην είναι εφικτός με την μέθοδο *competitive reconstruction opening*. Σε αντίθετη περίπτωση εάν ο αριθμός των συνεκτικών περιοχών γίνει μεγαλύτερος από N_{cc} υποδεικνύεται μερική επικάλυψη. Σε αυτή την περίπτωση με αφετηρία αυτές τις συνεκτικές περιοχές επεκτείνουμε με στόχο να καλύψουμε την μάσκα S_2 . Στο Σχήμα 2.2(ε) μπορούμε να δούμε το αποτέλεσμα του *competitive reconstruction opening* αλγορίθμου όπου σε κάθε βήμα επεκτείνουμε κάθε συνεκτική περιοχή εφαρμόζοντας conditional dilation και αφαιρούμε από όλες τις συνεκτικές περιοχές τα εικονοστοιχεία που ανήκουν σε παραπάνω από δύο περιοχές. Αυτό έχει ως αποτέλεσμα οι αρχικές συνεκτικές περιοχές να επεκτείνονται εσωτερικά της μάσκας S_2 έως ότου συναντηθούν μεταξύ τους.

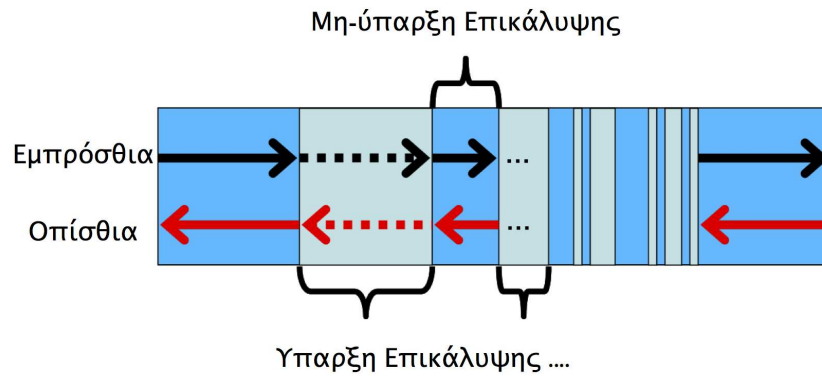
2.3 Ανίχνευση και Παρακολούθηση των Χεριών και του Κεφαλιού

Μετά την εφαρμογή της κατάτμησης στην μάσκα δέρματος S_2 εφαρμόζουμε ένα αλγόριθμο για την ανίχνευση και παρακολούθηση των χεριών και του κεφαλιού. Ο παρακάτω αλγόριθμος έχει ως αποτέλεσμα την εύρεση των αρθρωτών που περιλαμβάνει κάθε κατατμημένη περιοχή σε κάθε πλαίσιο και την εκτίμηση των παραμέτρων μιας έλλειψης για κάθε αρθρωτή που συμμετέχει σε επικάλυψη. Οι ελλείψεις αυτές αποτελούν μια αδρή εκτίμηση του σχήματος και της θέσης των αρθρωτών στις περιοχές όπου έχουμε επικαλύψεις. Η χρησιμοποίηση ελλείψεων βασίζεται στην υπόθεση ότι μια έλλειψη μπορεί προσεγγιστικά να περιγράψει το σχήμα του κεφαλιού και των χεριών ενός ανθρώπου [2]. Για την ανίχνευση και παρακολούθηση των αρθρωτών σε κάθε χρονικό πλαίσιο χρησιμοποιούμε διαφορετική προσέγγιση ανάλογα με την ύπαρξη επικάλυψης.

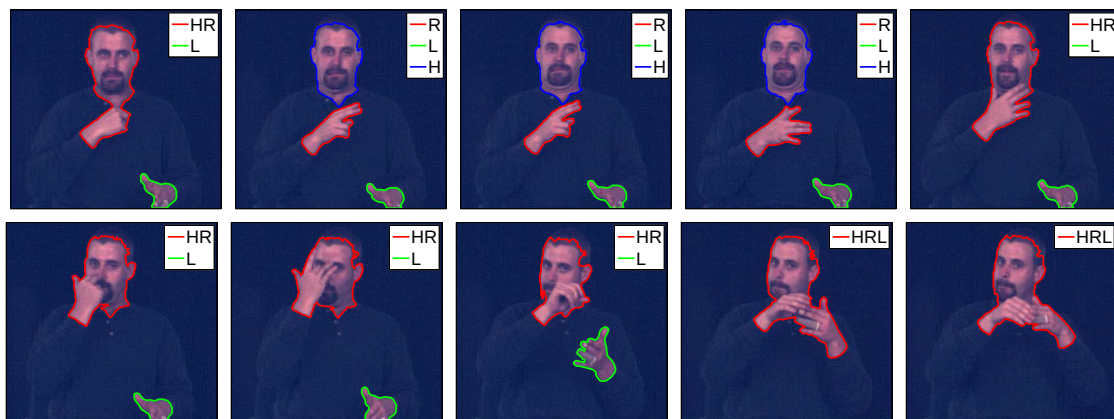
Μη ύπαρξη επικάλυψης: Το αποτέλεσμα της κατάτμησης της μάσκας δέρματος S_2 είναι 3 διαφορετικές συνεκτικές περιοχές. Αντιστοιχούμε το κεφάλι στην περιοχή με το μεγαλύτερο εμβαδόν υποθέτοντας ότι το εμβαδόν του κεφαλιού είναι πάντα μεγαλύτερο από των χεριών. Δεδομένου ότι τα δύο χέρια έχουν ανιχνευθεί στα προηγούμενα πλαίσια εφαρμόζουμε μια γραμμική πρόβλεψη της θέσης του κεντροειδούς για κάθε χέρι χρησιμοποιώντας τα τρία προηγούμενα πλαίσια. Οι συντελεστές πρόβλεψης υπολογίζονται χρησιμοποιώντας ένα απλό μοντέλο σταθερής επιτάχυνσης. Έτσι κάθε χέρι αντιστοιχίζεται στην συνεκτική περιοχή που βρίσκεται πιο κοντά στην εκτιμώμενη θέση του κεντροειδούς του.

Ύπαρξη επικάλυψης: Το αποτέλεσμα της κατάτμησης της μάσκας δέρματος S_2 είναι 1 ή 2 διαφορετικές συνεκτικές περιοχές. Σε αυτή την περίπτωση εφαρμόζουμε ένα συνδυασμό των δυο παρακάτω μεθόδων: 1) Εμπρόσθια-οπίσθια γραμμική εκτίμηση των παραμέτρων της έλλειψης για κάθε αρθρωτή (βλ. Σχήμα 2.3) και 2) Template matching μεταξύ διαδοχικών χρονικών πλαισίων εκμεταλλευόμενοι την εκ των προτέρων πληροφορία για το σχήμα και την υφή των αρθρωτών από το προηγούμενο/επόμενο πλαίσιο. Πιο συγκεκριμένα χρησιμοποιώντας τις προ-υπολογισμένες παραμέτρους των ελλείψεων που έχουμε ταιριάζει στους αρθρωτές στα τρία προηγούμενα πλαίσια εφαρμόζουμε μια εμπρόσθια γραμμική πρόβλεψη των παραμέτρων της έλλειψης για κάθε αρθρωτή για το τρέχον πλαίσιο. Λόγο της ευαισθησίας της γραμμικής πρόβλεψης σε σχέση με τον αριθμό των συνεχόμενων πλαισίων όπου έχουμε επικάλυψη, εφαρμόζουμε επιπλέον οπίσθια γραμμική πρόβλεψη (δες Σχήμα 2.3). Επίσης κάνοντας την υπόθεση ότι ένας αρθρωτής (κεφάλι ή χέρι) δεν μεταβάλλεται έντονα μεταξύ δύο διαδοχικών χρονικών πλαισίων εφαρμόζουμε template matching μεταξύ διαδοχικών χρονικών πλαισίων αποκτώντας μια επιπλέον εκτίμηση των αρθρωτών στο τρέχον πλαίσιο. Τελικά συνδυάζοντάς τις εκτιμήσεις από την εμπρόσθια-οπίσθια γραμμική πρόβλεψη και την template matching τεχνική αποκτάμε μια τελική εκτίμηση της θέσης των αρθρωτών.

Στα Σχήματα 2.4,2.5,2.6,2.7 έχουμε απεικονίσει τα αποτελέσματα από την ανίχνευση και παρακολούθηση των χεριών και κεφαλιού του νοηματοστή σε μια ακολουθία από πλαίσια η οποία περιλαμβάνει περιπτώσεις όπου έχουμε επικαλύψεις.



Σχήμα 2.3: Εναλλαγή πλαισίων με ύπαρξη ή μη επικαλύψεων: Σχηματική αναπαράσταση της εμπρόςθια-οπίσθια γραμμικής εκτίμησης



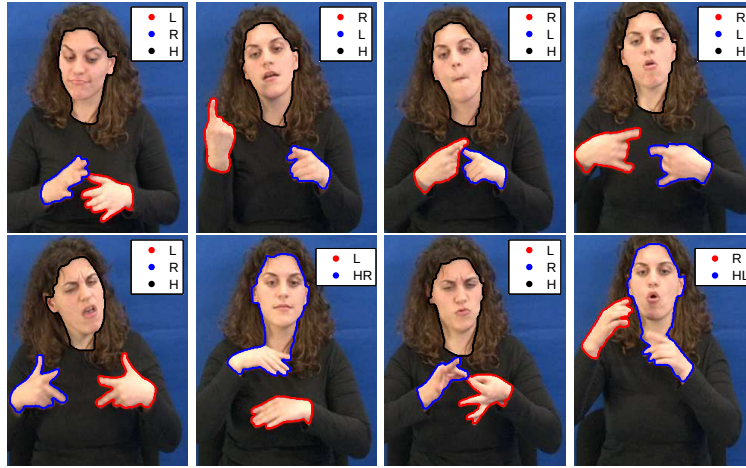
Σχήμα 2.4: Αποτέλεσμα της ανίχνευσης και παρακολούθησης των χεριών και του κεφαλιού του νοηματιστή και σε περιπτώσεις όπου έχουμε επικαλύψεις από την βάση δεδομένων BU400. Παράδειγμα πλαισίων με την μάσκα δέρματος και τους αρθρωτές που περιλαμβάνει κάθε κατατημένη περιοχή. Το H αντιστοιχεί στο κεφάλι, το L στο αριστερό χέρι και το R στο δεξί χέρι.



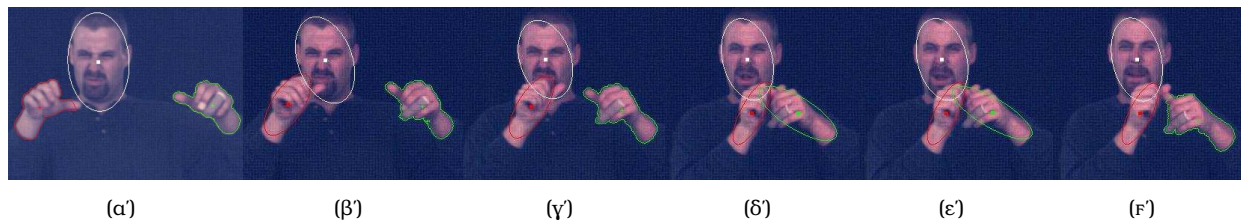
Σχήμα 2.5: Αποτέλεσμα της ανίχνευσης και παρακολούθησης των χεριών και κεφαλιού του νοηματιστή και σε περιπτώσεις όπου έχουμε επικαλύψεις από την βάση δεδομένων Dicta-Sign Corpus.

2.4 Μοντελοποίηση και εξαγωγή χαρακτηριστικών για τις χειρομορφές

Σε αυτή την ενότητα για λόγους συνέπειας θα παρουσιάσουμε το σύστημα Δυναμική Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος-Εμφάνισης (Dynamic Affine-invariant Shape-Appearance



Σχήμα 2.6: Αποτέλεσμα της ανίχνευσης και παρακολούθησης των χεριών και κεφαλιού του νοηματιστή και σε περιπτώσεις όπου έχουμε επικαλύψεις από την βάση δεδομένων Dicta-Sign Corpus.



Σχήμα 2.7: Αποτέλεσμα της ανίχνευσης και παρακολούθησης των χεριών και κεφαλιού του νοηματιστή και σε περιπτώσεις όπου έχουμε επικαλύψεις μαζί με τις ελλείψεις που έχουν ταιριάζει σε κάθε αρθρωτή στις περιπτώσεις που έχουμε επικαλύψεις στην βάση δεδομένων BU400.

Model) το οποίο αναπτύχθηκε στα πλαίσια τις διατριβής του Α. Ρούσσου [;, 1]. Το σύστημα Δυναμική Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος-Εμφάνισης προσφέρει μια αναπαράσταση των χειρομορφών. Επιπλέον επιτρέπει την παρακολούθηση και εξαγωγή χαρακτηριστικών σχετιζόμενα με το σχήμα και την εμφάνιση από τα χέρια ενός νοηματιστή. Τέλος αποτελεί ένα ευσταθές και ακριβές σύστημα ακόμα και κατά την διάρκεια επικαλύψεων.

Αναπαράσταση Χεριών με εικόνες Σχήματος-Εμφάνισης (ΣΕ): Πρωταρχικός στόχος είναι η μοντελοποίηση όλων των δυνατών διαφορετικών χειρομορφών που εμφανίζονται κατά την διάρκεια του νοηματισμού χρησιμοποιώντας δισδιάστατες εικόνες. Αυτές οι εικόνες έχουν μεγάλη διαφορετικότητα λόγω της ποικιλίας των χειρομορφών και της τρισδιάστατης πόζας. Επίσης το σύνολο των σημείων πάνω στο επίπεδο του χεριού που είναι ορατά από την κάμερα αλλάζει διαρκώς. Αυτό έχει ως αποτέλεσμα για τα πλαίσια αυτής της εφαρμογής την μη χρησιμοποίηση ενδεικτικών σημείων (landmark points) για την αναπαράσταση της δισδιάστατης χειρομορφής. Έτσι για την αναπαράσταση των χειρομορφών χρησιμοποιούμε την δυαδική μάσκα η οποία έχει προκύψει από την ανίχνευση του χρώματος δέρματος (δες εν. 2.2.1). Επιπλέον χρησιμοποιούμε την εμφάνιση των χεριών δηλαδή τις χρωματικές τιμές εσωτερικά της δυαδικής μάσκας. Αυτές οι τιμές εξαρτώνται από την υφή και σκίαση του χεριού με αποτέλεσμα να προσφέρουν πληροφορία για την τρισδιάστατη χειρομορφή.

Μοντελοποίηση της ποικιλίας των εικόνων ΣΕ: Οι ΣΕ εικόνες των χεριών, $f(x)$, μοντελοποιούνται χρησιμοποιώντας ένα γραμμικό συνδυασμό προκαθορισμένων εικόνων μεταβολής,

ακολουθούμενο από ένα αφινικό μετασχηματισμό :

$$f(W_{\mathbf{p}}(\mathbf{x})) \approx A_0(\mathbf{x}) + \sum_{i=1}^{N_c} \lambda_i A_i(\mathbf{x}), \mathbf{x} \in \Omega \quad (2.2)$$

όπου $A_0(\mathbf{x})$ είναι η εικόνα βάσης, $A_i(\mathbf{x})$ είναι N_c ιδιοεικόνες (eigenimages) που μοντελοποιούν την γραμμική μεταβολή, $\boldsymbol{\lambda} = (\lambda_1 \cdots \lambda_{N_c})$ είναι τα βάρη του γραμμικού συνδυασμού και $W_{\mathbf{p}}$ είναι ο αφινικός μετασχηματισμός με παραμέτρους $\mathbf{p} = (p_1 \cdots p_6)$ που αντιστοιχεί από τον χώρο του μοντέλου Ω στον χώρο της εικόνας. Από εδώ και στο εξής θα αναφερόμαστε στο προτεινόμενο μοντέλο ως *Μοντέλο Σχήματος-Εμφάνισης (ΜΣΕ)*. Ένα συγκεκριμένο ΜΣΕ χεριού καθορίζεται από την εικόνα βάσης $A_0(\mathbf{x})$ και τις ιδιοεικόνες $A_i(\mathbf{x})$ του γραμμικού συνδυασμού και τον αριθμό N_c . Τα διανύσματα $\mathbf{p}$ και $\boldsymbol{\lambda}$ είναι οι παράμετροι του μοντέλου που ταιριάζουν το δεδομένο μοντέλο σε μια εικόνα ΣΕ σε κάθε χρονικό πλαίσιο. Οι παράμετροι αυτοί θεωρούνται ως χαρακτηριστικά της πόζας και του σχήματος του χεριού αντίστοιχα.

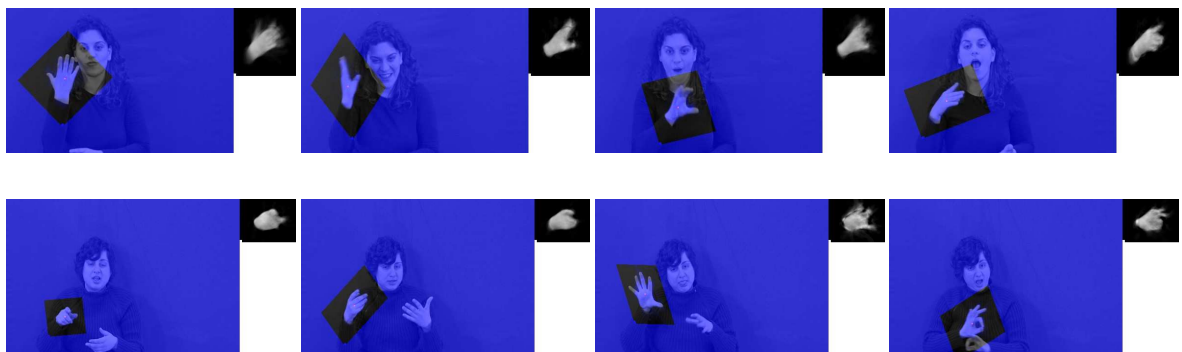
Εκπαίδευση του γραμμικού συνδυασμού του ΜΣΕ: Για την εκπαίδευση του μοντέλου Σχήματος-Εμφάνισης χρησιμοποιούμε ένα αντιπροσωπευτικό σύνολο από εικόνες χειρομορφών από χρονικά πλαίσια όπου το χέρι είναι πλήρως ορατό και δεν υπάρχουν επικαλύψεις. Δεδομένης αυτής της επιλογής το σύνολο εκπαίδευσης κατασκευάζεται από τις αντίστοιχες ΣΕ εικόνες $f_1 \cdots f_N$. Για την αφαίρεση της μεταβλητότητας που εξηγεί ο αφινικός μετασχηματισμός εφαρμόζουμε μια ημιαυτόματη διαδικασία αφινικής ευθυγράμμισης του συνόλου εκπαίδευσης. Στην συνέχεια οι εικόνες του γραμμικού συνδυασμού μαθαίνονται χρησιμοποιώντας PCA στο ευθυγραμμισμένο σύνολο.

Ομαλοποιημένο Ταίριασμα του ΜΣΕ με Στατική και Δυναμική Πρότερη Πληροφορία :

Για την αντιμετώπιση των επικαλύψεων χρησιμοποιήθηκε ένα ομαλοποιημένο ταίριασμα του ΜΣΕ το οποίο εκμεταλλεύεται πρότερη πληροφορία που σχετίζεται με την δυναμική της χειρομορφής. Πιο συγκεκριμένα προσθέτουμε στο μέσο τετραγωνικό σφάλμα ανακατασκευής E_{rec} , δύο όρους που αντιστοιχούν στην στατική και δυναμική πρότερη γνώση των παραμέτρων του ΜΣΕ $\boldsymbol{\lambda}$ και $\mathbf{p}$. Σε κάθε πλαίσιο n , βρίσκουμε τα βέλτιστα $\boldsymbol{\lambda} = \boldsymbol{\lambda}[n]$ και $\mathbf{p} = \mathbf{p}[n]$ που ελαχιστοποιούν την συνολική ενέργεια :

$$E(\boldsymbol{\lambda}, \mathbf{p}) = E_{rec}(\boldsymbol{\lambda}, \mathbf{p}) + w_S E_S(\boldsymbol{\lambda}, \mathbf{p}) + w_D E_D(\boldsymbol{\lambda}, \mathbf{p}) , \quad (2.3)$$

όπου w_S, w_D είναι θετικά βάρη που ελέγχουν την ισορροπία μεταξύ των τριών όρων. Οι όροι $E_S(\boldsymbol{\lambda}, \mathbf{p})$ και $E_D(\boldsymbol{\lambda}, \mathbf{p})$ αντιστοιχούν στην ενέργεια της στατικής και δυναμικής πρότερης πληροφορίας. Στο Σχήμα 2.8 έχουμε απεικονίσει τα αποτελέσματα από το ομαλοποιημένο ταίριασμα του κυρίαρχου (δεξιού) χεριού του ΜΣΕ. Παρατηρούμε ότι κατά την διάρκεια των επικαλύψεων έχουμε ακριβή παρακολούθηση του χεριού.



Σχήμα 2.8: Ομαλοποιημένο Ταίριασμα του ΜΣΕ για δύο διαφορετικούς νοηματιστές (O11A,012B) πρώτη και δεύτερη σειρά αντίστοιχα από την βάση δεδομένων Dicta-Sign Συνεχής ΕΝΓ. Σε κάθε αρχική εικόνα πλαισίου υπερθέτουμε την ανακατασκευή βασισμένοι στο μοντέλο, $A_0(W_p^{-1}(x)) + \sum \lambda_i A_i(W_p^{-1}(x))$. Στην πάνω δεξιά γωνία δείχνουμε κάθε φορά την ανακατασκευή, αλλά στο χώρο του μοντέλου ΣΕ $A_0(x) + \sum \lambda_i A_i(x)$, η οποία καθορίζει τα βέλτιστα βάρη.

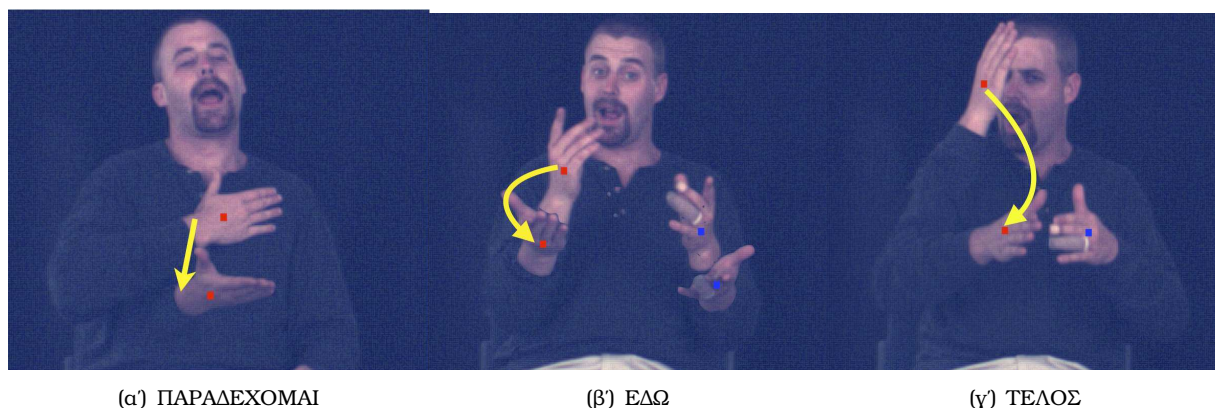
Κεφάλαιο 3

Μοντελοποίηση της Νοηματικής Γλώσσας με δεδομενοκεντρικές υπομονάδες

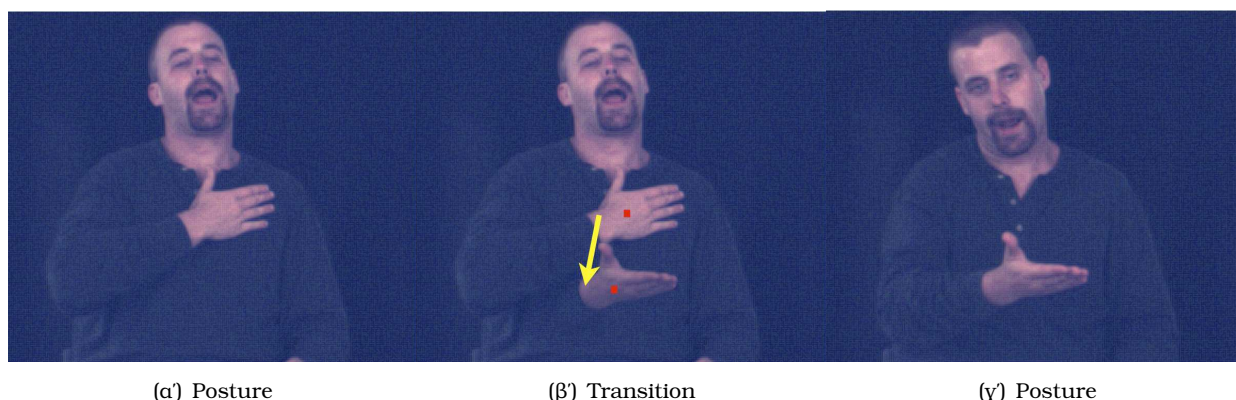
3.1 Εισαγωγή στην έννοια των υπομονάδων

Οι πρώτες προσπάθειες για την αναγνώριση νοηματικής γλώσσας αντιμετώπισαν απλά προβλήματα αναγνώρισης [58] παρόμοια με αυτά που είχαν αντιμετωπιστεί στην περιοχή της αναγνώριση φωνής λίγες δεκαετίες παλαιότερα. Τέτοια προβλήματα αφορούσαν αναγνώριση μεμονωμένων νοημάτων από ένα μικρό ή μεσαίο σε μέγεθος λεξιλόγιο, μερικών δεκάδων έως μια ή δύο εκατοντάδων νοημάτων, χρησιμοποιώντας μοντέλα στο επίπεδο των νοημάτων. Μια άτυπη και χονδροειδής αντιστοιχία ενός νοήματος στην νοηματική γλώσσα είναι η λέξη στην ομιλούμενη γλώσσα. Οι ομοιότητες της νοηματικής γλώσσας και της ομιλούμενης γλώσσας έχουν ως αποτέλεσμα αρκετές μέθοδοι οι οποίες αναπτύχθηκαν στα πλαίσια της αναγνώρισης φωνής να χρησιμοποιούνται για την αναγνώριση νοηματικής γλώσσας. Ωστόσο οι διαφορές των δύο γλωσσών οδηγούν σε διαφοροποιήσεις στην εφαρμογή των μεθόδων αυτών. Οι νοηματικές γλώσσες σε αντίθεση με τις προφορικές τείνουν να είναι μονοσυλλαβικές και πολυμορφικές [26]. Στο Σχήμα. 3.1 παρουσιάζονται 3 νοήματα στην αμερικάνικη νοηματική γλώσσα (ΠΑΡΑΔΕΧΟΜΑΙ, ΕΔΩ, ΤΕΛΟΣ). Τα 3 νοήματα συντίθενται από μια κίνηση ενώ τα υπόλοιπα χαρακτηριστικά της άρθρωσης όπως είδος/προσανατολισμός χειρομορφής, θέση των χεριών, χαρακτηριστικά του προσώπου κ.τ.λ μπορούν να μεταβάλλονται παράλληλα. Μια ακόμα σημαντική διαφορά της νοηματικής και προφορικής γλώσσας αφορά τις φωνητικές υπομονάδες που συνθέτουν ένα νόημα. Οι φωνητικές υπομονάδες έχουν τελείως διαφορετική φύση συγκρινόμενες με τις αντίστοιχες στην ομιλούμενη γλώσσα δηλαδή τα φωνήματα. Αυτή η διαφορά σχετίζεται με την παραλληλία η οποία υπάρχει κατά την διάρκεια της άρθρωσης της νοηματικής γλώσσας λόγω των πολλών ανεξάρτητων αρθρωτών. Ωστόσο υπάρχουν ενδείξεις για την ύπαρξη μιας δομής αλληλουχίας κατά την διάρκεια της άρθρωσης ενός νοήματος [49, 48, 18]. Η έννοια του φωνήματος/υπομονάδας στην νοηματική δεν είναι δεδομένη ούτε πλήρως καθορισμένη όπως στην ομιλούμενη γλώσσα. Παρόλα αυτά έχουν γίνει αρκετές γλωσσολογικές μελέτες και έχουν προταθεί διάφορα μοντέλα από τα οποία η προσέγγιση του Stokoe και των Liddell και Johnson [72, 48, 42] είναι ίσως τα πιο διαδεδομένα πάνω στα οποία βασίστηκε και η δική μας έρευνα.

Ο Stokoe το 1980 στο άρθρο [72] πρότεινε ότι οι φωνητικές υπομονάδες που αποτελούν τα δομικά στοιχεία που συνθέτουν κάθε νόημα είναι τριών ειδών (TAB, SIGS, DEZ). Πιο συγκεκριμένα κάθε νόημα μπορεί να σπάσει σε 3 βασικές παραμέτρους: την θέση των χεριών (TAB), το είδος της κίνησης τους (SIGS), και το είδος της χειρομορφής (DEZ). Ένα παράδειγμα παρουσιάζεται στο



Σχήμα 3.1: Παραδείγματα 3 νοημάτων της Αμερικάνικης Νοηματικής Γλώσσας (ΑΝΓ) τα οποία εμφανίζονται στην βάση δεδομένων BU400 Συνεχής ΑΝΓ. Τα παραδείγματα αυτά είναι ενδεικτικές προφορές ανάμεσα σε πολλές διαφορετικές προφορές που εμφανίζονται στην συγκεκριμένη βάση. Οι πολλαπλές αυτές προφορές οφείλονται στο συνεχή νοηματικό λόγο όπου φαινόμενα όπως συνάρθρωση εμφανίζονται, με αποτέλεσμα την διαφοροποίηση της εκτέλεσης ενός νοήματος και την εξάρτηση της από το συγκείμενο στο οποίο βρίσκεται.



Σχήμα 3.2: Διάσπαση του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ στις PDTS υπομονάδες από τις οποίες αποτελείται.

Σχήμα.3.1(α), το νόημα ΠΑΡΑΔΕΧΟΜΑΙ στην ΑΝΓ με βάση τον Stokoe αποτελείται από : 1) Θέση του κύριου χεριού (δεξιό) στο θώρακα, 2) Ευθεία κίνηση προς τα κάτω και έξω και 3) Χειρομορφή ανοιχτή η παλάμη κλειστά δάκτυλα και ανοιχτός ο αντίχειρας (Flat Hand -B,5-).

Στην συνέχεια το 1989 οι Liddell και Johnson πρότειναν ένα διαφορετικό μοντέλο για τα δομικά στοιχεία κάθε νοήματος το οποίο το ονόμασαν Movement-Hold μοντέλο [48]. Η βασική διαφορά μεταξύ αυτής της προσέγγισης και του Stokoe είναι πως το Movement-Hold μοντέλο ενσωματώνει την έννοια του χρόνου. Πιο συγκεκριμένα η νέα αυτή μοντελοποίηση λαμβάνει υπόψιν την φωνολογική παραλληλία αλλά και διαδοχή [49]. Το Movement-Hold μοντέλο αποτελείται από δύο ειδών υπομονάδες : 1) Κίνηση (Movement -M-), ύπαρξη μεταβολής τουλάχιστον μιας από τις 3 παραμέτρους. 2) Στάση (Hold -H-), μη-ύπαρξη μεταβολής καμίας από τις 3 παραμέτρους του Stokoe Παραδείγματος χάριν το νόημα ΠΑΡΑΔΕΧΟΜΑΙ στην ΑΝΓ Σχήμα.3.1(α) με βάση το Movement-Hold μοντέλο αναλύεται στην ακολουθία -H M H-.

Πιο πρόσφατα το 2011 οι Liddell και Johnson στο άρθρο [41] πρότειναν μια επέκταση του Movement-Hold μοντέλου που ονόμασαν Posture-Detention-Transition-Steady-Shift (PDTS). Σύμ-

φωνα με το μοντέλο αυτό οι υπομονάδες των νοημάτων χωρίζονται σε 4 κατηγορίες 1) Στάσεις (Postures), δηλαδή η μη-ύπαρξη μεταβολής κανενός αρθρώτη το δεδομένο χρονικό παράθυρο 2) Μικρές Στάσεις (Detentions) αντίστοιχες με τις στάσεις αλλά πολύ μικρότερης διάρκειας 3) Κινήσεις (Transitions), δηλαδή η ύπαρξη μεταβολής ενός τουλάχιστον αρθρώτη το δεδομένο χρονικό παράθυρο και 4) Σταθερής-Αλλαγής (Steady-Shift) αντίστοιχες με τις κινήσεις αλλά μικρότερης ταχύτητας. Ένα παράδειγμα των PDTs υπομονάδων που απαρτίζουν το νόημα ΠΑΡΑΔΕΧΟΜΑΙ παρουσιάζεται στο Σχήμα. 3.2. Όπως φαίνεται στο σχήμα το νόημα ΠΑΡΑΔΕΧΟΜΑΙ αποτελείται από την ακολουθία -P T P- δηλαδή από μια στάση στο θώρακα (Posture -P-) ακολουθούμενη από μια ευθεία κίνηση προς τα κάτω (Transition -T-) και τέλος από μια στάση στον ουδέτερο χώρο (Posture -P-).

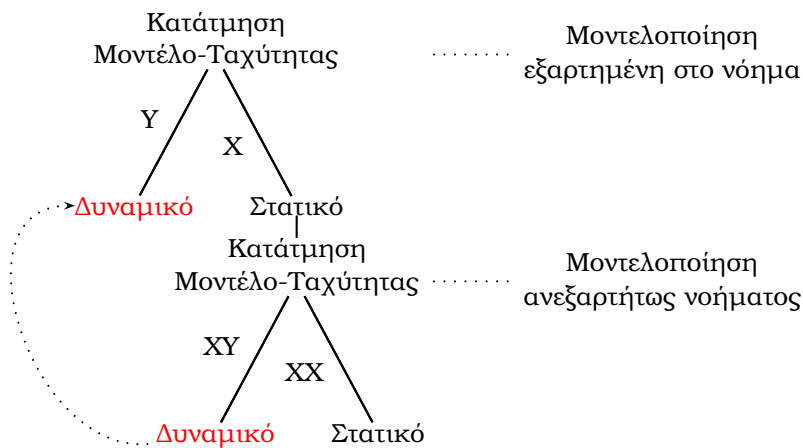
Όπως αναφέραμε και προηγουμένως η έννοια του φωνήματος/υπομονάδας στην νοηματική δεν είναι δεδομένη ούτε πλήρως καθορισμένη όπως στην ομιλούμενη γλώσσα. Επιπλέον η ανθρώπινη επισημείωση των νοηματικών βάσεων δεδομένων σε επίπεδο υπομονάδων -αντίστοιχη με αυτή που έχει γίνει στην ομιλούμενη γλώσσα - είναι αρκετά χρονοβόρα. Αυτό έχει ως αποτέλεσμα την άνθηση δεδομενοκεντρικών μεθόδων για την εύρεση και μοντελοποίηση των υπομονάδων της νοηματικής γλώσσας.

Στο παρόν κεφάλαιο θα παρουσιάσουμε μια νέα δεδομενοκεντρική μέθοδο για την εύρεση και μοντελοποίηση των υπομονάδων της νοηματικής γλώσσας την οποία ονομάζουμε Dynamic/Static Classification Modeling (DSCM). Η DSCM μέθοδος αποσκοπεί στην αυτόματη κατάτμηση, εύρεση και μοντελοποίηση δεδομενοκεντρικών υπομονάδων λαμβάνοντας υπόψη την φωνολογική παραλληλία αλλά και αλληλουχία [49, 42]. Οι υπομονάδες αυτές αποτελούν τα δομικά στοιχεία που συνθέτουν ένα νόημα. Παρόλο που δεν γίνεται άμεση χρήση γλωσσολογικής πληροφορίας η μέθοδος είναι εμπνευσμένη από την προσέγγιση των Liddell και Johnson.

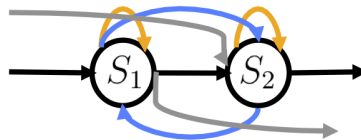
Οι ροές πληροφορίας που εκμεταλλευόμαστε σχετίζονται με την θέση και κίνηση του κυρίαρχου χεριού. Η θέση και η κίνηση είναι ανάμεσα στις πιο σημαντικές ροές πληροφορίας που αποτελούν ένα νόημα γλωσσικά και αντιληπτικά [72, 26, 18, 61]. Η κίνηση των χεριών ανήκει στα πιο ηχηρά μέρη των νοημάτων ενώ η θέση είναι από τα κυρίαρχα χαρακτηριστικά τα οποία αντιλαμβάνεται με ευκολία ο άνθρωπος [61] και είναι συμφωνικό τύπου. Από εδώ και πέρα θα αναφερόμαστε σε αυτές τις δύο ροές πληροφορίες ως ροές *κίνησης-θέσης*. Βασιζόμενοι στις ροές κίνησης-θέσης, προχωράμε στην αυτόματη δεδομενοκεντρική μοντελοποίηση των υπομονάδων της νοηματικής γλώσσας που αντιστοιχούν σε εσωτερικά τμήματα των νοημάτων. Το σύστημα αποτελείται από τα εξής : 1) Αυτόματη κατάτμηση και ταξινόμηση σε δυναμικά (dynamic) και στατικά (static) τμήματα σχετικά με την κίνηση και μη-κίνηση των χεριών εμπνευσμένη από το Movement-Hold μοντέλο. 2) Μοντελοποίηση των δυναμικών και στατικών υπομονάδων βασισμένη στην παραπάνω κατάτμηση και ταξινόμηση. Ανάλογα με το είδος της υπομονάδας (δυναμικό, στατικό) χρησιμοποιούμε διαφορετικές ροές πληροφορίας, αρχιτεκτονική και παραμέτρους. 3) Αυτόματη κατασκευή λεξικού υπομονάδων που υποδεικνύει τις υπομονάδες που συνθέτουν κάθε νόημα που περιέχεται στο λεξικό των νοημάτων.

3.2 Αυτόματη κατάτμηση σε Στατικά και Δυναμικά τμήματα

Μοντελοποίηση Ταχύτητας: Στόχος είναι η κατάτμηση και ταξινόμηση σε στατικά και δυναμικά τμήματα με βάση την ταχύτητα. Τα δυναμικά τμήματα θα αντιστοιχούν σε τμήματα στα οποία υπάρχει κίνηση των χεριών ενώ τα στατικά σε τμήματα που δεν παρατηρείται κίνηση. Η προσέγγιση αυτή είναι εμπνευσμένη από την γλωσσολογική μοντελοποίηση των Liddell και Johnson (PDTs) [41]. Η κατάτμηση βασίζεται στην εκπαίδευση ενός στατιστικού μοντέλου (statistical model-based approach). Αυτό δίνει την δυνατότητα να δουλεύουμε σε ένα πιθανοτικό πλαίσιο το οποίο καταστεί πολύ εύκολη την εφαρμογή του σε διαφορετικές βάσεις δεδομένων. Το χαρακτηριστικό που θα χρησιμοποιηθεί για αυτή την κατάτμηση και ταξινόμηση είναι η ταχύτητα του



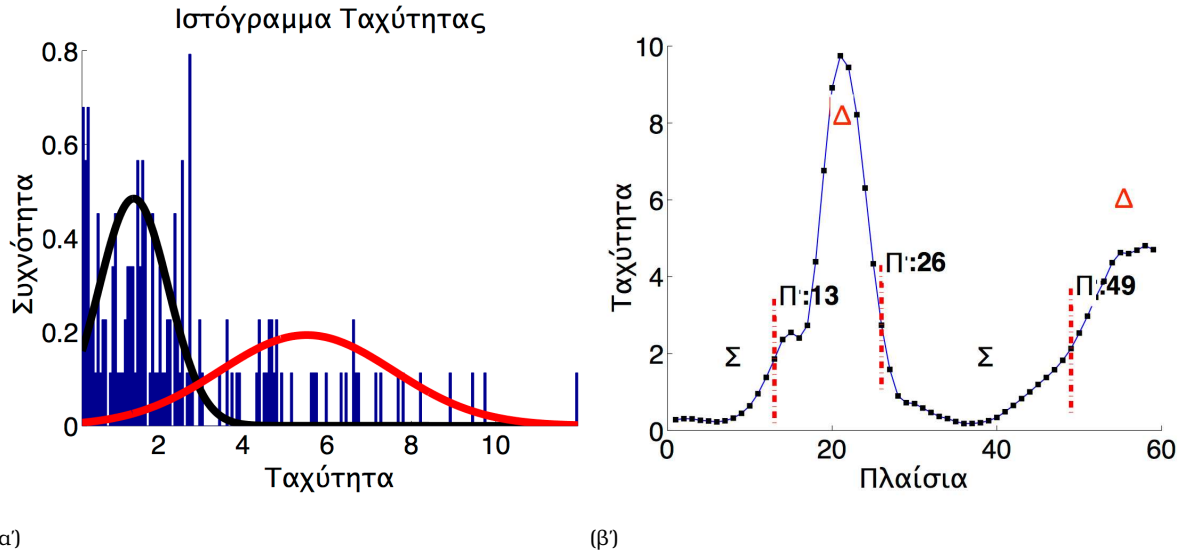
Σχήμα 3.3: Διάγραμμα το οποίο δείχνει την διαδικασία της αυτόματης κατάτμηση και ταξινόμηση σε στατικά και δυναμικά τμήματα. Στο πρώτο επίπεδο εκπαιδεύουμε ένα μοντέλο για κάθε νόημα το οποίο χρησιμοποιούμε για την κατάτμηση σε τμήματα με βάση την ταχύτητα (υψηλή -Y- και χαμηλή -X-). Στο δεύτερο επίπεδο η ίδια διαδικασία επαναλαμβάνεται χρησιμοποιώντας αυτή την φορά ένα μοντέλο το οποίο εκπαιδεύεται χρησιμοποιώντας μόνο τα στατικά τμήματα που έχουν εξαχθεί από το πρώτο επίπεδο. Στην συνέχεια επαναταξινομούμε όλα τα στατικά τμήματα με βάση την ταχύτητα (υψηλή -Y- και χαμηλή -X-).



Σχήμα 3.4: Το εργοδικό κρυφό μαρκοβιανό μοντέλο δύο καταστάσεων (2S-ERG) το οποίο χρησιμοποιήθηκε για την κατάτμηση και ταξινόμηση σε δυναμικά (Δ) και στατικά (Σ) τμήματα.

κυρίαρχου χεριού. Η επιτάχυνση μπορεί να χρησιμοποιηθεί ως πληροφορία δευτέρου επιπέδου. Η ταχύτητα του κυρίαρχου χεριού υπολογίζεται από το μέτρο του διανύσματος των παράγωγων των συντεταγμένων του κεντροειδούς του κυρίαρχου χεριού x, y . Το μοντέλο που χρησιμοποιούμε είναι ένα εργοδικό κρυφό μαρκοβιανό μοντέλο δύο καταστάσεων (2S-ERG βλέπε Σχ. 3.4). Αυτό επιτρέπει την είσοδο και έξοδο και από τις δύο καταστάσεις και την μετάβαση από οποιαδήποτε κατάσταση σε οποιαδήποτε άλλη.

Μοντελοποίηση εξαρτημένη στο νόημα: Εκπαιδεύουμε ένα εργοδικό Κρυφό Μαρκοβιανό Μοντέλο (βλ. Σχήμα.3.4) για κάθε νόημα (π.χ. ένα μοντέλο για όλες τις επαναλήψεις του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ της ΑΝΓ). Με τον τρόπο αυτό μοντελοποιούμε το στατιστικό προφίλ της ταχύτητας για κάθε νόημα. Κάθε κατάσταση του HMM μοντελοποιεί ένα διαφορετικό επίπεδο της ταχύτητας. Πιο συγκεκριμένα η πρώτη κατάσταση αντιστοιχεί σε χαμηλές (X) ενώ η δεύτερη σε υψηλές (Y) ταχύτητες. Αυτό γίνεται πιο κατανοητό εάν παρατηρήσουμε την κατανομή της ταχύτητας για όλες τις διαφορετικές επαναλήψεις για το νόημα ΠΑΡΑΔΕΧΟΜΑΙ μαζί με τις δύο γκαουσιανές κατανομές που αντιστοιχούν στις δύο καταστάσεις του μοντέλου για το νόημα ΠΑΡΑΔΕΧΟΜΑΙ (βλέπε Σχήμα. 3.5(α')). Μετά την εκπαίδευση των μοντέλων εφαρμόζουμε τον αλγόριθμο Viterbi σε κάθε επανάληψη για όλα τα νοήματα. Αυτό έχει ως αποτέλεσμα την εύρεση των σημείων κατάτμησης κάθε επανάληψης σύμφωνα με την πιο πιθανή ακολουθία καταστάσεων του μοντέλου. Ένα παράδειγμα κατάτμησης για μια επανάληψη του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ φαίνεται στο Σχήμα. 3.5(β').



Σχήμα 3.5: (α) Κατανομή της ταχύτητας (Ιστογράμμο) μαζί με τις δύο γκαουσιανες κατανομές που έχουν ταιριάξει οι οποίες αντιστοιχούν στις δύο διαφορετικές καταστάσεις του εργοδικού μοντέλου για το νόημα ΠΑΡΑΔΕΧΟΜΑΙ. (β) Σημεία κατάτμησης πάνω στο προφίλ της ταχύτητας για μια επανάληψη του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ παράλληλα με την αντιστοιχία των ετικετών Δυναμικό-/Στατικό τμήμα.

Στο Σχήμα. 3.6 φαίνονται τα πραγματικά πλαίσια των τμημάτων εφαρμόζοντας υποδειγματοληψία σε κάθε δυναμικό και στατικό τμήμα για την καλύτερη οπτικοποίηση. Όπως παρατηρούμε το νόημα ΠΑΡΑΔΕΧΟΜΑΙ αποτελείται από την εξής ακολουθία δυναμικών/στατικών υπομονάδων: – Σ Δ Σ Δ–. Το δεύτερο δυναμικό τμήμα αντιστοιχεί σε μια επένθεση κίνηση [79] λόγω του φαινομένου της συνάρθρωσης.

Μοντελοποίηση των Στατικών τμημάτων: Μετά την παραπάνω κατάτμηση και την ανάθεση της ετικέτας Δυναμικό/Στατικό τμήμα προχωράμε σε ένα βήμα επανεξέτασης της ταξινόμησης των στατικών τμημάτων η οποία απεικονίζεται σχηματικά στο δεύτερο επίπεδο του Σχήματος 3.3. Ακολουθούμε την ίδια διαδικασία όπως και παραπάνω με την διαφορά ότι 1) εφαρμόζουμε το βήμα της επανεξέτασης μόνο στα τμήματα που είχαν χαρακτηριστεί ως στατικά 2) εκπαιδεύουμε ένα εργοδικό HMM για όλα τα στατικά τμήματα ανεξαρτήτως του νοήματος από τα οποία έχουν προέλθει. Οι δύο καταστάσεις αυτού του μοντέλου αντιστοιχούν σε μια αναθεωρημένη διάκριση σε χαμηλή και υψηλή ταχύτητα (στατικό και δυναμικό τμήμα αντίστοιχα), των τμημάτων στα οποία είχε ανατεθεί η ετικέτα στατικό τμήμα στο πρώτο επίπεδο. Η επανεξέταση αυτή έχει ως αποτέλεσμα την ανίχνευση κινήσεων με πολύ χαμηλή ταχύτητα οι οποίες είχαν αντιστοιχηθεί ως στατικά τμήματα.

Αυτόματη vs. Ανθρώπινη κατάτμηση: Για την αξιολόγηση της προτεινόμενης μεθόδου αυτόματης κατάτμησης συγκρίνουμε το αποτέλεσμα με αυτό που έχει προκύψει από ανθρώπινη επισημείωση. Στο Σχήμα 3.7 παρουσιάζουμε τα αποτελέσματα κατάτμησης για μια επανάληψη του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ¹. Συγκρίνοντας τα αποτελέσματα παρατηρούμε ότι η αυτόματη κατάτμηση εφαρμόζοντας την προτεινόμενη μέθοδο είναι αρκετά κοντά στην κατάτμηση από τον άνθρωπο. Η προτεινόμενη μέθοδος για κατάτμηση έχει τα εξής πλεονεκτήματα: 1) Εύρεση των σημείων κατάτμησης και ταξινόμησης των τμημάτων σε δυναμικά/στατικά αφού έχουμε ενθυλακώσει στις καταστάσεις του μοντέλου κατάτμησης τα στατιστικά χαρακτηριστικά των δυναμικών και

¹Η ανθρώπινη επισημείωση έγινε από την Annelies Braffort και την ομάδα της στο CNRS-LIMSI.



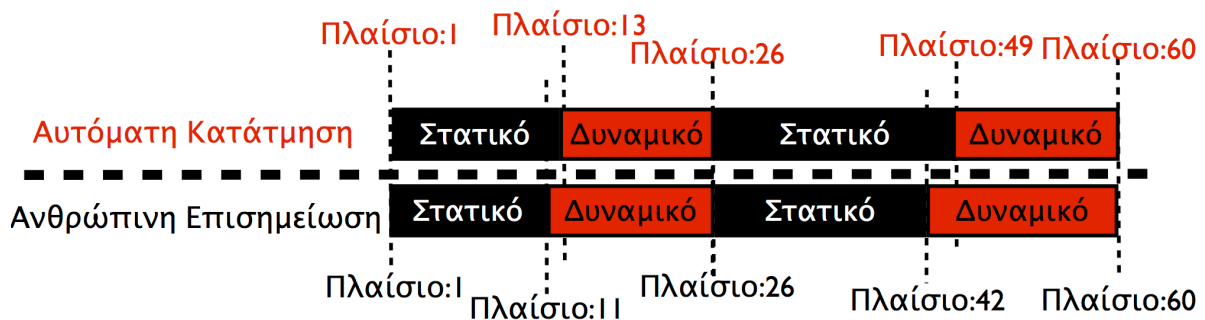
Σχήμα 3.6: Κατάτμηση σε δυναμικά και στατικά τμήματα για μια επανάληψη του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ της ΑΝΓ. Κάθε γραμμή αντιστοιχεί σε διαφορετικό τμήμα.

στατικών τμημάτων. 2) Δεν χρειάζεται να ορίσουμε καμία παράμετρο για κατοφλιωποίηση αφού το μοντέλο κατάτμησης εκπαιδεύεται αυτόματα ενθυλακώνοντας τα στατιστικά της βάσης δεδομένων προς κατάτμηση.

3.3 Ομαδοποίηση και Μοντελοποίηση των υπομονάδων

3.3.1 Κανονικοποίηση Χαρακτηριστικών

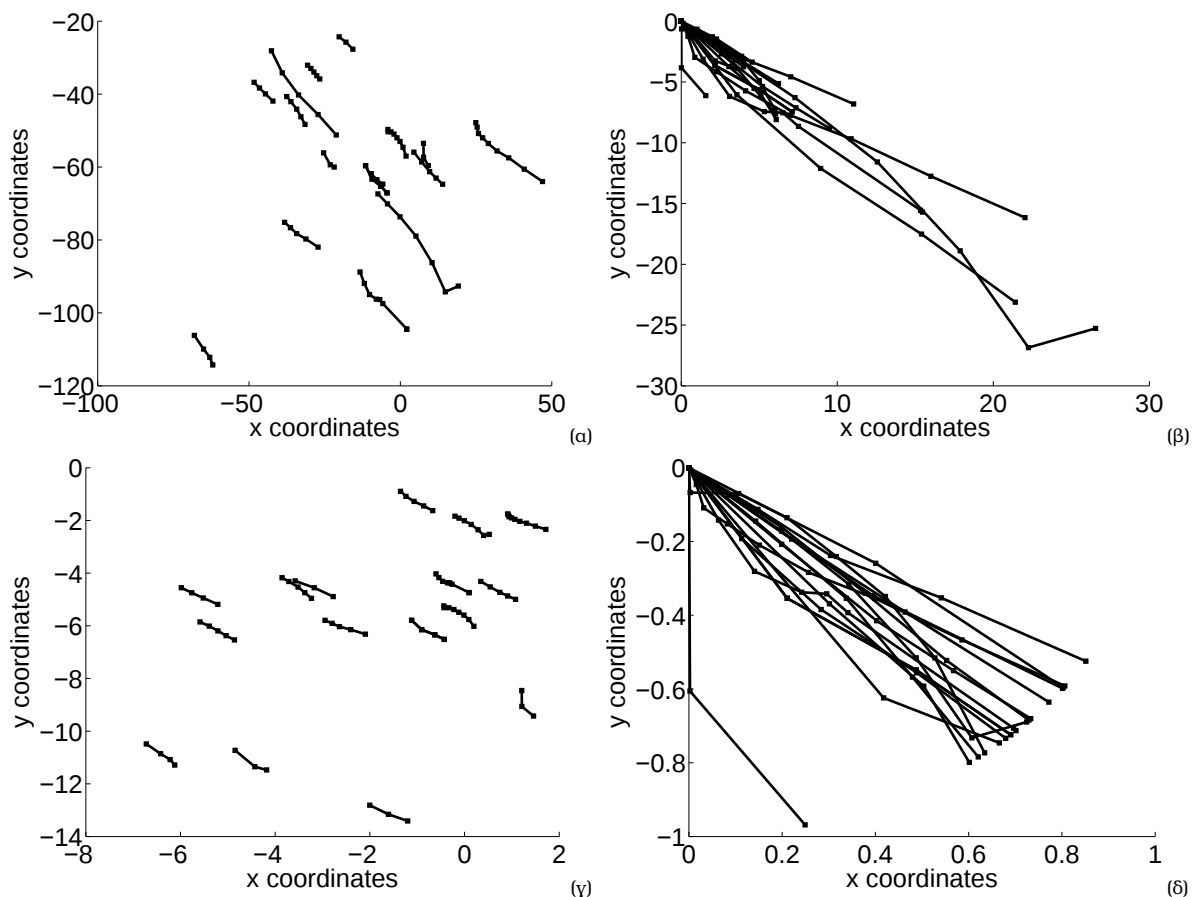
Αρχική Θέση: Στόχος μας είναι η μοντελοποίηση της δυναμικής των κινήσεων κατά την άρθρωση της νοηματικής γλώσσας. Το κύριο χαρακτηριστικό για την μοντελοποίηση των δυναμικών τμημάτων είναι η τροχιά (x, y δισδιάστατες συντεταγμένες). Σε κάθε κίνηση η ακολουθία θέσεων (τροχιά) εξαρτάται από την αρχική θέση στην οποία βρισκόταν το χέρι. Η μοντελοποίηση και



Σχήμα 3.7: Αυτόματη vs. Ανθρώπινη κατάτμηση και ταξινόμηση των τμημάτων για μια επανάληψη του νοήματος ΠΑΡΑΔΕΧΟΜΑΙ.

ομαδοποίηση των κινήσεων σύμφωνα με την τροχιά, θα οδηγούσε σε μοντέλα με αυξημένη μεταβλητότητα λόγω των πολλαπλών πιθανών αρχικών θέσεων για κάθε κίνηση. Για την αντιμετώπιση του προβλήματος της αυξημένης μεταβλητότητας κανονικοποιούμε τις τροχιές με βάση την αρχική τους θέση. Η κανονικοποίηση οδηγεί στην μοντελοποίηση των κινήσεων ανεξαρτήτως της θέσης τους. Στο Σχήμα 3.8(α) παρουσιάζονται οι τροχιές δυναμικών τμημάτων πάνω στο διδιάστατο νοηματικό χώρο. Στο Σχήμα 3.8(β) παρουσιάζονται οι ίδιες τροχιές μετά από κανονικοποίηση με βάση την αρχική θέση (Initial Position Normalized (Pn)).

Μέγεθος Κίνησης Κλίμακα -Scale (S):- Έναν επιπλέον παράγοντα μεταβλητότητας αποτελεί το μέγεθος της κίνησης. Έτσι προχωράμε σε μια ακόμα κανονικοποίηση των χαρακτηριστικών με βάση το μέγεθος της κίνησης (Scale Normalized (Sn)). Η κανονικοποίηση γίνεται διαιρώντας κάθε x, y με το μέγεθος της κίνησης (βλ. Σχήμα 3.8(γ)). Στο Σχήμα 3.8(δ) παρουσιάζουμε τις τροχιές των ίδιων δυναμικών τμημάτων εφαρμόζοντας κανονικοποίηση με βάση την αρχική θέση αλλά και το μέγεθος της κίνησης (Scale Position Normalized (SPn)). Στην συνέχεια θα δούμε ότι η χρήση αυτών των κανονικοποιημένων χαρακτηριστικών είναι περισσότερο αποτελεσματική για την ομαδοποίηση και εκπαίδευση των Κρυφών Μαρκοβιανών Μοντέλων καθώς περιέχουν πληροφορία για το είδος των κινήσεων ανεξαρτήτως της αρχικής θέσης και του μεγέθους της κίνησης.



Σχήμα 3.8: Οι τροχιές δυναμικών τμημάτων πάνω στο διδιάστατο χώρο νοηματοδότησης: (α) Χωρίς κανονικοποίηση (P) (β) Κανονικοποίηση με βάση την αρχική θέση (Pn) (γ) Κανονικοποίηση με βάση το μέγεθος της κίνησης (Sn) (δ) Κανονικοποίηση με βάση και την αρχική θέση και το μέγεθος της κίνησης (SPn).

3.3.2 Ομαδοποίηση των δυναμικών τμημάτων

Χρησιμοποιώντας το αποτέλεσμα της κατάτμησης όπως είδαμε στην ενότητα 3.2 προχωράμε στην αυτόματη ομαδοποίηση/συσταδοποίηση των δυναμικών τμημάτων. Για την επίτευξη αυτής της ομαδοποίησης χρειαζόμαστε ένα δείκτη ομοιότητας μεταξύ δύο ακολουθιών τιμών ενός χαρακτηριστικού. Ο αλγόριθμος Dynamic Time Warping (DTW) εφαρμόζεται για τον υπολογισμό ενός πίνακα ομοιότητας μεταξύ όλων των δυναμικών τμημάτων. Θεωρούμε δύο δυναμικά τμήματα $X = (X_1, X_2, \dots, X_{T_x})$ και $Y = (Y_1, Y_2, \dots, Y_{T_y})$ όπου T_x, T_y είναι ο αριθμός πλαισίων για κάθε τμήμα. Ορίζουμε το warping μονοπάτι $W = ((x_1, y_1), \dots, (x_N, y_N))$ όπου $1 \leq x_i \leq T_x$, $1 \leq y_i \leq T_y$, N είναι το μήκος του μονοπατιού και η σημειωγραφία για το ζευγάρι (x_i, y_i) σημαίνει ότι το πλαίσιο x_i του X τμήματος αντιστοιχίζεται στο πλαίσιο y_i του τμήματος Y . Το μετρικό $d(X_{x_i}, Y_{y_i})$ που χρησιμοποιούμε είναι η ευκλείδεια απόσταση. Στόχος του DTW είναι η εύρεση της μικρότερης απόστασης συσχετιζόμενη με το warping μονοπάτι:

$$D(X, Y) = \min_W \sum_{n=1}^N d(X_{x_i}, Y_{y_i}) \quad (3.1)$$

Στην συνέχεια ο πίνακας ομοιότητας μεταξύ όλων των δυναμικών τμημάτων προς ομαδοποίηση εισάγεται σε έναν αλγόριθμο agglomerative hierarchical clustering. Με αυτό τον τρόπο κατασκευάζουμε τις συστάδες από δυναμικά τμήματα λαμβάνοντας υπόψη την δυναμική φύση των τροχιών των δυναμικών τμημάτων. Κάθε μια από αυτές τις συστάδες αντιστοιχεί σε μια υπομονάδα η οποία στην συνέχεια θα μοντελοποιηθεί από ένα κρυφό μαρκοβιανό μοντέλο.

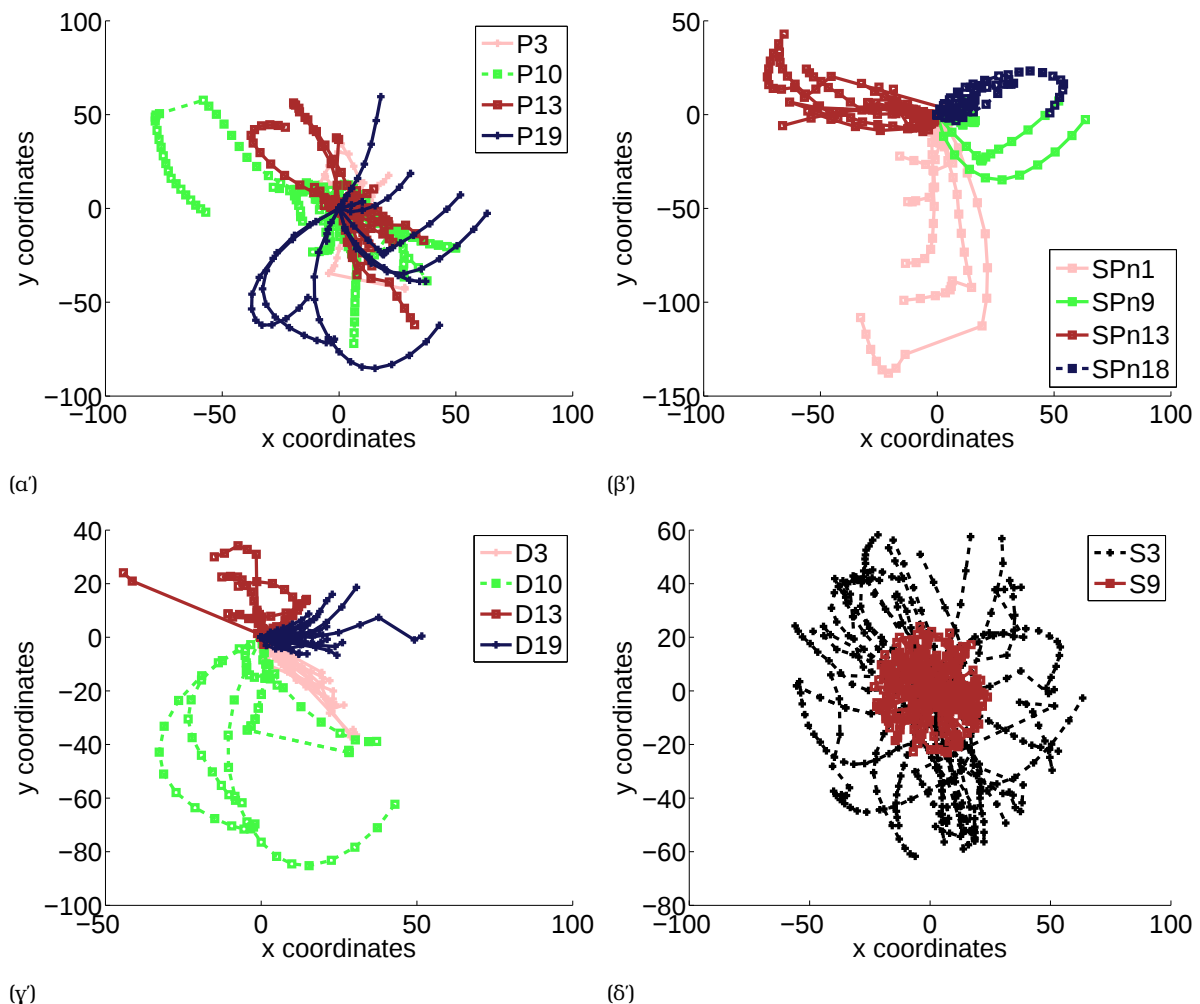
3.3.3 Δυναμικές υπομονάδες ανά χαρακτηριστικό

Στην συνέχεια διερευνούμε τα κατάλληλα χαρακτηριστικά για την συσταδοποίηση και μοντελοποίηση των δυναμικών τμημάτων. Το αποτέλεσμα της συσταδοποίησης προκαλεί διαίρεση του χώρου των χαρακτηριστικών. Κάθε συστάδα αποτελεί μια υπομονάδα.

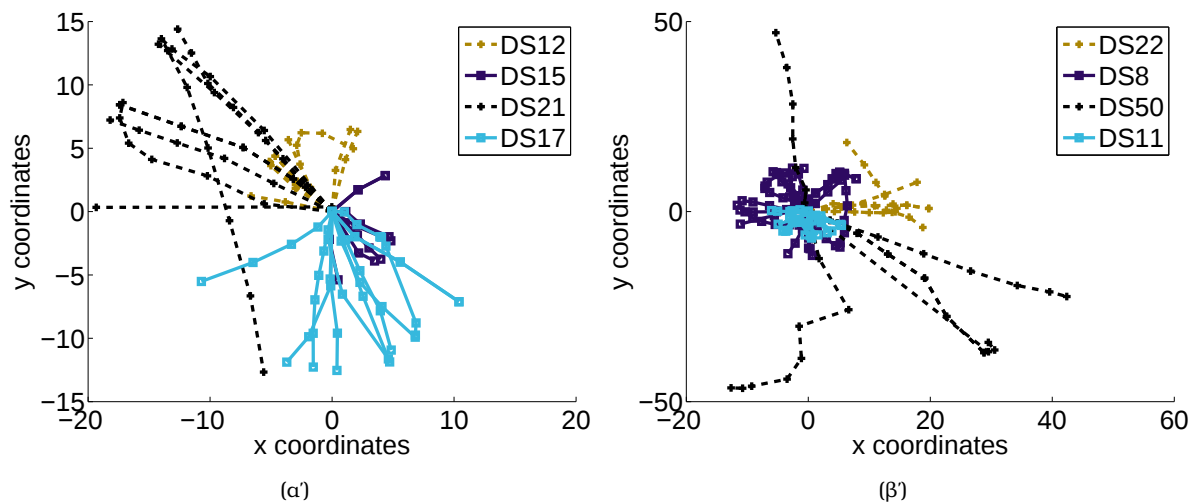
Τροχιά Κίνησης: Εφαρμόζοντας την κανονικοποίηση των τροχιών που είδαμε στην ενότητα 3.3.1 κάθε δυναμικό τμήμα αντιπροσωπεύεται από διάνυσμα χαρακτηριστικών SPn στο δισδιάστατο νοηματικό χώρο. Στο Σχήμα 3.9(β') βλέπουμε ενδεικτικές υπομονάδες που αυτές αντιστοιχούν σε συστάδες οι οποίες δημιουργήθηκαν εφαρμόζοντας ιεραρχική συσταδοποίηση και απεικονίζονται δισδιάστατο νοηματικό χώρο με διαφορετικά χρώματα. Για παράδειγμα η υπομονάδα 'SPn1' αντιστοιχεί σε καμπύλες κινήσεις με κατεύθυνση κάτω και προς τα αριστερά. Ένα παράδειγμα νοήματος στην ΑΝΓ όπου εμφανίζεται η υπομονάδα 'SPn1' είναι το νόημα ΤΕΛΟΣ (βλ. Σχήμα.3.1(γ)). Στο Σχήμα 3.9(α') βλέπουμε ενδεικτικά παραδείγματα από υπομονάδες που προέκυψαν εφαρμόζοντας συσταδοποίηση με χαρακτηριστικό την τροχιά των δυναμικών τμημάτων χωρίς κανονικοποίηση (P). Συγκρίνοντας τα δύο σχήματα 3.9(α') 3.9(β') βλέπουμε ότι υπάρχει πολύ μεγάλη μεταβλητότητα η οποία εισάγεται λόγω της ποικιλίας των αρχικών θέσεων και του μεγέθους των κινήσεων. Επίσης στις συστάδες που προέκυψαν χωρίς κανονικοποίηση είναι δύσκολο να αντιστοιχηθεί μια ενδεικτική κίνηση σε κάθε συστάδα. Σε αντίθεση στην περίπτωση των κανονικοποιημένων τροχιών κάθε συστάδα αντιστοιχεί σε κίνηση διαφορετικής κατεύθυνσης. Η παραπάνω παρατήρηση ήταν αναμενόμενη καθώς η τροχιά μια κίνησης περιλαμβάνει έμμεσα ισχυρή πληροφορία σχετικά με την κατεύθυνση της κίνησης.

Μέγεθος Κίνησης -κλίμακα- (S): Το μέγεθος της κίνησης, που χρησιμοποιήθηκε ως παράγοντας κανονικοποίησης στο σημείο αυτό θα αποτελέσει παράγοντα συσταδοποίησης των κινήσεων. Στο Σχήμα. 3.9(δ) έχουμε απεικονίσει με διαφορετικό χρώμα δύο υπομονάδες ('S9' και 'S3') χρησιμοποιώντας ως χαρακτηριστικό το μέγεθος τις κίνησης (S). Όπως παρατηρούμε η υπομονάδα 'S9' αντιστοιχεί σε κινήσεις μικρότερου μεγέθους από την υπομονάδα 'S3'. Ένα παράδειγμα όπου εμφανίζονται οι υπομονάδες 'S9' και 'S3' είναι τα νοήματα ΠΑΡΑΔΕΧΟΜΑΙ Σχήμα 3.1(α) και ΤΕΛΟΣ Σχήμα 3.1(γ) αντίστοιχα.

Κατεύθυνση (Direction (D)): Οι υπομονάδες που κατασκευάστηκαν με χαρακτηριστικό διάνυσμα την κατεύθυνση είναι αντίστοιχες με αυτές που παράχθηκαν με χαρακτηριστικό διάνυσμα το SPn. Κάθε υπομονάδα περιέχει κινήσεις με παρόμοια κατεύθυνση. Στο Σχήμα 3.9(γ') βλέπουμε ενδεικτικά τέσσερις υπομονάδες (με διαφορετικό χρώμα) που αντιστοιχούν σε κινήσεις με διαφορετική κατεύθυνση. Παραδείγματος χάριν η υπομονάδα 'D10' μοντελοποιεί καμπύλες κινήσεις με κατεύθυνση προς τα κάτω και δεξιά. Ένα παράδειγμα όπου η υπομονάδα 'D10' εμφανίζεται είναι το νόημα ΕΔΩ Σχήμα3.1(β).



Σχήμα 3.9: Οι τροχιές διαφορετικών δυναμικών τμημάτων πάνω στο διδιάστατο χώρο νοηματισμού μετά από κανονικοποίηση με βάση και την αρχική θέση. Το χρώμα των τροχιών ομαδοποιεί τα δυναμικά τμήματα ανάλογα με την υπομονάδα/συστάδα (subunit/cluster) που αντιστοιχούν. (α) Τροχιές των δυναμικών τμημάτων που ομαδοποιήθηκαν στις αντίστοιχες υπομονάδες/συστάδες χρησιμοποιώντας ως χαρακτηριστικό τις x, y συντεταγμένες χωρίς κανονικοποίηση (P). (β) Τροχιές των δυναμικών τμημάτων που ομαδοποιήθηκαν στις αντίστοιχες υπομονάδες/συστάδες χρησιμοποιώντας ως χαρακτηριστικό τις κανονικοποιημένες x, y συντεταγμένες με βάση την αρχική θέση και το μέγεθος της κίνησης (SPn). (γ) Τροχιές των δυναμικών τμημάτων που ομαδοποιήθηκαν στις αντίστοιχες υπομονάδες/συστάδες χρησιμοποιώντας ως χαρακτηριστικό διάνυσμα την κατεύθυνση των τροχιών (D). (δ) Τροχιές των δυναμικών τμημάτων που ομαδοποιήθηκαν σε δύο υπομονάδες/συστάδες χρησιμοποιώντας ως χαρακτηριστικό διάνυσμα το μέγεθος της κίνησης (S).



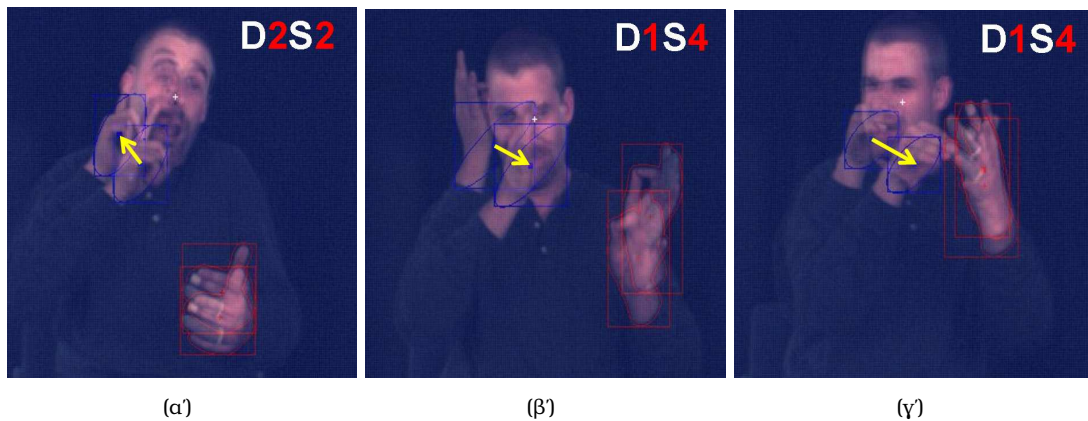
Σχήμα 3.10: Οι τροχιές των δυναμικών τμημάτων που αντιστοιχούν σε διαφορετικές υπομονάδες με βάση το χρώμα. Οι υπομονάδες βασίζονται ταυτόχρονα και στην κατεύθυνση αλλά και στο μέγεθος της κίνησης.

3.3.4 Δυναμικές υπομονάδες πολλαπλών ροών πληροφορίας

Στην προηγούμενη ενότητα παρουσιάσαμε την κατασκευή/εκπαίδευση δυναμικών υπομονάδων από τα δυναμικά τμήματα χρησιμοποιώντας μια ροή πληροφορίας κάθε φορά. Σε αυτή την ενότητα θα παρουσιάσουμε την κατασκευή υπομονάδων βασιζόμενοι σε πολλαπλές ροές πληροφορίας. Για την εκμετάλλευση πολλαπλών ροών πληροφορίας συνενώνουμε πολλαπλά διανύσματα χαρακτηριστικών. Συνενώνοντας π.χ. τα χαρακτηριστικά διανύσματα της κατεύθυνσης (D) και του μεγέθους της κίνησης (S) κατασκευάζουμε υπομονάδες οι οποίες βασίζονται ταυτόχρονα τόσο στην κατεύθυνση όσο στο μέγεθος της κίνησης. Ενδεικτικές υπομονάδες μπορούμε να δούμε στο Σχήμα 3.10 όπου απεικονίζουμε τις τροχιές των δυναμικών τμημάτων που περιλαμβάνουν, στον πραγματικό νοηματικό χώρο. Κάθε υπομονάδα αντιπροσωπεύει κινήσεις συγκεκριμένης τόσο κατεύθυνσης όσο και μεγέθους σε αντίθεση με τις single-cue υπομονάδες της προηγούμενης ενότητας. Πρέπει να αναφέρουμε ότι καθώς η συσταδοποίηση γίνεται στο κοινό χώρο χαρακτηριστικών της κατεύθυνσης και του μεγέθους της κίνησης κάποιες από τις υπομονάδες μπορεί να είναι αρκετά σύνθετες, γεγονός το οποίο εξαρτάται από την από κοινού κατανομή των πολλαπλών ροών χαρακτηριστικών. Μερικά παραδείγματα μπορούμε να δούμε το Σχήμα 3.10(α,β). Στο Σχήμα 3.11 βλέπουμε ενδεικτικά παραδείγματα υπομονάδων στο επίπεδο των πραγματικών πλαισίων υπερθέτοντας το αρχικό πλαίσιο στο τελικό και τοποθετώντας ένα βέλος το οποίο υποδεικνύει το είδος της κίνησης. Πιο συγκεκριμένα στο Σχήμα 3.11 απεικονίζουμε τις κινήσεις από τρία νοήματα τις ΑΝΓ: ΚΟΥΦΟΣ, ΑΠΟΦΑΣΙΖΩ, ΜΕΣΑ και τις αντίστοιχες υπομονάδες κατεύθυνσης-κλίμακας. Το Σχήμα 3.11(α) απεικονίζει την υπομονάδα κατεύθυνσης-κλίμακας D2S2 που αντιστοιχεί σε ευθεία κίνηση με κατεύθυνση D2 (πάνω-αριστερά) και κλίμακα S2 (μικρή). Τα Σχήματα 3.11(β,γ) απεικονίζουν την υπομονάδα κατεύθυνσης-κλίμακας D1S4 που αντιστοιχεί σε ευθεία κίνηση με κατεύθυνση D1 (κάτω-δεξιά) και κλίμακα S4 (μεσαία).

3.3.5 Ομαδοποίηση στατικών υπομονάδων

Σε αυτή την ενότητα θα παρουσιάσουμε την μοντελοποίηση των στατικών υπομονάδων ομαδοποιώντας τα στατικά τμήματα βασιζόμενοι στην αυτόματη κατάτμηση και ταξινόμηση σε δυναμικά και στατικά τμήματα που περιγράψαμε στην ενότητα 3.2. Για την μοντελοποίηση τους θα χρη-



Σχήμα 3.11: Παραδείγματα υπομονάδων κατεύθυνσης-κλίμακας στο επίπεδο των πραγματικών πλαισίων υπερθέτοντας το αρχικό στο τελικό πλαίσιο και τοποθετώντας ένα βέλος το οποίο υποδεικνύει το είδος της κίνησης. Οι τρεις δυναμικές υπομονάδες αντιστοιχούν σε κινήσεις που εμφανίζονται στα νοήματα ΚΟΥΦΟΣ, ΑΠΟΦΑΣΙΖΩ, ΜΕΣΑ αντίστοιχα.

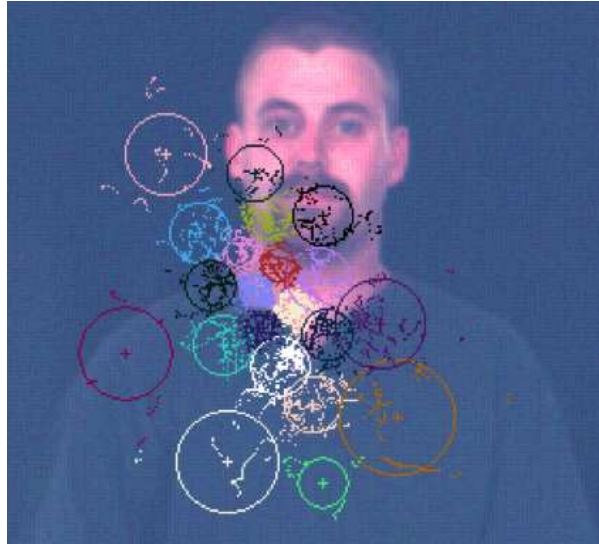
σιμοποιήσουμε *μόνο* τα πλαίσια που ανήκουν σε στατικά τμήματα. Για την εύρεση των περιοχών άρθρωσης και την αντιστοίχισή τους με τις στατικές υπομονάδες εφαρμόζουμε τον αλγόριθμο K-means με διάνυσμα χαρακτηριστικών την μη-κανονικοποιημένη θέση του κεντροειδούς του χεριού (P). Το σχήμα 3.12 απεικονίζει πάνω στο νοηματικό χώρο τις διαφορετικές στατικές υπομονάδες/συστάδες και το αντίστοιχο κεντροειδές τους. Κάθε συστάδα αντιστοιχεί σε μια στατική υπομονάδα η οποία μοντελοποιεί μια περιοχή άρθρωσης.

3.4 Φωνητικό Λεξικό με Δυναμικές/Στατικές Υπομονάδες

3.4.1 Συνδυασμός των Δυναμικών και Στατικών Υπομονάδων

Δεδομένης της μη-ύπαρξης προκαθορισμένου φωνητικού λεξικού, συνηθισμένο φαινόμενο σε σώματα δεδομένων νοηματικής γλώσσας, κατασκευάζουμε αυτόματα ένα δεδομενοκεντρικό φωνητικό λεξικό εφαρμόζοντας την προτεινόμενη μέθοδο DSCM. Μετά την κατάτμηση σε δυναμικά και στατικά τμήματα (εν. 3.2) και μοντελοποίηση των δυναμικών και στατικών υπομονάδων (εν. 3.3.5) συνδυάζουμε τις υπομονάδες για την κατασκευή ενός φωνητικού λεξικού. Για κάθε διαφορετική προφορά ενός νοήματος του λεξιλογίου δημιουργούμε μια νέα εγγραφή στο λεξικό συνενώνοντας τις υπομονάδες από τις οποίες αποτελείται κάθε νόημα. Έτσι παράγεται ένα λεξικό το οποίο αποτελείται από μια εγγραφή για κάθε διαφορετική προφορά ενός νοήματος όπως εμφανίζεται στο σώμα δεδομένων. Η αντιστοίχιση των νοημάτων με μια ακολουθία από υπομονάδες θα χρησιμοποιηθεί κατά την διάρκεια της αναγνώρισης. Κάθε υπομονάδα συμβολίζεται με ένα κωδικό ο οποίος αποτελείται από 3 χαρακτηριστικά: 1) Ο χαρακτηρισμός Δυναμική -Dynamic Movement (M)- ή Στατική -Static Movement (H)- υπομονάδα. 2) Το διάνυσμα χαρακτηριστικών που χρησιμοποιήθηκε στην μοντελοποίηση. Τα χαρακτηριστικά αυτά είναι η κατεύθυνση (D), η κανονικοποιημένη θέση του κεντροειδούς του χεριού (SPn), το μέγεθος της κίνησης -κλίμακα (S)- και η μη-κανονικοποιημένη θέση του κεντροειδούς του χεριού (P). 3) Ο δείκτης που αντιστοιχεί στην συστάδα μετά την εφαρμογή του αλγόριθμου ομαδοποίησης. Η ετικέτα μιας υπομονάδας κατασκευάζεται συνενώνοντας τα 3 παραπάνω σύμβολα. Για παράδειγμα η υπομονάδα “MD12” αντιστοιχεί σε μια δυναμική υπομονάδα (M) όπου χρησιμοποιήθηκε η κατεύθυνση (D) ως διάνυσμα χαρακτηριστικών και η συστάδα που μοντελοποιεί είναι η δωδέκατη.

Συνδυάζοντας τις δυναμικές και στατικές υπομονάδες δημιουργούμε το φωνητικό λεξικό. Ένα



Σχήμα 3.12: Χρήση του αλγορίθμου K-means για την ομαδοποίηση των στατικών τμημάτων και κβάντιση του δισδιάστατου νοηματικού χώρου. Κάθε στατική υπομονάδα απεικονίζεται με την χρήση διαφορετικού χρώματος.

παράδειγμα διαφορετικών λεξικών φαίνεται στον Πίνακα 3.1. Στα λεξικά αυτά χρησιμοποιήθηκε η μη-κανονικοποιημένη θέση (P) ως διάνυσμα χαρακτηριστικών για τις στατικές υπομονάδες και η κατεύθυνση (D) ή η κανονικοποιημένη θέση (SPn) για τις δυναμικές υπομονάδες. Κάθε διαφορετικός συνδυασμός των διανυσμάτων χαρακτηριστικών που χρησιμοποιήθηκαν απεικονίζεται σε διαφορετική γραμμή του πίνακα. Κάθε γραμμή του πίνακα αποτελείται από 1) Το νόημα όπως ΒΛΕΠΩ 2) Ένα δείκτη που αντιπροσωπεύει μια συγκεκριμένη επανάληψη του νοήματος και 3) Από την ακολουθία των στατικών/δυναμικών υπομονάδων που αντιστοιχεί στην συγκεκριμένη επανάληψη του νοήματος. Παραδείγματος χάριν η δεύτερη επανάληψη του νοήματος ΒΛΕΠΩ στην ΑΝΓ αντιστοιχεί στην ακολουθία 'MD12 HP16' ή 'MSP6 HP16' υπομονάδων ανάλογα με το διάνυσμα χαρακτηριστικών που χρησιμοποιήθηκε.

3.4.2 Πολλαπλές Προφορές Νοημάτων

Κατά την διάρκεια συνεχούς νοηματικού λόγου παρατηρείται μεγάλη μεταβλητότητα στην άρθρωση νοημάτων, με αποτέλεσμα πολλαπλές επαναλήψεις του ίδιου νοήματος να διαφέρουν αρκετά μεταξύ τους. Τις διαφορετικές εκτελέσεις του ίδιου νοήματος τις ονομάζουμε προφορές ενός νοήματος. Η εμφάνιση πολλαπλών προφορών ενός νοήματος σε ένα λεξικό μπορεί να οφείλεται είτε λόγω πραγματικά διαφορετικής εκτελέσεως είτε λόγω μικρής μεταβολής της εκτέλεσης εξαιτίας του φαινομένου της συνάρθρωσης στον συνεχή νοηματικό λόγο. Ενδεικτικά παραδείγματα των πολλαπλών προφορών ενός νοήματος βλέπουμε στον Πίνακα 3.1. Το νόημα ΒΛΕΠΩ αντιστοιχίζεται σε δύο διαφορετικές ακολουθίας υπομονάδων (2 πρώτες γραμμές). Η πρώτη ακολουθία υπομονάδων 'MSPn6 HP16' (πρώτη γραμμή του Πίνακα 3.1) αποτελείται από μια δυναμική υπομονάδα ακολουθούμενη από μια στατική. Ένα πραγματικό παράδειγμα εκτέλεσης της συγκεκριμένης προφοράς φαίνεται στην πρώτη γραμμή του Πίνακα 3.2. Σε αντίθεση η δεύτερη ακολουθία 'HP16 MSPn10' (δεύτερη γραμμή του Πίνακα 3.1) αποτελείται από μια στατική υπομονάδα ακολουθούμενη από μια δυναμική. Ένα πραγματικό παράδειγμα εκτέλεσης της συγκεκριμένης προφοράς φαίνεται στην πρώτη γραμμή του Πίνακα 3.3. Οι δύο παραπάνω ακολουθίες έχουν μια κοινή υπομονάδα, την 'HP16', ωστόσο οι δυναμικές τους υπομονάδες είναι διαφορετικές ('MSPn6' και 'MSPn10'). Η διαφορά στις δύο αυτές προφορές οφείλεται στο φαινόμενο της συνάρθρωσης αφού

Πίνακας 3.1: Παράδειγμα δύο φωνητικών λεξικών[†].

Νόημα :Επανάληψη Πλαίσιο (A/T)	D-P SPn-P		Παράδειγμα
ΒΛΕΠΩ :2 A T	MD12 MSPn6	HP16	Πίνακας 3.2
	20131	20134	
	20133	20141	
ΒΛΕΠΩ :5 A T	HP16	MD20 MSPn10	Πίνακας 3.3
	7007	7010	
	7009	7017	
ΕΝΑ :2 A T	MD12 MSPn6	HP16	Πίνακας 3.2
	8121	8124	
	8123	8127	
ΕΝΑ :24 A T	MD4 MSPn42	HP21	Πίνακας 3.3
	13219	13227	
	13226	13231	
ΕΔΩ :3 A T	MD10 MSPn8		Σχήμα 3.1(β)
	6983		
	6995		
ΤΕΛΟΣ :2 A T	MD24 MSPn1		Σχήμα 3.1(γ)
	6691		
	6707		

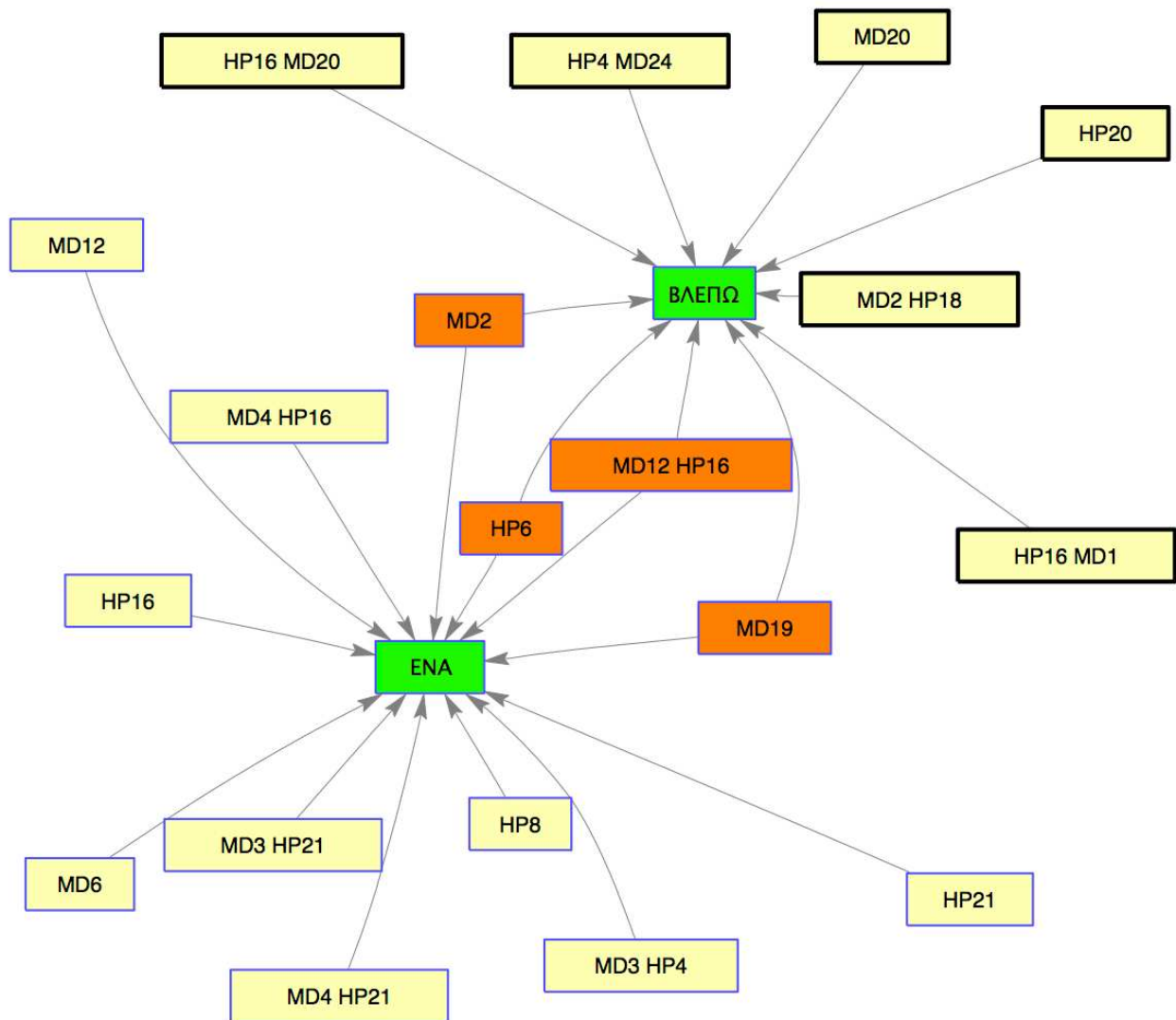
[†] Για κάθε νόημα απεικονίζουμε το όνομα του νοήματος, το αριθμό της επανάληψης, την ακολουθία των δυναμικών και στατικών υπομονάδων και επίσης τον αριθμό του πρώτου και το τελευταίου πλαισίου κάθε τμήματος.

και οι δύο δυναμικές υπομονάδες αντιστοιχούν σε επένθεση κίνηση η οποία εξαρτάται μόνο από το συγκείμενο της κάθε επανάληψης.

3.4.3 Αντιστοίχιση μιας ακολουθίας υπομονάδων σε πολλαπλά νοήματα

Λόγο της αυτόματης και δεδομενοκεντρικής κατασκευής του λεξικό είναι συνηθισμένο μια ακολουθία υπομονάδων να αντιστοιχεί σε πολλαπλά νοήματα. Μερικοί από τους πιο σημαντικούς λόγους για την εμφάνιση του παραπάνω φαινομένου είναι οι εξής: 1) Η μη επαρκής μοντελοποίηση των νοημάτων σε σχέση με τα διανύσματα χαρακτηριστικών που έχουν επιλεχθεί. Για παράδειγμα υπάρχουν περιπτώσεις όπου δύο νοήματα αποτελούνται από τις ίδιες κινήσεις-θέσεις των χεριών αλλά διαφέρουν στο τύπο της χειρομορφής, χαρακτηριστικό που δεν λαμβάνουμε υπόψη στο παρόν κεφάλαιο. Ένα παράδειγμα είναι τα νοήματα ΒΛΕΠΩ και ΕΝΑ ² τα οποία αποτελούνται από τις ίδιες κινήσεις-θέσεις των χεριών αλλά διαφέρουν στον τύπο της χειρομορ-

²Το νόημα ΕΝΑ σε αυτή τη συγκεκριμένη εκτέλεση περιλαμβάνει και μια δυναμική υπομονάδα η οποία αντιστοιχεί σε επένθεση κίνηση λόγω του φαινομένου της συνάρθρωσης.



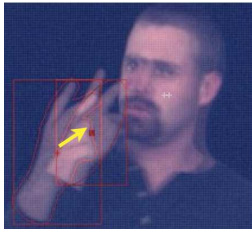
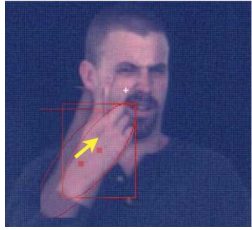
Σχήμα 3.13: Ακολουθίες υπομονάδων αντιστοιχίζονται σε δύο επίλεγμένα νοήματα τις ANTs (BΛEΠΩ και ENA).

φής (Πίνακας 3.2). 2) Ένας άλλος παράγοντας που επηρεάζει είναι ο αριθμός των συστάδων που έχει επιλεχθεί κατά την διάρκεια της εφαρμογής του αλγορίθμου ομαδοποίησης. Για παράδειγμα εάν χρησιμοποιήσουμε πολύ μικρό αριθμό συστάδων για να ομαδοποιήσουμε τις διαφορετικές κινήσεις τότε πολλαπλές κινήσεις θα αντιστοιχούνται στην ίδια υπομονάδα. 3) Άλλοι λόγοι σχετίζονται με την εξαγωγή χαρακτηριστικών όπως λάθη κατά την διάρκεια της παρακολούθησης των αρθρωτών, η έλλειψη τρισδιάστατων μετρήσεων συνηθισμένο σε αρκετά σώματα δεδομένων κ.α..

Ένα παράδειγμα του προαναφερθέντος προβλήματος για δύο νοήματα τις ANTs απεικονίζεται στον γράφο του Σχήματος 3.13³. Στο γράφο αυτό απεικονίζουμε δύο νοήματα (με πράσινο χρώμα) και τις ακολουθίες των υπομονάδων οι οποίες αντιστοιχίζονται σε αυτά. Ανάμεσα σε αυτές τις ακολουθίες υπάρχει ένα υποσύνολο το οποίο αντιστοιχίζεται και στα δύο νοήματα (πορτοκαλί χρώμα). Στον Πίνακα 3.2 απεικονίζουμε μια εκτέλεση για το κάθε νόημα BΛEΠΩ και ENA. Όπως παρατηρούμε και οι δύο εκτελέσεις αντιστοιχούν στην ίδια κοινή ακολουθία 'MD12 HP16'. Η ακολουθία αυτή αποτελεί την πιο συχνή προφορά και για τα δύο νοήματα στο συγκεκριμένο

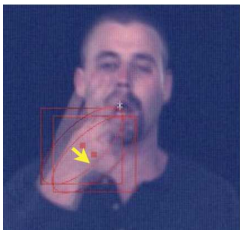
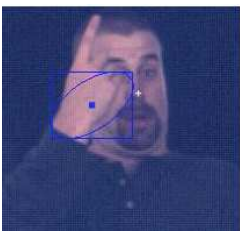
³Ο απεικονιζόμενος γράφος αποτελεί ένα υποσύνολο του συνολικού γράφου για λόγους οπτικοποίησης.

Πίνακας 3.2: Μια εκτέλεση των νοημάτων (ΕΝΑ και ΒΛΕΠΩ της ΑΝΓ) που αντιστοιχούν στην ακολουθία υπομονάδων ‘MD12 HP16’

Νόημα	Ακολουθία Υπομονάδων	
	MD12	HP16
ΒΛΕΠΩ		
ΕΝΑ		

σώμα δεδομένων. Παρόλα αυτά υπάρχει και ένα υποσύνολο ακολουθιών το οποίο είναι μοναδικό για κάθε νόημα. Στον Πίνακα 3.3 μπορούμε να δούμε μια εκτέλεση για κάθε νόημα η οποία αντιστοιχεί σε διαφορετική ακολουθία υπομονάδων.

Πίνακας 3.3: Μια εκτέλεση των νοημάτων (ΕΝΑ και ΒΛΕΠΩ της ΑΝΓ) που αντιστοιχούν σε διαφορετικές ακολουθίες υπομονάδων.

Νόημα	Ακολουθία Υπομονάδων	
	HP16	MD20
ΒΛΕΠΩ		
ΕΝΑ		

3.4.4 Κοινές Προφορές και Ομοιότητα Μεταξύ Νοημάτων

Για την ποσοτικοποίηση του φαινομένου της αντιστοιχίας μιας ακολουθίας σε πολλαπλά νοήματα κατασκευάζουμε μια νέα απόσταση η οποία ποσοτικοποιεί την ομοιότητα δύο νοημάτων σε σχέση με τον αριθμό των κοινών τους προφορών. Αρχικά μετράμε για το νόημα i τον αριθμό των εκτελέσεων $R(i, j)$ που αντιστοιχίζονται σε κάθε ακολουθία υπομονάδων j για ένα σύνολο από $i = 1, 2, \dots, N_G$ νοήματα και από $j = 1, 2, \dots, N_S$ ακολουθίες υπομονάδων. Στην συνέχεια κανονικοποιούμε τον πίνακα $R(i, j)$, $R_n(i, j) = R(i, j)/N_i$ με τον αριθμό των συνολικών εκτελέσεων για κάθε νόημα που έχουμε στο σώμα δεδομένων N_i . Τέλος ορίζουμε την απόσταση $d_s(m, n)$ μεταξύ δύο νοημάτων m, n ως:

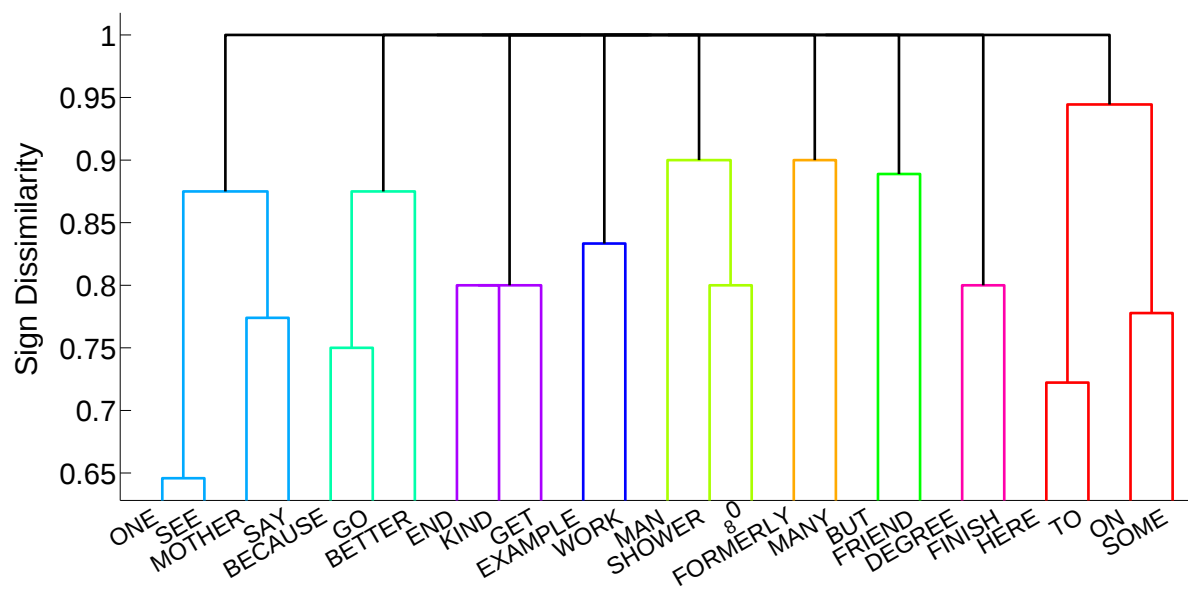
$$d_s(m, n) = 1 - \sum_{j=1}^{N_S} \min(R_n(m, j), R_n(n, j)) \quad (3.2)$$

Όταν απόσταση d_s μεταξύ δύο νοημάτων ισούται με μηδέν, αυτό υποδεικνύει ότι και τα δύο νοήματα αντιστοιχίζονται στις ίδιες ακριβώς ακολουθίες υπομονάδων και με την ίδια κατανομή στις εκτελέσεις. Σε αντίθετη περίπτωση όταν η απόσταση d_s ισούται με ένα τότε τα δύο νοήματα δεν έχουν καμία κοινή ακολουθία υπομονάδων μεταξύ τους. Στην συνέχεια ορίζουμε ένα νέο συντελεστή για την μέτρηση της διακριτότητας των νοημάτων μέσα σε ένα φωνητικό λεξικό. Αρχικά υπολογίζουμε την απόσταση d_s για όλα τα ζευγάρια των νοημάτων που έχουμε στο λεξικό μας. Στην συνέχεια ομαδοποιούμε τα νοήματα με ένα ιεραρχικό αλγόριθμο και κατασκευάζουμε το αντίστοιχο δένδρογραμμα. Μετά υπολογίζουμε το διάνυσμα $Z(i)$ το οποίο περιέχει τις αποστάσεις μεταξύ δύο νοημάτων ή ένα σύνολο από νοήματα που συγχωνεύτηκαν κατά την διάρκεια της ιεραρχικής ομαδοποίησης (linkage distances). Πιο συγκεκριμένα, το διάνυσμα Z περιέχει τις αποστάσεις για κάθε κόμβο i του δένδρογραμματος. Τέλος ορίζουμε τον συντελεστή διάκρισης νοημάτων (F_d) σε ένα λεξικό ως εξής:

$$F_d = \frac{1}{N_G - 1} \sum_{i=1}^{N_G-1} Z(i). \quad (3.3)$$

Στην ακραία περίπτωση όπου ο συντελεστής διάκρισης F_d είναι ίσος με ένα τότε κάθε νόημα που περιέχεται στο λεξικό είναι τελείως ανεξάρτητο από όλα τα άλλα νοήματα δηλαδή κάθε ακολουθία υπομονάδων αντιστοιχεί σε ένα μοναδικό νόημα. Στην άλλη ακραία περίπτωση όπου ο συντελεστής διάκρισης F_d είναι ίσος με μηδέν τότε όλα τα νοήματα αντιστοιχούν στις ίδιες ακολουθίες υπομονάδων δηλαδή δεν μπορούμε να διαχωρίσουμε κανένα νόημα. Εκτός από τον συντελεστή διάκρισης F_d παρουσιάζουμε άλλο ένα μέγεθος για την ποσοτικοποίηση της διακριτότητας ενός λεξικού το οποίο το ονομάζουμε *Ενεργά Νοήματα (Effective Signs)* και μετράει τον αριθμό των νοημάτων που είναι τελείως ανεξάρτητα από όλα τα άλλα. Ο αριθμός των Ενεργών Νοημάτων υπολογίζεται στο υψηλότερο επίπεδο του δένδρογραμματος μετρώντας τον αριθμό των κόμβων όπου η απόσταση τους είναι ένα. Ένα παράδειγμα δένδρογραμματος ⁴ απεικονίζεται στο Σχήμα 3.14. Στο κάτω επίπεδο του δένδρογραμματος παρατηρούμε το σύνολο των νοημάτων. Στο υψηλότερο επίπεδο παρατηρούμε τις συστάδες νοημάτων που είναι τελείως ανεξάρτητες από όλες τις άλλες και αντιστοιχούν στα Ενεργά Νοήματα.

⁴Χρησιμοποιήθηκαν 26 διαφορετικά νοήματα για λόγους οπτικοποίησης.



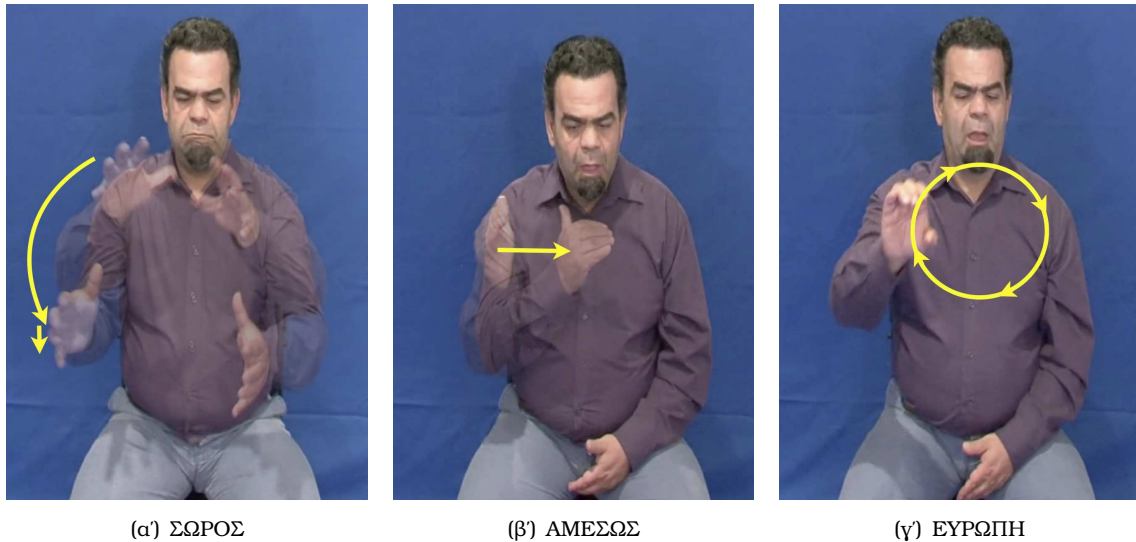
Σχήμα 3.14: Δενδρόγραμμα με βάση την διακριτότητα των νοημάτων (d_s).

Κεφάλαιο 4

Μοντελοποίηση της Νοηματικής Γλώσσας με φωνητικές υπομονάδες

Βασιζόμενοι στην πολυετή εμπειρία από τα συστήματα αναγνώρισης φωνής η επισημείωση σε φωνητικό επίπεδο αποτελεί κρίσιμο χαρακτηριστικό για την απόδοση των συστημάτων αναγνώρισης της ΝΓ. Ωστόσο η έννοια του φωνήματος στην ΝΓ δεν είναι δεδομένη και ούτε πλήρως καθορισμένη όπως στην ομιλούμενη γλώσσα παρότι έχουν προταθεί αρκετά γλωσσολογικά/φωνητικά μοντέλα για την ΝΓ τα τελευταία χρόνια. Επιπλέον η επισημείωση σε φωνητικό επίπεδο αποτελεί μια αρκετά χρονοβόρα διαδικασία. Τα παραπάνω έχουν ως αποτέλεσμα την κυριαρχία δεδομενοκεντρικών μεθόδων για την μοντελοποίηση της ΝΓ σε επίπεδο υπομονάδας [6, 28, 10, 33, 23, 90, 7]. Οι δεδομενοκεντρικές υπομονάδες έχουν το πλεονέκτημα ότι για την κατασκευή/εκπαίδευση τους δεν χρειάζονται επισημειώσεις σε φωνητικό επίπεδο. Χρησιμοποιούν αλγορίθμους αυτόματης ομαδοποίησης για την διαίρεση του χώρου των χαρακτηριστικών στις βασικές δεδομενοκεντρικές υπομονάδες. Το βασικό τους μειονέκτημα είναι ότι οι δεδομενοκεντρικές υπομονάδες που κατασκευάζονται αντιστοιχούν σε αυθαίρετα σύμβολα τα οποία δύσκολα μπορεί να κατανοήσει ένας άνθρωπος. Σε αντίθεση η κατασκευή υπολογιστικών φωνητικών υπομονάδων οι οποίες βασίζονται σε γλωσσολογικούς κανόνες και φωνητικές επισημειώσεις έχουν πλήρη αντιστοιχία με φωνητικά σύμβολα τα οποία έχουν κατασκευαστεί από τους γλωσσολόγους. Επιπλέον η χρησιμοποίηση φωνητικών υπομονάδων δίνουν την δυνατότητα αναγνώρισης νοημάτων τα οποία δεν υπάρχουν στο σώμα εκπαίδευσης αρκεί να είναι γνώση η αντιστοιχία με τις φωνητικές υπομονάδες.

Σε αυτό το κεφάλαιο θα παρουσιάσουμε μια μέθοδο για την αξιοποίηση γλωσσολογικής πληροφορίας με στόχο την μοντελοποίηση και αναγνώριση της ΝΓ. Η μέθοδος αυτή βασίζεται στα εξής: 1) Μετατροπή των επισημειώσεων επιπέδου HamNoSys [63] σε δομημένα φωνητικά σύμβολα και 2) Ενσωμάτωση των φωνητικών συμβόλων σε ένα σύστημα αναγνώρισης της ΝΓ. Η μέθοδος μας βασίζεται στην μετατροπή HamNoSys επισημειώσεων σε φωνητικά σύμβολα και στην ενσωμάτωση αυτών σε ένα πιθανοτικό πλαίσιο με στόχο την αναγνώριση της ΝΓ. Η παραπάνω μετατροπή αντιστοιχεί εννοιολογικά στην μετατροπή μιας λέξης από τον γραπτό λόγο στην προφορά της στον προφορικό λόγο. Για την μετατροπή των HamNoSys σε φωνητικά σύμβολα σύμφωνα με το PDTS [41] μοντέλο χρησιμοποιήσαμε την ερευνητική εργασία του C. Vogler [78]. Ωστόσο οι ακολουθίες των PDTS συμβόλων που παράγονται δεν περιέχουν πληροφορία σε σχέση με τον χρόνο. Για την αντιμετώπιση του παραπάνω προβλήματος συνδυάζουμε τα διανύσματα χαρακτηριστικών που έχουν εξαχθεί από το βίντεο νοηματισμού με τα αντίστοιχα φωνητικά σύμβολα για την εκπαίδευση υπολογιστικών φωνητικών υπομονάδων κάνοντας χρήση Κρυφών Μαρκοβιανών Μοντέλων. Έτσι επιτυγχάνουμε την αντιστοίχιση των φωνητικών συμβόλων με τα διανύσματα χαρακτηριστικών. Έτσι προκύπτουν υπολογιστικές φωνητικές υπομονάδες οι οποίες αντιστοιχούν σε PDTS σύμβολα και επιπλέον περιέχουν πληροφορία σε σχέση με τον χρόνο και την ακριβή αντιστοιχία τους με τα πραγματικά δεδομένα. Αυτό έχει ως αποτέλεσμα το σύστημα αναγνώρισης



Σχήμα 4.1: Παραδείγματα εκτέλεσης για τρία νοήματα της ΕΝΓ : ΣΩΡΟΣ, ΑΜΕΣΩΣ και ΕΥΡΩΠΗ.

να εξαγάγει PDTs σύμβολα κατά την διάρκεια της αναγνώρισης.

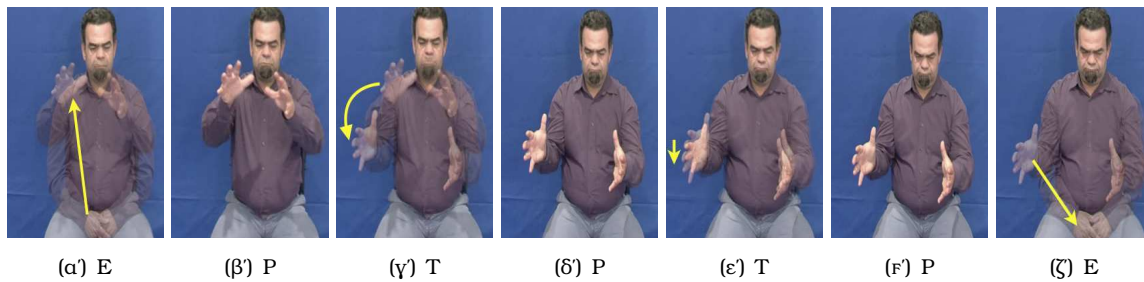
4.1 Μετατροπή των HamNoSys σε PDTs φωνητικά σύμβολα¹

Στην περιοχή της φωνητικής μοντελοποίησης για τους σκοπούς της αναγνώρισης ΝΓ υπήρξε μικρή πρόοδος μετά την ερευνητική εργασία των Vogler και Metaxas [82]. Ένας από τους βασικότερους λόγους ήταν η έλλειψη επισημειώσεων σε φωνητικό επίπεδο. Η έλλειψη αυτή οφειλόταν κυρίως στην αρκετά χρονοβόρα διαδικασία για την παραγωγή τους. Σε αυτή την ενότητα θα παρουσιάσουμε για λόγους συνέπειας επιγραμματικά την ερευνητική εργασία του C. Vogler [78] για την αυτόματη μετατροπή των HamNoSys σε PDTs σύμβολα τα οποία στην συνέχεια ενσωματώνονται στο σύστημα αναγνώρισης ΝΓ.

Σε αντιστοιχία με την δουλειά των Vogler και Metaxas η βασική φωνητική δομή ενός νοήματος είναι μια ακολουθία από τμήματα τα οποία μπορούν να μοντελοποιηθούν με βάση τα PDTs σύμβολα από την πρόσφατη δουλειά των Johnson's και Liddell's [41]. Το PDTs μοντέλο αποτελεί μια εξέλιξη του Movement-Hold μοντέλου που είχε προταθεί από τους ίδιους συγγραφείς στην δεκαετία του 90 [48]. Πιο συγκεκριμένα με βάση το PDTs μοντέλο, κάθε νόημα μπορεί να θεωρηθεί ότι αποτελείται από μια ακολουθία από PDTs σύμβολα. Τα PDTs σύμβολα χωρίζονται σε 4 κατηγορίες 1) Postures (P), δηλαδή η μη-ύπαρξη μεταβολής κανενός αρθρώτη το δεδομένο χρονικό παράθυρο 2) Detentions (D) αντίστοιχα με τα Postures αλλά πολύ μικρότερης διάρκειας 3) Transitions (T), δηλαδή η ύπαρξη μεταβολής ενός τουλάχιστον αρθρώτη το δεδομένο χρονικό παράθυρο και 4) Steady-Shift (SS), αντίστοιχα με τα Transitions αλλά μικρότερης σταθερής ταχύτητας. Σε αυτό το κεφάλαιο θα διαχωρίζουμε μόνο μεταξύ των P,D και T. Επιπλέον εισάγουμε μια τέταρτη κατηγορία η οποία αντιστοιχεί στις κινήσεις επένθεσης (E) [48] η οποία αντιστοιχεί στην κίνηση μεταξύ δύο θέσεων χωρίς να λαμβάνουμε υπόψη το είδος της κίνησης και συνήθως εμφανίζεται μεταξύ διαδοχικών νοημάτων ή επαναλαμβανόμενες κινήσεις. Ένα παράδειγμα της ακολουθίας των PDTs συμβόλων για το νόημα ΣΩΡΟΣ — E P T P T P E — φαίνεται στο Σχήμα 4.2 και τον Πίνακα 4.1.

Η φωνητική επισημείωση των νοημάτων είναι σε επίπεδο HamNoSys [63], ένα συμβολικό σύ-

¹Το σύστημα μετατροπής των HamNoSys σε PDTs αναπτύχθηκε από τον C.Vogler [78] στα πλαίσια του ευρωπαϊκού προγράμματος Dicta-Sign.



Σχήμα 4.2: Νόημα ΣΩΡΟΣ : Κατάτμηση σε χρονικά τμήματα με την ενσωμάτωση των PDTs φωνητικών συμβόλων. Έχουμε κάνει υπέρθεση του πρώτου και τελευταίο πλαισίου κάθε τμήματος. Τα PDTs φωνητικών συμβόλων που αντιστοιχούν σε transitions ή epenthesis συνοδεύονται από ένα βέλος που υποδεικνύει το είδος της κίνησης. Κάθε χρονικό τμήμα αντιστοιχίζεται σε ένα PDTs σύμβολο. Τα είδη των PDTs συμβόλων είναι Epenthesis (E), Posture (P), Transition (T).

σημα επισημείωσης το οποίο μπορεί να περιγράψει ένα νόημα με επαρκή λεπτομέρεια για την αυτόματη αναπαραγωγή του από ένα avatar. Πιο συγκεκριμένα εμπεριέχει πληροφορία σε σχέση με την χειρομορφή, τον προσανατολισμό της παλάμης, την θέση και την κίνηση των χεριών. Ωστόσο δεν περιλαμβάνει πληροφορία σχετικά με την χρονική δομή/ακολουθία των συμβόλων με αποτέλεσμα να μην είναι κατάλληλη αναπαράσταση για τα συστήματα αναγνώρισης ΝΓ. Η φιλοσοφία των HamNoSys είναι μινιμαλιστική δηλαδή αποφεύγεται ο πλεονασμός και εμμένει στην λεπτομερή αναπαράσταση ενός νοήματος με το μικρότερο αριθμό συμβόλων. Ένα παράδειγμα για το νόημα ΣΩΡΟΣ απεικονίζεται παρακάτω :

$$\text{.. } \text{⌘}_{17} \backslash \text{r} \ominus \text{⌘}^2 \text{ } \text{⌘}^2 \text{ } \text{⌘} \text{ } \text{⌘} \text{ } \text{⌘}$$

Η παραπάνω επισημείωση αναφέρει ότι στο συγκεκριμένο νόημα τα χέρια κινούνται συμμετρικά οπότε περιγράφει την κίνηση και θέση μόνο του δεξιού χεριού και το ότι τα δάκτυλα των δύο χεριών ακουμπάνε μεταξύ τους.

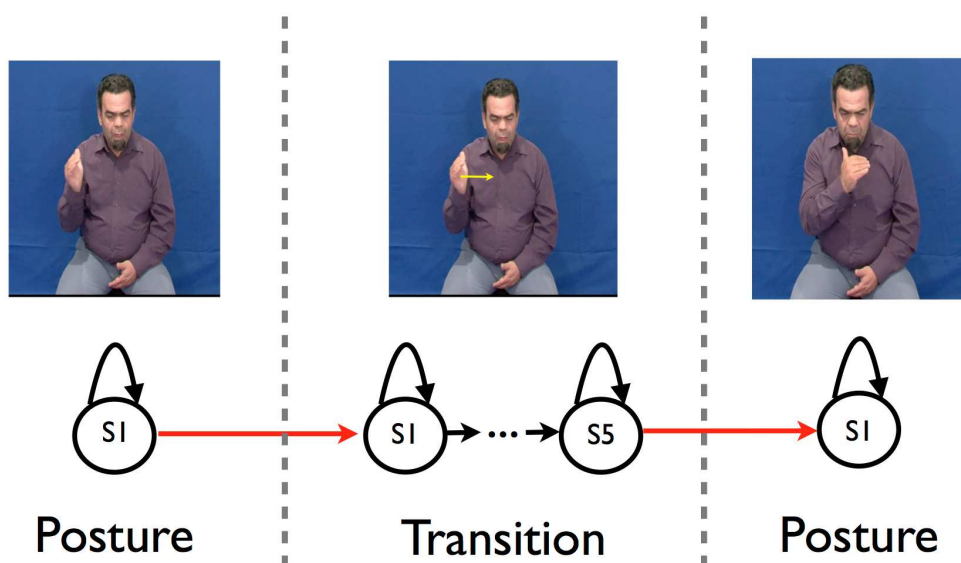
Για την μοντελοποίηση των νοημάτων σε ένα σύστημα αναγνώρισης ΝΓ είναι απαραίτητο για κάθε χρονικό τμήμα το σύνολο της πληροφορίας με βάση την PDTs μοντελοποίηση. Δηλαδή τα Ps και τα Ds πρέπει να περιέχουν την πληροφορία για την θέση και το είδος χειρομορφής όπως και τα Ts την πληροφορία για το είδος της κίνησης των χεριών. Η μέθοδος μετατροπής των HamNoSys σε ακολουθία PDTs συμβόλων ανακτά την συνεπαγόμενη πληροφορία που εμπεριέχεται στην αναπαράσταση με βάση τα HamNoSys και αναλύει κάθε νόημα σε μια ακολουθία χρονικών τμημάτων που αντιστοιχίζονται σε PDTs σύμβολα. Το βασικό βήμα είναι η συσσώρευση των διαφορικών, που περιγράφουν πως μια στάση ή κίνηση άλλαξαν σε σχέση με την προηγούμενη κατάσταση των αρθρώσεων. Στην συνέχεια τα διαφορικά αυτά εφαρμόζονται με συγκεκριμένη σειρά για την πλήρη ανάκτηση της πληροφορίας σε κάθε χρονική κατάσταση. Η παραπάνω διαδικασία εφαρμόζεται για κάθε κανάλι πληροφορίας ανεξάρτητα (χειρομορφής vs. θέσης, δεξί vs. αριστερό χέρι κ.α.). Ένα παράδειγμα της μετατροπής από HamNoSys σε ακολουθία PDTs συμβόλων για το νόημα ΣΩΡΟΣ απεικονίζεται στο Σχήμα 4.2 και στο Πίνακα 4.1

4.2 Ανίχνευση και μοντελοποίηση και εκπαίδευση των φωνητικών υπομονάδων

Σε αυτή την ενότητα θα παρουσιάσουμε την βασική συνεισφορά μας που σχετίζεται με την ενσωμάτωση των PDTs φωνητικών συμβόλων στο σύστημα αναγνώρισης ΝΓ. Οι δεδομενοκεντρικές υπομονάδες που παρουσιάσαμε στο προηγούμενο κεφάλαιο 3 προσαρμόζονται κατάλληλα στον

Πίνακας 4.1: PDTS φωνητικά σύμβολα για το νόημα ΣΩΡΟΣ της ΕΝΓ

Πλαίσια	Είδος	PDTS φωνητικά σύμβολα
1:12	E	rest-position — location-head
13:13	P	ocation-head
14:25	T	directedmotion, curve-r, direction-o, second-direction-do, tense-true
26:27	P	location-torso, side=right_beside
28:50	T	directedmotion, direction-dr, small
51:51	P	location-torso, side=right_beside_down
52:66	E	location-torso, side=right_beside_down — rest-position



Σχήμα 4.3: Αντιστοιχία των PDTS συμβόλων με τα HMM μοντέλα για το νόημα ΑΜΕΣΩΣ.

χώρο των χαρακτηριστικών αλλά δεν είναι άμεσα εκμεταλλεύσιμες από άλλους διεπιστημονικούς τομείς όπως η γλωσσολογική έρευνα, αυτόματη σύνθεση νοηματικής γλώσσας κ.α. Για τον λόγω αυτό προχωρήσαμε στην ενσωμάτωση γλωσσολογικής πληροφορίας για την κατασκευή υπολογιστικών φωνητικών υπομονάδων. Το συνολικό σύστημα αναγνώρισης απαρτίζεται από τις εξής συνιστώσες: 1) Μοντελοποίηση και εκπαίδευση των PDTS φωνητικών υπομονάδων, 2) Κατάτμηση και συγχρονισμός των PDTS συμβόλων με τα πραγματικά δεδομένα, 3) Κατασκευή ενός PDTS φωνητικού λεξικού 4) Αναγνώριση της ΝΓ.

4.2.1 Εκπαίδευση των PDTS φωνητικών υπομονάδων

Για κάθε PDTS σύμβολο χρησιμοποιώντας τα διανύσματα χαρακτηριστικών από τα πραγματικά δεδομένα εκπαιδεύουμε μια PDTS φωνητική υπομονάδα χρησιμοποιώντας Κρυφά Μαρκοβιανὰ Μοντέλα [43]. Ανάλογα με το είδος της υπομονάδας χρησιμοποιούμε και HMM διαφορετικής τοπολογίας. Πιο συγκεκριμένα για τις υπομονάδες που αντιστοιχούν σε PDTS σύμβολα, T και E χρησιμοποιούμε ένα HMM τριών και πέντε καταστάσεων αντίστοιχα με Bakis τοπολογία. Για τις υπομονάδες που αντιστοιχούν σε PDTS σύμβολα, P και D χρησιμοποιούμε ένα HMM ενός και τριών καταστάσεων left-right τοπολογίας. Επιπλέον σε κάθε κατάσταση όλων των HMM χρησιμοποιού-

με μια γκαουσιανή με διαγώνιο πίνακα συνδιακύμανσης. Η αρχικοποίηση των HMM μοντέλων γίνεται χρησιμοποιώντας την μέση τιμή και την τυπική απόκλιση του συνόλου των δεδομένων (flat-start [92]) και στην συνέχεια εφαρμόζουμε embedded training. Οι νέες υπολογιστικές υπομονάδες εμπεριέχουν στατιστική πληροφορία η οποία προήλθε κατά την εκπαίδευση τους χρησιμοποιώντας τα διανύσματα χαρακτηριστικών από τα πραγματικά δεδομένα και επίσης αντιστοιχούν ένα προς ένα με PDTs σύμβολα. Στο Σχήμα 4.3 απεικονίζουμε την αντιστοιχία μεταξύ των HMM και των PDTs συμβόλων για το νόημα ΑΜΕΣΩΣ της ΕΝΓ.

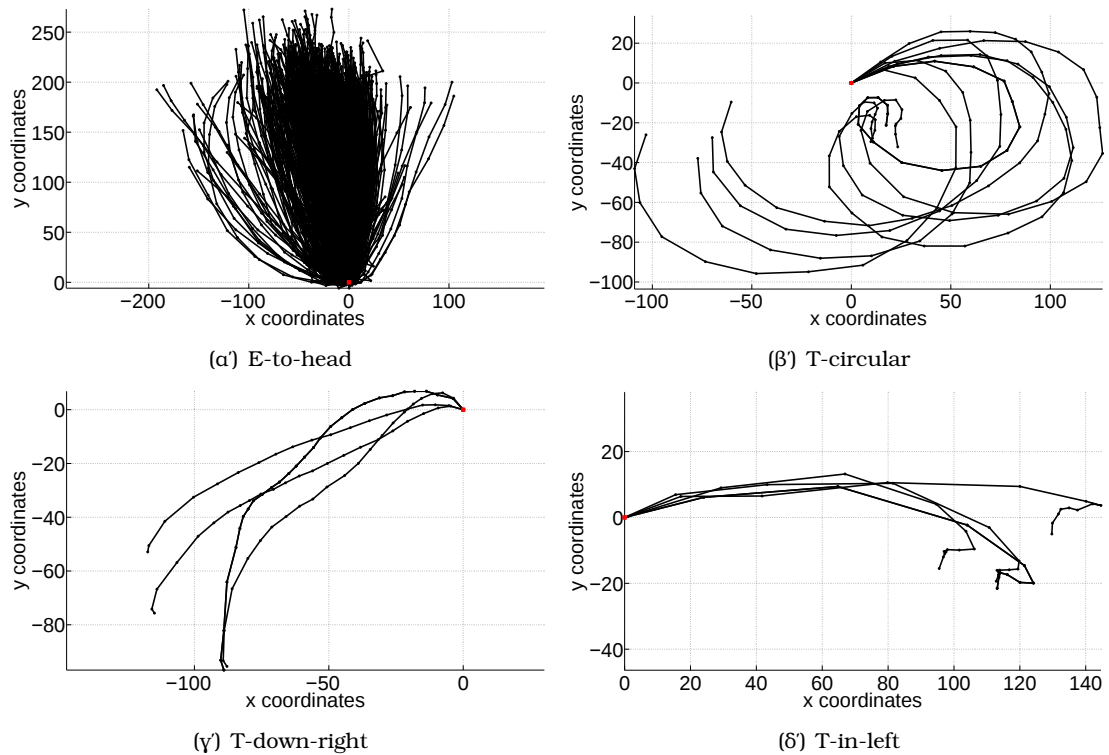
4.2.2 Αντιστοίχιση των PDTs συμβόλων με τα πραγματικά δεδομένα

Οι ακολουθίες των PDTs συμβόλων που παράγονται, μια για κάθε νόημα από σύστημα μετατροπής των HamNoSys σε PDTs (εν. 4.1), δεν περιέχουν πληροφορία σε σχέση με τον χρόνο. Με άλλα λόγια δεν γνωρίζουμε την ακριβή αντιστοιχία τους με τα πραγματικά δεδομένα. Για την αντιστοίχιση αυτή συνδυάζουμε τα ήδη εκπαιδευμένα HMM μοντέλα των PDTs υπομονάδων τις ακολουθίες των PDTs συμβόλων και τα διανύσματα χαρακτηριστικών των πραγματικών δεδομένων. Συνενώνουμε τα HMM μοντέλα δημιουργώντας ένα δίκτυο από HMMs Σχήμα 4.3 και εφαρμόζουμε τον αλγόριθμο Viterbi για κάθε διάνυσμα χαρακτηριστικών και της αντίστοιχης ακολουθίας των PDTs συμβόλων. Το αποτέλεσμα του αλγόριθμου Viterbi είναι η ακολουθία από PDTs σύμβολα μαζί με την χρονική τους αντιστοιχία με το διάνυσμα χαρακτηριστικών των πραγματικών δεδομένων. Εφαρμόζοντας την παραπάνω διαδικασία για κάθε νόημα προκύπτει ένα φωνητικό λεξικό με PDTs σύμβολα μαζί με την χρονική τους αντιστοιχία στα πραγματικά δεδομένα (σημεία κατάτμησης).

Στο Σχήμα 4.2 απεικονίζουμε ένα παράδειγμα της κατάτμησης των πραγματικών δεδομένων και της αντιστοίχιας των PDTs συμβόλων με το διάνυσμα χαρακτηριστικών για το νόημα ΣΩΡΟΣ. Κάθε εικόνα αντιστοιχεί σε ένα PDTs σύμβολο. Για τους σκοπούς της οπτικοποίησης χρησιμοποιήσαμε την εξής παραδοχή: (1) Για τα τμήματα που αντιστοιχούν σε T και E υπρεθέτουμε το αρχικό και τελικό πλαίσιο και απεικονίζουμε την τροχιά της κίνησης με ένα βέλος από την αρχική στην τελική θέση. (2) Για τα τμήματα που αντιστοιχούν σε P και D απεικονίζουμε μόνο το αρχικό πλαίσιο αφού τα χέρια παραμένουν σταθερά. Επιπλέον στον Πίνακα 4.1 απεικονίζουμε την ακολουθία των PDTs συμβόλων για το νόημα ΣΩΡΟΣ μαζί με τα σημεία κατάτμησης κάθε τμήματος.

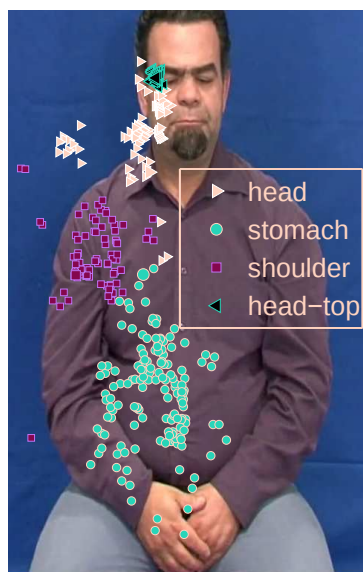
4.2.3 Αποτελέσματα Φωνητικών Υπομονάδων

Στο Σχήμα 4.4 βλέπουμε παραδείγματα από τις φωνητικές υπομονάδες που αντιστοιχούν σε κίνηση (T και E) απεικονίζοντας τις τροχιές των κινήσεων πάνω στο νοηματικό χώρο. Η αντιστοιχία τους με τα PDTs σύμβολα φαίνεται στον Πίνακα 4.2. Στο Σχήμα 4.4(α') βλέπουμε τα δεδομένα εκπαίδευσης από μια συχνά εμφανιζόμενη υπομονάδα η οποία αντιστοιχεί σε επένθεση κίνηση (E-to-head). Η υπομονάδα αυτή μοντελοποιεί κινήσεις που ξεκινούν από το σημείο ηρεμίας και καταλήγουν στο κεφάλι. Στο Σχήμα 4.4(β') απεικονίζουμε τα δεδομένα εκπαίδευσης για την υπομονάδα (T-circular) που μοντελοποιεί κυκλικές κινήσεις. Ένα παράδειγμα νοήματος που περιέχει αυτή την υπομονάδα είναι το νόημα ΕΥΡΩΠΗ (Σχήμα 4.1(γ')). Στα Σχήματα 4.4(γ') και 4.4(δ') απεικονίζουμε τα δεδομένα εκπαίδευσης των υπομονάδων T-down-right και T-in-left που μοντελοποιούν ευθείες κινήσεις με κατεύθυνση προς τα δεξιά-κάτω και αριστερά αντίστοιχα. Δύο ενδεικτικά νοήματα όπου εμφανίζονται οι δύο υπομονάδες είναι ο ΣΩΡΟΣ και ΑΜΕΣΩΣ αντίστοιχα (Σχήματα 4.1(α'), 4.1(β')). Στο Σχήμα 4.5 απεικονίζουμε με διαφορετικά χρώματα τα δεδομένα εκπαίδευσης (σε πραγματικές διαστάσεις συντεταγμένες) για 4 υπομονάδες που μοντελοποιούν 4 διαφορετικά P. Πιο συγκεκριμένα αυτά είναι τα P-head, P-stomach, P-shoulder και P-head-top που αντιστοιχούν σε θέσης του δεξιού χεριού στο κεφάλι, στομάχι, ώμος και πάνω από το κεφάλι του νοηματιστή αντίστοιχα.



Σχήμα 4.4: Φωνητικές υπομονάδες που αντιστοιχούν σε Ts και Es. (α) Η E-to-head υπομονάδα αντιστοιχεί σε επένθεση κίνηση E, (β-δ) Οι T-circular, T-down-right, T-in-left υπομονάδες αντιστοιχούν σε μεταβάσεις (Ts). Απεικονίζουμε τις τροχιές στο πραγματικό νοηματικό χώρο εφαρμόζοντας κανονικοποίηση με βάση την αρχική τους θέση. Το κόκκινο σημάδι υποδεικνύει την αρχή της τροχιάς. Στο Πίνακα 4.2 απεικονίζουμε τα αντίστοιχα PDTs σύμβολα.

Παρατηρούμε ότι υπάρχουν φωνητικές υπομονάδες που είναι αρκετά συμπαγής (μικρή διασπορά) λόγω του μικρού αριθμού των δεδομένων εκπαίδευσης σε αντίθεση με άλλες οι οποίες είναι λιγότερο συμπαγής (μεγάλη διασπορά) όπου στα δεδομένα εκπαίδευσης εμπεριέχονται ακραίες τιμές. Για παράδειγμα η υπομονάδα P-head-top είναι αρκετά συμπαγής σε αντίθεση με την υπομονάδα P-head η οποία είναι λιγότερο. Άλλο ένα παράδειγμα είναι η υπομονάδα E-rest-head η οποία εμφανίζεται σε πληθώρα αριθμό νοημάτων σε αντίθεση με την υπομονάδα T-circular η οποία εμφανίζεται αρκετά πιο σπάνια.



(α')

Σχήμα 4.5: Φωνητικές υπομονάδες που αντιστοιχούν σε Ps. Απεικονίζουμε στον νοηματικό χώρο με διαφορετικό χρώμα τα δεδομένα εκπαίδευσης τεσσάρων υπομονάδων που αντιστοιχούν σε 4 διαφορετικά P (P-forehead, P-stomach, P-shoulder, P-head-top).

Πίνακας 4.2: Παραδείγματα των φωνητικών υπομονάδων και των νοημάτων στα οποία εμφανίζονται. '*' αντιστοιχεί σε πολλαπλά νοήματα.

Φων. Υπομ.	Νόημα	Τύπος	PDTS σύμβολο
E-to-head	*	E	rest-position — location-head
T-circular	ΕΥΡΩΠΗ	T	circularmotion, axis=i
T-down-right	ΣΟΡΩΣ	T	directedmotion, direction=dr, small
T-in-left	ΑΜΕΣΩΣ	T	directedmotion, direction=il, fast=true, halt=true
P-forehead	*	P	location=forehead
P-stomach	*	P	location=stomach
P-shoulder	*	P	location=shouldertop, side=right-beside
P-head-top	*	P	location=head-top

Κεφάλαιο 5

Αναγνώριση Νοηματικής Γλώσσας

5.1 Αναγνώριση Χειρομορφών

5.1.1 Σώμα Δεδομένων και Επισημείωση των Χειρομορφών

Σε αυτή την ενότητα περιγράφουμε τα δεδομένα που χρησιμοποιήθηκαν στα πειραματικά αποτελέσματα και την διαδικασία επισημείωσης των χειρομορφών. Οι παράμετροι που επισημειώθηκαν σχετίζονται με το είδος και την πόζα των χειρομορφών.

Δεδομένων και Προδιαγραφές: Το σώμα δεδομένων BU400 Συνεχής ANΓ [25] αποτελείται από 843 προτάσεις, 406 διαφορετικά νοήματα και 4 διαφορετικούς νοηματιστές. Το φόντο του βίντεο είναι ομοιόμορφο, οι εικόνες έχουν ανάλυση 648×484 εικονοστοιχεία και η καταγραφή έχει γίνει σε 60 πλαίσια ανά δευτερόλεπτο. Στα πειράματα ταξινόμησης των χειρομορφών που θα παρουσιάσουμε στην συνέχεια, χρησιμοποιήσαμε τα δεδομένα από την εμπρόσθια κάμερα, για ένα νοηματιστή και της ιστορία 'Accident'.

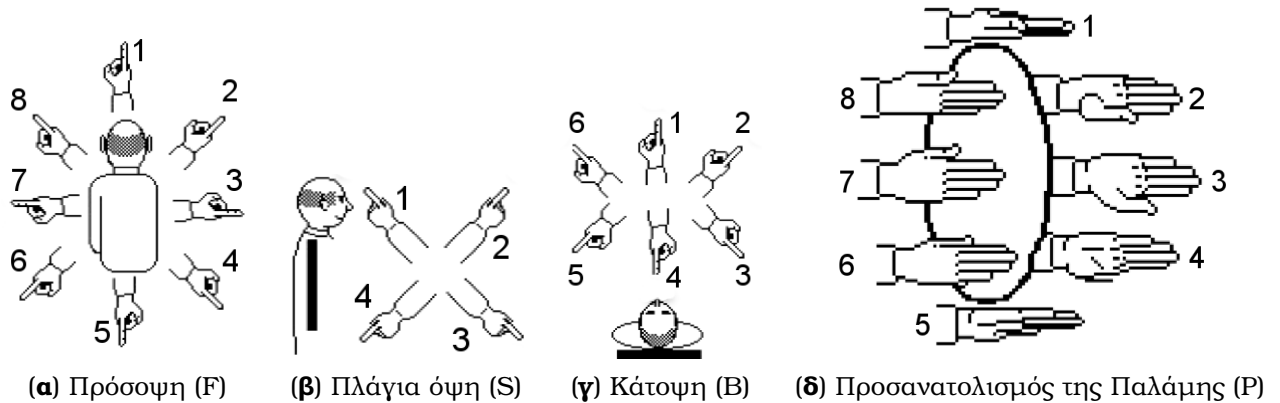
Παράμετροι Επισημείωσης Χειρομορφών: Οι παράμετροι οι οποίοι πρέπει να καθοριστούν στην επισημείωση των χειρομορφών για την επιβλεπόμενη εκπαίδευση των μοντέλων είναι: το είδος της χειρομορφής ανεξαρτήτως πόζας και η τρισδιάστατη πόζα, δηλαδή ο προσανατολισμός της χειρομορφής στον τρισδιάστατο χώρο. Για την επισημείωση του είδους της χειρομορφής ακολουθήσαμε την σύμβαση που παρουσιάζεται στο τεχνικό κείμενο [56]. Για την επισημείωση της τρισδιάστατης πόζας χρησιμοποιήσαμε τα HamNoSys σύμβολα [34]. Πιο συγκεκριμένα οι παράμετροι που επισημειώθηκαν είναι οι εξής:

1. **Είδος Χειρομορφής -Handshape identity (HSId)-** όπου αντιστοιχεί στο είδος της χειρομορφής ('A', 'B', 'I', 'C' κ.α.), ανεξαρτήτως πόζας. Μερικά παραδείγματα απεικονίζονται στον Πίνακα 5.1.
2. **Τρισδιάστατη πόζα** η οποία αποτελείται από τις εξής παραμέτρους (Σχήμα. 5.1):
 - (α') *Κατεύθυνση εκτεταμένων δακτύλων (Extended Finger Direction)* η οποία καθορίζει τον προσανατολισμό της χειρομορφής στον τρισδιάστατο χώρο και αποτελείται από την κατεύθυνση της χειρομορφής σε σχέση με τρία διαφορετικά επίπεδα: 1) Πρόσοψη -Front view (F)-, 2) Κάτοψη -Bird's view (B)- και 3) Πλάγια όψη -Side view (S)-.
 - (β') *Προσανατολισμός της Παλάμης (P)* ο οποίος ορίζεται σε σχέση με την κάτοψη όπως απεικονίζεται στο Σχήμα 5.1(δ).

Επιλογή Δεδομένων και Κλάσεις: Στα πειράματα που ακολουθούν έγινε επισημείωση των χειρομορφών έτσι ώστε να περιλαμβάνονται πολλαπλά είδη χειρομορφών και ποικιλία στην τρισδιάστατη πόζα. Πιο συγκεκριμένα δημιουργήσαμε τρεις βάσεις δεδομένων:

HSId	1	4	5	BL	cS
					
					

Πίνακας 5.1: Είδος Χειρομορφής -Handshape identity (HSId)- αντιστοιχεί στο είδος της χειρομορφής ανεξαρτήτως πόζας : 5 ενδεικτικά παραδείγματα.



Σχήμα 5.1: Παράμετροι Τρισδιάστατης πόζας : (α-γ) Κατεύθυνση εκτεταμένων δακτύλων : (α) Πρόσωση, (β) Πλάγια όψη, (γ) Κάτοψη και (δ) Προσανατολισμός της Παλάμης. Παρατηρούμε ότι έχουμε τροποποιήσει τα αντίστοιχα σχήματα από του άρθρου [34] ορίζοντας αριθμητικές παραμέτρους για κάθε διαφορετικό προσανατολισμό.

1. **DS-1**, αποτελείται από 1430 δείγματα χειρομορφών, χωρίς επικάλυψη, οι οποίες αντιστοιχούν σε 18 διαφορετικά είδη χειρομορφών.
2. **DS-1-extend**, αποτελείται από 3000 δείγματα χειρομορφών, χωρίς επικάλυψη, οι οποίες αντιστοιχούν σε 24 διαφορετικά είδη χειρομορφών.
3. **DS-2**, αποτελείται από 4962 δείγματα χειρομορφών, με και χωρίς επικάλυψη, οι οποίες αντιστοιχούν σε 42 διαφορετικά είδη χειρομορφών.

Στον Πίνακα 5.2 παρουσιάζουμε ενδεικτικά παραδείγματα από χειρομορφές που έχουν επισημειωθεί σε σχέση με το είδος τους και τον τρισδιάστατο προσανατολισμό τους.

5.1.2 Πειραματικά Αποτελέσματα

Σε αυτή την ενότητα παρουσιάζουμε τα πειραματικά αποτελέσματα βασιζόμενοι σε ένα στατιστικό σύστημα για την ταξινόμηση χειρομορφών. Αυτό βασίζεται στα εξής : 1) Στην εξαγωγή χαρακτηριστικών που σχετίζονται με τις χειρομορφές εφαρμόζοντας τη Δυναμική Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος-Εμφάνισης που περιγράφηκε στην ενότητα 2.4, 2) Στην επισημείωση των παραμέτρων που χαρακτηρίζουν μια χειρομορφή που περιγράφηκε στην ενότητα 5.1.1 και 3) Στην επιλογή των δεδομένων και κλάσεων (βλ. εν. 5.1.1). Στην συνέχεια περιγράφουμε τα πειράματα που υλοποιήθηκαν όπου εξαρτώνται από την βάση δεδομένων, τις κλάσεις και τη μέθοδο για την εξαγωγή των χαρακτηριστικών.

HSId	1	1	4	4	5Σ	5	5	5	A	A	BL	BL	BL	BL	
3D hand pose	F	8	1	7	6	1	7	8	1	8	8	8	7	8	8
	S	0	0	0	3	1	0	2	2	0	2	0	0	0	0
	B	0	0	0	6	4	0	1	1	0	6	0	0	0	0
	P	1	8	3	1	3	3	1	5	3	2	2	3	3	4
# insts.	14	24	10	12	27	38	14	19	14	31	10	15	23	30	
exmpls.															

HSId	BL	CUL	F	F	U	UL	V	Y	b1	c5	c5	cS	cS	fO2	
3D hand pose	F	8	7	7	1	7	7	8	8	7	8	8	7	8	8
	S	2	0	0	2	0	0	0	0	0	0	0	0	2	0
	B	6	0	0	1	0	0	0	0	0	0	6	6	6	0
	P	4	3	3	3	2	3	2	2	3	3	1	3	3	1
# insts.	20	13	23	13	10	60	16	16	10	17	18	10	34	12	
exmpls.															

Πίνακας 5.2: Παραδείγματα από διαφορετικά είδη χειρομορφών και οι αντίστοιχες παραμέτρους επισημείωσης. ‘# insts.’ αντιστοιχεί στον αριθμό των εμφανίσεων στην βάση δεδομένων. Σε κάθε περίπτωση ένα ενδεικτικό πραγματικό παράδειγμα απεικονίζεται το οποίο αντιστοιχεί στο συγκεκριμένο είδος χειρομορφής με τον συγκεκριμένο τρισδιάστατο προσανατολισμό.

Πειραματικό Πρωτόκολλο και σύγκριση με άλλες μεθόδους

Για τα πειράματα που υλοποιήσαμε εφαρμόσαμε διασταυρωμένη επικύρωση (cross-validation). Χωρίσαμε με τυχαίο τρόπο το σύνολο των δεδομένων σε 2 υποσύνολα δεδομένων, ένα για την εκπαίδευση των μοντέλων και ένα για την αξιολόγηση τους επαναλαμβάνοντας την παραπάνω διαδικασία 5 φορές. Τα υποσύνολα δεδομένων για την εκπαίδευση των μοντέλων και για την αξιολόγηση αποτελείτο από το 60% και 40% του συνόλου των δεδομένων αντίστοιχα. Ο αριθμός των εμφανίσεων κάθε κλάσης είναι κατά μέσο όρο 50 και διακυμαίνεται από 10 έως 300 ανάλογα με το πείραμα. Για την μοντελοποίηση χρησιμοποιήσαμε GMMs με μια γκαουσιανή κατανομή και διαγώνιο πίνακα συνδιακύμανσης. Τα GMMs αρχικοποιήθηκαν ομοιόμορφα και ο αλγόριθμος Baum-Welch χρησιμοποιήθηκε για την εκπαίδευση τους [92]. Δεν χρησιμοποιήσαμε πιο σύνθετους ταξινομημένες λόγω του ότι μας ενδιαφέρει η αξιολόγηση των εξαγόμενων χαρακτηριστικών για τις χειρομορφές και όχι η αξιολόγηση ενός πιο σύνθετου ταξινομητή. Επιπλέον τα GMMs ταιριάζουν με το συνολικό HMM-based σύστημα για την αναγνώριση νοηματικής γλώσσας [80, 1]. Τα πειράματα που υλοποιήσαμε χαρακτηρίζονται από την βάση δεδομένων, από το είδος εξάρτησης των κλάσεων σε σχέση με τις παραμέτρους επισημείωσης -Class Dependency (CD)- και από την μέθοδο εξαγωγής χαρακτηριστικών που χρησιμοποιήθηκε.

1. **Βάση Δεδομένων (DS):** Υλοποιήσαμε πειράματα σε τρεις βάσεις δεδομένων DS-1, DS-1-extend και DS-2 (βλέπε εν. 5.1.1).
2. **Class Dependency (CD):** Το είδος εξάρτησης των κλάσεων σε σχέση με τις παραμέτρους επισημείωσης της χειρομορφής (βλ.Πίνακα 5.3). Η εξάρτηση ή μη-εξάρτηση των μοντέλων μας π.χ. στην παράμετρο ‘Πρόσωση’ (F) αντιστοιχεί: είτε στην εκπαίδευση μοντέλων χειρομορφής ανεξαρτήτως του προσανατολισμού των δεδομένων σε σχέση με το επίπεδο της

Ετικέτα Εξάρτησης Κλάσεων	Παράμετροι Επισημείωσης				
	Πρόσοψη (F)	Πλάγια όψη (S)	Κάτοψη (B)	Παλάμη (P)	HSId
D-HFSBP	E	E	E	E	E
D-HSBP	*	E	E	E	E
D-HBP	*	*	E	E	E
D-HP	*	*	*	E	E
D-H	*	*	*	*	E

Πίνακας 5.3: Είδος εξάρτησης των κλάσεων σε σχέση με τις παραμέτρους επισημείωσης της χειρομορφής. Κάθε γραμμή αντιστοιχεί σε διαφορετική εξάρτηση. Εξάρτηση ή μη-εξάρτηση σε μια παράμετρο σημειώνεται με 'E' ή '*' αντίστοιχα. Για παράδειγμα στην τρίτη σειρά (D-HBP) τα μοντέλα εξαρτώνται από τις παραμέτρους [HSId,B,P].

Πρόσοψης είτε στην εκπαίδευση διαφορετικών μοντέλων χειρομορφών ανάλογα με των προσανατολισμού των δεδομένων σε σχέση με το επίπεδο της Πρόσοψης. Με άλλα λόγια στη μια ακραία περίπτωση (D-HFSBP) εκπαιδεύονται διαφορετικά μοντέλα χειρομορφών για κάθε διαφορετικό συνδυασμό των παραμέτρων επισημείωσης των χειρομορφών. Στην άλλη ακραία περίπτωση (D-H) εκπαιδεύονται διαφορετικά μοντέλα χειρομορφών για κάθε είδος χειρομορφής ανεξαρτήτως πόζας.

3. **Μέθοδος Εξαγωγής Χαρακτηριστικών:** Οι μέθοδοι που χρησιμοποιήθηκαν για την εξαγωγή χαρακτηριστικών είναι οι εξής:

- (α) **Αφινικά Αναβληθώδη Μοντελοποίηση Σχήματος- Εμφάνισης (Aff-SAM):** Αντιστοιχεί στην προτεινόμενη μέθοδο η οποία περιγράφηκε στην ενότητα 2.4.
- (β) **Μοντελοποίηση Σχήματος- Εμφάνισης με Απευθείας Μετασχηματισμούς Ομοιότητας (DS-SAM):** Αποτελεί μια απλοποιημένη εκδοχή της Aff-SAM μεθόδου με τις εξής διαφορές: **1)** Αντικατάσταση των αφινικών μετασχηματισμών που είχαν ενσωματωθεί στο μοντέλο ΜΣΕ από μετασχηματισμούς ομοιότητας (similarity transforms). **2)** Αντικατάσταση του ομαλοποιημένου ταιριάσματος του ΜΣΕ (βλέπε εν. 2.4) με απλή εκτίμηση (χωρίς βελτιστοποίηση) των παραμέτρων του μετασχηματισμού ομοιότητας χρησιμοποιώντας το κεντροειδές, το εμβαδόν και τον προσανατολισμό του χεριού. Στις περιπτώσεις όπου έχουμε επικαλύψεις το ταιρίασμα του μοντέλου εφαρμόζεται πάνω στην εικόνα ΜΣΕ η οποία περιέχει τους αρθρωτές που βρίσκονται σε επικάλυψη (όπως το άλλο χέρι ή το κεφάλι) χωρίς την χρησιμοποίηση της στατικής και δυναμικής πρότερη γνώσης των παραμέτρων του ΜΣΕ. Η παραπάνω μέθοδος είναι παρόμοια με την προσέγγιση του άρθρου [8].
- (γ) **Μοντελοποίηση Σχήματος- Εμφάνισης με Απευθείας Μετασχηματισμούς Μετατόπισης και Κλιμάκωσης (DTS-SAM):** Αποτελεί μια περαιτέρω απλοποιημένη εκδοχή της Aff-SAM μεθόδου με τις εξής διαφορές: **1)** Αντικατάσταση των αφινικών μετασχηματισμών που είχαν ενσωματωθεί στο ΜΣΕ από μετασχηματισμούς μετατόπισης και κλιμάκωσης. **2)** Αντικατάσταση του ομαλοποιημένου ταιριάσματος του ΜΣΕ μοντέλου με άμεση εκτίμηση των παραμέτρων μετατόπισης και κλιμάκωσης χρησιμοποιώντας το τετράγωνο με κατακόρυφες και οριζόντιες πλευρές που εφάπτεται στην δυαδική μάσκα του χεριού. Επιπλέον και σε αυτή την μέθοδο δεν γίνεται χρήση της στατικής και δυναμικής πρότερη γνώσης των παραμέτρων του ΜΣΕ. Η παραπάνω μέθοδος είναι παρόμοια με την προσέγγιση των άρθρων [19, 85].
- (δ) **Περιγραφητές Φουριέρ (FD):** Βασίζονται στην εξαγωγή των συντελεστών Φουριέρ από το περίγραμμα που περικλείει το χέρι μετά από κατάλληλες κανονικοποιήσεις που τους

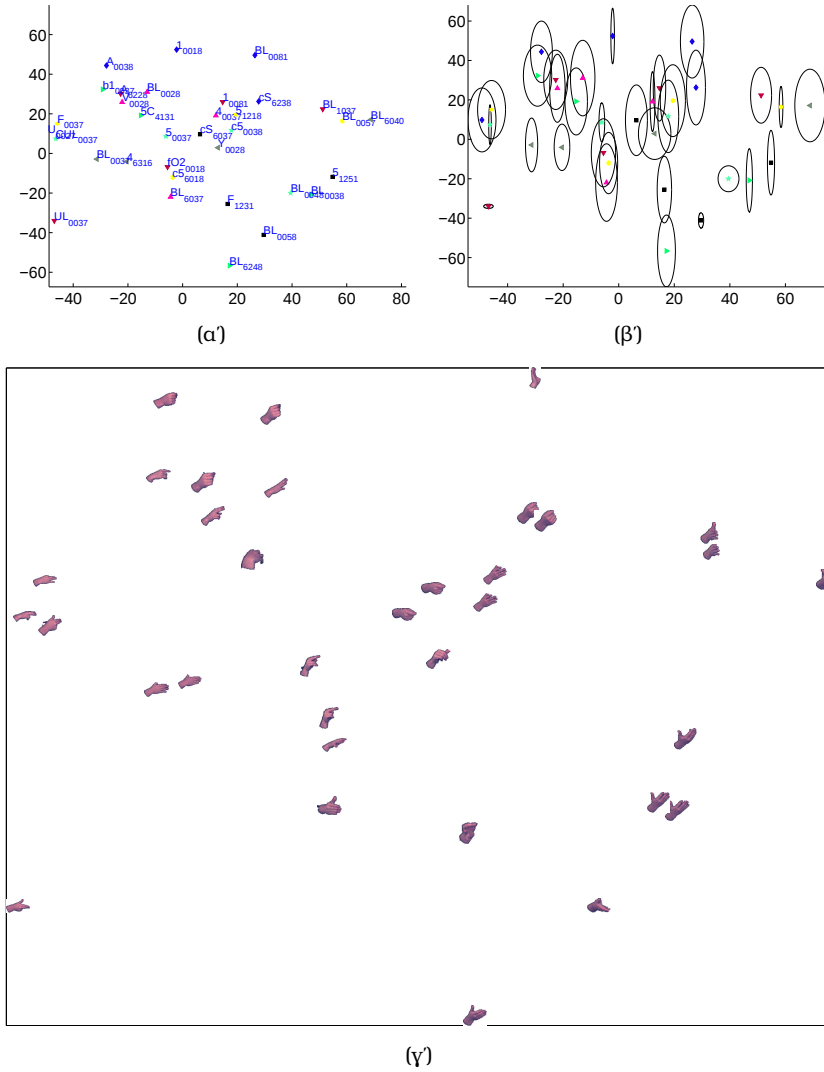
κάνουν αναλλοίωτους σε μετατόπιση κλιμάκωση και περιστροφή [14, 15]. Για λόγους μείωσης της διάστασης των χαρακτηριστικών κρατάμε τους πρώτους N_{FD} συντελεστές. Υλοποιήσαμε πολλαπλά πειράματα μεταβάλλοντας των αριθμό N_{FD} καταλήγοντας σε $N_{FD} = 30$ όπου είχαμε τα καλύτερα ποσοστά αναγνώρισης.

- (ε) **Ροπές Hu (M)**: Αποτελούνται από επτά μετρήσεις που εξαρτώνται από τις κεντρικές ροπές ενός δυαδικού σχήματος και είναι αναλλοίωτες ως προς τους μετασχηματισμούς ομοιότητας του σχήματος [39].
- (ϛ) **Γεωμετρικά Χαρακτηριστικά (RB)**: Αποτελούνται από τα γεωμετρικά χαρακτηριστικά του χεριού τα οποία είναι το εμβαδόν, η εκκεντρότητα, ο βαθμός συμπίκνωσης, μήκη του ελάσσονος και μείζονος άξονα του χεριού [1].

Για την σύγκριση των Aff-SAM χαρακτηριστικών χρησιμοποιήσαμε τις πέντε παραπάνω μεθόδους τις οποίες τις χωρίζουμε σε δύο κατηγορίες. Η πρώτη κατηγορία αντιστοιχεί στις baseline μεθόδους οι οποίες είναι οι FD, M και RB. Η δεύτερη αντιστοιχεί στις πιο προχωρημένες μεθόδους DS-SAM και DTS-SAM οι οποίες αποτελούν απλοποιημένες εκδοχές της Aff-SAM.

Αξιολόγηση του Χώρου Χαρακτηριστικών

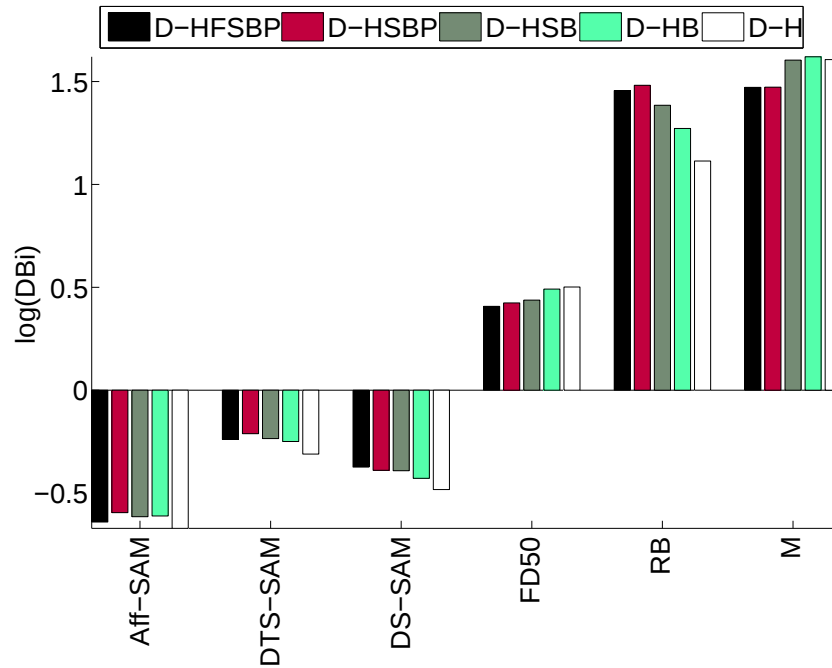
Σε αυτή την ενότητα θα ασχοληθούμε με την ανάλυση και αξιολόγηση του χώρου χαρακτηριστικών για την μέθοδο εξαγωγής χαρακτηριστικών Aff-SAM. Για την οπτικοποίηση του χώρου χαρακτηριστικών προβάλλουμε τα διανύσματα χαρακτηριστικών στο επίπεδο $\lambda_1 - \lambda_2$ όπου λ_1 και λ_2 αντιστοιχούν στις δύο πρώτες ιδιοκατευθύνσεις. Στο Σχήμα 5.2 έχουμε απεικονίσει τρεις διαφορετικούς τρόπους οπτικοποίησης των εκπαιδευμένων μοντέλων ανά κλάση για το πείραμα όπου κάθε μοντέλο είναι εξαρτημένο σε όλες τις παραμέτρους επισημείωσης (D-HFSBP Πίνακας 5.3). Στο Σχήμα 5.2(α) απεικονίζουμε τα κεντροειδή των μοντέλων μαζί με την ετικέτα της κάθε κλάσης. Στο Σχήμα 5.2(β) απεικονίζουμε τα ίδια κεντροειδή μαζί με τις ελλείψεις οι οποίες αντιπροσωπεύουν τις αντίστοιχες πυκνότητες πιθανότητες των μοντέλων. Τέλος στο Σχήμα 5.2(γ) απεικονίζουμε μια ενδεικτική εικόνα χειρομορφής για κάθε κλάση η οποία αντιστοιχεί στο διάνυσμα χαρακτηριστικών το οποίο βρίσκεται πλησιέστερα στο κεντροειδές του μοντέλου. Παρατηρούμε ότι μοντέλα που αντιπροσωπεύουν παρόμοιες χειρομορφές βρίσκονται κοντά πάνω στον χώρο των χαρακτηριστικών. Σε αντίθεση μοντέλα που αντιπροσωπεύουν πολύ διαφορετικές χειρομορφές βρίσκονται μακριά. Ο απεικονιζόμενος χώρος χαρακτηριστικών της μεθόδου Aff-SAM διαχωρίζει αρκετά χειρομορφές διαφορετικού είδους σε σύγκριση με τους χώρους χαρακτηριστικών από τις υπόλοιπες μεθόδους εξαγωγής χαρακτηριστικών. Για να υποστηρίξουμε το παραπάνω επιχείρημα ποσοτικοποιήσαμε τον διαχωρισμού διαφορετικών χειρομορφών για όλες τις μεθόδους εξαγωγής χαρακτηριστικών. Χρησιμοποιήσαμε τον δείκτη Davies-Boulding [21], ο οποίος ποσοτικοποιεί την ποιότητα και διακριτική ικανότητα ενός χώρου χαρακτηριστικών. Το Σχήμα 5.3 απεικονίζει τις μετρήσεις για κάθε μέθοδο εξαγωγής χαρακτηριστικών μεταβάλλοντας την εξάρτηση των μοντέλων στις παραμέτρους προσανατολισμού της χειρομορφής όπως περιγράφηκε στην ενότητα 5.1.2 και στον Πίνακα 5.3. Παρατηρούμε ότι οι δείκτες DB για την μέθοδο Aff-SAM είναι μικρότεροι σε σχέση με τις άλλες μεθόδους. Υποδεικνύοντας ότι ο χώρος χαρακτηριστικών χρησιμοποιώντας την Aff-SAM μέθοδο είναι περισσότερο συμπαγής και έχει μεγαλύτερη διακριτική ικανότητα σε διαφορετικού είδους χειρομορφές. Επιπλέον παρατηρούμε ότι οι δείκτες DB παραμένουν σταθεροί στην μεταβολή της εξάρτησης των μοντέλων από τις παραμέτρους του προσανατολισμού υποδεικνύοντας ότι τα εκπαιδευμένα μοντέλα αντιμετωπίζουν επιτυχώς την μεταβλητότητα στην πόζα των χειρομορφών.



Σχήμα 5.2: Χώρος χαρακτηριστικών του πειράματος D-HFSBP όπου τα μοντέλα εξαρτώνται από όλες τις παραμέτρους επισημείωσης χρησιμοποιώντας την μέθοδο Aff-SAM. Για λόγους οπτικοποίησης προβάλλουμε τα διανύσματα χαρακτηριστικών των εκπαιδευμένων μοντέλων στο επίπεδο $\lambda_1 - \lambda_2$ όπου λ_1 και λ_2 αντιστοιχούν στις δύο πρώτες ιδιοκατευθυνσεις. (α) Τα κεντροειδή των μοντέλων μαζί με τις αντίστοιχες ετικέτες που υποδεικνύουν το είδος της χειρομορφής και του προσανατολισμού της ($[HSId]_{FSP}$.) (β) Τα κεντροειδή των μοντέλων και οι ελλείψεις που δείχνουν την διασπορά τους. (γ) Πραγματικές κομμένες εικόνες χειρομορφών για κάθε κεντροειδή.

5.1.3 Πειράματα Ταξινόμησης

Για την αξιολόγηση των αποτελεσμάτων ταξινόμησης εφαρμόσαμε διασταυρωμένη επικύρωση. Σε κάθε πείραμα απεικονίζουμε την μέση τιμή και τυπική απόκλιση των πειραματικών αποτελεσμάτων. Στον Πίνακα 5.8 παρουσιάζουμε μια σύνοψη των μέσων αποτελεσμάτων από όλα τα διαφορετικά πειράματα που υλοποιήσαμε στις τρεις διαφορετικές βάσεις δεδομένων που περιγράψαμε προηγουμένως. Οι μέσοι όροι των αποτελεσμάτων υπολογίστηκαν για τα πέντε διαφορετικά σύνολα δεδομένων εκπαίδευσης και αξιολόγησης (cross-validation) και για τις διαφορετικές εξαρτήσεις των κλάσεων από τις παραμέτρους επισημείωσης του προσανατολισμού της χειρομορφής (όπου έχει υπάρξει μεταβολή). Στην περίπτωση 'DS-1' ο μέσος όρος αποτελεσμάτων που απεικονί-



Σχήμα 5.3: Davies-Bouldin Index (DBi) σε λογαριθμική κλίμακα (άξονας y) για όλες τις μεθόδους εξαγωγής χαρακτηριστικών μεταβάλλοντας την εξάρτηση των κλάσεων στις παραμέτρους επισημείωσης.

Πείραμα	Μέθοδος	Σώμα Δεδομένων	Εξ. Κλάσεων	Επικ.	Αποτελέσματα		
					#HSIds	Avg. Acc. %	Std.
DS-1	Aff-SAM	DS-1	Πίνακας 5.3	✗	18	93.7	1.5
	DS-SAM			✗	18	93.4	1.6
	DTS-SAM			✗	18	89.2	1.9
DS-1-extend	Aff-SAM	DS-1-extend	'D-H'	✗	24	77.2	1.6
	DS-SAM			✗	24	74	2.3
	DTS-SAM			✗	24	67	1.4
DS-2	Aff-SAM	DS-2	Πίνακας 5.3	✓	42	74.9	0.9
	DS-SAM			✓	42	66.1	1.1
	DTS-SAM			✓	42	62.7	1.4

Πίνακας 5.4: Σύνοψη των πειραματικών αποτελεσμάτων, μεταβάλλοντας το σώμα δεδομένων, την εξάρτηση των κλάσεων και την μέθοδο εξαγωγής των χαρακτηριστικών. Όπου η Εξ. Κλάσεων αντιστοιχεί στην εξάρτηση των κλάσεων σε σχέση με της παραμέτρους επισημείωσης του προσανατολισμού της χειρομορφής. Επικ. υποδεικνύει την ύπαρξη ή όχι επικαλύψεων στις χειρομορφές που περιέχονται στο σώμα δεδομένων. # HSIds αντιστοιχεί στον αριθμό του είδους των διαφορετικών χειρομορφών που περιέχονται στο σώμα δεδομένων. Avg. Acc. είναι ο μέσος όρος αναγνώρισης και Std είναι η τυπική απόκλιση.

ζεται είναι για όλες τις διαφορετικές περιπτώσεις εξάρτησης των κλάσεων όπως περιγράφονται στον Πίνακα 5.3. Στην περίπτωση 'DS-1-extend' χρησιμοποιούμε εξάρτηση κλάσεων D-H, με στόχο να αυξήσουμε την μεταβλητότητα κάθε κλάσεις.

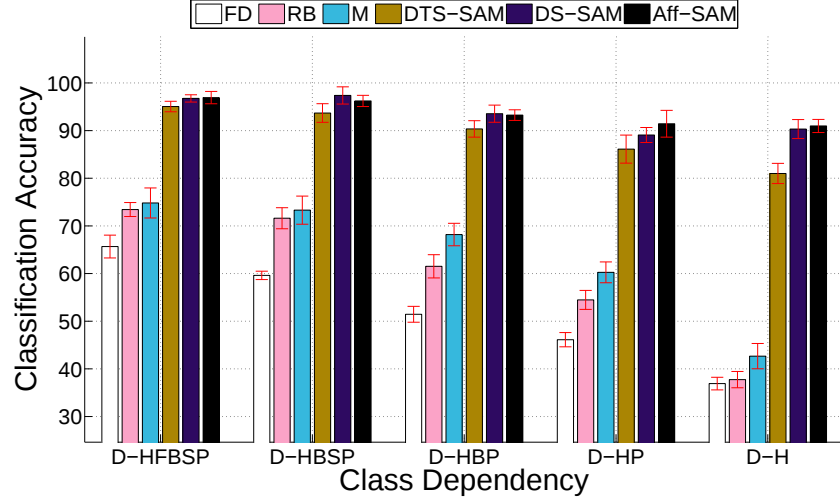
Σώμα Δεδομένων DS-1: Στο Σχήμα 5.4 συγκρίνουμε τις διαφορετικές μεθόδους για την εξαγωγή χαρακτηριστικών, μεταβάλλοντας την εξάρτηση των μοντέλων στις παραμέτρους επισημείωσης του προσανατολισμού της χειρομορφής (άξονας x). Χρησιμοποιήσαμε την βάση δεδομένων *DS-1* η οποία αποτελείται από 18 διαφορετικά είδη χειρομορφών από περιπτώσεις όπου δεν έχουμε επικαλύψεις. Ο αριθμός των κλάσεων μεταβάλετε ανάλογα με το είδος εξάρτησης στις παραμέτρους επισημείωσης του προσανατολισμού και απεικονίζεται στον Πίνακα 5.5 για κάθε πείραμα. Στο ένα άκρο ('D-HFBSP') εκπαιδεύουμε ένα GMM μοντέλο για κάθε διαφορετικό συνδυασμό των παραμέτρων επισημείωσης του προσανατολισμού της χειρομορφής. Έτσι έχουμε ένα μοντέλο για κάθε τρισδιάστατη πόζα και άρα τον ίδιο αριθμό κλάσεων (34 διαφορετικές κλάσεις). Στο άλλο άκρο ('D-H') εκπαιδεύουμε ένα GMM μοντέλο για κάθε διαφορετικό είδος χειρομορφής ανεξαρτήτως της τρισδιάστατης πόζας. Έτσι καταλήγουμε σε 18 διαφορετικά μοντέλα και κλάσεις. Παρατηρώντας το Σχήμα 5.4 βλέπουμε ότι η μέθοδος Aff-SAM υπερτερεί των μεθόδων FD, RB, M και DTS-SAM. Τα αποτελέσματα των μεθόδων Aff-SAM και DS-SAM είναι αρκετά αντίστοιχα. Για το παραπάνω ευθύνεται το πρόβλημα ταξινόμησης το οποίο είναι αρκετά εύκολο: μικρός αριθμός από HSId, μικρή μεταβλητότητα της τρισδιάστατης πόζας και η μη-ύπαρξη επικαλύψεων. Μια επιπλέον παρατήρηση είναι ότι το αποτέλεσμα ταξινόμησης της μεθόδου Aff-SAM επηρεάζεται ελάχιστα από την μεταβολή της εξάρτησης των κλάσεων στις παραμέτρους επισημείωσης του προσανατολισμού. Το παραπάνω ισχυροποιεί την παρατήρηση που έγινε στην προηγούμενη ενότητα ότι η Aff-SAM μέθοδο αντιμετωπίζει επιτυχώς την μικρή μεταβλητότητα στην τρισδιάστατη πόζα των χειρομορφών. Μια σύνοψη των αποτελεσμάτων μπορούμε να δούμε στην πρώτη γραμμή του πίνακα 5.8 (DS-1).

Σώμα Δεδομένων DS-1-extend: Στην συνέχεια παρουσιάζουμε αποτελέσματα χρησιμοποιώντας την βάση δεδομένων *DS-1-extend*. Η βάση δεδομένων είναι μια επέκταση της βάσης DS-1 η οποία αποτελείται από 24 είδη χειρομορφών (HSIds) και μεγάλη μεταβλητότητα της τρισδιάστατης πόζας. Δηλαδή εκπαιδεύουμε μοντέλα τα οποία είναι ανεξάρτητα της τρισδιάστατης πόζας. Η εξάρτηση των μοντέλων στις παραμέτρους επισημείωσης αντιστοιχεί στην D-H περίπτωση. Στην δεύτερη γραμμή του πίνακα 5.8 απεικονίζουμε τα αποτελέσματα της ταξινόμησης για τις μεθόδους Aff-SAM, DS-SAM και DTS-SAM. Παρατηρούμε ότι η μέθοδος Aff-SAM υπερτερεί και των δύο μεθόδων DS-SAM και DTS-SAM επιτυγχάνοντας αύξηση του ποσοστού αναγνώρισης 3.2% και 10% αντίστοιχα. Παρατηρούμε ότι σε αρκετά πιο δύσκολα προβλήματα ταξινόμησης (μεγαλύτερο αριθμό από HSId και μεγαλύτερη μεταβλητότητα της τρισδιάστατης πόζας) η Aff-SAM υπέρτερη των μεθόδων DS-SAM και DTS-SAM.

Σώμα Δεδομένων DS-2: Στην συνέχεια παρουσιάζουμε αποτελέσματα χρησιμοποιώντας την βάση δεδομένων *DS-2*. Η βάση δεδομένων αποτελείται από 42 είδη χειρομορφών (HSIds) με μεγάλη μεταβλητότητα στην τρισδιάστατη πόζα και επιπλέον περιέχονται περιπτώσεις όπου έχουμε επικαλύψεις. Υλοποιήσαμε πειράματα για τις τρεις μεθόδους Aff-SAM, DS-SAM και DTS-SAM μεταβάλλοντας την εξάρτηση των κλάσεων στις παραμέτρους επισημείωσης του προσανατολισμού. Ο αριθμός των κλάσεων για κάθε περίπτωση απεικονίζονται στον πίνακα 5.6. Στο Σχήμα 5.5 παρατηρούμε ότι η μέθοδος Aff-SAM υπερτερεί εμφανώς των μεθόδων DS-SAM και DTS-SAM επιτυγχάνοντας αύξηση του ποσοστού αναγνώρισης 10% κατά μέσο όρο σε όλες τις περιπτώσεις. Το παραπάνω υποδεικνύει ότι η μέθοδος Aff-SAM επιτυγχάνει ικανοποιητικά αποτελέσματα ταξινόμησης ακόμα και κατά την διάρκεια επικαλύψεων. Παρατηρούμε ότι δεν απεικονίζουμε τα αποτελέσματα των μεθόδων FD, M και RB επειδή οι συγκεκριμένες μέθοδοι δεν μπορούν να αντιμετωπίσουν τις περιπτώσεις όπου έχουμε επικαλύψεις έτσι τα αποτελέσματα ταξινόμησης είναι χαμηλά. Άλλη μια παρατήρηση είναι ότι κάνοντας τα μοντέλα μας ανεξάρτητα της τρισδιάστατης πόζας (D-H) το ποσοστό ταξινόμησης μειώνεται. Η μείωση αυτή οφείλεται στην μεγάλη μεταβλητότητα της τρισδιάστατης πόζας, με αποτέλεσμα το πρόβλημα της ταξινόμησης γίνεται αρκετά δύσκολο. Σε αυτό το πείραμα το εύρος της μεταβλητότητας είναι αρκετά μεγαλύτερο από αυτό που

Πειρ.	Πόζα & Εξάρτηση των Κλάσεων				
	D-HFSBP	D-HSBP	D-HBP	D-HP	D-H
#	34	33	33	31	18

Πίνακας 5.5: Περιπτώσεις χειρομορφών χωρίς επικαλύψεις. Αριθμός κλάσεων για κάθε πείραμα στην βάση δεδομένων DS-1.



Σχήμα 5.4: Πειράματα ταξινόμησης για περιπτώσεις χωρίς επικαλύψεις στην βάση δεδομένων DS-1. Πειραματικά αποτελέσματα μεταβάλλοντας την εξάρτηση των κλάσεων στις παραμέτρους επισημείωσης του προσανατολισμού [H,F,B,S,P] (άξονας x) και της μεθόδου εξαγωγής χαρακτηριστικών (legend). Ο αριθμός των κλάσεων για κάθε πείραμα απεικονίζεται στον πίνακα 5.5.

μπορεί να αντιμετωπιστεί με επιτυχία από τους αφινικούς μετασχηματισμούς του Aff-SAM μοντέλου. Επιπλέον κατά την διάρκεια επικαλύψεων εμφανίζονται λάθη στην οπτική μοντελοποίηση για την εκτίμηση της χειρομορφής η οποία βρίσκεται σε επικάλυψη τα οποία μεταφέρονται και στην ταξινόμηση χειρομορφών.

Πειρ.	Πόζα & Εξάρτηση των Κλάσεων				
	D-HFSBP	D-HSBP	D-HBP	D-HP	D-H
#	100	88	83	72	42

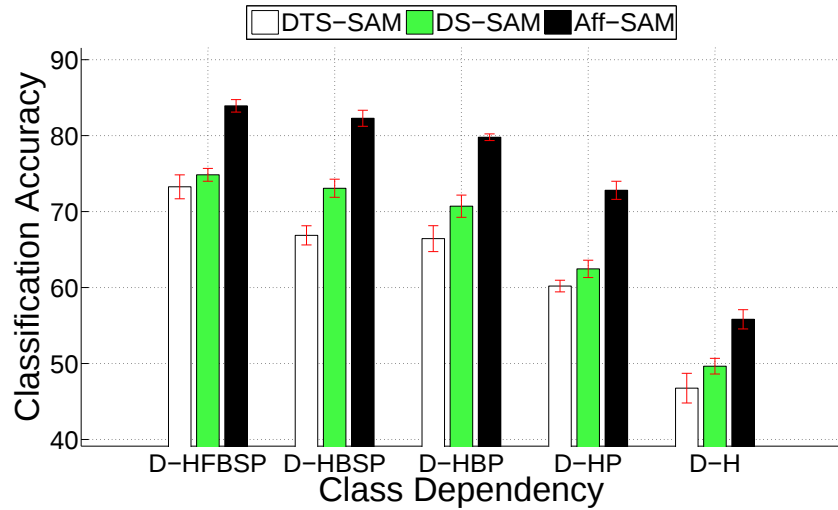
Πίνακας 5.6: Περιπτώσεις χειρομορφών με και χωρίς επικαλύψεις. Αριθμός κλάσεων για κάθε πείραμα στην βάση δεδομένων DS-2.

5.2 Αναγνώριση Μεμονωμένων Νοημάτων

5.2.1 Δεδομενοκεντρικές Υπομονάδες

Σύνοψη Συστήματος Αναγνώρισης ΝΓ με την μέθοδο DSCM

Σε αυτή την ενότητα θα κάνουμε μια σύνοψη του συστήματος αναγνώρισης νοηματικής γλώσσας με την χρήση της μεθόδου DSCM η οποία περιγράφηκε εκτενώς στο κεφάλαιο 3. Ως εισόδους έχουμε τις μετρήσεις από το σύστημα εξαγωγής χαρακτηριστικών (κεφ. ::), την χρονική κατάτμηση σε επίπεδο νοήματος και τις αντίστοιχες ετικέτες που υποδεικνύουν το νόημα που εκτελέστηκε.



Σχήμα 5.5: Πειράματα ταξινόμησης με και χωρίς επικαλύψεις στην βάση δεδομένων DS-2. Πειραματικά αποτελέσματα μεταβάλλοντας την εξάρτηση των κλάσεων στις παραμέτρους επισημείωσης του προσανατολισμού [H,F,B,S,P] (άξονας x) και της μεθόδου εξαγωγής χαρακτηριστικών (legend). Ο αριθμός των κλάσεων για κάθε πείραμα απεικονίζεται στον πίνακα 5.6.

1. Δυναμική και Στατική Κατάτμηση και Ταξινόμηση:

- 1α. Κατάτμηση και ταξινόμηση κάθε τμήματος σε δυναμικό ή στατικό μοντελοποιώντας το προφίλ της ταχύτητας για κάθε νόημα (εν. 3.2).
- 1β. Αναθεώρησης της ταξινόμησης στα στατικά τμήματα (εν. 3.2).

2. Δυναμικές και Στατικές Υπομονάδες, Εκπαίδευση, Λεξικό

- 2α. Δυναμικές Υπομονάδες: Ιεραρχική συσταδοποίηση των δυναμικών τμημάτων συνδυάζοντας τον αλγόριθμο DTW (εν. 3.3.5).
- 2β. Στατικές Υπομονάδες: Συσταδοποίηση των στατικών τμημάτων εφαρμόζοντας τον αλγόριθμο K-means (εν. 3.3.5).
- 2γ. Λεξικό: Συνδυασμός των στατικών και δυναμικών υπομονάδων για την κατασκευή λεξικού σε επίπεδο υπομονάδας (εν. 3.4).
- 2δ. Εκπαίδευση Υπομονάδων: Χρησιμοποίηση left-right HMM τριών καταστάσεων για τις δυναμικές υπομονάδες και GMM για τις στατικές υπομονάδες.

3. Αναγνώριση με βάση τις δυναμικές/στατικές υπομονάδες δυο σταδίων

- 3α. Κατάτμηση των δεδομένων αξιολόγησης σε δυναμικά/στατικά τμήματα χρησιμοποιώντας το ήδη εκπαιδευμένο εργοδικό HMM 2 καταστάσεων (εν. 3.2) και το κριτήριο μέγιστης πιθανοφάνειας.
- 3β. Ταξινόμηση των δυναμικών/στατικών τμημάτων στις δυναμικές/στατικές υπομονάδες με βάση το κριτήριο μέγιστης πιθανοφάνειας.
- 3γ. Αντιστοίχιση κάθε ακολουθίας υπομονάδων στο αντίστοιχο νόημα χρησιμοποιώντας το λεξικό το οποίο αποτελείται από μια εγγραφή για κάθε διαφορετική προφορά ενός νοήματος έτσι όπως έχει εμφανιστεί στην βάση δεδομένων.

Πίνακας 5.7: Τριάντα ενδεικτικά νοήματα της ANΓ μαζί με τον αριθμό επαναλήψεων τους.

<i>REALLY</i>	56	<i>SAY</i>	31	<i>BUT</i>	25	<i>ONE</i>	24	<i>ON</i>	18	<i>HERE</i>	17
<i>SEE</i>	16	<i>COP</i>	13	<i>LOOK</i>	12	<i>TO</i>	12	<i>FRIEND</i>	12	<i>MOTHER</i>	11
<i>MANY</i>	10	<i>MAN</i>	10	<i>GET</i>	9	<i>GO</i>	7	<i>BETTER</i>	7	<i>BECAUSE</i>	7
<i>CIGARETTE</i>	6	<i>SOME</i>	6	<i>EXAMPLE</i>	6	<i>WEATHER</i>	5	<i>END</i>	5	<i>FORMERLY</i>	5
<i>FINISH</i>	5	<i>DEGREE</i>	5	<i>80</i>	5	<i>SHOWER</i>	5	<i>KIND</i>	5	<i>WORK</i>	5

Πίνακας 5.8: Σύνοψη των πειραμάτων χρησιμοποιώντας διαφορετικές μεθόδους και παραμέτρους η οποίες περιγράφονται αναλυτικά στην ενότητα 5.2.1. Οι ορισμοί των ακρωνύμιων[†] κάθε μεθόδου περιγράφονται στην ενότητα 5.2.1.

Πειρ.Μετ.	Μέθοδος	Κατμ.	DSCM	#Υπ	#N	Χαρ.	M.O. Av. Νοήματος %	M.O. DF
#Υπ	SU-DSC	2S-ERG	✓	{21-61}	94	SPn-P	88.6	0.80
	SU-DSC	2S-ERG	✓				81.2	0.84
Χαρ.	SU-noDSC	2S-ERG	✗	Πίν. 5.9	94	Πίν. 5.9	70.5	0.81
	SU-Segm [28]	3S-LR	✗				69.9	0.82
	SU-Frame [6]	✗	✗				59.2	0.93
	SU-DSC	2S-ERG	✓			SPn-P	87.6	0.82
#N	SU-noDSC	2S-ERG	✗	46	{25,50	SPn-P	73.6	0.8
	SU-Segm [28]	3S-LR	✗		70,94}	SPn-P	75.4	0.8
	SU-Frame [6]	✗	✗			SPn-P	57.2	0.93

[†] Η στήλη Κατμ. αντιστοιχεί στον τύπο HMM που χρησιμοποιήθηκε για την κατάτμηση. '2S-ERG' αντιστοιχεί σε εργο-δικό HMM 2 καταστάσεων και '3S-LF' αντιστοιχεί σε left-right HMM 3 καταστάσεων. Πειρ.Μετ. είναι η συντομογραφία της πειραματικής μεταβολής. Μεταβολή των αριθμό των υπομονάδων (# Υπ.) μεταβολή των χαρακτηριστικών που χρησιμοποιήθηκαν (Χαρ.) και μεταβολή του αριθμού των νοημάτων (# N).

Πειραματικές Προδιαγραφές

Δεδομένα: Στα πειράματα που θα παρουσιάσουμε στην συνέχεια έχει χρησιμοποιηθεί η βάση δεδομένων BU400 Συνεχής ANΓ [25]. Πιο συγκεκριμένα έχουν επιλεγθεί 6 βίντεο/ιστορίες τις οποίες έχει διηγηθεί ένα νοηματιστής¹. Χρησιμοποιώντας το σύνολο των νοημάτων που εμφανίζονται στις παραπάνω 6 ιστορίες κατασκευάζουμε 4 διαφορετικά σύνολα δεδομένων με πλήθος νοημάτων 25, 50, 70 και 94. Κάθε ένα από αυτά τα σύνολα δεδομένων χρησιμοποιείται ανεξάρτητα σε διαφορετικό πείραμα. Για την εκπαίδευση επιλέγουμε τυχαία το 60% του συνόλου των δεδομένων και το υπόλοιπο 40% χρησιμοποιείται για την αξιολόγηση. Ο παραπάνω διαχωρισμός σε δεδομένα εκπαίδευσης και δεδομένα αξιολόγησης εφαρμόζεται στις επαναλήψεις κάθε νοήματος ξεχωριστά έτσι ώστε να εμφανίζονται όλα τα νοήματα στα δεδομένα εκπαίδευσης. Σε όλα τα πειράματα έχουμε εφαρμόσει διασταυρωμένη επικύρωση: με άλλα λόγια σε κάθε πείραμα εφαρμόζουμε τον διαχωρισμό σε δεδομένα εκπαίδευση και αξιολόγησης 3 φορές. Έτσι επαναλαμβάνουμε το ίδιο πείραμα 3 φορές χρησιμοποιώντας κάθε φορά διαφορετικά σύνολα δεδομένων για εκπαίδευση και αξιολόγηση. Το τελικό αποτέλεσμα προκύπτει από των μέσο όρο των επιμέρους αποτελεσμάτων. Ένα ενδεικτικό υποσύνολο των νοημάτων που περιέχονται στις 6 ιστορίες μαζί με τον αριθμό επαναλήψεων τους βλέπουμε στον πίνακα 5.7. Ο μέσος αριθμός επαναλήψεων ανά

¹Τα ονόματα των βίντεο είναι τα εξής: accident, biker_buddy, boston_la, football, lapd_story and siblings.

νόημα είναι 13 με μέγιστο αριθμό επαναλήψεων 137 και ελάχιστο 5.

Σημειογραφία Διανυσμάτων Χαρακτηριστικών: Τα διανύσματα χαρακτηριστικών που χρησιμοποιούμε σχετίζονται με την θέση-κίνηση το κυρίαρχου χεριού και είναι τα εξής: Κατεύθυνση (D), Κανονικοποιημένη Τροχιά Κίνησης (SPn), Μέγεθος Κίνησης -κλίμακα- (S) και Θέση (P). Η χρησιμοποίηση πολλαπλών χαρακτηριστικών διανυσμάτων στα πειράματα (βλ. Σχήμα 5.7), κωδικοποιείται με το σύμβολο ‘-’ (π.χ. A-B). Πιο συγκεκριμένα όταν χρησιμοποιείται η μέθοδος DSCM, το A-B υποδεικνύει ότι το διάνυσμα χαρακτηριστικών A θα χρησιμοποιηθεί στις δυναμικές υπομονάδες και το B στις στατικές. Σε αντίθεση όταν η μέθοδος DSCM δεν χρησιμοποιείται (σε όλες τις άλλες μεθόδους εκτός της SU-DSC) τότε συνενώνουμε τα διανύσματα χαρακτηριστικών A και B και χρησιμοποιούμε το νέο πολυδιάστατο διάνυσμα χαρακτηριστικών και για τις στατικές αλλά και για τις δυναμικές υπομονάδες.

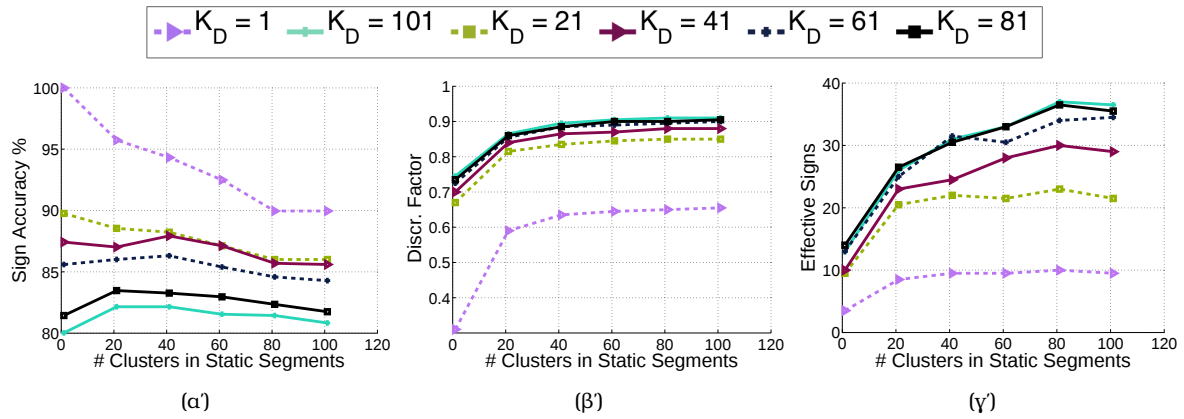
Κριτήρια Αξιολόγησης: Τα αποτελέσματα που παρουσιάζονται περιλαμβάνουν το ποσοστό αναγνώρισης σε επίπεδο νοήματος (Sign Accuracy), τον συντελεστή διάκρισης νοημάτων (F_d) και τον αριθμό των ενεργών νοημάτων που επεξηγήσαμε εκτενώς στην ενότητα 3.4.4.

Πειραματικά Αποτελέσματα

Τα πειράματα που υλοποιήθηκαν μπορούν να κατηγοριοποιηθούν ως εξής:

1. *Αριθμός Υπομονάδων:* Μεταβολή του αριθμού των σύστάδων που χρησιμοποιήθηκαν κατά την διάρκεια της συσταδοποίησης. Ο αριθμός αυτός αποτελείται από τον αριθμό των δυναμικών και στατικών υπομονάδων.
2. *Σύγκριση με Άλλες Μεθόδους:* Σύγκριση της προτεινόμενης μεθόδου SU-DSC με άλλες από την διεθνή βιβλιογραφία. Οι μέθοδοι προς σύγκριση αναπτύχθηκαν με σκοπό να περιλαμβάνουν τα βασικά χαρακτηριστικά των μεθόδων [6, 28] έτσι ώστε να ταιριάζουν στο παρόν πλαίσιο. Επιπλέον γίνεται σύγκριση των διαφορετικών αρχιτεκτονικών που χρησιμοποιήθηκαν για την κατάτμηση των νοημάτων σε υπομονάδες.
3. *Μέγεθος Λεξικού:* Μεταβολή του αριθμού των νοημάτων προς αναγνώριση με στόχο την διερεύνηση της επίδρασης της αύξησης του λεξιλογίου στην αναγνώριση.
4. *Διανύσματα Χαρακτηριστικών και Συνδυασμοί:* Για όλες τις διαφορετικές μεθόδους παρουσιάζουμε αποτελέσματα με διαφορετικά διανύσματα χαρακτηριστικών. Επίσης διερευνάται η επίδραση των κανονικοποιημένων διανυσμάτων χαρακτηριστικών και η ενσωμάτωση πολλαπλών ροών πληροφορίας.

Μεταβολή του αριθμού των υπομονάδων: Μια σύνοψη των αποτελεσμάτων παρουσιάζεται στην πρώτη γραμμή (με ετικέτα #Υπ) του πίνακα 5.8. Για κάθε τύπο υπομονάδας (δυναμική ή στατική) χρησιμοποιούμε διαφορετικό αριθμό υπομονάδων. Επιπλέον ανάλογα με το διάνυσμα χαρακτηριστικών, χρησιμοποιούμε και διαφορετικό αριθμό υπομονάδων. Το παραπάνω αποτελεί ένα πλεονέκτημα της προτεινόμενης μεθόδου SU-DSC. Έχουμε την δυνατότητα να χρησιμοποιούμε διαφορετικό αριθμό υπομονάδων ανάλογα με το διάνυσμα χαρακτηριστικών και τον τύπο της υπομονάδας (δυναμική ή στατική). Στα πειραματικά αποτελέσματα που ακολουθούν πρέπει να προσμετρήσουμε δύο αντικρουόμενες τάσεις. Από την μία, όσο μικρότερο αριθμό υπομονάδων χρησιμοποιούμε τόσο πιο εύκολο γίνεται το πρόβλημα της ταξινόμησης κάθε τμήματος στην σωστή υπομονάδα με αποτέλεσμα να αυξάνουν τα αποτελέσματα ταξινόμησης. Ωστόσο ο συντελεστής διακρίσιμότητας των νοημάτων μειώνεται δραστικά κάνοντας έντονη την εμφάνιση του φαινομένου της αντιστοίχισης μιας ακολουθίας σε πολλαπλά νοήματα (εν. 3.4.4). Από την άλλη, όσο μεγαλύτερο αριθμό υπομονάδων χρησιμοποιούμε ο συντελεστής διακρίσιμότητας αυξάνει. Ωστόσο το πρόβλημα της ταξινόμησης των τμημάτων στις σωστές υπομονάδες δυσκολεύει με αποτέλεσμα να μειώνεται το

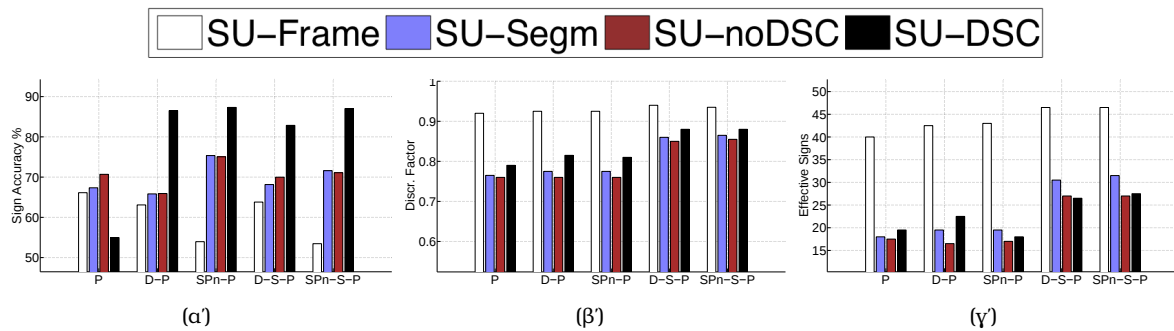


Σχήμα 5.6: Πειραματικά αποτελέσματα χρησιμοποιώντας 94 διαφορετικά νοήματα μεταβάλλοντας τον αριθμό των υπομονάδων που χρησιμοποιήθηκαν

αποτέλεσμα ταξινόμησης. Στην συνέχεια παρουσιάζουμε αποτελέσματα μεταβάλλοντας των αριθμό των υπομονάδων που χρησιμοποιήθηκαν. Οι παράμετρος αυτή επηρεάζει δραστικά τα αποτελέσματα αναγνώρισης και τον συντελεστή διακριτότητας. Για την ανάδειξη της παραπάνω επιρροής υλοποιήσαμε πειράματα μεταβάλλοντας ταυτόχρονα των αριθμό συστάδων που χρησιμοποιήθηκαν και στις δυναμικές υπομονάδες αλλά και στις στατικές όπως φαίνεται στο Σχήμα 5.6. Παρατηρούμε ότι όταν χρησιμοποιούμε μια συστάδα και στις στατικές αλλά και δυναμικές υπομονάδες το ποσοστό αναγνώρισης είναι 100%. Δεδομένο ότι έχουμε να επιλέξουμε ανάμεσα σε μία στατική και μία δυναμική υπομονάδα όλα τα τμήματα θα τα ταξινομηθούν στην σωστή υπομονάδα. Ωστόσο λόγω της χρήσης μίας υπομονάδας για τις στατικές και δυναμικές υπομονάδες η διάκριση μεταξύ διαφορετικών νοημάτων δεν είναι εφικτή. Όλα τα νοήματα αντιστοιχούν σε μία ακολουθία υπομονάδων με αποτέλεσμα την δραστική μείωση του συντελεστή διακριτότητα νοημάτων Σχήμα 5.6(β'). Επιπλέον παρατηρούμε ότι όταν ο αριθμός των συστάδων είναι πολύ μεγάλος (81 και 100 συστάδες για τις δυναμικές και στατικές υπομονάδες αντίστοιχα) ο συντελεστή διακριτότητας νοημάτων αυξάνει Σχήμα 5.6(β'). Ωστόσο ο χώρος των χαρακτηριστικών κατακερματίζεται, τα μοντέλα μας υπέρ-εκπαιδεύονται με αποτέλεσμα να μην έχουν την δυνατότητα γενίκευσης και έτσι το ποσοστό ταξινόμησης μειώνεται. Από τα παραπάνω συμπεραίνουμε ότι πρέπει να προσμετρήσουμε και τις δύο αντικρουόμενες τάσεις που σχετίζονται με την γενίκευση από τη μία και διακριτότητα των μοντέλων από την άλλη. Στα αποτελέσματα που ακολουθούν επιλέγουμε αριθμό υπομονάδων που βρίσκεται σε ένα μέσο εύρος τιμών. Πιο συγκεκριμένα για τις δυναμικές υπομονάδες επιλέγουμε 24 συστάδες, αριθμός που αποτελεί μια πρακτική μέση λύση για την προσμέτρηση των δύο αντικρουόμενων τάσεων. Επιπλέον παρόλο που η επιλογή του αριθμού αυτού δεν βασίζεται σε άμεση γλωσσολογική γνώση αντιστοιχεί έμμεσα σε ένα κβαντισμό του αριθμού των κινήσεων: 8 υπομονάδες για κάθε είδος κινήσεων (ευθείες, καμπύλες και περίπλοκες κινήσεις). Επιπλέον χρησιμοποιούμε 4 συστάδες για την μοντελοποίηση του χώρου της κλίμακας (μικρές, μεσαίες, μεγάλες και πολύ μεγάλες κινήσεις). Τέλος χρησιμοποιούμε 22 συστάδες για τις στατικές υπομονάδες οι οποίες αντιστοιχούν σε τυχαίες θέσης άρθρωσης σχετιζόμενες με το αποτέλεσμα του αλγόριθμο συσταδοποίησης. Οι αριθμοί των παραπάνω συστάδων χρησιμοποιήθηκαν στα πειραματικά αποτελέσματα που απεικονίζονται στον πίνακα 5.9. Για παράδειγμα στα πειράματα που αντιστοιχούν στην πρώτη στήλη του πίνακα 5.9 χρησιμοποιήσαμε 46 υπομονάδες. Σε αντίθεση με τα πειράματα που αντιστοιχούν στην δεύτερη στήλη του πίνακα 5.9 όπου χρησιμοποιήσαμε 118 υπομονάδες. Σε πειράματα όπου συγκρίνουμε διαφορετικές μεθόδους χρησιμοποιούμε τον ίδιο αριθμό υπομονάδων.

Πίνακας 5.9: Αντιστοίχιση του διανύσματος χαρακτηριστικών που χρησιμοποιήθηκε με τον αριθμό υπομονάδων που επιλέχθηκε.

Διάνυσμα Χαρακτηριστικών και Συνδυασμοί	P	D-S-P
	SPn-P	SPn-S-P
	D-P	
# Υπομονάδες	46	118



Σχήμα 5.7: Πειραματικά αποτελέσματα χρησιμοποιώντας 94 διαφορετικά νοήματα μεταβάλλοντας τις μεθόδους και τα διανύσματα χαρακτηριστικών που χρησιμοποιήθηκαν.

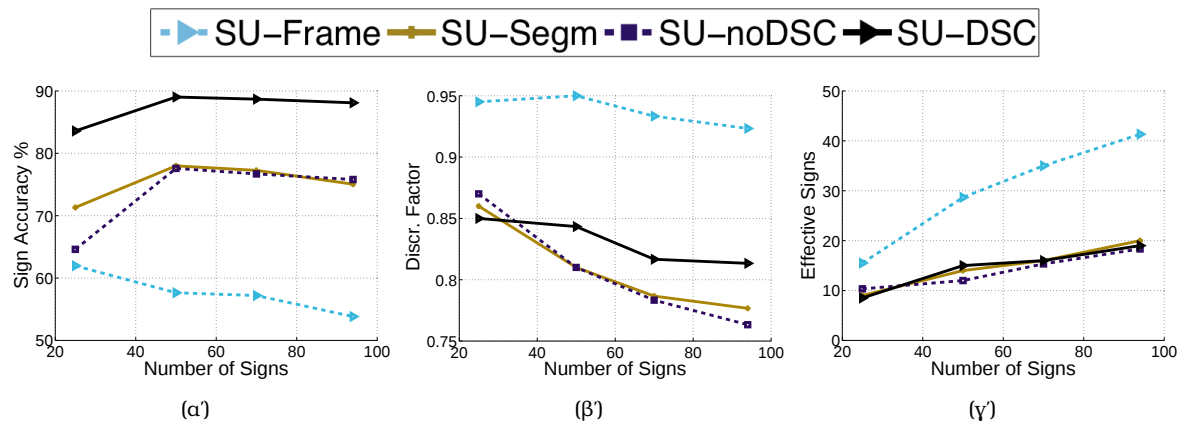
Σύγκριση με άλλες μεθόδους:

Συγκρίνουμε την προτεινόμενη μέθοδο SU-DSC με άλλες μεθόδους που έχουν εμφανιστεί στην διεθνή βιβλιογραφία. Αυτές συνοψίζονται στις εξής:

- 1) Η μέθοδος SU-noDSC αποτελεί μια απλοποιημένη εκδοχή της SU-DSC. Χρησιμοποιεί την ίδια τοπολογία HMM για την κατάτμηση (εργοδικό HMM 2 καταστάσεων -2S-ERG-). Ωστόσο δεν διαχωρίζει τα τμήματα σε δυναμικά και στατικά όπως η SU-DSC. Με αυτόν το τρόπο αξιολογούμε την συνεισφορά του διαχωρισμού μεταξύ στατικών και δυναμικών τμημάτων.
- 2) Η μέθοδος SU-Segm [28] χρησιμοποιεί μια διαφορετική τοπολογία HMM για την κατάτμηση (left-right HMM 3 καταστάσεων -3S-LR-). Επιπλέον δεν διαχωρίζει τα τμήματα σε δυναμικά και στατικά αντίστοιχα όπως η SU-noDSC.
- 3) Η μέθοδος SU-Frame [6] δεν χρησιμοποιεί αλγόριθμο κατάτμησης αλλά βασίζεται στην συσταδοποίηση μεμονωμένων πλαισίων και όχι τμημάτων όπως όλες οι προηγούμενες μέθοδοι. Για την κατασκευή των υπομονάδων εφαρμόζεται ο αλγόριθμός K-means σε μεμονωμένα πλαίσια. Κάθε συστάδα που δημιουργείται αντιστοιχίζεται σε μια υπομονάδα.

Οι τρεις παραπάνω μέθοδοι προσαρμόστηκαν έτσι ώστε να ταιριάζουν στο παρόν πειραματικό πλαίσιο με στόχο να είναι δίκαιη η σύγκρισή τους με την προτεινόμενη μέθοδο. Συγκρίνοντας την μέθοδο SU-DSC με τις υπόλοιπες παρατηρούμε τα παρακάτω. Πρώτον η μέθοδος SU-DSC υπερτερεί των υπόλοιπων στα ποσοστά αναγνώρισης σε επίπεδο νοήματος Σχήμα 5.7(α'). Παρατηρώντας τον συντελεστή διάκρισης νοημάτων βλέπουμε ότι η μέθοδος SU-Frame επιτυγχάνει μεγαλύτερη διακριτοποίηση νοημάτων σε σχέση με τις υπόλοιπες μεθόδους. Ωστόσο αυτό οφείλεται στον κατακερματισμό που υφίσταται κάθε νόημα λόγω της συσταδοποίησης σε επίπεδο πλαισίου. Όμως τα μοντέλα των υπομονάδων που κατασκευάζονται είναι υπερ-εκπαιδευμένα με αποτέλεσμα το ποσοστό αναγνώρισης νοημάτων μειώνεται δραστικά. Σε αντίθεση εφαρμόζοντας ένα αλγόριθμο

κατάτμηση πρώτα και στην συνέχεια συσταδοποίηση σε επίπεδο τμημάτων τα φαινόμενα του κατακερματισμού και της υπερ-εκπαίδευσης των μοντέλων δεν εμφανίζονται με αποτέλεσμα το ποσοστό αναγνώρισης νοημάτων αυξάνει. Τέλος συγκρίνοντας την μέθοδο SU-DSC με την SU-noDSC παρατηρούμε ότι το ποσοστό αναγνώρισης νοημάτων αλλά και ο συντελεστής διακριτοποίησης τους αυξάνουν. Έτσι συμπεραίνουμε ότι η διάκριση σε δυναμικά και στατικά τμήματα και η χρησιμοποίηση διαφορετικού είδους μοντελοποίηση επιδρά θετικά στην αναγνώριση νοημάτων. Μια σύνοψη των αποτελεσμάτων απεικονίζεται στην δεύτερη σειρά (ετικέτα Χαρ.) του πίνακα 5.8.



Σχήμα 5.8: Πειραματικά αποτελέσματα μεταβάλλοντας τον αριθμό νοημάτων στο λεξικό.

Μεταβολή του αριθμού νοημάτων: Στην συνέχεια συγκρίνουμε την προτεινόμενη μέθοδο SU-DSC με τις υπόλοιπες μεταβάλλοντας τον αριθμό των νοημάτων στο λεξικό. Τα πειράματα έχουν πραγματοποιηθεί χρησιμοποιώντας το διάνυσμα χαρακτηριστικών SPn-P (βλέπε Σχήμα 5.8). Επιπλέον έχουμε χρησιμοποιήσει τον ίδιο αριθμό υπομονάδων για όλες τις διαφορετικές μεθόδους. Παρατηρούμε ότι η μέθοδος SU-DSC υπερτερεί όλων των υπόλοιπων μεθόδων σε ποσοστό αναγνώρισης νοημάτων για όλες τις περιπτώσεις. Ακόμα βλέπουμε ότι το ποσοστό αναγνώρισης νοημάτων για την SU-DSC μέθοδο παραμένει ανεπηρέαστο από την αύξηση των νοημάτων στο λεξικό. Χρησιμοποιώντας την μέθοδο SU-Frame βλέπουμε ότι παρότι ο συντελεστής διάκρισης νοημάτων είναι μεγαλύτερος από τις άλλες μεθόδους Σχήμα 5.8(β') τα αποτελέσματα αναγνώρισης νοημάτων είναι χαμηλότερα, ιδίως όσο το μέγεθος του λεξιλογίου αυξάνει. Αυτό οφείλεται στον κατακερματισμό και την υπερ-εκπαίδευση των μοντέλων, φαινόμενο που δεν υφίστανται οι υπόλοιπες μέθοδοι. Παίρνοντας τον μέσο όρο αποτελεσμάτων μεταβάλλοντας τον αριθμό των νοημάτων στο λεξικό (τρίτη γραμμή με ετικέτα #N στον πίνακα 5.8) η μέθοδος SU-DSC επιτυγχάνει ποσοστό αναγνώρισης νοημάτων 14%, 12.2% και 30.4% μεγαλύτερο από τις μεθόδους SU-noDSC, SU-Segm και SU-Frame αντίστοιχα. Μια σύνοψη των αποτελεσμάτων απεικονίζεται στην τρίτη γραμμή (ετικέτα #N) στον πίνακα 5.8.

Χαρακτηριστικά διανύσματα και Συνδυασμοί: Στην συνέχεια αξιολογούμε τα διανύσματα χαρακτηριστικών και τους συνδυασμούς τους. Αρχικά παρατηρούμε ότι η κανονικοποίηση της τροχιάς της κίνησης σε σχέση με την αρχική θέση και την κλίμακα (SPn-P vs. P διανύσματα) επηρεάζει θετικά τα ποσοστά αναγνώρισης νοημάτων. Χρησιμοποιώντας τα διανύσματα χαρακτηριστικών SPn-P και D-P με την μέθοδο SU-DSC επιτυγχάνουμε μεγαλύτερα ποσοστά αναγνώρισης από ότι με το διάνυσμα χαρακτηριστικών P (βλ. Σχήμα 5.7(α')). Αυτό επιβεβαιώνει την υπόθεσή μας ότι η προτεινόμενη μέθοδος δεν ενδείκνυται στην χρησιμοποίηση του διανύσματος χαρακτηριστικών P στις δυναμικές υπομονάδες. Συγκρίνοντας την SU-DSC με την SU-noDSC μέθοδο χρησιμοποιώντας τα διανύσματα χαρακτηριστικών SPn-P ή D-P παρατηρούμε ότι η SU-DSC μέθοδος υπερτερεί αρκετά και στο ποσοστό αναγνώρισης νοημάτων αλλά και στον συντελεστή διάκρισης νοημάτων Σχήματα 5.7(α'), 5.7(β'). Το παραπάνω υποδεικνύει ότι ο διαχωρισμός σε δυναμικά και

στατικά τμήματα και η χρησιμοποίηση των διανυσμάτων χαρακτηριστικών SPn και D *μόνο* στα δυναμικά τμήματα και του P *μόνο* στα στατικά τμήματα αποτελεί κρίσιμο χαρακτηριστικό για την απόδοση του συστήματος αναγνώρισης. Παίρνοντας το μέσο όρο των αποτελεσμάτων μεταβάλλοντας τα διανύσματα χαρακτηριστικών που χρησιμοποιήθηκαν (τρίτη σειρά με ετικέτα Χαρ. του πίνακα 5.8) παρατηρούμε την αύξηση του ποσοστού αναγνώρισης νοημάτων χρησιμοποιώντας την μέθοδο SU-DSC κατά 11% συγκρινόμενη με την SU-noDSC και κατά 21% σε σχέση με την SU-Frame. Επιπλέον χρησιμοποιώντας τα διανύσματα χαρακτηριστικών 'SPn-P' ή 'D-P' τα αποτελέσματα είναι αντίστοιχα. Το παραπάνω ήταν αναμενόμενο αφού το διάνυσμα χαρακτηριστικών SPn ενσωματώνει την πληροφορία της κατεύθυνσης της κίνησης (βλέπε Σχήμα 3.9(β)). Τέλος η ενσωμάτωση πολλαπλών ροών πληροφορίας στις δυναμικές υπομονάδες κάνοντας χρήση της μεθόδου SU-DSC π.χ. 'SPn-S-P' ή 'D-S-P' συγκρινόμενη με τα διανύσματα χαρακτηριστικών 'SPn-P' ή 'D-P' αντίστοιχα βλέπουμε μια μικρή μείωση του ποσοστού αναγνώρισης νοημάτων. Ωστόσο ο συντελεστή διακριτότητας και ο αριθμός των ενεργών νοημάτων αυξάνουν (βλέπε Σχήματα 5.7(β), 5.7(γ)) αφού αποκτάμε την δυνατότητα να διαχωρίζουμε τις κινήσεις και με βάση της κλίμακας τους και άρα διάκριση μεταξύ περισσότερων νοημάτων.

5.2.2 Φωνητικές Υπομονάδες

Σύνοψη Συστήματος Αναγνώρισης ΝΓ με την χρήση PDTS φωνητικών υπομονάδων

Παρακάτω κάνουμε μια σύνοψη του συστήματος αναγνώρισης ΝΓ χρησιμοποιώντας PDTS φωνητικές υπομονάδες η οποία περιγράφηκε εκτενώς στο κεφάλαιο 4. Ως είσοδο έχουμε τις μετρήσεις από το σύστημα εξαγωγής χαρακτηριστικών (κεφ. 2) και ένα φωνητικό λεξικό σε επίπεδο HamNoSys.

1. Μετατροπή των επισημειώσεων επιπέδου HamNoSys σε δομημένα PDTS φωνητικά σύμβολα χρησιμοποιώντας την ερευνητική εργασία του C.Vogler [78].
2. Εκπαίδευση υπολογιστικών φωνητικών υπομονάδων κάνοντας χρήση Κρυφών Μαρκοβιανών Μοντέλων.
3. Αντιστοίχιση των φωνητικών συμβόλων με τα διανύσματα χαρακτηριστικών.
4. Εξαγωγή PDTS φωνητικών συμβόλων κατά την διάρκεια της αναγνώρισης.
5. Αντιστοίχιση των PDTS φωνητικών ακολουθιών με τα νοήματα βάση του PDTS λεξικού για την αναγνώριση σε επίπεδο νοήματος.

Δεδομένα Πειραματισμού

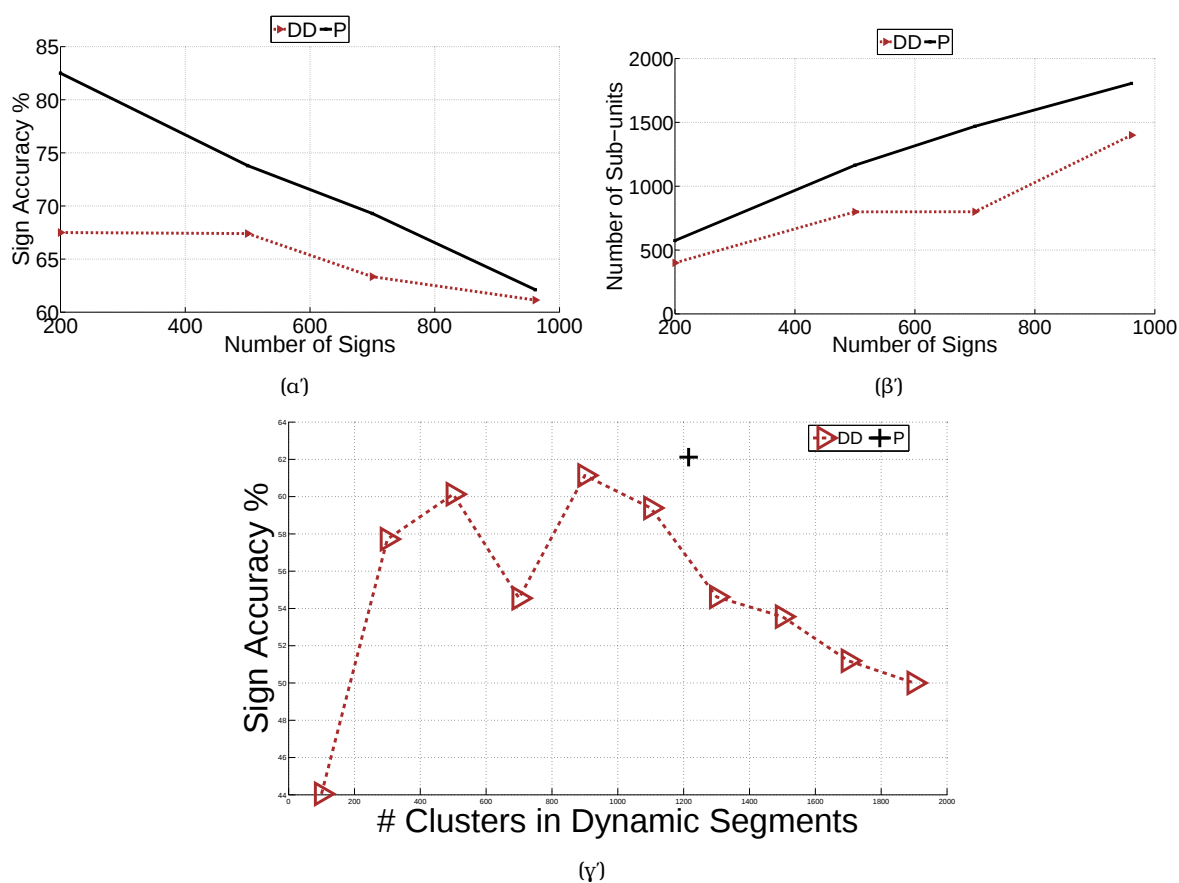
Η βάση δεδομένων που χρησιμοποιήθηκε είναι η Dicta-Sign ENΓ Λείμματα η οποία αποτελείται από την εκτέλεση 1046 μεμονωμένων νοημάτων της ENΓ πέντε φορές από δύο νοηματιστές. Στα πειραματικά αποτελέσματα που θα παρουσιάσουμε χρησιμοποιήσαμε 961 νοήματα από τα 1046 από ένα νοηματιστή. Περίπου μισά από τα 85 νοήματα τα οποία απορρίφθηκαν αντιστοιχούσαν στην ίδια ακολουθία από PDTS σύμβολα με άλλα νοήματα και τα υπόλοιπα απορρίφθηκαν λόγω λαθών κατά την διαδικασία εξαγωγής των διανυσμάτων χαρακτηριστικών ή λάθη στην επισημείωση των σημείων κατάτμησης των διαφορετικών επαναλήψεων. Για την εκπαίδευση χρησιμοποιήσαμε τέσσερις επαναλήψεις για κάθε νόημα ενώ για την αξιολόγηση χρησιμοποιήσαμε μία.

Πειραματικά Αποτελέσματα

Αρχικά συγκρίνουμε τις δύο μεθόδους κατασκευής υπομονάδων για την αναγνώριση νοηματικής γλώσσας που περιγράφηκαν στα κεφάλαια 3,4.

1. Δεδομενοκεντρικές υπομονάδες -Data-driven Subunits (SU-DD)-: Η κατασκευή τους βασίζεται στην μέθοδο DSCM και δεν περιλαμβάνει γλωσσική πληροφορία ή φωνητικές επισημειώσεις.
2. Φωνητικές υπομονάδες -Phonetic Subunits (SU-P)-: Βασίζονται σε πρότερη γλωσσική/φωνητική πληροφορία και σε επισημειώσεις σε επίπεδο HamNoSys.

Επιπλέον εξετάζουμε την επιρροή του μεγέθους του λεξιλογίου στο ποσοστό αναγνώρισης και για τις δύο παραπάνω μεθόδους. Ακόμα για την SU-DD μέθοδο εξετάζουμε το ποσοστό αναγνώρισης μεταβάλλοντας τον αριθμό των δυναμικών και στατικών υπομονάδων. Ο αριθμός των υπομονάδων για την μέθοδο SU-P παραμένει σταθερό και εξαρτάται από τα PDS σύμβολα.



Σχήμα 5.9: Σύγκριση των μεθόδων SU-DD και SU-P. (α) Αναγνώριση Νοημάτων (β) Αριθμός υπομονάδων. Στα σχήματα (α,β) ο άξονας x αντιστοιχεί στην μεταβολή του αριθμού των νοημάτων που χρησιμοποιήθηκαν. Στο σχήμα (γ) ο αριθμός των νοημάτων είναι 961. Για την SU-DD μέθοδο απεικονίζουμε την επίδραση της μεταβολής του αριθμού των δυναμικών υπομονάδων (x άξονας). Ο αριθμός των στατικών υπομονάδων είναι σταθερός και ίσος με 500. Σε αντίθεση ο αριθμός των υπομονάδων για την SU-P μέθοδο είναι προκαθορισμένος και σταθερός.

Στο Σχήμα 5.9(α') συγκρίνουμε τις μεθόδους SU-DD και SU-P μεταβάλλοντας τον αριθμό των νοημάτων στο λεξικό. Για την SU-DD μέθοδο χρησιμοποιούμε τον αριθμό των υπομονάδων που

αντιστοιχεί στο καλύτερο αποτέλεσμα. Παρατηρούμε ότι αυξάνοντας τον αριθμό των νοημάτων τα αποτελέσματα και των δύο μεθόδων μειώνονται. Το παραπάνω είναι αναμενόμενο αφού το πρόβλημα αναγνώρισης γίνεται δυσκολότερο. Επιπλέον βλέπουμε ότι η SU-P μέθοδος υπερτερεί της SU-DD σε όλες τις περιπτώσεις. Ωστόσο παρατηρούμε ότι αυξάνοντας ο αριθμός των νοημάτων επηρεάζει δραστικά την SU-P σε αντίθεση με την SU-DD μέθοδο η οποία επηρεάζεται λιγότερο. Το παραπάνω οφείλεται στο ότι όσο αυξάνουμε τον αριθμό των νοημάτων τα PDTS σύμβολα αυξάνουν και άρα οι φωνητικές υπομονάδες (βλέπε Σχήμα 5.9(β')) με αποτέλεσμα το πρόβλημα της αναγνώρισης δυσκολεύει. Σε αντίθεση η SU-DD μέθοδος όσο ο αριθμός των νοημάτων αυξάνει περισσότερα δεδομένα εισάγονται στο σύστημα και έτσι επωφελείται ο αλγόριθμος συσταδοποίησης και άρα η κατασκευή των υπομονάδων. Μια άλλη οπτική γωνία παρουσιάζεται στο Σχήμα 5.9(γ') όπου απεικονίζονται τα αποτελέσματα αναγνώρισης για 961 νοήματα μεταβάλλοντας τον αριθμό των δυναμικών υπομονάδων για την SU-DD μέθοδο. Ο αριθμός των στατικών υπομονάδων παραμένει σταθερός και ίσος με 500. Σε αντίθεση ο αριθμός των υπομονάδων για την SU-P μέθοδο εξαρτάται από τον αριθμό των PDTS συμβόλων και άρα είναι προκαθορισμένος και σταθερός. Παρατηρούμε ότι η μέθοδος SU-P υπερτερεί της μεθόδου SU-DD για όλες τις τιμές του αριθμού των δυναμικών υπομονάδων. Από τα παραπάνω συμπεραίνουμε ότι χρησιμοποιώντας την μέθοδο SU-P σε πειράματα αναγνώρισης 961 μεμονωμένων νοημάτων οδηγούμαστε σε μια αύξηση 2% το λιγότερο και 7% κατά μέσο όρο σε σχέση με την SU-DD μέθοδο.

Κεφάλαιο 6

Σύνοψη Προόδου και Μελλοντική Έρευνα

6.1 Ερευνητική Συνεισφορά

Η διδακτορική μας έρευνα κινείται στις περιοχές της Αναγνώρισης Προτύπων και της Όρασης Υπολογιστών. Η παρούσα έρευνα επικεντρώνεται στην αυτόματη επεξεργασία βίντεο συνεχούς νοηματικής γλώσσας, στην εξαγωγή χαρακτηριστικών, στην μοντελοποίηση και τελικά στην αναγνώριση συνεχούς νοηματικής γλώσσας. Σε αυτά τα πλαίσια αναπτύσσονται μέθοδοι για την οπτική επεξεργασία και εξαγωγή χαρακτηριστικών από βίντεο συνεχούς νοηματικής γλώσσας σχετιζόμενα με τους αρθρωτές όπως τα χέρια και το κεφάλι. Επιπλέον αναπτύσσονται στατιστικές μέθοδοι για την μοντελοποίηση και αναγνώριση της νοηματικής γλώσσας. Οι συγκεκριμένες ερευνητικές μας συνεισφορές μπορούν να συνοψισθούν στα ακόλουθα σημεία :

- Αναπτύξαμε ένα σύστημα εξαγωγής χαρακτηριστικών σχετιζόμενα με τους αρθρωτές που συμμετέχουν κατά την διάρκεια άρθρωσης της νοηματικής γλώσσας. Βασιστήκαμε στην ανίχνευση περιοχών χρώματος δέρματος κάνοντας χρήση ενός πιθανοτικού μοντέλου χρώματος. Στην μορφολογική επεξεργασία και κατάτμηση των εξαγόμενων μασκών δέρματος και συνδυάσαμε εμπρόσθια/οπίσθια γραμμική πρόβλεψη και template matching για την επίλυση επικαλύψεων.
- Αναπτύξαμε ένα σύστημα για την μοντελοποίηση και αναγνώριση ΝΓ με δεδομενοκεντρικές υπομονάδες. Βασιστήκαμε στην αυτόματη κατάτμηση και ταξινόμηση κάθε τμήματος σε δυναμικό ή στατικό μοντελοποιώντας το προφίλ της ταχύτητας. Επιπλέον εφαρμόσαμε αλγορίθμους συσταδοποίησης για τα στατικά και δυναμικά τμήματα κατασκευάζοντας τις στατικές και δυναμικές υπομονάδες. Τέλος συνδυάσαμε τις στατικές και δυναμικές υπομονάδες για την κατασκευή λεξικού σε επίπεδο υπομονάδας.
- Αναπτύξαμε ένα σύστημα για την μοντελοποίηση και αναγνώριση ΝΓ κάνοντας χρήση Γλωσσικής και Φωνητικής Πληροφορίας. Πιο συγκεκριμένα μετατρέψαμε τις HamNoSys επισημειώσεις σε PDTS φωνητικά σύμβολα χρησιμοποιώντας την προ-επεξεργασία από την εργασία του C.Vogler [78]. Στην συνέχεια εκπαιδεύσαμε υπολογιστικές φωνητικές υπομονάδες κάνοντας χρήση Κρυφών Μαρκοβιανών Μοντέλων και τέλος αντιστοιχήσαμε τα PDTS φωνητικά σύμβολα με τα διανύσματα χαρακτηριστικών.

Με την μέχρι τώρα ερευνά μας συνεισφέραμε στην περιοχή της αναγνώρισης νοηματικής γλώσσας. Οι επιδράσεις της έρευνας αυτής και των εφαρμογών της αναμένεται να έχουν διεπιστημονικό χαρακτήρα, όπως για παράδειγμα στη γλωσσολογία και ανάλυση των νοηματικών γλωσσών καθώς

και την αυτόματη επεξεργασία και επισημείωση μεγάλων σωμάτων βίντεο νοηματικών γλωσσών. Επιπλέον η εφαρμογή τους στην αναγνώριση χειρονομιών θα οδηγήσει στην εξέλιξη και βελτίωση της επικοινωνίας ανθρώπου-μηχανής κάνοντας την πιο φυσική και άμεση.

6.2 Κατευθύνσεις Μελλοντικής Έρευνας

Κάποιες σχετικές κατευθύνσεις μελλοντικής έρευνας, τις οποίες έχουμε ήδη ξεκινήσει και σχεδιάζουμε να προχωρήσουμε κατά το χρόνο που υπολείπεται για την ολοκλήρωση της διδακτορικής μας έρευνας, είναι οι εξής:

- Συνδυασμός των ροών πληροφορίας και από τα δύο χέρια του νοηματιστή. Κατασκευή στατιστικών μοντέλων σε επίπεδο υπομονάδας τα οποία θα μοντελοποιούν την πληροφορία θέσης/κίνησης και από τα δύο χέρια του νοηματιστή.
- Σύμμειξη των ροών πληροφορίας της χειρομορφής και της θέσης/κίνησης των χεριών. Στην μέχρι τώρα έρευνα μας εξετάσαμε την ροή πληροφορίας της θέσης/κίνησης των χεριών με στόχο την αναγνώριση νοηματικής γλώσσας. Επιπλέον διερευνήσαμε την ταξινόμηση διαφορετικών χειρομορφών χρησιμοποιώντας τα εξαγόμενα χαρακτηριστικά από το σύστημα Δυναμική Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος-Εμφάνισης. Στην συνέχεια της έρευνας μας θα προχωρήσουμε στην ενσωμάτωση της ροής πληροφορίας για την χειρομορφή στο συνολικό σύστημα αναγνώρισης νοηματικής γλώσσας.
- Αναγνώριση συνεχούς νοηματικής γλώσσας. Στην παρούσα εργασία ασχοληθήκαμε με την αναγνώριση μεμονωμένων νοημάτων. Στόχος για την συνέχεια της διδακτορικής μας διατριβής είναι η επέκταση του συστήματος για την αναγνώριση συνεχούς νοηματικής γλώσσας.
- Αναγνώριση νοηματικής γλώσσας από πολλαπλούς νοηματιστές. Μέχρι τώρα ασχοληθήκαμε με την κατασκευή και εκπαίδευση στατιστικών μοντέλων προσαρμοσμένα σε ένα νοηματιστή. Στην συνέχεια της έρευνας μας θα προχωρήσουμε στην επέκταση των παραπάνω στατιστικών μοντέλων με στόχο την δυνατότητα γενίκευσης και άρα την αναγνώρισης νοηματικής γλώσσας από πολλαπλούς νοηματιστές.

Βιβλιογραφία

- [1] U. Agris, J. Zieren, U. Canzler, B. Bauer, and K. F. Kraiss. Recent developments in visual sign language recognition. *Universal Access in the Information Society*, 6:323–362, 2008.
- [2] A. Argyros and M. Lourakis. Real time tracking of multiple skin-colored objects with a possibly moving camera. In *Proc. Europ. Conf. on Computer Vision*, 2004.
- [3] V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, A. Stefan, A. Thangali, H Wang, and Q Yuan. Large lexicon project: American sign language video corpus and sign language indexing/retrieval algorithms. In *Proc. LREC Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 2010.
- [4] B. Bauer and H. Hienz. Relevant features for video-based continuous sign language recognition. In *Proc. Int'l Conf. on Automatic Face & Gesture Recognition*, pages 440–445. IEEE, 2000.
- [5] B. Bauer, H. Hienz, and K.-L. Kraiss. Video-based continuous sign language recognition using statistical methods. *icpr*, 2000.
- [6] B. Bauer and K. F. Kraiss. Towards an automatic sign language recognition system using subunits. In *Proc. of Int'l Gesture Workshop*, pages 64–75, 2001.
- [7] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 259–302, 1986.
- [8] H. Birk, T.B. Moeslund, and C.B. Madsen. Real-time recognition of hand alphabet gestures using principal component analysis. In *Proc. SCIA*, 1997.
- [9] R. Bowden and M. Sarhadi. A nonlinear model of shape and motion for tracking fingerspelt american sign language. *Image and Vision Computing*, 20:597–607, 2002.
- [10] R. Bowden, D. Windridge, T. Kadir, A. Zisserman, and M. Brady. A linguistic feature vector for the visual interpretation of sign language. In *Proc. Europ. Conf. on Computer Vision*, 2004.
- [11] P. Buehler, M. Everingham, and A. Zisserman. Learning sign language by watching TV (using weakly aligned subtitles). In *Proc. Conf. on Computer Vision & Pattern Recognition*, pages 2961–2968, June 2009.
- [12] P. Buehler, M. Everingham, and A Zisserman. Exploiting signed tv broadcasts for automatic learning of british sign language. In *Proc. LREC Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 2010.
- [13] J. Cai and A. Goshtasby. Detecting human faces in color images. *Image and Vision Computing*, 18:63–75, 1999.

- [14] F.-S. Chen, C.-M. Fu, and C.-L. Huang. Hand gesture recognition using a real-time tracking method and hidden markov models. *Image and Visual Computing*, 21(8):745–758, 2003.
- [15] S. Conseil, S. Bourennane, and L. Martin. Comparison of Fourier descriptors and Hu moments for hand posture recognition. In *EUSIPCO*, 2007.
- [16] H. Cooper and R. Bowden. Sign language recognition using linguistically derived sub-units. In *Proc. LREC Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 57–61, 2010.
- [17] T.F. Cootes, C.J. Taylor, D. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995.
- [18] D. Corina and W. Sandler. On the nature of phonological structure in sign language. *Phonology*, 10:165–207, 2008.
- [19] Y. Cui and J. Weng. Appearance-based hand sign recognition from intensity image sequences. *Computer Vision and Image Understanding*, 78(2):157–176, 2000.
- [20] N. Dalal and B. Triggs. Histogram of oriented gradients for human detection. In *Proc. Conf. on Computer Vision & Pattern Recognition*, 2005.
- [21] L. Davies, David and W. Bouldin, Donald. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1:224 – 227, April 1979.
- [22] J.-W. Deng and H.T. Tsui. A novel two-layer PCA/MDA scheme for hand posture recognition. In *International Conference on Pattern Recognition*, volume 1, pages 283–286, 2002.
- [23] L. Ding and A. Martinez. Modelling and recognition of the linguistic components in american sign language. *Image and Visual Computing*, 27(12):1826 – 1844, 2009.
- [24] A.C. Downton and H. Drouet. Model-based image analysis for unconstrained human upper-body motion. In *International Conference on Image Processing and Its Applications*, pages 274–277, Apr. 1992.
- [25] P. Dreuw, C. Neidle, V. Athitsos, S. Sclaroff, and H. Ney. Benchmark databases for video-based automatic sign language recognition. In *Proc. LREC*, May 2008.
- [26] K. Emmorey. *Language, cognition, and the brain: insights from sign language research*. Erlbaum, 2002.
- [27] G. Fang, W. Gao, and D. Zhao. Large-vocabulary continuous sign language recognition based on transition-movement models. *IEEE Trans. on Systems, Man and Cybernetics*, 37(1):1–9, 2007.
- [28] G. Fang, X. Gao, W. Gao, and Y. Chen. A novel approach to automatically extracting basic units from chinese sign language. In *Proc. Int’l Conf. on Pattern Recognition*, 2004.
- [29] A. Farhadi and D. Forsyth. Aligning ASL for Statistical Translation Using a Discriminative Word Model. In *Proc. Conf. on Computer Vision & Pattern Recognition*, pages 1471–1476. IEEE Computer Society, 2006.
- [30] H. Fillbrandt, S. Akyol, and K.-F. Kraiss. Extraction of 3D hand shape and posture from images sequences from sign language recognition. In *Proc. Int’l Workshop on Analysis and Modeling of Faces and Gestures*, pages 181–186, 2003.

- [31] W. Gao, G. Fang, D Zhao, and Y. Chen. Transition movement models for large vocabulary continuous sign language recognition. In *Proc. Int'l Conf. on Automatic Face & Gesture Recognition*, volume 0, page 553, CA, USA, 2004.
- [32] K. Grobel and M. Assan. Isolated sign language recognition using hidden markov models. In *IEEE Trans. on Systems, Man and Cybernetics*, volume 1, pages 162–167. IEEE, 1997.
- [33] J. Han, G. Awad, and A. Sutherland. Modelling and segmenting subunits for sign language recognition based on hand motion analysis. *Pat. Rec. Lett.*, 30(6):623–633, 2009.
- [34] T. Hanke. HamNoSys Representing sign language data in language resources and language processing contexts. In *Workshop on the Representation and Processing of Sign Languages on the occasion of the Fourth International Conference on Language Resources and Evaluation, Lisbon, Portugal*, 2004.
- [35] H.J.A.M. Heijmans. *Morphological Image Operators*. Acad. Press, Boston, 1994.
- [36] K. Hienz, H. Grobel and G. Offner. Real-time hand-arm motion analysis using a single video camera. In *International Conference on Automatic Face and Gesture Recognition*, pages 323–327, 1996.
- [37] E.-J. Holden and R. Owens. Visual sign language recognition. In *Proc. Int'l Workshop on Theoretical Foundations of Computer Vision*, pages 270–287, 2000.
- [38] Eun-Jung Holden, Gareth Lee, and Robyn Owens. Australian sign language recognition. *Machine Vision and Application*, 16(5):312–320, 2005.
- [39] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179–187, February 1962.
- [40] C.-L. Huang and S.-H. Jeng. A model-based hand gesture recognition system. *Machine Vision and Application*, 12(5):243–258, 2001.
- [41] R. E. Johnson and S. K. Liddell. A segmental framework for representing signs phonetically. *sign language studies*. 11(3), 2011.
- [42] R. E. Johnson and S. K. Liddell. Toward a phonetic representation of signs, i: Sequentiality and contrast. *sign language studies*. 11(2), 2011a.
- [43] B. H. Juang and L. R. Rabiner. The segmental k-means algorithm for estimating parameters of hidden markov models. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 38(9):1639–1641, Sep 1990.
- [44] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3):1106–1122, Mar. 2007.
- [45] D. Kelly, J. Mc Donald, and C. Markham. Weakly supervised training of a sign language recognition system using multiple instance learning density matrices. *IEEE Trans. on Systems, Man and Cybernetics*, (99):1–16, 2010.
- [46] W. Kong and S. Ranganath. Sign language phoneme transcription with rule-based hand trajectory segmentation. *Journal Signal Processing Systems*, 59:211–222, 2010.
- [47] R.H. Liang and M. Ouhyoung. A real-time continuous gesture recognition system for sign language. In *Proc. Int'l Conf. on Automatic Face & Gesture Recognition*, pages 558–567. IEEE, 1998.

- [48] S. K. Liddell and R. E. Johnson. American sign language: The phonological base. *Sign Language Studies*, 64:195 – 277, 1989.
- [49] Scott K. Liddell. Think and believe: Sequentiality in american sign language. *Language*, 60(2):372–399, 1984.
- [50] RP. Lippmann. Speech recognition by machines and humans. *Speech Communication*, 1997.
- [51] S. Liwicki and M. Everingham. Automatic recognition of fingerspelled words in British sign language. In *Proc. of CVPR4HB*, 2009.
- [52] P. Maragos. *The Image and Video Processing Handbook*, chapter Morphological Filtering for Image Enhancement and Feature Detection. Elsevier, 2nd edition, 2005.
- [53] D. McNeill. Hand and mind: what gestures reveal about thought. *University of Chicago Press*, 1992.
- [54] S. Nayak, S. Sarkar, and B. Loeding. Unsupervised modeling of signs embedded in continuous sentences. *Proc. Conf. on Computer Vision & Pattern Recognition*, 0:81, 2005.
- [55] S. Nayak, S. Sarkar, and B. Loeding. Automated extraction of signs from continuous sign language sentences using iterated conditional modes. *Proc. Conf. on Computer Vision & Pattern Recognition*, 0:2583–2590, 2009.
- [56] C. Neidle. *SignStream Annotation: Addendum to Conventions used for the American Sign Language Linguistic Research Project Report No. 13 American Sign Language Linguistic Research Project*. 2007.
- [57] E.-J. Ong and R. Bowden. A boosted classifier tree for hand shape detection. In *International Conference on Automatic Face and Gesture Recognition*, pages 889–894, 2004.
- [58] S. Ong and S. Ranganath. Automatic sign language analysis: A survey and the future beyond lexical meaning. *ieeetpami*, 27(6):873–891, 2005.
- [59] S. C. W. Ong and S. Ranganath. A new probabilistic model for recognizing signs with systematic modulations. In *Proceedings of the 3rd international conference on Analysis and modeling of faces and gestures*, pages 16–30, 2007.
- [60] S.C.W. Ong, S. Ranganath, and Y.V. Venkatesh. Understanding gestures with systematic variations in movement dynamics. *Pattern recognition*, 39(9):1633–1648, 2006.
- [61] E. Orfanidou, R. Adam, J.M. McQueen, and G. Morgan. Making sense of nonsense in British Sign Language (BSL): The contribution of different phonological parameters to sign recognition. *Memory & cognition*, 37:302, 2009.
- [62] V. Pitsikalis, S. Theodorakis, and P. Maragos. Data-driven sub-units and modeling structure for continuous sign language recognition with multiple cues. In *Proc. LREC Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 2010.
- [63] S. Prillwitz, R. Leven, H. Zienert, R. Zienert, T. Hanke, and J. Henning. HamNoSys. Version 2.0. *Int’l Studies on SL and Communication of the Deaf*, 1989.

- [64] A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Hand tracking and affine shape-appearance handshape sub-units in continuous sign language recognition. In *Proc. Workshop on Sign, Gesture and Activity (SGA), ECCV*, Sep. 2010.
- [65] A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Dynamic affine-invariant shape-appearance model for handshape feature extraction and classification in sign language videos. *Journal of Machine Learning Research*, (under review).
- [66] T. Shanableh, K. Assaleh, and M. Al-Rousan. Spatio-temporal feature-extraction techniques for isolated gesture recognition in arabic sign language. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 37(3):641–650, 2007.
- [67] J. Sherrah and S. Gong. Resolving visual uncertainty and occlusion through probabilistic reasoning. In *British Machine Vision Conference*, pages 252–261, 2000.
- [68] P. Soille. *Morphological Image Analysis: Principles & Applications*. Springer, 2004.
- [69] T. Starner and A. Pentland. Real-time american sign language recognition from video using hidden markov models. *Computational Imaging and Vision*, 9:227–244, 1997.
- [70] T. Starner, J. Weaver, and 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, Dec. 1998.
- [71] T Starner, J Weaver, and A Pentland. Real-time american sign language recognition using desk and wearable computer based video. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20:1371–1375, 1998.
- [72] WC Stokoe. Sign language structure. *Annual Review of Anthropology*, 9(1):365–390, 1980.
- [73] T. Supalla and E. Newport. How many seats in a chair ? the derivation of nouns and verbs in american sign language. *Understanding Language Through Sign Language Research*, pages 91–132, 1978.
- [74] G.J. Sweeney and A.C. Downton. Towards appearance-based multi-channel gesture recognition. In *Gesture Workshop*, pages 7–16, 1996.
- [75] N. Tanibata, N. Shimada, and Y. Shirai. Extraction of hand features for recognition of sign language words. In *Proceedings of the International Conference on Vision Interface*, pages 391–398, 2002.
- [76] J. Terrillon, M. Shirazi, H. Fukamachi, and S. Akamatsu. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *Int’l Conf. on AFGR*, pages 54–61, 2000.
- [77] S. Theodorakis, V. Pitsikalis, I. Rodomagoulakis, and P. Maragos. Recognition with raw canonical phonetic movement and handshape subunits on videos of continuous sign language. In *Proc. Int’l Conf. on Image Processing*, 2012.
- [78] C. Vogler. Extraction of segmental phonetic structures from hamnosys annotations. Technical Report D4.2, Institute for Language and Speech Processing, Greece, January 2011.
- [79] C. Vogler and D. Metaxas. Adapting hidden markov models for asl recognition by using three-dimensional computer vision methods. In *International Conference on Systems, Man, Cybernetics*, volume 1, pages 156–161, 1997.

- [80] C. Vogler and D. Metaxas. Parallel hidden markov models for american sign language recognition. In *International Conference on Computer Vision*, volume 1, pages 116–122, 1999.
- [81] C. Vogler and D. Metaxas. A framework for recognizing the simultaneous aspects of american sign language. *Computer Vision and Image Understanding*, 81:358–384, March 2001.
- [82] C. Vogler and D. Metaxas. A framework for recognizing the simultaneous aspects of american sign language. *Comp. Vis. Image Understanding*, 81(3):358–384, 2001.
- [83] C. Vogler and D. Metaxas. Handshapes and movements: Multiple-channel american sign language recognition. In *Gesture Workshop*, pages 247–258, 2003.
- [84] M.B. Waldron and S. Kim. Isolated asl sign recognition system for deaf persons. *IEEE Trans. on Rehabilitation Engineering*, 3(3):261–271, 1995.
- [85] Y. Wu and T.S. Huang. View-independent recognition of hand postures. In *Proc. Conf. on Computer Vision & Pattern Recognition*, volume 2, pages 88–94, 2000.
- [86] H. D. Yang, S. Sclaroff, and S. W. Lee. Sign language spotting with a threshold model based on conditional random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31:1264–1277, 2009.
- [87] M.-H. Yang, N. Ahuja, and M. Tabb. Extraction of 2d motion trajectories and its application to hand gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1061–1074, Aug. 2002.
- [88] R. Yang, S. Sarkar, and B. Loeding. Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(3):462–477, 2010.
- [89] R.D. Yang, S. Sarkar, and B. Loeding. Enhanced level building algorithm for the movement epenthesis problem in sign language recognition. In *Proc. Conf. on Computer Vision & Pattern Recognition*, pages 1–8, 2007.
- [90] P. Yin, T. Starner, H. Hamilton, I. Essa, and J. Rehg. Learning the basic units in american sign language using discriminative segmental feature selection. In *Proc. ICASSP*, pages 4757–4760, 2009.
- [91] P. Yin, T. Starner, H. Hamilton, I. Essa, and J.M. Rehg. Learning the basic units in American Sign Language using discriminative segmental feature selection. In *Proc. Int’l Conf. on Acoustics, Speech and Signal Processing*, pages 4757–4760, Taiwan, April 2009.
- [92] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Woodland, and P. Valtchev. *The HTK Book*. Entropic Ltd., 1999.
- [93] L.G. Zhang, Y. Chen, G. Fang, X. Chen, and W. Gao. A vision-based sign language recognition system using tied-mixture density hmm. In *Proc. Int’l Conf. on Multimodal interfaces*, pages 198–204. ACM, 2004.
- [94] J. Zieren and K.F. Kraiss. Robust person-independent visual sign language recognition. *Pattern Recognition and Image Analysis*, pages 333–355, 2005.

- [95] J. Zieren, N. Unger, and S. Akyol. Hands tracking from frontal view for vision-based gesture recognition. In *24th DAGM Symposium*, pages 531–539, 2002.

Παράρτημα Α΄

Κατάλογος Δημοσιεύσεων του Συγγραφέα

Αποτελέσματα της μέχρι τώρα έρευνας στην διδακτορικής μας έρευνας έχουν δημοσιευθεί σε διεθνώς αναγνωρισμένα περιοδικά και συνέδρια με κριτή. Ακολουθεί πλήρης κατάλογος των σχετικών δημοσιεύσεων. Ηλεκτρονικά ανάτυπα είναι διαθέσιμα από την ιστοσελίδα :

<http://cvsp.cs.ntua.gr/sth>.

Δημοσιεύσεις σε Διεθνή Περιοδικά με Κριτές

1. A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Dynamic affine-invariant shape-appearance model for handshape feature extraction and classification in sign language videos. *Journal of Machine Learning Research*, (under review).
2. S. Theodorakis and V. Pitsikalis, and P. Maragos. Dynamic vs. static sub-unit modeling for continuous sign language recognition with movement-position cues. *IEEE Trans. on Systems, Man and Cybernetics*, (under revision).

Δημοσιεύσεις σε Διεθνή Συνέδρια με Κριτές

1. S. Theodorakis, V. Pitsikalis, I. Rodomagoulakis, and P. Maragos. Recognition with raw canonical phonetic movement and handshape subunits on videos of continuous sign language. In *Proc. Int'l Conf. on Image Processing*, 2012.
2. V. Pitsikalis, S. Theodorakis, C. Vogler, and P. Maragos. Advances in phonetics-based sub-unit modeling for transcription alignment and sign language recognition. In *Proc. CVPR Workshop Gesture Recognition*, 2011. **(Best Paper Award)**
3. I. Rodomagoulakis, S. Theodorakis, V. Pitsikalis, and P. Maragos. Experiments on global and local active appearance models for analysis of sign language facial expressions. In *Proc. Gesture Workshop*, 2011.
4. S. Theodorakis, V. Pitsikalis, and P. Maragos. Advances in dynamic-static integration of movement and handshape cues for sign language recognition. In *Proc. Gesture Workshop*, 2011.

5. A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Hand tracking and affine shape-appearance handshape sub-units in continuous sign language recognition. In *Proc. Workshop on Sign, Gesture and Activity (SGA), ECCV*, Sep. 2010.
6. A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Affine-invariant modeling of shape-appearance images applied on sign language handshape classification. In *Proc. Int'l Conf. on Image Processing*, Sep. 2010.
7. V. Pitsikalis, S. Theodorakis, and P. Maragos. Data-driven sub-units and modeling structure for continuous sign language recognition with multiple cues. In *Proc. LREC Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 2010.
8. S. Theodorakis, V. Pitsikalis, and P. Maragos. Model-level data-driven sub-units for signs in videos of continuous sign language. In *Proc. IEEE Int'l Conference on Acoustics, Speech, and Signal Processing (ICASSP-2010), Dallas, Texas*, 2010.
9. S. Theodorakis, A. Katsamanis, and P. Maragos. Product-hmms for automatic sign language recognition. In *Proc. IEEE Int'l Conference on Acoustics, Speech, and Signal Processing (ICASSP-2009), Taipei, Taiwan, Apr. 2009*, 2009.

Τεχνικές Αναφορές (δημόσια διαθέσιμες)

1. P. Maragos, V. Pitsikalis, S. Theodorakis, and A. Roussos. Initial report on hmm model training and temporal sign segmentation. Technical Report D2.1, National Technical University of Athens, Greece, January 2010.
2. P. Maragos, V. Pitsikalis, S. Theodorakis, A. Roussos, and I. Rodomagoulakis. Progress report on multimodal fusion. Technical Report D2.2, National Technical University of Athens, Greece, February 2011.
3. P. Maragos, V. Pitsikalis, S. Theodorakis, Rodomagoulakis I., and Antonakos E. Report on integrated continuous sign recognition. Technical Report D2.3, National Technical University of Athens, Greece, February 2012.